Linear Algebra: Complete Notes

August 8, 2023

Systems of Linear Equations

A **linear equation** with variables $x_1, \ldots, x_n$ can be written:

$$a_1x_1 + \cdots + a_nx_n = b$$

Where b and the coefficients a_c are real or complex numbers typically known in advance.

A **system of linear equations**, or a **linear system**, is a collection of one more linear equations with the same variables (e.g. $x_1, \ldots, x_n$).

E.g...

1. $2x_1 - 2x_2 + 3x_3 = 8$
2. $x_2 - 4x_3 = -7$
3. $x_3 = 2$

A **solution** *of the system* is a list of numbers $(s_1, \ldots, s_n)$ that give a valid answer to each equation when utilizing the values of s_c in place of each x_c ($s_1 \to x_1$, and so on).

The set of all possible solutions is the **solution set** *of the linear system.*

Two linear systems are **equivalent** if they have the same solution set.

A linear system is **consistent** if it has one or infinite solutions.

A linear system is **inconsistent** if it has no solution.

If a system has each variable aligned in columns, the matrix is known as a **coefficient matrix**.

If the coefficient matrix has the constant of the equation in the last column, then it is known as the **augmented matrix**.

There are three basic operations, **elemental row operations**, for simplifying a linear system:

1. **Replacement**: Replacing an equation with the addition of it and a multiple of another.
2. **Interchange**: Interchanging two equations.
3. **Scaling**: Multiplying all terms of an equation with by a constant $k \neq 0$.

Two matrices are **row-equivalent** if there exists a sequence of *elemental row operations* that transform one matrix to the other.

If the *augmented matrices* of two linear systems are row-equivalent, then the two systems *have the same solution set.*

This is to say that every set of solutions such as $(s_1, \ldots, s_n)$, however many exist, solve equally both matrices.

A matrix is in **staggered** or **staircase** form if it has these three properties:

1. No row with all 0's for its coefficients lies above a row with non-zero coefficients.

2. Every **principal entry**, or first non-zero number, of a row is *at least* one column to the right of the principal entry in the row above.

3. Every number in the same column below a *principal entry* is 0.

A *staggered* matrix is in ***reduced*** **staircase form** if...

1. The principle entry of every file is 1.

2. Every principal entry is the only entry $\neq 0$ in its column.

To elaborate on the above rules...

- This means that every entry in a row right up until its principal entry *must be* 0 in both staircase forms.

- Staircase form *can* have principal entries not equal to 1; reduced staircase form *requires* each principal entry to be 1 or 0.

- Staircase form *can* have non-zero numbers above its principal entries; reduced staircase forms *requires* every number above a principal entry to be 0.

Every matrix is *row-equivalent* to *one and only one* reduced staircase matrix.

If a staircase matrix U is row-equivalent with a matrix A, then U is **staircase form** of A.

If U is a reduced staircase matrix, then U is *the* reduced staircase form of A.

A **pivot position** in a matrix A is a location in A which corresponds with a principal entry 1 in the staircase reduced form of A.

A **pivot column** is simply a column of A that contains a *pivot position.*

These are the steps of row-reduction for solving a linear system:

1. Write the augmented matrix of the system.
2. Use the row-reduction algorithm to obtain the staircase form.
3. Determine if the system is consistent. If it is not, terminate this process.
4. Bring the matrix to reduced staircase form.
5. Write the resulting system of equations.
6. Rewrite each non-zero equation so that its single *basic variable* is expressed in terms of any of the *free variables* that appear in the equation.

Variables that correspond with the *pivot columns* are known as **basic variables**.

Variables that correspond with with *non-pivot* columns are known as **free variables**.

A **pivot** is a number $\neq 0$ in a *pivot position.*

Every different assignment of a free variable determines a unique solution to the linear system.
Therefore, if there is a free variable, the system has infinite solutions.

In this case where we have the following equations...

- $1x_1 + 0x_2 + 5x_3 = 1$

- $0x_1 + 1x_2 + 1x_3 = 4$
- $0x_1 + 0x_2 + 0x_3 = 0$

...we find that $x_1 = 1 - 5x_3$ and $x_2 = 4 - x_3$. Because x_1 and x_2 can take on any needed values to equate to $1 + 5x_3$ and $4 - x_3$, x_3 is free (*unbound* by other variables) to take on any arbitrary value; the other variables will adjust for it.

Note that x_1, x_2 are *bound* and can not take on any arbitrary value. They are bound to take on needed values determined by the unbound x_3.

Free variables act as **parameters**. Solving a system means to find a *parametric description* of the solution set or to determine that the solution set *is empty*.

A linear system is *consistent* if and only if the column most right of the augmented matrix is **not** a pivot column.

In other words, the matrix is consistent if it does not have any row of type $[0 \ldots 0 = b_c]$ where $b_c \neq 0$ *after reduction*.

If a linear system is consistent, the solution set contains either a unique solution (*without* free variables) or infinite solutions (*with*).

Vector Equations

A matrix of a single column is a **column vector** or simply a **vector**.

Example of vectors include $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} .2 \\ .3 \\ 11 \end{bmatrix}, \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$.

$\mathbb{R}^2$ indicates the set of all vectors with two entries.

$\mathbb{R}$ represents all real numbers. The exponent indicates the number of entries.

Two vectors are equal if and only if the corresponding entries are equal. For example...

$$\begin{bmatrix} 4 \\ 7 \end{bmatrix} \neq \begin{bmatrix} 7 \\ 4 \end{bmatrix}$$

A vector $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \cdot c = \begin{bmatrix} c \cdot 3 \\ c \cdot -1 \end{bmatrix}$.

If $c = 5$, we then have $\begin{bmatrix} 15 \\ -5 \end{bmatrix}$.

In the above case, c is a **scalar**. Unlike vectors which are **bolded** in notation, scalars are typically written in *italics*.

$\mathbb{R}^2$ is the set of all points in the plane.

The sum of vectors $\mathbf{u} + \mathbf{v}$ can be visualized geometrically as forming a parallelogram in which its other vertices, apart from $\mathbf{u} + \mathbf{v}$, are $\mathbf{u}, \mathbf{v}, \mathbf{0}$.

In $\mathbb{R}^n$ where n is a positive integer, it denotes the collection of all lists of n real numbers.

The zero vector is a vector of all 0's denoted $\mathbf{0}$, $\vec{0}$ or simply 0. The number of 0 entries present within the vector is incidated by context.

The equality of vectors in $\mathbb{R}^n$ is evaluated, and operations of scaling and addition are performed, entry by entry.

The **algebraic properties of vectors in** $\mathbb{R}^n$ are the following:

1. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
3. $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$
4. $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$ where $-\mathbf{u} \to (-1)\mathbf{u}$
5. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
6. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
7. $c(d\mathbf{u}) = (cd)(\mathbf{u})$
8. $1\mathbf{u} = \mathbf{u}$

Given the vectors $\mathbf{v_1}, \ldots, \mathbf{v_p}$ and the scalars $c_1, \ldots, c_p$, a combination $\mathbf{y}$ is defined:

$$\mathbf{y} = c_1\mathbf{v_1} + \cdots + c_p\mathbf{v_p}$$

Note that $\mathbf{y}$ is a **linear combination** of vectors $\mathbf{v_c}$ with **weights** c_c, and together their equation represents a *vector equation*.

Each entry of $\mathbf{y}$, y_i is represented:

$$y_i = c_1(\mathbf{v_1})_i + \cdots + c_p(\mathbf{v_p})_i$$

If $\mathbf{v_1}, \ldots, \mathbf{v_p}$ is in $\mathbb{R}^n$, then the set of all linear combinations of those vectors is denotated:

$$\text{Gen}\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$$

The above is the subset of $\mathbb{R}^n$ generated by $\mathbf{v_1}, \ldots, \mathbf{v_p}$. In other words, it is the set of all possible scalar combinations of the given vectors.

In *other* other words, it is the set of all vectors that can be written in the form $c_1\mathbf{v_1} + c_2\mathbf{v_2} + \cdots + c_p\mathbf{v_p}$ with scalars $c_1, \ldots, c_p$.

A **vector equation** of the form $x_1\mathbf{a_1} + \cdots + x_n\mathbf{a_n} = \mathbf{b}$ (as seen just before) has the same solution set as the linear system whose augmented matrix is $[\mathbf{a_1} \ldots \mathbf{a_n} = \mathbf{b}]$.

Matrices are simply alternative notation for vector equations.

Note that $\mathbf{b}$ represents the rightmost column of the augmented matrix and contains the values which the equations of the coefficient matrix are equal to or are *supposed to be* equal to.

It is such that $\mathbf{b}$ can be generated through a linear combination of $\mathbf{a_1}, \ldots, \mathbf{a_n}$ if and only if there exists a *solution to the linear system* which corresponds to the matrix.

If $\mathbf{v}$ is a vector $\neq 0$ in $\mathbb{R}^3$, then $\text{Gen}\{\mathbf{v}\}$ is the set of all the scalar multiples of $\mathbf{v}$ and $\text{Gen}\{\mathbf{v}\}$ forms a straight line extending infinitely between 0 and $\mathbf{v}$ and beyond, positively and negatively.

$\text{Gen}\{\mathbf{u}, \mathbf{v}\}$, where $\mathbf{u} \neq 0 \neq \mathbf{v}$ and $\mathbf{u} \neq \mathbf{v}$, forms a plane with the two lines that pass through $\mathbf{u}, 0$ and $\mathbf{v}, 0$.
It includes all possible scalar combinations of $\mathbf{u}$ and $\mathbf{v}$.

Let $A_{m \times n}$ be a matrix with columns $\mathbf{a_1}, \ldots, \mathbf{a_n}$, and let $\mathbf{x}$ exist in $\mathbb{R}^n$. The product is denoted $A\mathbf{x}$, the linear combination of columns of A with weights from the entries of vector $\mathbf{x}$.

$$A\mathbf{x} = [\mathbf{a_1} \dots \mathbf{a_n}] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a_1} + \cdots + x_n\mathbf{a_n}$$

As can be seen by the notation, each column of A is represented as a vector $\mathbf{a_c}$. Likewise each value of $\mathbf{x}$ is a scalar x_c.

If $A_{m\times n}$ is a matrix with columns $\mathbf{a_1}, \dots, \mathbf{a_n}$ and if $\mathbf{b}$ is in $\mathbb{R}^m$, then the matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solution set as the vector equation...

$$x_1\mathbf{a_1} + \cdots + x_n\mathbf{a_n} = \mathbf{b}$$

And thus the above could be written all as...

$$A\mathbf{x} = \underbrace{\begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix}}_{m\times n} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{n\times 1} = \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}}_{m\times 1} = x_1\mathbf{a_1} + \cdots + x_n\mathbf{a_n} = \mathbf{b}$$

In the above, m could be of any length, and what would change is that each $x_c\mathbf{a_c}$ would become of the new $m \times 1$ size, and $\mathbf{b}$ as well.

See that each x_c exclusively pairs to each column $\mathbf{a_c}$ and multiplies it.

See also that $x_1\mathbf{a_1} + \cdots + x_n\mathbf{a_n}$ has the same solution set as the linear system of equations whose augmented matrix is:

$$[\mathbf{a_1} \dots \mathbf{a_n} = \mathbf{b}]$$

The equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Let $A_{m\times n}$ be a matrix. Then the following are either all true or all false:

1. For every $\mathbf{b}$ in $\mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ has at least one solution. (Note that $\mathbf{b}$ is a matrix $m \times 1$ and lies in the codomain of A.)

2. Every $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A.

3. The columns of A generate $\mathbb{R}^m$.

4. A has a pivot position in every row.

See the following properties:

1. If every row has a pivot, the linear system $A\mathbf{x} = \mathbf{b}$ has ≥ 1 solution for every $\mathbf{b}$.

2. If every column has a pivot, $A\mathbf{x} = \mathbf{b}$ has ≤ 1 solution.

3. If both every row and every column has a pivot, then A *has to be a* ***square*** matrix and $A\mathbf{x} = \mathbf{b}$ has a **single, unique** solution for every $\mathbf{b}$.

See the examples below which correspond with the above properties:

$$A_{3\times 4} = \underbrace{\begin{bmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \end{bmatrix}}_{\geq 1 \text{ solution}}, A_{4\times 3} = \underbrace{\begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \\ 0 & 0 & 0 \end{bmatrix}}_{\leq 1 \text{ solution}}, A_{3\times 3} = \underbrace{\begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}}_{= 1 \text{ solution}}$$

For a $m \times n$ matrix, *there can only be* $\leq \min(m, n)$ *pivots.*

- If there were more pivots than m, there would be rows with more than one pivot, which is impossible.

- If there were more pivots than n, there would be columns with more than one pivot. This is also impossible.

Recall that a pivot (position) requires a reduced staircase form matrix and corresponds with each principal entry of 1. Each principal entry *must* be to the right of the principal entry above.

Also recall that a reduced staircase form matrix can have values $\neq 1, \neq 0$ so long as they are not principal entries or in the column of a principal entry.

E.g., the below matrix is in reduced staircase form despite having two potentially non-zero, non-one values:

$$\begin{bmatrix} 1 & 0 & c_1 & 0 & b_1 \\ 0 & 1 & c_2 & 0 & b_2 \\ 0 & 0 & 0 & 1 & b_3 \end{bmatrix}$$

Matrices of *inconsistent systems* will have a pivot position in the augmented column to the effect of...

$$\begin{bmatrix} 1 & 0 & \dots & \dots & \dots & b_1 \\ 0 & 1 & \dots & \dots & \dots & b_2 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Realize that $0x_1 + \cdots + 0x_n = 0 \neq 1$.

Matrices of **independent systems** have a pivot position in every column of the coefficient matrix. The number of pivots in this case *is equal to* the number of variables being solved for, as each column is a variable.

Matrices of **dependent systems** have fewer pivots than columns. There exist fewer pivots than variables, indicating there are *free variables*.

Note there these two distinct ways of annotating a matrix-vector mulitiplication:

1. $$\begin{bmatrix} 1 & 2 & 4 \\ 2 & -1 & -4 \\ 2 & 1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

2.
$$x\begin{bmatrix}1\\2\\2\end{bmatrix}+y\begin{bmatrix}2\\-1\\1\end{bmatrix}+z\begin{bmatrix}4\\-4\\5\end{bmatrix}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$$

The results above are exactly equivalent to this result below:

1. $1x + 2y + 4z = b_1$
2. $2x - 1y - 4z = b_2$
3. $2x + 1y + 5z = b_3$

The solution to the above (independent) linear system, which is a system of three distinct planes, is the tiny point of intersection between all of them.

If a linear system were *dependent* (and in $\mathbb{R}^3$), then there is the possiblility the solution could be a line or even a plane.
However, the solution to a linearly independent system will always be an exact point.

When thinking of the problem $A\mathbf{x} = \mathbf{b}$ as one of scalars of vectors, the question becomes of which scale factors are required to modify the column vectors so that the their "head-to-tail" sum reachs the exact point of the vector defined by $\mathbf{b}$.

Note that unless two planes are parellel, they will *always* intersect.

To span a space $\mathbb{R}^n$, you need to have n linearly independent vectors. Thus $\mathbb{R}^3$ can not be spanned (or *generated*) by 2 vectors (no matter how many entries they have).

Similarly, for n vectors to be independent, their length must be at least n as well. That is to say that $\mathbb{R}^n$ could not be spanned by n vectors of size

$(n-1) \times 1$.

Vector-Vector and Matrix-Vector Operations

If a vector $\mathbf{u}$ and a vector $\mathbf{v}$ are of equal length n, then $\mathbf{u}\mathbf{v}^T$ and $\mathbf{v}\mathbf{u}^T$ result in matrices of $n \times n$ length.

1. $$\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \underbrace{\begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}}_{\mathbf{v}^T} = \begin{bmatrix} u_1v_1 & \dots & u_1v_n \\ \vdots & \ddots & \vdots \\ u_nv_1 & \dots & u_nv_n \end{bmatrix} = \mathbf{u}\mathbf{v}^T$$

2. $$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \underbrace{\begin{bmatrix} u_1 & \dots & u_n \end{bmatrix}}_{\mathbf{u}^T} = \begin{bmatrix} v_1u_1 & \dots & v_1u_n \\ \vdots & \ddots & \vdots \\ v_nu_1 & \dots & v_nu_n \end{bmatrix} = \mathbf{v}\mathbf{u}^T$$

Note that $\mathbf{u}\mathbf{v}^T \neq \mathbf{v}\mathbf{u}^T$. However $\mathbf{u^T}\mathbf{v} = \mathbf{v^T}\mathbf{u}$ as both result in dot produts, or 1×1 matrices, of the two vectors.

$$\underbrace{\begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}}_{1\times n} \underbrace{\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}}_{n\times 1} = u_1v_1 + \cdots + u_nv_n = [c] = c = \mathbf{v}^T\mathbf{u} = \mathbf{u}^T\mathbf{v}$$

The **row-vector rule for calculating $\boldsymbol{A}\mathbf{x}$** says that if multiplication is defined, the i^{th} entry in $A\mathbf{x}$ is the sum of the products of the corresponding entries of row i of A and vector $\mathbf{x}$.

In demonstration, if $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$, and if we then have $A\mathbf{x} = [\mathbf{a_1}x_1 + \mathbf{a_2}x_2]$, we find that…

1. $(A\mathbf{x})_1 = A_{1,1}x_1 + A_{1,2}x_2 = 1 \cdot 5 + 2 \cdot 2 = \mathbf{9}$

2. $(A\mathbf{x})_2 = A_{2,1}x_1 + A_{2,2}x_2 = 3 \cdot 5 + 4 \cdot 2 = \mathbf{23}$

3. We can find the same results, however, by solving ordinarily so...

$$\mathbf{a_1}x_1 + \mathbf{a_2}x_2 = \begin{bmatrix}1\\3\end{bmatrix} \cdot 5 + \begin{bmatrix}3\\4\end{bmatrix} \cdot 2 = \begin{bmatrix}5\\15\end{bmatrix} + \begin{bmatrix}4\\8\end{bmatrix} = \begin{bmatrix}\mathbf{9}\\\mathbf{23}\end{bmatrix}$$

Notice the same 9's and 23's.

Let $A_{m \times n}$ be a matrix, $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$, and c be a scalar. Then:

1. $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$
2. $A(c\mathbf{u}) = c(A\mathbf{u})$

The following example demonstrates the above principles:

1. If $A \begin{bmatrix}1\\-1\\2\end{bmatrix} = \begin{bmatrix}1\\2\\3\end{bmatrix}$, $A \begin{bmatrix}2\\0\\2\end{bmatrix} = \begin{bmatrix}2\\4\\6\end{bmatrix}$, and $A\mathbf{x} = \begin{bmatrix}3\\6\\9\end{bmatrix}\dots$
2. ...then it follows that $A \begin{bmatrix}2\\0\\2\end{bmatrix} + A \begin{bmatrix}1\\-1\\2\end{bmatrix} = \begin{bmatrix}3\\6\\9\end{bmatrix} = A \begin{bmatrix}3\\-1\\4\end{bmatrix}$
3. But likewise $1.5 \cdot A \begin{bmatrix}2\\0\\2\end{bmatrix} = 1.5 \cdot \begin{bmatrix}2\\4\\6\end{bmatrix} = \begin{bmatrix}3\\6\\9\end{bmatrix} = A \begin{bmatrix}3\\0\\3\end{bmatrix} = A \begin{bmatrix}3\\-1\\4\end{bmatrix}$
4. Note that two distinct solutions for $\mathbf{x}$ exist, and because there is more than one, there are infinite solutions.

The Homogeneous Equation

A system of equations is **homogeneous** if it can be written in the form $A\mathbf{x} = \mathbf{0}$, where $A_{m \times n}$ is a matrix and $\mathbf{0}$ is in $\mathbb{R}^m$. Such a system always has at least 1 solution ($\mathbf{x} = \mathbf{0}$).

This *zero solution* is known as the **trivial solution**. For an equation $A\mathbf{x} = \mathbf{0}$, if a **non-trivial solution** exists in which $\mathbf{x} \neq \mathbf{0}$, then the system must be linearly dependent and have infinite solutions for $A\mathbf{x} = \mathbf{0}$.

I.e., the null space must have at least one vector and can be described as $\operatorname{Nul} A = \operatorname{Gen}\{\mathbf{x}\ldots\}$.

The *homogeneous equation* $A\mathbf{x} = \mathbf{0}$ has a *non-trivial solution* if and only if it has ≥ 1 free variable.

If $A\mathbf{x} = \mathbf{b}$ is consistent for some $\mathbf{b}$ and $\mathbf{p}$ is a solution, then the solution set for the equation is the set of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v_h}$ where $\mathbf{v_h}$ is any solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Thus this means that the solution to a system is the combined solution sets of *both* the given vector equation *and* the homogeneous equation.

This is to say that a particular solution $\mathbf{p}$ can have any single solution to the homogeneous equation added to it while still holding true.

The indexed set of vectors $\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ in $\mathbb{R}^n$ is **linearly independent** if the vector equation $x_1\mathbf{v_1} + \cdots + x_p\mathbf{v_p} = \mathbf{0}$ only has, as its solution, the trivial solution.

In other words, the *columns* of a matrix are linearly independent if and only if the equation $A\mathbf{x} = \mathbf{0}$ only has the trivial solution.

A set of two vectors $\{\mathbf{v_1}, \mathbf{v_2}\}$ is linearly *dependent* if at least one of the vectors is *a multiple of the other.*

If a set contains a larger number of vectors than number of entries *in* each vector, then the set is *linearly dependent.* I.e., f a set has n vectors but each vector has $n-1$, it is linearly dependent.

E.g....

$$\{\mathbf{w_1} = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{w_2} = \begin{bmatrix}0\\1\\0\end{bmatrix}, \mathbf{w_2} = \begin{bmatrix}0\\0\\1\end{bmatrix} \mathbf{w_4} = \begin{bmatrix}-0.5\\1\\.8\end{bmatrix}\}$$

Any set which contains the zero vector is linearly dependent.

Linear Transformations

A **transformation** (also known as a **function** or **map**) T of $\mathbb{R}^n \to \mathbb{R}^m$, is a rule that asigns every vector $\mathbf{x}$ in $\mathbb{R}^n$ a vector $T(\mathbf{x})$ in $\mathbb{R}^m$.

The set of $\mathbb{R}^n$ is the **domain** of the transformation T and the set $\mathbb{R}^m$ is the **codomain** of T.

The above can be annotated $T : \mathbb{R}^n \to \mathbb{R}^m$, meaning that T transforms items in $\mathbb{R}^n$ to items in $\mathbb{R}^m$. Vectors move from the domain to the codomain.

For $\mathbf{x}$ in $\mathbb{R}^n$ and $T(\mathbf{x})$ in $\mathbb{R}^m$, the transformation of its original form to its final form, is the **image** of $\mathbf{x}$. The set of all *images* of $T(\mathbf{x})$ is the **range** of T.

The range is a subset and less than or equal to the codomain.

Occasionally a matrix transformation of the matrix-mulitiplication type will be denoted $\mathbf{x} \rightarrowtail A\mathbf{x}$.

See that $\mathbf{x} \rightarrowtail A\mathbf{x}$ is the same as $A\mathbf{x}$ and $T(\mathbf{x})$; it is simply a difference of notation.

Note the following about domain, codomain, and range:

- The domain of T is $\mathbb{R}^n$ when the matrix A has n columns.
- The codomain of T is $\mathbb{R}^m$ when every column of A has m entries. (This is to say when A has m rows across each column.)
- The range of T is all linear combinations of the columns of A because every image $T(\mathbf{x})$ is of the form $A\mathbf{x}$.
- Note that range then is a *subset* (or *subspace*) of the codomain.

A transformation T is said to be **linear** if…

1. $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}$ in the domain of T.
2. $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars c and for all $\mathbf{u}$ in the domain of T.

If T is a *linear transformation*, then $T(\mathbf{0}) = \mathbf{0}$.

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then there exists a unique matrix A such that…

$$\forall \mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) = A\mathbf{x}$$

This is to say that there exists some matrix A that describes the transformation in total and equivalently. Note that $\mathbf{x}$ is in $\mathbb{R}^n$ because there are n columns and each entry in $\mathbf{x}$ matches to a column.

If A is a $m \times n$ matrix whose j^{th} column is the column vector $T(\mathbf{e_j})$ where $\mathbf{e_j}$ is the j^{th} column of the **identity matrix** in $\mathbb{R}^n$, then the transformation matrix A would be defined $A = [T(\mathbf{e_1}) \dots T(\mathbf{e_n})]$.

The *identity matrix* is a matrix comprised of all 0's apart from 1's which compose the diagonal and its dimensionality is $n \times n$.

$$I_{1\times 1} = I_1 = [1],\ I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ I_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is **over** (*onto*) $\mathbb{R}^m$ if every $\mathbf{b}$ in the *codomain* ($\mathbb{R}^m$) is the image of at least one $\mathbf{x}$ in the *domain* ($\mathbb{R}^n$).

In more obvious terms, if and only if the columns of A generate $\mathbb{R}^m$, is it over.

A transformation $T : \mathbb{R}^n \to R^m$ is **one-to-one** if every $\mathbf{b}$ in the codomain is the image of *at most one* $\mathbf{x}$ in the domain.

In more obvious terms, if and only if all columns of matrix A are linearly independent, the transformation is one-to-one.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a *linear transformation*. Then T is *one-to-one* if and only if $T(\mathbf{x}) = \mathbf{0}$ has the trivial solution.

To emphasize the above's importance, let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and A its **standard matrix**. Then again. . .

1. T maps $\mathbb{R}^n$ over $\mathbb{R}^m$ if and only if the columns of A generate $\mathbb{R}^m$.

2. T is one-to-one if and only if the columns of A are linearly independent.

Further Matrix Operations

In matrix notation where $A = [\mathbf{a_1} \ldots \mathbf{a_n}]$, $a_{i,j}$ represents the i^{th} entry of the j^{th} vector.

This is to say that i represents rows and j represents columns.

E.g., $a_{2,3}$, or $A_{2,3}$, indicates the entry at row 2 and column 3.

The **principal diagonal** of a matrix refers to the diagonal (top-left to bottom-right) line formed by all $a_{i,j}$ where $i = j$ (or all $a_{i,i}$).

A **diagonal matrix** is a matrix whose non-diagonal entries are all 0.

A matrix whose entries are all 0 is a **zero matrix** or **null matrix**.

Two matrices are **equal** if they have the same number of rows and columns and their corresponding columns and entries are equal.

If A and B are matrices of $m \times n$, then the sum of $A + B$ is a matrix of the same size whose columns are the sums of the columns of the addends.

$A + B$ is **only defined** when the matrixes are the same size.

If r is a scalar and A a matrix, the multiple scalar rA is the matrix whose columns are r times the columns of A.

$$-A = (-1)A \text{ and } A - B = A + (-1)B$$

Let A, B, C be matrices of the same size. Let r, s be scalars.

1. $A + B = B + A$
2. $(A + B) + C = A + (B + C)$
3. $A + 0 = A$
4. $r(A + B) = rA + rB$
5. $(r + s)A = rA + sA$

6. $r(sA) = (rs)A$

We see that...

$\mathbf{x}$ times $B = B\mathbf{x}$, and $B\mathbf{x}$ times $A = A(B\mathbf{x}) = \mathbf{x}$ times $AB = (AB)\mathbf{x}$

The **file-column rule for calculating a matrix** AB is that the entry $ab_{i,j}$ is defined as the sum of the products of the corresponding entries of row i of A and column j of B.

$$(AB)_{i,j} = a_{i,1}b_{1,j} + ... + a_{i,n}, b_{n,j}$$

The equation above clarifies that during the process of finding products to sum, the entire i^{th} row of A is multiplied sequentially along with the entire j^{th} column of B.

Thus $(AB)_{1,1}$ calls for multiplications between all of A's entries from the top-left corner straight right (row fixed) and all B's entries from the top-left corner straight down (column fixed).

In better notation:

$$(AB)_{i,j} = \sum_{\mathbf{k}=1}^{n} a_{i,\mathbf{k}} \cdot b_{\mathbf{k},j}$$

Properties of matrix multiplication are illustrated below with $A_{m \times n}$ where B and C are of sizes such that the below operations are defined:

1. $A(BC) = (AB)C$ (Associative Law)
2. $A(B + C) = AB + AC$ (Distributive Law)
3. $(B + C)A = BA + CA$ (Distributive Law)
4. $r(AB) = (rA)B = A(rB)$ for any scalar r
5. $I_mA = A = AI_n$ (Identity for Multiplication of Matrices)

The last property above illustrates that for any matrix of any size, there always exists a I that may multiply it on its left and a I that may multiply it on its right. See...

$$
\begin{aligned}
I_3 A_{3\times 4} &= A_{3\times 4} I_4 = A \\
I_3 A_{3\times 3} &= A_{3\times 3} I_3 = A \\
\rightarrow AI &= IA = A
\end{aligned}
$$

Note the following **warnings**:

1. In general, $AB \neq BA$.
2. Laws of cancellation do not apply to matrix mulitiplication. If $AB = AC$, it is generally not certain that $B = C$. Only if A is invertible can it be ascertained that $B = C$.
3. If a product AB is the **zero matrix**, you generally can not conclude that $A = 0$ or that $B = 0$.

If A is a matrix of $n \times n$, and k is a positive integer, then A^k denotes the product of k copies of A: $A^k = A_1 A_2 \ldots A_k$.

Given a matrix $A_{m\times n}$, the **transpose** of A is $n \times m$ and is denotated A^T. The *transpose's* columns form based on the corresponding rows of A.

The *key* ***formula*** *for the transpose of a matrix is so:*

$$A^T = [a_{j,i}] \text{ where } A = [a_{i,j}]$$

Every element $a_{i,j}$ moves to a location defined by $a_{j,i}$.

Note that when $m \neq n$, the transpose of a matrix $m \times n$ becomes $n \times m$.

$$A = \begin{bmatrix} z & e & t & a \\ c & e & r & o \end{bmatrix} \rightarrow A^T = \begin{bmatrix} z & c \\ e & e \\ t & r \\ a & o \end{bmatrix}$$

Assume that A and B are matrices whose sizes are adequate for following sums and products:

1. $(A^T)^T = A$
2. $(A+B)^T = A^T + B^T$
3. $(rA)^T = r(A^T)$ for any scalar r
4. $(AB)^T = B^T A^T$

Note the following regarding the last property: If A were 3×4 and B were 4×5, then the product of the tranpose would be $B_{5\times4}A_{4\times3}$, which is defined.

Conversely, the intuitive transposed product $A^T B^T$ would be $A_{4\times3}B_{5\times4}$, which is not defined.

The Inverse of a Matrix

A matrix $A_{n\times n}$ is **invertible** if there exists a matrix $C_{n\times n}$ such that:

$$CA = I, AC = I$$

This *unique* inverse of A is denoted A^{-1}.

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

See that order of multiplication between a matrix and its inverse does not matter in the above cases, and both orders resolve in the identity matrix.

If A^{-1} is a matrix, then its inverse could be denoted $(A^{-1})^{-1}$, or just A.

Flipping the notation around, we could write, A^{-1} to be B, and thus its inverse is B^{-1}, the original A.

I.e., one inverse can not be said to be the originator or proprietor of the other.

Let C be the inverse matrix. C is determined uniquely by A such that if B *were* another inverse of A, then it follows:

$$B = BI = B(AC) = (BA)C = IC = C = A^{-1}$$

A matrix that can't be inverted is called a **singular matrix**. A matrix that can is known as a **non-singular matrix**.

Likewise, if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ad - bc \neq 0$, then A is invertible and...

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

If $ad - bc = 0$, the matrix is not invertible.

Similarly, if the *determinant*, which is what is measured by the above equation, is ever $= 0$ for any square matrix, then the matrix *does not* have an inverse.

Conversely, if a square matrix ever has a determinant $\neq 0$, then it is *guaranteed* to have an inverse.

It is **very important** to realize that all *non-square* matrices are **not invertible**.

For this reason, all invertible matrices are $n \times n$, and therefore $\mathbf{b}$, while usually said to be in $\mathbb{R}^m$, is now said to be in $\mathbb{R}^n$.

If A is an invertible matrix of $n \times n$, then ...

$$\forall \mathbf{b} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b} \text{ has a unique solution, } \mathbf{x}, \text{ such that } A^{-1}\mathbf{b} = \mathbf{x}$$

This is analogous to the logic that if $f(x) = y$, then $f^{-1}(y) = x$. Through the inverse function, the output returns the input.

If A is a $n \times n$ matrix and B is its inverse, then $AB = I, BA = I$, and therefore the following holds:

$$AB = [A\mathbf{b}_1 \; A\mathbf{b}_2 \; \ldots \; A\mathbf{b}_n] = [\mathbf{e}_1 \; \mathbf{e}_2 \; \ldots \; \mathbf{e}_n] = I$$
$$BA = [B\mathbf{a}_1 \; B\mathbf{a}_2 \; \ldots \; B\mathbf{a}_n] = [\mathbf{e}_1 \; \mathbf{e}_2 \; \ldots \; \mathbf{e}_n] = I$$

A matrix $A_{n\times n}$ is invertible if and only if A is row-equivalent to I_n. In that case any sequence of *elemental row operations* that reduces A to I_n also transforms I_n to A^{-1}.

Remember that **row equivalence** is determined by the ability to change one marix into another through elementary row operations.

The **algorithm for determining** A^{-1} follows:

1. File-reduce the augmented matrix $[A|I]$.
2. If A is row-equivalent to I, then $[A|I]$ is equivalent by rows to $[I|A^{-1}]$.
3. Otherwise, A does not have an inverse.

The second step above may be thought of as I in $[A|I]$ being built, step by step, into the matrix that summarizes all changes required to take A to I. And thus AA^{-1} takes A to I.

If A is invertible, A^{-1} is also invertible and...

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

If A and B are invertible and of $n \times n$, then so is AB such that...

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

Note the same pattern of distribution as in $(AB)^T = B^T A^T$.

If A is is invertible, then so is A^T and...

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

The **theorem of the invertible matrix** states that if A is a matrix of $n \times n$, then all the following are either all true or all false:

1. A has an inverse.
2. A^T has an inverse.
3. A is row-equivalent to $I_{n\times n}$. (Or more simply, I_n.)
4. A has n pivot positions.
5. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
6. The columns of A form a linearly independent set.
7. The linear transformation $\mathbf{x} \rightarrowtail A\mathbf{x}$ is one-to-one.
8. $A\mathbf{x} = \mathbf{b}$ has at least one solution for all $\mathbf{b}$ in $\mathbb{R}^n$.
9. The columns of A generate $\mathbb{R}^n$.
10. $\mathbf{x} \rightarrowtail A\mathbf{x}$ maps $\mathbb{R}^n$ over $\mathbb{R}^n$
11. $\exists C_{n\times n} : CA = I$
12. $\exists D_{n\times n} : AD = I$

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible if there exists $S : \mathbb{R}^n \to \mathbb{R}^n$ like so:

$$\forall \mathbf{x} \in \mathbb{R}^n : S(T(\mathbf{x})) = \mathbf{x}$$
$$\forall \mathbf{x} \in \mathbb{R}^n : T(S(\mathbf{x})) = \mathbf{x}$$

Because a linear transformation $T(\mathbf{x})$ can be equally well represented by $A\mathbf{x}$, all the above properties of the invertible matrix can be just as well applied.

An **elemental (or *elemental*) matrix** is obtained by performing a single elemental row operation over an identity matrix.

If a matrix $A_{m \times n}$ receives a single elemental row operation, then the resulting matrix could be written as EA where $E_{m \times m}$ is created upon performing the same row operation on I_m.

All elementary matrices E are invertible. The inverse of E is the elemental matrix of the same type that transforms E back to I.

PALU Factorization

A matrix $A_{n \times n}$ is **diagonal** if $A_{i,j} = 0$ for $i \neq j$.

The matrix $D = \text{diag}(d_1, \ldots, d_n)$ denotes a diagonal matrix where $D_{i,i} = d_i$.

Both the *identity* and *null* matrices are diagonal matrices.

If D and F are both diagonal matrices and α is a scalar...

1. $D + F$ and αF are both diagonal.
2. DA is the matrix obtained by multiplying row i of A by d_i from $i = 1$ to $i = n$. (This is to say, by multiplying each row of A by the diagonal's solitary value from its comparable row.)
3. AD is the matrix obtained by multiplying column i of A by d_i as i grows from 0 to n.
4. $FD = DF = \text{diag}(d_1 f_1, \ldots, d_n f_n)$

5. D has an inverse if and only if $d_i \neq 0$ for all i and $D^{-1} = \text{diag}(\frac{1}{d_1}, \ldots, \frac{1}{d_n})$

There exist both *superior* and *inferior* triangle-shaped matrices in which all values either above or below the diagonal are 0.

The matrix $A_{n\times n}$ is a **superior triangle** if $\forall i > j : A_{i,j} = 0$. All diagonal matrices are inherently *superior triangles.*

Assume U, V are superior triangles and α is a scalar. The properties of a superior triangle follow:

1. $U + V$ and αU are superior triangles.
2. UV is a superior triangle. If U, V have ones in their diagonals, then UV also has ones in its diagonal.
3. U has an inverse if and only if $U_{i,i} \neq 0$ for i from 1 up to n. (This is to say that the diagonal of U does not contain any zeroes.)
4. U^{-1} is, if it exists, a superior triangle. If U has ones in its diagonal, then so does U^{-1}.

The matrix $A_{n\times n}$ is a **inferior triangle** if $\forall i < j : A_{i,j} = 0$ (or in other words, when A^T is a superior triangle).

All diagonal matrices are inherently *inferior triangles* as well.

Note...

- If A is a superior triangle, A^T is an *inferior* triangle.
- If A is an inferior triangle, A^T is a *superior* triangle.

Assume L, M are inferior triangles of $n \times n$ and α is any scalar. Then these properties follow:

1. $L + M$ and αL are inferior triangles.

2. LM is an inferior triangle. If L, M have ones in their diagonal, so does their product.

3. L has an inverse if and only if $l_{i,i} \neq 0$ with i from 0 up to n. (Again, this is to say if there are no 0's in its diagonal.)

4. If L^{-1} exists, it is also an inferior triangle. If L has ones in its diagonal, so does its inverse.

A **factorization** of a matrix A is an equation that expresses A as a product of ≥ 2 more matrices.

LU Factorization functions such that rather than $A\mathbf{x} = \mathbf{b}$...

$$U\mathbf{x} = \mathbf{y} \text{ and } L\mathbf{y} = \mathbf{b} \text{ and so therefore } A\mathbf{x} = L(U\mathbf{x}) = \mathbf{b}$$

L and U should be thought of literally as **lower** and **upper**, as in they both represent lower and upper diagonal matrices, respectively.

If A can be reduced to staircase form U using only row replacements that add a multiple of one row to another row ***below* it**, then there exist unitary elemental inferior triangle matrices $E_1, \ldots E_p$ such that...

$$E_p \ldots E_1 A = U$$

Note that the above implies a multiplication of the elementary matrices. Order matters.

From the above, we continue...

$$A = (E_p \ldots E_1)^{-1} U = LU \text{ where } L = (E_p \ldots E_1)^{-1}$$

Let A be $m \times n$. The factorization $A = LU$ is obtained upon bringing the matrix A to the staircase form U by *exclusively using the elemental row operation* ***addition of a multiple of one row to another***.

$A = LU$ expresses each row of A as a linear combination of the rows of U.

Below represents an intuitive representation for a 3×3 matrix:

$$A = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ * & 1 & 0 \\ * & * & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}}_{U} \text{ where } * \in \mathbb{R}$$

Note that $l_{i,j}$ is algorithmically derived by...

$$l_{i,j} = \frac{\text{[element that is deleted/overriden/nullified]}}{\text{[pivot]}}$$

This is to say that when an element in A ($a_{i,j}$) is brought to 0 in U, the corresponding element in A is equal to $a_{i,j}$ divided by the pivot element.

- When $A = LU$, the equation $A\mathbf{x} = \mathbf{b}$ is written as $L(U\mathbf{x}) = \mathbf{b}$
- Writing $\mathbf{y}$ in place of $U\mathbf{x}$, we can find $\mathbf{x}$ by solving the pair of equations $L\mathbf{y} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{y}$.
- First we clear $\mathbf{y}$ from $L\mathbf{y} = \mathbf{b}$ such that $L^{-1}\mathbf{b} = \mathbf{y}$.
- Second we solve $U\mathbf{x} = \mathbf{y}$, where $U^{-1}\mathbf{y} = \mathbf{x}$, to obtain $\mathbf{x}$.

$A = LU$ can not always be created in certain cases where there is **forced row-exchange**. In this case **PALU factorization** is obtained where $PA = LU$.

The matrix P is a **permutation matrix** defined as the *identity matrix* with its rows *interchanged*.

The matrix PA is simply the matrix A with the modifying *permutation* matrix P that exchanges A's rows. See the example below.

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} 3 & 3 & 4 \\ 3 & 2 & 1 \\ 1 & 0 & -9 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 3 & 2 & 1 \\ 3 & 3 & 4 \\ 1 & 0 & -9 \end{bmatrix}}_{PA}$$

The factorization $PA = LU$ is obtained by bringing the matrix A to staircase form U exclusively using the elemental row operations...

1. Subtraction of a muliple of one row from another.
2. Exchanging two rows.

Further, the elemental row operation of replacing one row i with a multiple of itself **can not be performed**.

P may be thought of as tracking the forced row-exchanges as they occur while factoring A to LU.

A *forced row-exchange* occurs when a matrix A that is being reduced to U but can not be resolved to a upper staircase form U through procedural row subtractions.

See that $P(A\mathbf{x} = \mathbf{b}) = (PA\mathbf{x} = Pb)$. Thus this implies the following:

1. $LU\mathbf{x} = P\mathbf{b}$ (Note that our LU, by design and neccesity, already accounts for the permutation effect.)
2. $L\mathbf{y} = P\mathbf{b}$
3. $U\mathbf{x} = \mathbf{y}$

See the following...

1. $L\mathbf{y} = P\mathbf{b} \to \mathbf{y} = L^{-1}P\mathbf{b}$
2. $U\mathbf{x} = \mathbf{y} \to U\mathbf{x} = L^{-1}P\mathbf{b} \to \mathbf{x} = U^{-1}L^{-1}P\mathbf{b}$

Determinants

For 2×2 matrices such as $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the **determinant** is calculated...

$$\det A = ad - bc$$

The determinant of a 3×3 matrix is calculated so:

$$\triangle = \det A = a_{1,1} \cdot \det \begin{bmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{bmatrix} - a_{1,2} \cdot \det \begin{bmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{bmatrix} + a_{1,3} \cdot \det \begin{bmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{bmatrix}$$

Alternatively, it may be annotated so:

$$\text{where } A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \ldots$$

$$\det A = a(ei) - a(hf) - [b(di) - b(gf)] + c(dh) - c(ge) =$$

$$\det A = a[ei - hf] - b[di - gf] + c[dh - ge]$$

It is calculated like a cross product where a, b, c are analogous to $\hat{\imath}, \hat{\jmath}, \hat{k}$.

For $n \geq 2$, the *determinant* de $A_{n \times n} = a_{i,j}$ is the sum of n terms of the form $\pm a_{1,j} \cdot \det A_{1,j}$, with alternating plus and minus signs, where the entries $a_{1,1}, \ldots, a_{1,n}$ are the first row of A.

In notation, this looks like...

$$\det A = a_{1,1} \det A_{1,1} - a_{1,2} \det A_{1,2} + \cdots + (-1)^{1+n} a_{1,n} \det A_{1,n}$$

$$detA = \sum_{j=1}^{n} (-1)^{1+j} a_{1,j} \det A_{1,j}$$

Given $A = [a_{i,j}]$, the **cofactor** (i, j) of A is the number $C_{i,j}$ defined by...

$$C_{i,j} = (-1)^{i+j} \det A_{i,j}$$

With that equation, then...

$$\det A = a_{1,1}C_{1,1} + a_{1,2}C_{1,2} + \cdots + a_{1,n}C_{1,n}$$

The determinant of a $A_{n\times n}$ can be calculated through development of cofactors lengthwise from any row.

$$\det A = a_{i,1}C_{i,1} + \cdots + a_{i,n}C_{i,n}$$

Likewise, through any column...

$$\det A = a_{1,j}C_{1,j} + \cdots + a_{n,j}C_{n,j}$$

The positivity/negativity of the cofactor (i, j) depends on the position of $a_{i,j}$ in the matrix regardless of the sign of $a_{i,j}$. The factor $(-1)^{i+j}$ generates the following pattern:

$$\begin{bmatrix} + & - & + & \dots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$$

If A is a triangular matrix, either *superior* or *inferior*, then $\det A$ is the product of the entries on the principal diagonal of A.

Be aware that the determinant of the product of two matrices is the same as the product of the independent determinants of the matrices. Thus...

$$\det EA = (\det E)(\det A)$$

Let A be a square matrix, then...

1. If a multiple of a row of A is added to another row to produce a matrix B, then $\det B = \det A$.

2. If two files of A are interchanged to produce B, then $\det B = -\det A$.

3. If a row of A is scaled by k to produce B, then $\det B = k \times \det A$. Notice that scaling a single row affects the entire determinant by that multiple k.

Let there exist the square matrix A reduced to staircase form U through replacements and exchanges of rows. Note that this is *always possible* in this scenario. If there are r interchanges, then...

$$\det A = (-1)^r \det U$$

If A is invertible, the entries of U, $u_{i,i}$, are all pivots. This is because $A \sim I_n$ and the entries $u_{i,j}$ have not been scaled to 1 as they would be in reduced staircaise form.

Note that if there exists at least one $u_{i,i} = 0$, then the product $u_{1,1} \dots u_{n,n} = 0$ and therefore so is the determinant.

- When A is invertible, $\det A = (-1)^r \cdot$ [the product of pivots of U]
- When A is not invertible, $\det A = 0$.

Note that the staircase form U is not unique because it is not completely row-reduced and its pivots are not unique. Only the product of its pivots is unique. However, the pivot-product could both be positive and negative; this is why interchanges must be tracked.

Note the following properties about determinants:

1. A square matrix A is invertible if and only if $\det A \neq 0$.
2. Transposing a matrix does not change its determinant. See that if $A_{n\times n}$ is a matrix, then $\det A^T = \det A$.
3. If A and B are matrices of $n \times n$ (i.e., the same square size), then $\det AB = (\det A)(\det B)$.
4. Likewise, even if $AB \neq BA$, it is always true that $\det AB = \det BA$.

5. If A is invertible, $\det A^{-1} = \frac{1}{\det A}$. (That is to say that the determinant of the inverse is equal to the inverse of the determinant of original.)

6. Thus, $\det(BAB^{-1}) = \det A$.

Regarding the last step, see...

$$\det(BAB^{-1}) = \det B \cdot \det A \cdot \det B^{-1} = \det B \cdot \frac{1}{\det B} \cdot A = \det A$$

Miscellaneous Notes and Review

Matrices do not have defined division because they are not always divisible.

However, a matrix when multiplied by its own inverse matrix results in the identity matrix.

This is the closest thing there is to standard arithmetic in which $x \cdot \frac{1}{x} = x \cdot x^{-1} = 1$. That is to say that x times its reciprocal equals 1.

1. Note the inverse of a two-by-two matrix where $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$:

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

2. Now note how A multiplied by A^{-1} results in I:

$$A^{-1}A = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{bmatrix} =$$

$$\frac{1}{ad - bc} \begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix} = \begin{bmatrix} \frac{ad-bc}{ad-bc} & 0 \\ 0 & \frac{ad-bc}{ad-bc} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

As said before, if $\det A \neq 0$, then A has an inverse.

Just as well, if the determinant $\neq 0$, then the vectors comprising the matrix are linearly independent.

If the determinant $= 0$, then the matrix does not have an inverse.

Just as well, if the determinant $= 0$, then the vectors comprising the matrix are linearly dependent.

If we have the equation $XA = B$, then the following $X = \frac{B}{A}$ does not hold. However, $XAA^{-1} = BA^{-1} = XI = X$ does. This is to say that $X \neq \frac{B}{A}$ but that $X = BA^{-1}$.

BA^{-1} can then be calculated assuming that the matrices have appropriate dimensions.

The **theorem of existence and uniquity** indicates that if a row exists in the form $[0 \dots 0 = b_c]$, then the system is inconsistent.

A set of vectors $\{v_1, \dots, v_n\}$ are linearly dependent if and only if one of the vectors is in the span of the others.

Any such vector within the span of other vectors may be removed without affecting the set's span.

If the $\mathbf{0}$ vector is in the set, the set is automatically linearly dependent.

The **column space** is defined as Gen$\{\mathbf{a_1} \ldots \mathbf{a_n}\}$, all of the combinations of the columns of matrix A.

This is to say *column space* is the set of all possible $\mathbf{b}$ for which the system $A\mathbf{x}$ has a solution.

When $n > m$, there are more columns than rows, and this means more variables than needed to describe $\mathbf{b}$ in $\mathbb{R}^m$, and therefore there exist free variables.

Likewise, when $m > n$, that is to say there are more rows than columns, there is too much dimensionality in $\mathbb{R}^m$ to be generated by n. This means that not every $\mathbf{b} \in \mathbb{R}^m$ has a solution.

In other words, when the codomain is smaller than the domain, the codomain is spanned, and when the domain is smaller than the codomain, the codomain can not be spanned.

The **null space** of a matrix A is defined as the set of all possible solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

When A has n columns, the *null space* belongs to $\mathbb{R}^n$ and the null space is a subspace of $\mathbb{R}^n$.

A **base** of a subspace H of $\mathbb{R}^n$ is a linearly independent set in H that generates H.

The set $\{\mathbf{e_1}, \ldots \mathbf{e_n}\}$ with $\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, $\mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$, y $\mathbf{e_n} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$, forms the standard base for $\mathbb{R}^n$.

This is to say that the standard base for $\mathbb{R}^3$, for example, is defined as...

$$\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Notice that these simple bases above share the same pattern as forms the *identity matrix.*

To find the null space, you set the matrix equal to $\mathbf{b} = \mathbf{0}$, and then you reduce it. Afterwards, you can rewrite it as a vector equation and use Gen$\{\ldots\}$ notation.

If the equation Gen$\{\mathbf{a_1} \ldots \mathbf{a_n}\}$ describes the column space of a matrix, then in the case that any $\mathbf{a_c}$ is not unique, it may be removed from Gen$\{\mathbf{a_1} \ldots \mathbf{a_n}\}$ without effect.

Note that the pivot columns of a staircase reduced matrix B of an *original* matrix A do not form the base of A's column space.

Instead it is that the pivot columns of the reduced staircase matrix B inform which columns of the original matrix A, in their original state, form the column space.

Review of Midterm 1 Difficulties

Problem 1

Given the equations $x + ky = 1$ and $kx + y = 1$, find which values of k create a unique solution, no solution, and infinite solutions.

First, we need to transform the matrices to augmented matrix form to satisfy the rubric:

$$\begin{bmatrix} 1 & k & 1 \\ k & 1 & 1 \end{bmatrix} \xrightarrow{f_1 \cdot k} \begin{bmatrix} k & k^2 & k \\ k & 1 & 1 \end{bmatrix}$$

By multipling row 1 by k, we are given a form of row 1 suitable for subtraction.

$$\begin{bmatrix} k & k^2 & k \\ k & 1 & 1 \end{bmatrix} \xrightarrow{f_2 - f_1} \begin{bmatrix} k & k^2 & k \\ 0 & 1 - k^2 & 1 - k \end{bmatrix}$$

When $k = 1$, we find that our matrix becomes...

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

This indicates a singular pivot row and column and a consistent solution. Because y is associated with a *non-pivot column*, we know that y is a free variable.

See that $x = 1 - y$ and that y is free. Thus when $k = 1$, we have infinite solutions.

When $k = -1$, we find that...

$$\begin{bmatrix} k & k^2 & k \\ 0 & 1 - k^2 & 1 - k \end{bmatrix} \xrightarrow{k \leftarrow (-1)} \begin{bmatrix} -1 & (-1)^2 & -1 \\ 0 & 1 - (-1)^2 & 1 - (-1) \end{bmatrix} = \begin{bmatrix} -1 & 1 & -1 \\ 0 & 0 & 2 \end{bmatrix}$$

Because $0x + 0y \neq 2$, which can be seen in the second row, the system is inconsistent.

We then find that when $1 \neq k \neq -1$, the system has two pivots, a pivot for every column and row, and thus has a single unique solution for a given k. That is two say there is *one* unique set of x, y for every $k \neq 1, k \neq -1$.

Note that k can not take on multiple values simultaneously, and thus there *are not* infinite solutions for the above.

Problem 3

The main takeway from this is that if a third vector is not in the span of two other vectors, and given that the other two vectors are *not* parallel (i.e., they are linearly independent), then the combination of the third vector with the original two creates a linearly independent set.

Problem 6

Given the reduced staircase form matrix $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$, whose columns are $\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3}, \mathbf{a_4}$, and the knowledge that $\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3}, \mathbf{a_4} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, then we know the following:

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

That is to say that a particular solution to the equation $A\mathbf{x} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ is $\mathbf{p} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

This particular solution could then be added with a preexisting solution set for the homogeneous equation to determine the general solution to equation.

Note that general solution is all possible solutions which create the **0** vector (and thus do not affect a particular solution) in addition to a particular solution.

Problem 7

If we have the following...

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} A = I$$

We know that the matrix which reduces A to I is A's inverse. Thus the brace below which covers every matrix to the left of A defines its inverse:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}_{A^{-1}} A = I$$

Note that we can then find A's value through repeatedly dividing the component elemental matrices of its inverse. See that if $A^{-1}A$ is written $\underbrace{E_x E_y E_z}_{A^{-1}} A = I$, then we can perform...

$$(E_x)^{-1} E_x E_y E_z A = (E_x)^{-1} I$$

$$E_y E_z A = (E_x)^{-1} I$$

And following the above pattern, we arrive at...

$$A = (E_z)^{-1} (E_y)^{-1} (E_x)^{-1} I$$

Or, in the case of our matrix above, then we have the following:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{4} \end{bmatrix}^{-1} I$$

We can find the inverses of these matrices by row-reducing them through the algorithm $[E|I]$ or we can intuitively understand them.

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

1. This is a simple swap between row 1 and row 3. To re-swap any two rows, you repeat the operation. Thus, it is its own inverse.

2. This is an addition of row 2 to row 3. Thus, we need to substract row 2 from row 3.

3. This is a scaling of row 3 by $-\frac{1}{4}$. Thus we need to scale row 3 by the fraction's reciprocal, -4.

4. This is the identity matrix and can be ignored. It has no effect.

We then find the final value of A in terms of these three matrices to be...

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

The Rule of Cramer

The Rule of Cramer says that if $A_{n \times n}$ is invertible, then for any $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution $\mathbf{x}$, as in $A\mathbf{x} = \mathbf{b}$ has its entries given by the following formula:

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, \text{ where } i = 1, \ldots, n$$

Note that this notation means that $\mathbf{b}$ is replacing the i^{th} column, and $\det A_i(\mathbf{b})$ is the determinant of the entire A matrix with its i^{th} column replaced by $\mathbf{b}$.

Thus if $A = \begin{bmatrix} 3 & -2 \\ -5 & 4 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 6 \\ 8 \end{bmatrix}$, then we have the following:

1. We define $A_j(\mathbf{b})$ for all $j\ldots$

$$A_1(\mathbf{b}) = \begin{bmatrix} 6 & -2 \\ 8 & 4 \end{bmatrix}, A_2(\mathbf{b}) = \begin{bmatrix} 3 & 6 \\ -5 & 8 \end{bmatrix}$$

2. Then we find x_1, x_2 via...

$$\frac{\det(A_1(\mathbf{b}))}{\det A} = x_1, \ \frac{\det(A_2(\mathbf{b}))}{\det A} = x_2$$

Finally, our $\mathbf{b}$ as in $A\mathbf{x} = \mathbf{b}$ is given by $\mathbf{b} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 27 \end{bmatrix}$

Note that j^{th} column of A^{-1} is an $\mathbf{x}$ that satisfies the equation $A\mathbf{x} = \mathbf{e_j}$.

In other terms, $A \cdot \mathbf{a_j^{-1}} = \mathbf{e_j}$. I.e., a j^{th} column of A^{-1} when multiplied by A gives the corresponding j^{th} column of the identity matrix.

See the following example:

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

2. $$\begin{bmatrix} 2 & 4 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ \frac{3}{4} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \text{ and } \begin{bmatrix} 2 & 4 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Extrapolating Cramer's Rule, we see that...

$$\{\text{entry } (i,j) \text{ of } A^{-1}\} = x_i = \frac{\det A_i(\mathbf{e_j})}{\det A}$$

Remember $\det A_i(\mathbf{e}_j) = (-1)^{i+j} \det A_{j,i} = C_{j,i}$ where $C_{j,i}$ is a cofactor of A. Note that $A_{j,i}$ denotes the submatrix formed by eliminating row j and column i.

Thus the above could be rewritten $x_i = \frac{C_{j,i}}{\det A}$.

The source of the j is the column from which the vector $\mathbf{x}$ is derived, which could then be used in place to specify that...

$$x_i = \underbrace{(\mathbf{a_j^{-1}})_i}_{A^{-1}_{i,j}} = \frac{C_{j,i}}{\det A}$$

As an example...

1. If $A = \begin{bmatrix} 2 & 4 \\ 3 & 4 \end{bmatrix}$, then an $\mathbf{x}$ could be $\mathbf{a_1^{-1}}$, the first column of the inverse matrix.
2. Then $\mathbf{x}$ takes the form $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ such that $x_1 = \frac{C_{1,1}}{\det A} = \frac{(-1)^{1+1} \cdot 4}{-4} = -1$.
3. Likewise, $x_2 = \frac{C_{1,2}}{\det A} = \frac{(-1)^{(1+2)} \cdot 3}{-4} = \frac{-3}{-4} = \frac{3}{4}$.
4. From that then, $\mathbf{x}$ (or $\mathbf{a_1^{-1}}$) $= \begin{bmatrix} -1 \\ \frac{3}{4} \end{bmatrix}$.

Therefore, because each individual component of the matrix could be written in terms of cofactors C and a division of the determinant, an alternative formula for the inverse *as a whole* can be derived.

$$A^{-1} = \frac{1}{\det A} \cdot \underbrace{\begin{bmatrix} C_{1,1} & \dots & C_{n,1} \\ \vdots & \ddots & \vdots \\ C_{1,n} & \dots & C_{n,n} \end{bmatrix}}_{\text{adjugate/adjoint}}$$

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

The **adjugate** or **adjoint matrix** is defined as the *transpose* of *specically **the cofactor matrix***.

Thus the above could be rewritten as $A^{-1} = \frac{1}{\det A} C^T$.

The matrix of cofactors would ordinarily be written as...

$$\begin{bmatrix} C_{1,1} & \dots & C_{1,n} \\ \vdots & \ddots & \vdots \\ C_{n,1} & \dots & C_{n,n} \end{bmatrix}$$

If A is a matrix 2×2, then the area of its parallelogram defined by its columns is $|\det A|$. Likewise, if A is 3×3, then the area of the parallelepiped defined by its columns is also $|\det A|$.

It is easy to intuit the sense of this in terms of a rectangle R in which its matrix of coordinates is $\begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$ such that it its reaches out from the origin 5 across the x-axis and 10 across the y-axis. Then $\det R = 5 \cdot 10 - (0 \cdot 0) = 50$.

If two vectors $\mathbf{a}_1, \mathbf{a}_2$ do not equal 0, then for any scalar c, $\det[\mathbf{a_1}\ \mathbf{a_2}] = \det[\mathbf{a}_1\ (\mathbf{a}_2 + c\mathbf{a}_1)]$.

This is an extension of the property of determinants that the addition of a multiple of one row to another does not affect the determinant. Note a column vector is a row vector for a transpose, and tranposes maintain the same determinant.

Geometrically, this is because the distance of the base, defined by $\mathbf{0}$ and $\mathbf{a}_1$, and the height, defined by the parallel line that runs through $\mathbf{a_2}$ across the line formed by $\mathbf{0}, \mathbf{a}_1$ does not change when adding a scalar value of $\mathbf{a_1}$ to $\mathbf{a_2}$.

Note that translating (moving/shifting) a parallelogram does not change its area.

If T is a linear transformation and S is a set in the domain of T, then $T(S)$ is the set of images of points in S.

The area of $T(S)$ can be calculated so:

$$\text{area}(T(S)) = |\det A| \cdot \text{area}(S)$$

If $S = \{\mathbf{b_1} + \mathbf{b_2}\}$, then...

$$T(S) = T(\mathbf{b}_1 + \mathbf{b}_2) =$$
$$T(\mathbf{b}_1) + T(\mathbf{b}_2) =$$
$$T(S) = A\mathbf{b}_1 + A\mathbf{b}_2$$

Where A is the transformation matrix as in $T(\mathbf{x}) = A\mathbf{x}$. This can then be rewritten as the product of two matrices:

$$T(S) = AB \text{ where } B = [\mathbf{b}_1 \ \mathbf{b}_2]$$

Thus, the area$(T(S)) = \text{area}(AB) = |\det A| \cdot |\det B|$.

Following from the previous logic that the area of a parallelogram remains the same when one vector is modified by a a multiple of the other vector...

$$\text{area}(T(\mathbf{p} + S)) = \text{area}(T(\mathbf{p})) + \text{area}(T(S)) = \text{area}(T(S))$$

Where $\mathbf{p}$ is a vector and S is a parallelogram. Note that the area of a vector $= 0$.

Vector Spaces and Subspaces

A **vector space** is a non-empty space V of objects called *vectors* that are defined by the two operations *sum* and *multiplication by real scalars (real numbers)*.

Vector spaces are subject to the following 10 rules valid for vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V and for all scalars v, d.

1. $(\mathbf{u} + \mathbf{v}) \in V$
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
4. $\exists \mathbf{0} \in V : \mathbf{u} + \mathbf{0} = \mathbf{u}$
5. $\forall \mathbf{u} \exists (-\mathbf{u}) : \mathbf{u} + (-\mathbf{u}) = \mathbf{0} \wedge \{\mathbf{u}, -\mathbf{u}\} \in V$
6. $c \cdot \mathbf{u} = c\mathbf{u} \in V$
7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$
10. $1\mathbf{u} = \mathbf{u}$

Note that for every $\mathbf{u}$ in V and scalar c...

- $0\mathbf{u} = \mathbf{0}$
- $c\mathbf{0} = \mathbf{0}$
- $-\mathbf{u} = (-1)\mathbf{u}$

Examples of vector spaces are the set $\mathbb{S}$ of double-infinite series of numbers where we have, for example, $\{y_k\} = \{\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots\}$.

Likewise, another example is the set $\mathbb{P}_n$. The n of the set $\mathbb{P}$ informs the maximal power of the polynomials. It includes all polynomials in the form of...

$$\mathbf{p}(t) = a_0 + a_1 t + \cdots + a_n t^n$$

Another example is the set of all defined real-value functions in $\mathbb{D}$.

A **vector subspace** is an adequate subset of vectors of a larger vector space. In the case of subspaces, only three of the 10 axioms must be satisfied.

A *subspace* of a *vector space* V is a subset H of V that has these properties:

1. $\mathbf{0} \in H$
2. H is closed under the sum of vectors. I.e., $\forall \mathbf{u} \forall \mathbf{v} \in H : (\mathbf{u} + \mathbf{v}) \in \mathbf{H}$.
3. H is closed under the multiplication by scalars. I.e., $\forall c \forall \mathbf{u} \in H : c\mathbf{u} \in H$.

The above properties guarantee that a subspace H in V is *itself* a *vector subspace.*

Every subspace is a vector space. Likewise, every vector space is a subspace.

The set that contains only the zero vector in the vector space V is the **zero subspace** and is written $\{\mathbf{0}\}$.

Note that $\mathbb{P}$ is a subspace of all the real-value functions defined in $\mathbb{R}$ ($\mathbb{D}$). Likewise, for every $n \geq 0$, $\mathbb{P}_n$ is a subspace of $\mathbb{P}$.

However, the vector space $\mathbb{R}^2$ is not a subset of the vector space $\mathbb{R}^3$ because $\mathbb{R}^2$ is not even a subset of $\mathbb{R}^3$. The set below...

$$\left\{ \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} : s, t \in \mathbb{R} \right\}$$

...looks and acts like $\mathbb{R}^2$ but is logically distinct. All of the vectors of $\mathbb{R}^3$ have three entries and all of the vectors of R^2 have two entries.

Both a plane in $\mathbb{R}^3$ and a line in $\mathbb{R}^2$ that do not pass through the origin are not subspaces of their respective vector spaces. This is because they do not contain the *zero vector* ($\mathbf{0}$).

Recall that the term **linear combination** refers to any sum of scalar multiples of vectors, and $\text{Gen}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is the set of all vectors that can be written as linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$.

Every subspace of $\mathbb{R}^3$ that is distinct from $\mathbb{R}^3$ itself is represented by the generator $\text{Gen}\{\mathbf{v}_1, \mathbf{v}_2\}$ and forms a plane that intersects the origin. If this subspace formed more than a plane, than it would form a hyperplane and encompass $\mathbb{R}^3$ and thus not be a subset.

If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are in a vector space V, then their generator, $\text{Gen}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subspace of V known as a **generated subspace** by $\mathbf{v}_1, \ldots, \mathbf{v}_p$.

Given a subspace H of V, a set generator for H is a set $\mathbf{v}_1, \ldots, \mathbf{v}_p$ in H such that $H = \text{Gen}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

Using the above, we can show that $H = \{(a - 3b, b - a, a, b) : a, b \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^4$. First we place the set into vector form...

$$H = \begin{bmatrix} a - 3b \\ b - a \\ a \\ b \end{bmatrix}$$

By factoring out a and b, we can see that...

$$H = a \underbrace{\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{v}_1} + b \underbrace{\begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_2}$$

We are then left with a vector equation that can be simplified to Gen$\{\mathbf{v}_1, \mathbf{v}_2\}$.

We can test if a set H, defined as comprising all the points in $\mathbb{R}^2$ of the form $\langle 3s, 2+5s \rangle$, is a subspace like so...

1. Note that every point in this set satisfies the equation $\mathbf{i} = 3s$ and $\mathbf{j} = 2 + 5s$.
2. To test if it closed under multiplication, we take any vector inside, such as where $s = 3$, which returns $\langle 9, 17 \rangle$.
3. Then we multiply by an arbitrary scalar integer, such as 2: $2 \cdot \langle 9, 17 \rangle = \langle 18, 34 \rangle$.
4. Then we test if the new vector still holds for the set's constraints...
$$3s = 18 \to s = 6$$
$$2 + 5s = 34 \to s = \frac{32}{5}$$
5. We can see that $6 \neq \frac{32}{5}$, and thus we know the set is not a subspace because it is not closed under mulitiplication.

The **null space** of $A_{m \times n}$ is a subspace of $\mathbb{R}^n$ because every valid null vector has n entries.

Given A, then $A \cdot \mathbf{x}$, where $\mathbf{x}$ is $n \times 1$, results in a $m \times 1$ column vector, usually denoted $\mathbf{b}$.

In the case of the null space, each $\mathbf{x}$ originating from within it results in $\mathbf{0}$ when multiplied by A.

The **column space** of A is a subspace of $\mathbb{R}^m$ as each column vector of A, $\mathbf{a_c}$, has m entries.

The column space, which is a subspace of $\mathbb{R}^m$, is equal to the entire space of $\mathbb{R}^m$ if and only if $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{R}^m$.

Linear transformations follow the same rules in vector spaces, where $\mathbf{x} \in V \to T(\mathbf{x}) \in W$, such that...

- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
- $T(c\mathbf{u}) = cT(\mathbf{u})$

The *null space* of a transformation is also known as the **nucleus**.

The **range** of a transformation $T : V \to W$ is all the vectors of the form $T(\mathbf{x})$ for some $\mathbf{x} \in V$.

Being that $T(\mathbf{x})$ can be modeled by $A\mathbf{x}$ for some matrix A, the nucleus and the image of T are simply the null space and column space of A.

Bases

Recall that an indexed set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent only if the only solution to $\mathbf{0}$ is trivial.

It is linearly dependent is there is a non-trivial solution. In that case, there is a **relation of linear dependence** between $\mathbf{v}_1, \ldots, \mathbf{v}_p$.

See that $\{t, \sin t, \cos 2t, \sin t \cos t\}$ is an example of a linearly independent set of functions defined in $\mathbb{R}$.

1. This is to say that for all t, there is no linear combination of weights to make the equation, $c_1\mathbf{v_1} + \cdots + c_p\mathbf{v_p} = \mathbf{0}$, valid.
2. See that for $t = 0$, $a \cdot t + b \cdot \sin t + c \cdot \cos 2t + d \cdot \sin t \cos t = \mathbf{0}$ becomes...

$$a \cdot 0 + b \cdot 0 + c \cdot 1 + d \cdot 0 = \mathbf{0} \to c = 0$$

3. As c is equal to 0, we have found our required weight of c (for one value of t) that could lead to a non-trivial linear combination.
4. We can now substitute that $c = 0$ in to simplify our equation, effectively ignoring its associated function:

$$a \cdot t + b \cdot \sin t + d \cdot \sin t \cos t = \mathbf{0}$$

5. By setting $t = 2\pi$, we find $a \cdot 2\pi + b \cdot 0 + d \cdot 0 = \mathbf{0} \to a = 0$
6. Continuning with $b \cdot \sin t + d \cdot \sin t \cos t = \mathbf{0}$, when $t = \frac{\pi}{2}$, we find that $b \cdot 1 + d \cdot 0 \ = \mathbf{0} \to b = 0$.
7. For $t = \frac{\pi}{4}$, we see $d \cdot (\frac{\sqrt{2}}{2})^2 = \mathbf{0} \to d = 0$.
8. Thus, although we have not explored all values of t, in the few we have explored, we have already determined that a combination of all 0's is required to satisfy the homogeneous equation, and therefore it is impossible that there exists a non-trivial solution for all of t (i.e., even more values of t).

An indexed set of two or more vectors, with $\mathbf{v_1} \neq 0$, is linearly dependent if there exists some $\mathbf{v_j}$, where $j > 1$, that is a linear combination of the vectors before it.

If H is a vector subspace of V, a set of indexed vectors $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_p}\}$ in V is a **base** of H if...

1. $\mathcal{B}$ is a linearly independent set.
2. The subspace generated by $\mathcal{B}$ is equal to H, which is to say...

$$H = \text{Gen}\{\mathbf{b_1}, \dots, \mathbf{b_p}\}$$

If S is a set such that $S = \{\mathbf{v_1}, \dots, \mathbf{v_p}\} \in H$ y $H = \text{Gen}\{\mathbf{v_1}, \dots, \mathbf{v_p}\}$, then...

- If one of the vectores of S, $\mathbf{v_c}$, is a linear combination of the remaining vectors in S, then the set formed by S upon eliminating that vector $\mathbf{v_c}$ still forms H.
- If $H \neq \mathbf{0}$, some subset of S is a base for H. In other words, every set that is not comprised of only the zero vector has a base.

If $\mathcal{B} = \{\mathbf{b_1}, \dots, \mathbf{b_n}\}$ is a base for a vector space V, $\forall \mathbf{x} \in V$, there exists a unique set of scalars $\{c_1, \dots, c_n\}$ such that $\mathbf{x} = c_1\mathbf{b_1} + \cdots + c_n\mathbf{b_n}$.

In other words, there exists only one solution to every $\mathbf{x}$ for the given base.

This equation, $\mathbf{x} = c_1\mathbf{b_1} + \cdots + c_n\mathbf{b_n}$, is known as **the coordinates of x with respect to the base $\mathcal{B}$** (otherwise known as the **$\mathcal{B}$-coordinates of x**).

It can be denoted as the following vector:

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

The vector is the **vector of $\boldsymbol{x}$-coordinates with respect to $\mathcal{B}$**, also known as the **vector of $\mathcal{B}$ of $\boldsymbol{x}$**.

Note that $\mathbf{x} \rightarrowtail [\mathbf{x}]_{\mathcal{B}}$ is a coordinate transformation determined by $\mathcal{B}$.

If $[\mathbf{x}]_\mathcal{B} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, this implies $\mathbf{x} = 3\mathbf{b_1} + 1\mathbf{b_2}$.

See that $[\mathbf{x}]_\mathcal{B}$ describes the way to reach some universal vector $\mathbf{x}$ in terms of $\mathcal{B}$.

Implicitly, every $\mathbf{x}$ prior informed how to reach $\mathbf{x}$ in terms of the expected standard base. Thus prior $\mathbf{x}$'s could be written $[\mathbf{x}]_\mathcal{E}$.

And in that regard, $\mathcal{E} = \{\mathbf{e_1}, \mathbf{e_2}\}$ for $\mathbb{R}^2$ where $\mathbf{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

$\mathbf{P}_\mathcal{B}$ is called a **matrix of change of coordinates *of* $\mathcal{B}$** to the standard base in $\mathbb{R}^n$. Where $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$, $P_\mathcal{B} = [\mathbf{b_1} \ldots \mathbf{b_n}]$. $P_\mathcal{B}$ is the matrix of the base. When this *matrix of change of coordinates* transforms a vector in its respective language, such as $P_\mathcal{B}[\mathbf{x}]_\mathcal{B}$, the vector is terms of the standard base is found.

$$P_\mathcal{B}[\mathbf{x}]_\mathcal{B} = \mathbf{x}$$

This thus implies that $P_\mathcal{B}^{-1}\mathbf{x} = [\mathbf{x}]_\mathcal{B}$.

If $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ is a base for a vector space V, then it the transformation $\mathbf{x} \rightarrowtail [\mathbf{x}]_\mathcal{B}$ is one-to-one from V to $\mathbb{R}^n$.

The linearity of the transformation of coordinates extends to linear combinations as well. If $\mathbf{u_1}, \ldots, \mathbf{u_p}$ are in V and $c_1, \ldots, c_p$ are scalars, then...

$$\underbrace{[c_1\mathbf{u_1} + \cdots + c_p\mathbf{u_p}]_\mathcal{B}}_{\text{vector in } \mathcal{B}\text{-coordinates}} = c_1[\mathbf{u_1}]_\mathcal{B} + \cdots + c_p[\mathbf{u_p}]_\mathcal{B}$$

This is to say that a linear combination of vectors and scalars (resulting in a single vector) in $\mathcal{B}$-coordinates is the same as the linear combination of those vectors indiviudally in $\mathcal{B}$-coordinates with the scalars outside.

The Dimension of a Vector Space

If a vector space V has the base $\mathcal{B} = \{\mathbf{b_1} \dots \mathbf{b_n}\}$, then any set in V that has more than n vectors *must be linearly dependent.*

If a vector space V has a base of n vectors, then every base of V must consist of exactly n vectors. If there were more, then as above, there would be extraneous vectors; if there were fewer, then the entire space of V could not be generated and a base could not be formed.

- If V is generated by a finite set, then it has **finite dimension** and the dimension, $\dim V$, is the number of vectors in a base for V.
- If V is not generated by a finite set, then V has **infinite dimension**.
- If a vector space is defined by $\{\mathbf{0}\}$, then it has 0 dimension.

If V is a vector space of dimension p where $p \geq 1$, then any linearly independent set of exactly p elements in V is automatically a base for V.

If A is $m \times n$, then every row of A has n entries such that each row is a vector in $\mathbb{R}^n$.

The set of all linear combinations of the row vectors is known as the **row space** and is denoted, in Spanish, $\text{Fila}\, A$.

As every row has n entries, the *row space* is a subspace of $\mathbb{R}^n$.

As the rows of a matrix A are equivalent to the columns of A^T, the column space of A^T is equivalent to the row space of A.

$$\text{Col}\, A^T \equiv \text{Fila}\, A$$

If two matrices A and B are row-equivalent, then their row spaces are equal.

If B is in staircase form, the rows of B different from $\mathbf{0}$ form a base for the row space of both A and B.

Note, although A and B share the same row space generators, they *do not* share the same column space generators.

Note that the **range** of A is the *dimension* of the column space of A.

Be aware that the range and dimension are not equivalent to subspace's n or m as seen in $\mathbb{R}^n$ or $\mathbb{R}^m$.

The columns could be a subspace of $\mathbb{R}^{10}$ and yet A could still have a range and dimensionality of 1.

Dimensionality is always $\leq \min(m, n)$ and describes how greatly matrices can describe the spaces their vectors inhabit.

The dimensions of the column space and row space of a matrix $A_{m\times n}$ are equal. This common dimension, the *range* of A, is also equal to the number of *pivot positions* in A. Therefore...

$$\operatorname{range} A + \dim \operatorname{Nul} A = n$$

The above equivalently says that the number of pivot positions and non-pivot positions is equal to the total number of columns.

Additionally, more rules arise regarding an invertible matrix $A_{n\times n}$:

1. The columns of A form a base of $\mathbb{R}^n$.
2. $\operatorname{Col} A = \mathbb{R}^n$

3. $\dim \operatorname{Col} A = n$
4. $\operatorname{range} A = n$
5. $\operatorname{Nul} A = \{\mathbf{0}\}$
6. $\dim \operatorname{Nul} A = 0$

Note that the dimension of a space is calculated as the number of linearly independent vectors that form it.

Change of Base

If $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ and $\mathcal{C} = \{\mathbf{c_1}, \ldots, \mathbf{c_n}\}$ are both valid bases for a vector space V, then there exists a matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ of $n \times n$ such that...

$$[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$

This is to say that there exists some matrix that converts vectors encoded in $\mathcal{B}$ to vectors encoded in $\mathcal{C}$.

The columns of $P_{\mathcal{C}\leftarrow\mathcal{B}}$ are the vectors of base $\mathcal{B}$ encoded in $\mathcal{C}$.

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = [[\mathbf{b_1}]_{\mathcal{C}} \ldots [\mathbf{b_n}]_{\mathcal{C}}]$$

Thus, the following holds:

$$[\mathbf{b_1}]_{\mathcal{C}} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad [\mathbf{b_2}]_{\mathcal{C}} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

And thus too we can see:

$$[\mathbf{c_1}\,\mathbf{c_2}] \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \mathbf{b_1} \qquad [\mathbf{c_1}\,\mathbf{c_2}] \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \mathbf{b_2}$$

Because the above are linear systems which can be solved through row reduction of $[\mathbf{c_1}\,\mathbf{c_2}]$, we can solve for both columns simultaneously:

$$[\mathbf{c_1}\,\mathbf{c_2}|\mathbf{b_1}\,\mathbf{b_2}] \sim [I|P_{\mathcal{C}\leftarrow\mathcal{B}}]$$

The above is the formula for finding the change-of-base matrix and generalizes to larger matrices than 2×2.

Likewise as $P_{\mathcal{C}\leftarrow\mathcal{B}}$ translates from $\mathcal{B}$ to $\mathcal{C}$, there exists a matrix to convert from $\mathcal{C}$ back to $\mathcal{B}$...

$$(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = P_{\mathcal{B}\leftarrow\mathcal{C}}$$

$$(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1}[\mathbf{x}]_{\mathcal{C}} = [\mathbf{x}]_{\mathcal{B}}$$

If there exists both $\mathcal{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ and the *standard base* $\mathcal{E} = \{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$ and both are in $\mathbb{R}^n$, then $[\mathbf{b_1}]_{\mathcal{E}} = \mathbf{b_1}$. The same applies for other vectors in $\mathcal{B}$. As a result, $P_{\mathcal{E}\leftarrow\mathcal{B}} = P_{\mathcal{B}}$.

For every $\mathbf{x}$ in $\mathbb{R}^n$ where $\mathcal{B}, \mathcal{C}$ are valid bases...

- $P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{E}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$
- $P_{\mathcal{C}}[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{E}\leftarrow\mathcal{C}}[\mathbf{x}]_{\mathcal{C}} = \mathbf{x}$
- $P_{\mathcal{C}}^{-1}\mathbf{x} = P_{\mathcal{E}\leftarrow\mathcal{C}}^{-1}\mathbf{x} = P_{\mathcal{C}\leftarrow\mathcal{E}}\mathbf{x} = [\mathbf{x}]_{\mathcal{C}}$
- $[\mathbf{x}]_{\mathcal{C}} = P_{\mathcal{C}}^{-1}\mathbf{x} = P_{\mathcal{C}}^{-1}P_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$

Eigenvectors and Eigenvalues

An **eigenvector** is defined as a vector $\mathbf{x}$ for which some matrix A, the multiplication $A\mathbf{x}$ is equal to $\lambda\mathbf{x}$.

Essentially, it is a vector which is stretched by A.

If $A\mathbf{x}$ does not result in a scalar value of $\mathbf{x}$, $\mathbf{x}$ is not an eigenvector for the given matrix A.

An **eigenvalue** of A is a value λ that exists for some given matrix A. In the case that $A\mathbf{x} = 2\mathbf{x}$, 2 is an eigenvalue.

To test if an eigenvalue exists, then the following equations are used:

$$\begin{aligned}&A\mathbf{x} = \lambda\mathbf{x}\\&A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}\\&(A - \lambda I)\mathbf{x} = \mathbf{0}\end{aligned}$$

$$A - \lambda I = \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda \end{bmatrix} = \begin{bmatrix} (a_{1,1} - \lambda) & \dots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & (a_{n,n} - \lambda) \end{bmatrix}$$

The resulting matrix should be linearly dependent if λ is a valid eigenvalue.

Note that while row-reducing the above matrix and finding its null space will provide a base for any associated eigenvectors to the given λ, it will not provide eigenvalues.

The eigenvalues of a triangular matrix are trivially easy to find as they are the entries along its principal diagonal.

If A has a eigenvalue of 0, then $A\mathbf{x} = 0\mathbf{x} = \mathbf{0}$ has a nontrivial solution (and therefore the null space is greater than $\{\mathbf{0}\}$), this indicates that the matrix A is not invertible.

If $\mathbf{v_1} \dots \mathbf{v_n}$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1 \dots \lambda_n$, then the set $\{\mathbf{v_1} \dots \mathbf{v_n}\}$ is linearly independent.

Finding all the possible eigenvalues of a matrix A is equivalent to finding all λ in the equation $(A - \lambda I)\mathbf{x}$ for which the matrix is linearly dependent, which is to say for which the determinant is equal to 0.

This can be done by simply taking the determinant and factoring the resultant equation:

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

Remember that the determinant of A is equivalent to the product of its pivots in upper-staircase form U multiplied by $(-1)^r$ where r is the number of row interchanges, but only if $\exists A^{-1}$.

$$\det A = \det U \text{ where } \exists A^{-1} = [(u_{1,1})(u_{2,2}) \ldots (u_{n,n})] \cdot (-1)^r$$

Additional rules are provided for the invertible matrix where A is $n \times n$ and invertible:

1. $0 \notin \{\lambda_1 \ldots \lambda_n\}$ for A.
2. $\det A \neq 0$.

The determinant of a 3×3 matrix in $\mathbb{R}^3$ is the volume of the paralellepiped formed by its three column vectors. If any of the vectors is linearly dependent, the volume is 0. Thus, geometrically, a plane of any size has no volume, and a line of any size has no area.

Two matrices A, B are **similar** if the following equations are possible to be satisfied:

$$PAP^{-1} = B$$
$$P^{-1}BP = A$$

Similar matrices have the same *characteristic polynomial* (the λ equation found after simplifying the determinant of the *characteristic equation* $(A - \lambda I)$) and thus same eigenvalues. See the below proof:

1. $B \sim A \to B = P^{-1}AP$
2. To find the characteristic equation for B, we find $\det(B - \lambda I)$.
3. From the above we can rewrite it: $(\underbrace{P^{-1}AP}_{B}) - \lambda(\underbrace{P^{-1}P}_{I})$.
4. Because λ is a constant, we can rearrange it and then factor:

$$P^{-1}AP - P^{-1}\lambda P =$$

$$P^{-1}(AP - \lambda P)$$
$$P^{-1}(A - \lambda)P$$

5. From this, we may take the determinant to see:

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

$$\det P^{-1} \cdot \det(A - \lambda) \cdot \det P =$$
$$\det P \cdot \det P^{-1} \cdot \det(A - \lambda) \cdot =$$
$$\det(A - \lambda) \to$$
$$\det(A - \lambda) = \det(B - \lambda)$$

Note that matrices have have the same eigenvalues and yet not be similar.

Also note that similarity is *not* the same as row-equivalence.

For calculating eigenvalues, you will often need to use the quadratic formula:

$$\lambda = \frac{1}{2a}[-b \pm \sqrt{b^2 - 4ac}]$$

Diagonalization

A matrix is **diagonalizable** if it can be factored to take the form $A = PDP^{-1}$ where D is a diagonal matrix.

For $k \geq 1\ldots$

$$D^k = \begin{bmatrix} (d_{1,1})^k & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & (d_{n,n})^k \end{bmatrix}$$

$$A^k = PD^kP^{-1}$$

A matrix $A_{n\times n}$ is diagonalizable if and only if A has n linearly independent eigenvectors which can be used to compose P.

In the case which there are sufficient eigenvectors, eigenvalues are used to form the diagonal in D. If δ represents an eigenvector, then we have:

$$PD = [\lambda_1\delta_{\mathbf{1}}\ \lambda_2\delta_{\mathbf{2}}\ldots\lambda_n\delta_{\mathbf{n}}]$$

Remember that $AB = [A\mathbf{b_1}\ldots A\mathbf{b_n}]$. Because the columns of D apart from each unique λ_c are comprised of 0's, PD's $P\mathbf{d_c}$'s ultimately resolve to $\mathbf{p_c}\lambda_c = \lambda_c\delta_{\mathbf{c}}$.

The steps to diagonalize a matrix are as follows:

1. Determine the eigenvalues of $A_{n\times n}$ via $\det(A - \lambda I)$.
2. Find n linearly independent eigenvectors by finding the null spaces of A (with the discovered eigenvalues) via $A - \lambda I = \mathbf{0}$.
3. Construct $P = [\delta_1\ldots\delta_n]$ using the eigenvectors found above.
4. Construct D using the λ's in the respective order of P's associated eigenvectors. (*This is to say that if λ_1 was used to find eigenvector $\delta_{\mathbf{1}}$ and $\delta_{\mathbf{1}}$ was placed in column* 10 *of P, then λ_1 must be placed in column* 10 *of D.*)

5. Check that $AP = PD$ and that P is invertible by calculating P^{-1}.

A $n \times n$ matrix of n **distinct** eigenvalues is *always* diagonalizable.

Note the following:

- The dimension of the eigenspace (also known as the **geometric mulitiplicity**) for every λ of a matrix is less than or equal to the algebraic multiplicity of λ.
- A $n \times n$ matrix is diagonalizable if and only if the sum of the dimensions of the eigenspaces (the sum of its geometric multiplicities) is equal to n.
- The above only occurs when the characteristic polynomial completely factorizes to *linear factors* (factors of the form $ax+b$) and the dimension of the eigenspace for λ is equal to the algebraic multiplicity of λ.
- If the above is true and the given matrix is diagonalizable, then every base of each λ-associated eigenspace together form a base of eigenvectors in $\mathbb{R}^n$ for the given matrix.

Eigenvectors and Linear Transformations

If $V \in \mathbb{R}^n$ and $W \in \mathbb{R}^m$, and we have the transformation $T : V \to W$, then to associate a matrix with the transformation T, bases $\mathcal{B}, \mathcal{C}$ are selected for V, W (respectively) such that for any vector $\mathbf{x} \in V$, $[\mathbf{x}]_{\mathcal{B}} \in \mathbb{R}^n$ y $[T(\mathbf{x})]_{\mathcal{C}} \in \mathbb{R}^m$.

This is to say that for a vector in V, there is that same vector in $\mathcal{B}$-coordinates in $\mathbb{R}^n$ and there is the same *transformed* vector in $\mathcal{C}$-coordinates in $\mathbb{R}^m$.

Note that $\{\mathbf{b_1} \dots \mathbf{b_n}\} = \mathcal{B}$. This implies that $\mathbf{x}$ is defined as some linear combination of the base $\mathcal{B}$ such that $\mathbf{x} = r_1\mathbf{b_1} + \cdots + r_n\mathbf{b_n}$. See that...

$$\mathbf{x} = r_1\mathbf{b_1} + \cdots + r_n\mathbf{b_n} = [\mathbf{b_1} \ldots \mathbf{b_n}]\, \langle r_1 \ldots r_n \rangle^T = \underbrace{P_{\mathcal{E}\leftarrow\mathcal{B}}}_{P_\mathcal{B}} \underbrace{\begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}}_{[\mathbf{x}]_\mathcal{B}} = \mathbf{x}$$

And so from the above equation, we see that...

$$[\mathbf{x}]_\mathcal{B} = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

(*The reason the scalars are represented with* r *instead of* c *is arbitrary but meant to avoid confusion with base* $\mathcal{C}$.)

Therefore, then...

$$\begin{gathered} T(\mathbf{x}) = T(r_1\mathbf{b_1} + \cdots + r_n\mathbf{b_n}) = \\ T(r_1\mathbf{b_1}) + \cdots + T(r_n\mathbf{b_n}) = \\ r_1 T(\mathbf{b_1}) + \cdots + r_n T(\mathbf{b_n}) \end{gathered}$$

This implies that...

$$\underbrace{[T(\mathbf{x})]_{\mathcal{E}\leftarrow\mathcal{C}}}_{\text{column vector}} = [T(\mathbf{x})]_\mathcal{C} = r_1[T(\mathbf{b_1})]_\mathcal{C} + \cdots + r_n[T(\mathbf{b_n})]_\mathcal{C}$$

Because r_c are scalars and $[T(\mathbf{b_c})]_\mathcal{C}$ are column vectors, we can create matrix M and a valid multiplication between the matrix M and our vector $[\mathbf{x}]$ of scalars r_c such that:

$$[T(\mathbf{x})]_\mathcal{C} = M[\mathbf{x}]_\mathcal{B} \text{ where } M = [[T(\mathbf{b_1})]_\mathcal{C} \ldots [T(\mathbf{b_n})]_\mathcal{C}]$$

The above matrix M is called **the matrix for T with respect to bases $\mathcal{B}, \mathcal{C}$**. It is composed of the base vectors of $\mathcal{B}$ transformed and written in $\mathcal{C}$-coordinates.

In the case where $\mathcal{B}, \mathcal{C}$ share the same space V and T is an identity transformation $T(\mathbf{x}) = \mathbf{x}$ (this is to say that the vector remains itself; it identifies the same point in space), then M is actually just a matrix of change of coordinates.

When W is equal to V and $\mathcal{C}$ coincides with $\mathcal{B}$, M is the **matrix for** T **with respect to** $\mathcal{B}$, or simply the $\mathcal{B}$**-matrix for** T, which is denoted $[T]_{\mathcal{B}}$.

$$\forall \mathbf{x} \in V \text{ where } T : V \to V : [T(\mathbf{x})]_{\mathcal{B}} = [T]_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$

So in the above case, what differs is that M is replaced by $[T]_{\mathcal{B}}$. Essentially, the equation defines the transformation of $\mathbf{x}$ through T in $\mathcal{B}$-coordinates as the transformation matrix in $\mathcal{B}$-coordinates multiplying the $\mathbf{x}$ vector also in $\mathcal{B}$-coordinates.

If $A = PDP^{-1}$ (where D is a diagonal $n \times n$ matrix) and if $\mathcal{B}$ is a base for $\mathbb{R}^n$ formed by the columns of P, then D is the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \rightarrowtail A\mathbf{x}$.

To see this, let $\mathcal{B} = \{\mathbf{b_1} \dots \mathbf{b_n}\}$ and $P = [\mathbf{b_1} \dots \mathbf{b_n}]$. Then P is the matrix of change of coordinates $P_{\mathcal{B}}$, and thus we see that $P[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$ and $[\mathbf{x}]_{\mathcal{B}} = P^{-1}\mathbf{x}$.

If $T(\mathbf{x}) = A\mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$, then...

$$\begin{aligned}[T]_{\mathcal{B}} =& [[T(\mathbf{b_1})]_{\mathcal{B}} \ \dots \ [T(\mathbf{b_n})]_{\mathcal{B}}] = \\ & \underbrace{[[A\mathbf{b_1}]_{\mathcal{B}} \ \dots \ [A\mathbf{b_n}]_{\mathcal{B}}]}_{\text{definition of } T(\mathbf{x})} = \\ & [P^{-1}A\mathbf{b_1} \ \dots \ P^{-1}A\mathbf{b_n}] = \\ & P^{-1}A[\mathbf{b_1} \ \dots \ \mathbf{b_n}] = \\ & P^{-1}AP\end{aligned}$$

Given that $A = PDP^{-1}$, we see:

$$[T]_{\mathcal{B}} = P^{-1}AP = P^{-1}PDP^{-1}P = IDI = D$$

Because the above does not rely on D being diagonal, it means that for any similar matrix C such that $A = PCP^{-1}$, C is the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \rightarrowtail A\mathbf{x}$ where the $\mathcal{B}$-matrix is formed by the columns of P ($P_{\mathcal{E}\leftarrow\mathcal{B}}$).

Complex Eigenvalues

A $n \times n$ matrix always indicates a characteristic polynomial of grade n with n roots when counting multiplicities *and* when allowing *complex* roots.

A complex root is signified when, through manual factoring or typically the quadratic formula, the square root of some negative integer is involved.
E.g., $\lambda = \frac{1}{c_3}[c_1 \pm \sqrt{-c_2}]$.

- The imaginary number $\mathbf{i}$ represents $\sqrt{-1}$. Note that $i^2 = -1 \neq 1$. This is because the rule $\sqrt{a}\sqrt{b} = \sqrt{ab}$ does not hold when both $a, b < 0$.
- Also note that any number n where $0 \neq n \neq 1$ can be separated from the invisible -1 in $\sqrt{-n} = \sqrt{n(-1)}$.
- If $n = 64$ in the case of $\sqrt{-64}$, then we have $\sqrt{64(-1)} = \sqrt{64}\sqrt{-1} = 8\sqrt{-1} = 8i$.

The same process of finding the eigenvalues and then subtracting them from A to find the eigenvectors is used but with a slight modification. Instead of the use of row-reduction, a choice of how to assign one of the varibles x_1 determines how to assign the other (in the case of 2×2 matrices).

A system of equations such as the following may commonly occur:

$$(-0.3 + 0.6i)x_1 - (0.6)x_2 = 0$$
$$(0.75)x_1 + (0.3 + 0.6i)x_2 = 0$$

In such a case, it is correct to select an x_1 or x_2 that makes the system easier, such as $x_2 = 5$. Once one value is selected, the other inherently follows. Note that the solutions to x_1 and x_2 become the values in the complex vector $\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

A **complex conjugate** $\overline{z}$ of a complex number z is defined as that same number z with its imaginary component sign-flipped.

Should $z = a+bi$, $\overline{z} = a-bi$. The product of a complex number $a \pm bi$ and its conjugate is the real number $(a-bi)(a+bi) = a^2 - b^2(i^2) = a^2 - b^2(-1) = a^2 + b^2$.

Note the following properties where the overline (as seen in $\overline{\mathbf{z}}$) indicates an element that has had its complex entries changed to their conjugates:

$$\begin{aligned} \overline{r\mathbf{x}} &= \overline{r}\,\overline{\mathbf{x}} \\ \overline{B\mathbf{x}} &= \overline{B}\,\overline{\mathbf{x}} \\ \overline{BC} &= \overline{B}\,\overline{C} \\ \overline{rB} &= \overline{r}\,\overline{B} \end{aligned}$$

And if $A_{n\times n}$ is a matrix with real values, then $\overline{A\mathbf{x}} = \overline{A}\overline{\mathbf{x}} = A\overline{\mathbf{x}}$. If λ is an eigenvalue of A and $\overline{\mathbf{x}}$ is its corresponding eigenvector in $\mathbb{C}^n$, then:

$$A\overline{\mathbf{x}} = \overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$$

This is to say that whenever $A \in \mathbb{R}$, any of its complex eigenvectors and eigenvalues present themselves in pairs.

If $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with $a, b \in \mathbb{R}$, then:

1. The eigenvalues of C are $\lambda = a \pm bi$.

2. $r = |\lambda| = \sqrt{a^2 + b^2}$

3. $C = r\begin{bmatrix} a/r & -b/r \\ b/r & a/r \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} r\cos\phi & -r\sin\phi \\ r\sin\phi & r\cos\phi \end{bmatrix}$

A factorization for $A = PCP^{-1}$, where $\lambda = a - bi$, $b \neq 0$, and $\mathbf{v}$ is an eigenvector in $\mathbb{C}^2$, has the folllowing definitions:

$$P = [\mathrm{Re}(\mathbf{v})\ \mathrm{Im}(\mathbf{v})],\ C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

They can be used in combination with the above properties to discover the angle ϕ that describes the rotation of a transformation.

Dot Product, Length, and Orthogonality

The **dot product**, also known as the **scalar product** and **inner product**, of two matrices **u, v** is the sum of the products of their correspondingly indexed components.

It can be defined so:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^{\mathbf{T}}\mathbf{v} = \begin{bmatrix} u_1 & \dots & u_n \end{bmatrix}\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + ... + u_nv_n$$

Note that $\mathbf{v}^{\mathbf{T}}\mathbf{u}$ would produce the same result and that the dot product is commutative.

The following properties hold true for any two vectors and scalar in $\mathbb{R}^n$:

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
3. Likewise, $(c\mathbf{u})\mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$. *(Which is to say that scalars' placement do not affect the dot product.)*

4. $\mathbf{u} \cdot \mathbf{u} \geq 0 \land (\mathbf{u} \cdot \mathbf{u} = 0 \iff \mathbf{u} = \mathbf{0})$

The **length**, also known as the **norm** or **magnitude**, of a vector is given by the following equation:

$$\underbrace{||\mathbf{v}||}_{\text{norm of } \mathbf{v}} = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + \cdots + v_n^2}$$

Thus we can also see that $||\mathbf{v}||^2 = \mathbf{v} \cdot \mathbf{v}$.

Any constant factor c of a vector $\mathbf{t}$ can be factored out such that if $\mathbf{t} = \langle 10, 20, 30 \rangle$, we could find $c = 10, \mathbf{v} = \langle 1, 2, 3 \rangle$. Thus when calculating $||\mathbf{t}||$, we see that...

$$\begin{aligned} ||\mathbf{t}|| =& ||c\mathbf{v}|| = \sqrt{c^2 v_1^2 + \cdots + c^2 v_n^2} = \sqrt{c^2 (v_1^2 + \cdots + v_n^2)} \\ =& \sqrt{c^2} \sqrt{(v_1^2 + \cdots + v_n^2)} = |c| \sqrt{(v_1^2 + \cdots + v_n^2)} \\ \rightarrow& ||c\mathbf{v}|| = |c| ||\mathbf{v}|| \end{aligned}$$

A **unit vector** is a vector of length 1 and can be obtained by taking any vector and dividing it by its own length such that $\frac{\mathbf{u}}{||\mathbf{u}||}$ represents a unit vector for any $\mathbf{u}$.

To check that a vector is *unitary*, simply calculate if $||\mathbf{u}||^2 = 1$. That is to say, simply square and sum the components.

For any pair of vectores $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$, their distance is defined as...

$$\text{dist}(\mathbf{u}, \mathbf{v}) = ||\mathbf{u} - \mathbf{v}||$$

If two vectors $\mathbf{u}, \mathbf{v}$ are orthogonal, then the distances between a fixed point of $\mathbf{u}$ paired with alternatively with the head or tail of $\mathbf{v}$, should be identical. This only occurs in $\mathbb{R}^2$ when one vector is exactly $90°$ away in its orientation from the other and thus is perpendicular.

Otherwise, if one of the vectors were turned $90° \pm x$ where $x \neq 0$, the once-perpendicular vector would skew to one side, creating less distance between one pair of points and more distance between the converse pair of points.

Formulaically, this means the vectors are orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$.
Similarly, two vectores $\mathbf{u}, \mathbf{v}$ are orthogonal if and only if...

$$||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$$

Both the above are because $||\mathbf{u} - (-\mathbf{v})||^2$ and $||\mathbf{u} - \mathbf{v}||^2$ resolve to $||\mathbf{u}||^2 + ||\mathbf{v}||^2 - 2\mathbf{u} \cdot \mathbf{v}$ and $||\mathbf{u}||^2 + ||\mathbf{v}||^2 + 2\mathbf{u} \cdot \mathbf{v}$, respectively.

Thus, $2\mathbf{u} \cdot \mathbf{v} = -2\mathbf{u} \cdot \mathbf{v} \to \mathbf{u} \cdot \mathbf{v} = 0 \to \mathbf{u} \perp \mathbf{v}$. Likewise then for the original equations, we see $||\mathbf{u} - (-\mathbf{v})||^2 = ||\mathbf{u} - \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$.

If a vector $\mathbf{z}$ is orthogonal to all vectors in the subspace W, then $\mathbf{z}$ is said to be **completely orthogonal** to W.

Thus in $\mathbb{R}^3$ where W is a plane, $\mathbf{z}$ can be thought of as a line that pierces W and is orthogonal to all $\mathbf{w} \in W$.

The subspace *completely orthogonal* to W can be defined as L where $L = W^\perp$ and passes through the origin. Likewise, the inverse holds as $W = L^\perp$ and is completely orthogonal to L. Every vector that spreads from $\mathbf{0}$ outward in W is orthogonal to every vector that spreads outward from the origin in L.

A vector $\mathbf{x} \in W^\perp \iff \forall \mathbf{w} \in W : \mathbf{x} \perp \mathbf{w}$.

Note that where W can be represented by an $m \times n$ matrix, $W^\perp$ is a subspace of $\mathbb{R}^n$.

For a matrix $A_{m\times n}$, the following holds:

- $\text{Fila}(A)^\perp = \text{Nul}(A)$
- Thus, by the rule that $(W^\perp)^\perp = W$, $\text{Fila}(A) = \text{Nul}A^\perp$.
- $\text{Col}(A)^\perp = \text{Nul}(A^T)$
- $\text{Col}(A) = \text{Nul}(A^T)^\perp$

This is to say that the row space is completely orthogonal to the null space of A, and the column space of A is completely orthogonal to the nullspace of its transpose. Thus, $\text{Fila}(A)^\perp = \text{Nul}(A) \rightarrow \text{Fila}(A) \perp \text{Nul}(A)$.

Remember that $\text{Fila}(A) \equiv \text{Col}(A^T)$ and so...

$$\text{Col}(A^T)^\perp = \text{Nul}A$$

Note then that transposing one argument in the above equation transposes the other. Likewise, Remember that the rows are a subspace of $\mathbb{R}^n$ and columns are a subspace of $\mathbb{R}^m$.

The reason why this is true is that every vector in the null space is a vector of n entries, and each null vector when multiplied through the matrix is essentially dotted with each row vector (as every entry in each column is affected by one scalar from the null space vector).

See the following illustration where $\mathbf{n}$ is a vector from the null space:

$$a_{1,1}n_1 + a_{1,2}n_2 + \cdots + a_{1_n}n_n = 0 \rightarrow \mathbf{a_1^T n} = 0 \rightarrow \forall c : \mathbf{a_c^T n} = 0$$

Orthogonal Sets

A set of vectors $\{\mathbf{u_1} \dots \mathbf{u_p}\} \in \mathbb{R}^n$ is a **orthogonal set** if every vector in the set is orthogonal to every other vector in the set.

Thus is it is impossible to select any $\mathbf{u_{c_1}}, \mathbf{u_{c_2}}$ such that $\mathbf{u_{c_1}} \cdot \mathbf{u_{c_2}} \neq 0$.

If $S = \{\mathbf{u_1}, \dots, \mathbf{u_p}\}$ is an orthogonal set of vectors distinct from $\mathbf{0}$, then S forms a linearly independent base for the subspace generated by S.

Note that an **orthogonal base** for a subspace $W \in \mathbb{R}^n$ is both a base for W and an orthogonal set.

If $\{\mathbf{u_1}, \dots, \mathbf{u_p}\}$ is an orthogonal base for W in $\mathbb{R}^n$, then every $\mathbf{y}$ in W (whose definition is $\mathbf{y} = c_1\mathbf{u_1} + \dots + c_p\mathbf{u_p}$), has its weights c_j given by...

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u_j}}{\mathbf{u_j} \cdot \mathbf{u_j}} \text{ where } j \in \{1, \dots, p\}$$

And thus...

$$\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u_1}}{\mathbf{u_1} \cdot \mathbf{u_1}}\mathbf{u_1} + \dots + \frac{\mathbf{y} \cdot \mathbf{u_p}}{\mathbf{u_p} \cdot \mathbf{u_p}}\mathbf{u_p}$$

For a given distinct vector $\mathbf{u}$, any given vector $\mathbf{y}$ can be written as a composition of $\mathbf{u}$ and a vector orthogonal to $\mathbf{u}$.

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

In the above, $\hat{\mathbf{y}} = \alpha\mathbf{u}$ for any α and $\mathbf{z}$ is a vector orthogonal to $\mathbf{u}$. Thus, $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$ and $(\mathbf{y} - \hat{\mathbf{y}}) \perp \mathbf{u}$. See that...

$$\mathbf{z} \perp \mathbf{u} \iff (\alpha = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}) \wedge (\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u})$$

The vector $\hat{\mathbf{y}}$ is said to be the **orthogonal projection** of $\mathbf{y}$ over $\mathbf{u}$, and $\mathbf{z}$ is the component of $\mathbf{y}$ that is orthogonal to $\mathbf{u}$.

Note that $\mathbf{u}$ could be multiplied by any scalar c and the equations would still hold.

$$\hat{\mathbf{y}} = \text{proj}_L \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$$

Because the projection is determined by the subspace (line) L generated between $\mathbf{0}$ and $\mathbf{u}$, $\text{proj}_L \mathbf{y}$ is called the **orthogonal projection of $\mathbf{y}$ over L**. However, the meaning does not differ whether it is "over L" or "over $\mathbf{u}$".

In the case where $W = \mathbb{R}^2 = \text{Gen}\{\mathbf{u_1}, \mathbf{u_2}\}$ and $\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u_1}}{u_1 \cdot u_1} \mathbf{u_1} + \frac{\mathbf{y} \cdot \mathbf{u_2}}{u_2 \cdot u_2} \mathbf{u_2}$, the first term of $\mathbf{y}$ is the projection of $\mathbf{y}$ over $\mathbf{u_1}$ and the second term is the projection of $\mathbf{y}$ over $\mathbf{u_2}$. So $\mathbf{y}$ is the sum of its projections over the orthogonal axes determined by the base of $\mathbb{R}^2$.

A **orthonormal set** is a set of completely mutually orthogonal vectors that are also all unitary. If W is a subspace generated by such a set, then the set is an **orthonormal base** for W.

The standard base is an example of an orthonormal set.

A matrix $U_{m \times n}$ has orthonormal columns if and only if $U^T U = I$. Thus that equality can be used as a test for orthonormality.

Note too that a matrix $A^T A$ is invertible if and only if A has linearly independent columns.

If $U_{m \times n}$ indeed has orthnormal columns, then the following properties are true:

1. $||U\mathbf{x}|| = ||\mathbf{x}||$
2. $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$

3. $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0 \iff \mathbf{x} \cdot \mathbf{y} = 0$

An **orthogonal matrix** is a square, invertible matrix U such that $U^{-1} = U^T$, and it has orthonormal columns.

Note that *any* square matrix with orthonormal columns is an orthogonal matrix.

Likewise, any orthonormal matrix additionally, as a "byproduct", has orthonormal rows as well.

Orthogonal Projections

Given a vector $\mathbf{y}$ and a subspace $W \in \mathbb{R}^n$, there exists some vector $\hat{\mathbf{y}} \in W$ such that $\hat{\mathbf{y}}$ is the unique vector in W for which $\mathbf{y} - \hat{\mathbf{y}} \perp W$ and such that $\hat{\mathbf{y}}$ is the vector in W most close to $\mathbf{y}$.

A vector $\mathbf{y}$ in $\mathbb{R}^n$ can always be represented as a linear combination of the vectors $\mathbf{u_1}, \ldots, \mathbf{u_n} \in \mathbb{R}^n$. This linear combination can be grouped such that $\mathbf{y} = \mathbf{z_1} + \mathbf{z_2}$ where $\mathbf{z_1}$ is some group of some $\mathbf{u_c}$'s and $\mathbf{z_2}$ is some group of the remaining $\mathbf{u_c}$'s.

Thus if there exists a vector $\mathbf{y}$ in the orthogonal base $\{\mathbf{u_1}, \ldots, \mathbf{u_5}\} \in \mathbb{R}^5$, and there exists a subspace of that space, say $W = \text{Gen}\{\mathbf{u_1}, \mathbf{u_2}\}$, then $\mathbf{y}$ can be written as a combination of the vectors in W and $\mathbb{R}^5 \setminus W$.

$$\mathbf{y} = \underbrace{c_1\mathbf{u_1} + c_2\mathbf{u_2}}_{\mathbf{z_1}} + \underbrace{c_3\mathbf{u_3} + c_4\mathbf{u_4} + c_5\mathbf{u_5}}_{\mathbf{z_2}}$$

In the above example, $\mathbf{z_2} \in \mathbb{R}^5 \setminus W$ (or $\text{Gen}\{\mathbf{u_3}, \mathbf{u_4}, \mathbf{u_5}\}$) and $\mathbf{z_1} \in W$.

In this case, any vector in $\mathbf{z_1}$ dotted with $\mathbf{z_2}$ is always 0, as is $\mathbf{z_1} \cdot \mathbf{z_2} = 0$. For that reason, $\mathbf{z_2} \in W^{\perp}$.

If W is a subspace of $\mathbb{R}^n$, then for all $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ where $\hat{\mathbf{y}} \in W$ y $\mathbf{z} \in W^{\perp}$.

As with the case in the section before, $\hat{\mathbf{y}}$ and $\mathbf{z}$ are defined the same way but extrapolated over dimensions:

$$\hat{\mathbf{y}} = \text{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u_1}}{\mathbf{u_1} \cdot \mathbf{u_1}} \mathbf{u_1} + \cdots + \frac{\mathbf{y} \cdot \mathbf{u_p}}{\mathbf{u_p} \cdot \mathbf{u_p}} \mathbf{u_p}$$

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$$

The orthogonal projection of the vector $\mathbf{y}$ is the sum of its projections over the unidimensional subspaces which are mutually orthogonal.

If $\mathbf{y} \in W$ and $W = \text{Gen}\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$, and also $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is an orthogonal base, then $\text{proj}_W \mathbf{y} = \mathbf{y}$

If W is a subspace for $\mathbb{R}^n$ and $\mathbf{y}$ is any vector in $\mathbb{R}^n$ and also $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ over W, then $\hat{\mathbf{y}} \in W$ is the point most close to $\mathbf{y}$:

$$\forall \mathbf{v} \neq \hat{\mathbf{y}} : ||\mathbf{y} - \hat{\mathbf{y}}|| < ||\mathbf{y} - \mathbf{v}||$$

If $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is an orthonormal base for $W \in \mathbb{R}^n$, then...

$$\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u_1})\mathbf{u_1} + \cdots + (\mathbf{y} \cdot \mathbf{u_p})\mathbf{u_p}$$

Note the difference is that we do not need to divide each summand by its length, as the bases are already normalized.

And so if there exists U such that $U = [\mathbf{u_1} \ldots \mathbf{u_p}]$, then the projection of $\mathbf{y}$, $\hat{\mathbf{y}}$ becomes ...

$$\forall \mathbf{y} \in \mathbb{R}^n : \text{proj}_W \mathbf{y} = UU^T \mathbf{y}$$

Note $UU^T\mathbf{y}$ is equivalent to the elaborated form immediately before.

The Gram-Schmidt Process

The Gram-Schmidt Process is a algorithm for obtaining an orthogonal or orthonormal base for any subspace different from $\{\mathbf{0}\}$ in $\mathbb{R}^n$.

Given a base $\{\mathbf{x_1}, \dots, \mathbf{x_p}\}$ for W in $\mathbb{R}^n$, the algorithm works as follows:

$$\begin{aligned}
&\mathbf{v_1} = \mathbf{x_1} \\
&\mathbf{v_2} = \mathbf{x_2} - \frac{\mathbf{x_2} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} \\
&\mathbf{v_3} = \mathbf{x_3} - \frac{\mathbf{x_3} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} - \frac{\mathbf{x_3} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_2}}\mathbf{v_2} \\
&\vdots \\
&\mathbf{v_p} = \mathbf{x_p} - \frac{\mathbf{x_p} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} - \frac{\mathbf{x_p} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_1}}\mathbf{v_2} - \dots - \frac{\mathbf{x_p} \cdot \mathbf{v_{(p-1)}}}{\mathbf{v_{(p-1)}} \cdot \mathbf{v_{(p-1)}}}\mathbf{v_{(p-1)}}
\end{aligned}$$

So the set $\{\mathbf{v_1}, \dots, \mathbf{v_p}\}$ becomes an equal set to $\{\mathbf{x_1}, \dots, \mathbf{x_p}\}$ but is orthogonal. Further, any equally indexed vectors can be removed from the two sets and the two sets will remain equal.

The base could then be made orthonormal by manually normalizing each vector using division over their lengths.

If $A_{m\times n}$ is a matrix with linearly independent columns, then A can be factorized as $A = QR$ where $Q_{m\times n}$ has the columns which form a orthonormal base for $\operatorname{Col} A$ and where R is an invertible superior triangle matrix that contains the positive integers that inform how to convert Q's columns into A's original non-orthonormal columns.

This is possible because each vector $\mathbf{x_k}$ is composed:

$$\mathbf{x_k} = r_{1k}\mathbf{u_1} + \dots + r_{kk}\mathbf{u_k} + 0 \cdot \mathbf{u_{k+1}} + \dots + 0 \cdot \mathbf{u_n}$$

Note that $\mathbf{r_k}$ is a given column vector.

$$\mathbf{r_k} = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Finally, then, $\mathbf{x_k} = Q\mathbf{r_k}$ for $k \in \{1, \dots, n\}$, and $R = [\mathbf{r_1} \dots \mathbf{r_n}]$.

$$A = [\mathbf{x_1} \dots \mathbf{x_n}] = [Q\mathbf{r_1} \dots Q\mathbf{r_n}] = QR$$

Note that $Q^TQ = I$ due to it being an orthonormal matrix. Because $A = QR$, we have it that $Q^TA = R$. Thus we have an easy way of calculating R after having calculated Q through the Process of Gram-Schmidt.

Problems of Least Squares

If $A_{m\times n}, \mathbf{b} \in \mathbb{R}^m$, the **solution of least squares** $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ where:

$$\forall \mathbf{x} \in \mathbb{R}^n : ||\mathbf{b} - A\hat{\mathbf{x}}|| \leq ||\mathbf{b} - A\mathbf{x}||$$

That is simply to say that $\hat{\mathbf{x}}$ minimizes the distance between the vector outside of the column space of A moreso than any other vector within the column space.

Note that the “least squares” portion of the solution’s name refers to the fact that distance is calculated by *squaring* and summing the entries of a vector (before, of course, rooting it).

The vector $\hat{\mathbf{b}}$, defined as $\text{proy}_{\text{Col}(A)}\mathbf{b}$, is the vector projection of $\mathbf{b}$ over the column space of A and does have a consistent solution such that $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.

Note then in that case, $\mathbf{b} - A\hat{\mathbf{x}}$ is the orthogonal component of $\mathbf{b}$ to A such that, when summed with the non-orthogonal component, a solution can be described:

$$(\mathbf{b} - A\hat{\mathbf{x}}) + \hat{\mathbf{b}} = \mathbf{b}$$

The set of all possible least squares solutions to $A\mathbf{x} = \mathbf{b}$ is equal to the non-empty set of solutions to $A^T A\mathbf{x} = A^T\mathbf{b}$.

When A is $m \times n$, we see:

$$A^T_{n\times m} A_{m\times n} \mathbf{x}_{n\times 1} = A^T_{n\times m} \mathbf{b}_{m\times 1}$$

$$(A^T A)_{n\times n} \mathbf{x}_{n\times 1} = (A^T \mathbf{b})_{n\times 1}$$

Thus the above can then be solved via two routes:

- Solving for the inverse, if one exists, and calculating $\mathbf{x} = (A^T A)^{-1} A^T\mathbf{b}$.
- Using row-reduction techniques on the matrix $[A^T A | A^T\mathbf{b}]$.

The following statements about a matrix $A_{m\times n}$ are logically equivalent and thus all true or all false:

1. The equation $A\mathbf{x} = \mathbf{b}$ has a *unique* least squares solution for every $\mathbf{b}$ in $\mathbb{R}^m$.
2. The columns of A are linearly independent.
3. The matrix $A^T A$ is invertible.

Thus when any one of the above is known to be true, we know that $\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b}$.

Note then that when determining if the above solution can be used when examining a matrix, the easist method is to check for columnal independence.

Note that the *distance* between $\mathbf{b}$ and $A\hat{\mathbf{x}}$ is known as the **error of least squares** of the approximation.

Diagonalization of Symmetrical Matrixes

A matrix A is **symmetric** if a matrix $A^T = A$; i.e., its entries not along the diagonal present in pairs.

If A is symmetric, then any two of its eigenvectors from different eigenspaces are orthogonal. This is to say that the different eigenspaces are mutually orthogonal.

$A_{n\times n}$ is orthogonally diagonalizable if there exists an orthonormal matrix P (where $P^{-1} = P^T$) and a diagonal matrix D such that $A = PDP^{-1} = PDP^T$.

The above diagonalization requires n orthonormal, linearly independent eigenvectors.

Because $A = PDP^T$, we see then that:

$$A^T = (PDP^T)^T = P^{TT}D^TP^T = PDP^T = A$$

Thus, by virtue of the fact that an orthogonally diagonalizable matrix requires a P such that $P^{-1} = P^T$, we see that this then implies that any valid matrix with a diagonalization of said P is ultimately its own transpose.

Therefore, **all** matrices that are **orthogonally diagonalizable** are *also* **symmetric** and all matrices that are symmetric are orthogonally diagonalizable.

Spectral Theorem

The set of eigenvalues of a matrix A is sometimes denoted the **spectrum** of A. The **spectral theorem** suggests the following for a symmetric matrix A:

1. A has n real eigenvalues, counting the multiplicities.

2. The dimension of the eigenspace (or *geometric multiplicity*) for every eigenvalue is equal to the multiplicity of the eigenvalue as the root of the characteristic equation.

3. The eigenspaces are mutually orthogonal in the sense that all eigenvectores that correspond to different eigenvalues are orthogonal.

4. A is orthogonally diagonalizable.

If $A = PDP^{-1}$ and the columns of P are orthonormal eigenvectors $\mathbf{u_1}, \ldots, \mathbf{u_n}$ of A and the corresponding eigenvalues are in the diagonal matrix D, then $P^{-1} = P^T$.

This is because P is orthonormal with linearly independent columns.

$$A = PDP^T = [\mathbf{u_1} \ldots \mathbf{u_n}] \begin{bmatrix} \lambda_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u_1^T} \\ \vdots \\ \mathbf{u_n^T} \end{bmatrix}$$

$$A = [\lambda_1 \mathbf{u_1} \ldots \lambda_n \mathbf{u_n}] \begin{bmatrix} \mathbf{u_1^T} \\ \vdots \\ \mathbf{u_n^T} \end{bmatrix}$$

$$A = \lambda_1 \mathbf{u_1} \mathbf{u_1^T} + \cdots + \lambda_n \mathbf{u_n} \mathbf{u_n^T}$$

The representation of A as $\lambda_1 \mathbf{u_1} \mathbf{u_1^T} + \cdots + \lambda_n \mathbf{u_n} \mathbf{u_n^T}$ is called the **spectral decomposition** of A because A is divided into parts determined by its spectrum (i.e., its eigenvalues). Each term in the decomposition is a $n \times n$ matrix of range 1.

Every matrix $\mathbf{u_c} \mathbf{u_c^T}$ is a matrix of projection in that for every $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u_c} \mathbf{u_c^T} \mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ over the subspace generated by $\mathbf{u_c}$.

Matrix-Matrix Multiplication

The reason that $A = \lambda_1 \mathbf{u_1} \mathbf{u_1^T} + \cdots + \lambda_n \mathbf{u_n} \mathbf{u_n^T}$ holds is because the column-row expansion of two matrices $A_{m \times n}, B_{n \times p}$ functions as so:

$$[\mathbf{a_1} \dots \mathbf{a_n}] \begin{bmatrix} \mathbf{b_1^T} \\ \vdots \\ \mathbf{b_n^T} \end{bmatrix} = \mathbf{a_1} \mathbf{b_1^T} + \cdots + \mathbf{a_n} \mathbf{b_n^T}$$

Each entry $\mathbf{a_c} \mathbf{b_c^T}$ represents its own matrix. When summed, they create a sum of matrices.

To see why this is the case, see that a given entry (i, j) in $\mathbf{a_c} \mathbf{b_c^T}$, is the product of $A_{i,c}$ and $B_{c,j}$ and obtained from $\mathbf{a_c}$ and $\mathbf{b_c^T}$, respectively.

See the below example where $\mathbf{a_2} = \begin{bmatrix} b \\ d \end{bmatrix}, \mathbf{b_2^T} = \begin{bmatrix} y & z \end{bmatrix}$:

$$\mathbf{a_2} \mathbf{b_2^T} = [\begin{bmatrix} b \\ d \end{bmatrix} y \ \begin{bmatrix} b \\ d \end{bmatrix} z] = \begin{bmatrix} (b)(y) & (b)(z) \\ (d)(y) & (d)(z) \end{bmatrix} = \begin{bmatrix} A_{1,2} B_{2,1} & A_{1,2} B_{2,2} \\ A_{2,2} B_{2,1} & A_{2,2} B_{2,2} \end{bmatrix}$$

$$A \underbrace{1, 2}_{c=1,2} \cdot B \underbrace{2, 1}_{2,c=1} = (\mathbf{a_2})_1 + (\mathbf{b_2})_1 = (b)(y)$$

To get the final (i, j) entry in the product matrix, the above is calculated for c across all posiblities and all findings are summed.

Each entry of the matrix product is given as $a_{i,1} b_{1,j} + \cdots + a_{i,n} b_{n,j}$.

To elaborate on the above example, if we are given that $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $B = \begin{bmatrix} w & x \\ y & z \end{bmatrix}$, then the nondescribed component of their product is $\mathbf{a_1} \mathbf{b_1^T}$:

$$\mathbf{a_1} \mathbf{b_1^T} = [\begin{bmatrix} a \\ c \end{bmatrix} w \ \begin{bmatrix} a \\ c \end{bmatrix} x] = \begin{bmatrix} (a)(w) & (a)(x) \\ (c)(w) & (c)(x) \end{bmatrix}$$

The final result is this:

$$AB = \mathbf{a_1 b_1^T} + \mathbf{a_2 b_2^T} = [\begin{bmatrix} a \\ c \end{bmatrix} w \quad \begin{bmatrix} a \\ c \end{bmatrix} x] + [\begin{bmatrix} b \\ d \end{bmatrix} y \quad \begin{bmatrix} b \\ d \end{bmatrix} z]$$

$$AB = \begin{bmatrix} aw & ax \\ cw & cx \end{bmatrix} + \begin{bmatrix} by & bz \\ dy & dz \end{bmatrix} = \begin{bmatrix} aw + by & ax + bz \\ cw + dy & cx + dz \end{bmatrix}$$

Quadratic Form

A **quadratic form** in $\mathbb{R}^n$ is a function Q defined over $\mathbb{R}^n$ whose value is a vector $\mathbf{x}$ of $\mathbb{R}^n$ that can be calculated through an expression of the form $Q(\mathbf{x}) = \mathbf{x^T} A \mathbf{x}$, where A is $n \times n$ and symmetric. Note that A is known as the **matrix of the quadratic form**.

The above expression can be written as so:

$$\mathbf{x^T} A \mathbf{x} = Q(\mathbf{x}) = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} (A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix})$$

Where $Q(\mathbf{x}) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1 x_2 + 8x_2 x_3$, its A matrix can be reverse-engineered by observing the pattern of coefficients in the equation.

The coefficients of the squared values go along the diagonal in fashion corresponding to the index of x. Essentially, the c in x_c tells you to place the coefficient of x_c in $A_{c,c}$.

The coefficients which are shared by two x_c's are divided amongst them, and the indices indicate which two places the coefficient's halves go. E.g., $-x_1 x_2$ calls for -0.5 placed in $A_{1,2}$ and $A_{2,1}$.

Thus the correct $\mathbf{x^T} A \mathbf{x}$ equation for the above equation is the following:

$$Q(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 5 & -1/2 & 0 \\ -1/2 & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

If $\mathbf{x}$ represents a vector variable in $\mathbb{R}^n$, then a **change of variable** is an equation of the form $\mathbf{x} = P\mathbf{y}$, and thus $\mathbf{y} = P^{-1}\mathbf{x}$, where P is A's orthonormal eigenvector-based matrix and $\mathbf{y}$ is a new vector variable in $\mathbb{R}^n$.

Here, $\mathbf{y}$ is the vector of coordinates of $\mathbf{x}$ with respect to the base of $\mathbb{R}^n$ determined by the columns of P.

The following can be deduced by replacing $\mathbf{x}$:

$$\mathbf{x^T}A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y^T}P^T AP\mathbf{y} = \mathbf{y^T}(P^T AP)\mathbf{y}$$

And because $A = PDP^T$, $D = P^T AP$, and thus we have the simplification...

$$\mathbf{x^T}A\mathbf{x} = \mathbf{y^T}D\mathbf{y}$$

Where A is a symmetric matrix of $n \times n$, there exists a change of orthogonal variable where $\mathbf{x} = P\mathbf{y}$, that converts the quadratic form $\mathbf{x^T}A\mathbf{x}$ into $\mathbf{y^T}D\mathbf{y}$ *without cross products.*

A quadratic form Q is ...

- ... **positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- ... **negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- ... **indefinite** if $Q(\mathbf{x})$ takes positive and negative values.
- ... **positive semidefinite** if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$.
- ... **negative semidefinite** if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$.

Because in quadratic form all coefficients are defined by the eigenvalues of the matrix (as shown by a change of variable to $\mathbf{y}$ in which all variable terms are squared and there are no cross products and those variable terms have the diagonal's entries as coefficients), the eigenvalues themselves can be used to determine how the quadratic is defined.

However, eigenvalues are costly to compute and take a lot of time. Instead, a matrix can be brought to staircase form, and if any of its pivots are negative, then it the matrix is not positively defined.

For a matrix $A_{n \times n}$, a quadratic form $\mathbf{x}^{\mathbf{T}} A \mathbf{x}$ is…

1. …positive definite if and only if all eigenvalues of A are positive.
2. …negative definite if and only if all eigenvalues of A are negative.
3. …indefinite if and only if A has eigenvalues that are both positive and negative.

To find the maximum and minimum values of a quadratic form $Q(\mathbf{x})$ (where $\mathbf{x}^{\mathbf{T}} A \mathbf{x} = 1$), as is formalized below …

1. $m = \min\{\mathbf{x}^{\mathbf{T}} A \mathbf{x} : ||x|| = 1\}$
2. $M = \max\{\mathbf{x}^{\mathbf{T}} A \mathbf{x} : ||x|| = 1\}$

…we exploit the fact that m and M are defined as the smallest and largest eigenvalues of A, respectively.

Thus $\mathbf{x}^{\mathbf{T}} A \mathbf{x}$ is M when $\mathbf{x}$ is the unitary eigenvector associated with the largest eigenvalue.

Likewise, $\mathbf{x}^{\mathbf{T}} A \mathbf{x}$ is m when $\mathbf{x}$ is the unitary eigenvector associated with the smallest eigenvalue.

When we have a given quadratic form such as $9x_1^2+4x_2^2+3x_3^2$ with restiction $\mathbf{x^T}A\mathbf{x} = 1$, we can find the maximum and minimum values easily by looking at the maximum and minimum coefficients of the squares. If the quadratic form has cross product, we need to calculate the eigenvalues manually.

Note that there is always a quadratic form without cross products.

Singular Value Decomposition

A **singluar value decomposition** allows us to measure the quantities by which A stretches or compresses eigenvectors.

If A is $m \times n$, then A^TA is orthogonally diagonalizable. If $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$ is an orthonormal base for $\mathbb{R}^n$ that consists of the eigenvectors of A^TA, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues associated with A^TA, then we find that...

$$||A\mathbf{v_c}||^2 = \underbrace{(A\mathbf{v_c})^T A\mathbf{v_c}}_{\text{squared dot product}} = \mathbf{v_c^T} A^T A\mathbf{v_c} = \mathbf{v_c^T}(\lambda_c \mathbf{v_c}) = \lambda_c \underbrace{(\mathbf{v_c^T}\mathbf{v_c})}_{||\mathbf{v_c}||^2=1} = \lambda_c$$

For this reason, all the eigenvalues of A^TA are not negative.

The **singular values** of A are the square roots of the eigenvalues of A^TA and they are denoted $\sigma_1, \ldots, \sigma_n$. I.e., each eigenvalue of A^TA denoted λ_c is actually σ_c^2 and $\sigma_c = \sqrt{\lambda_c}$.

The singular values of A are actually the lengths of the vectors $A\mathbf{v_1}, \ldots, A\mathbf{v_n}$ as $\sigma_c = \sqrt{\lambda_c} = ||A\mathbf{v_c}||$ (where $\mathbf{v_c}$ is a unitary base eigenvector).

Note that the singular values and eigenvalues in any matrices in decomposition are written in *decreasing* order, left to right.

The maximum and minimum length vectors $\mathbf{x_c}$ of a matrix A under the restriction that $||\mathbf{x_c}|| = 1$ are given by the normalized eigenvectors encountered in $A^T A$, $\mathbf{v_c}$, multiplied by A. Therefore the maximum and minimum lengths are provided by those vectors norms', and those vectors norms' are *simply their associated singular values.*

The above then implies that σ_1 indicates the maximum length and σ_r indicates the minimum.

If the following hold...

- There exists $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$ as an orthonormal base for $\mathbb{R}^n$ and it consists of eigenvectors of $A^T A$.
- Those eigenvectors are arranged such that their correspondening eigenvalues, which are of $A^T A$, satisfy $\lambda_1 \geq \cdots \geq \lambda_n$.
- A has r singular values different from 0.

...then $\{A\mathbf{v_1}, \ldots, A\mathbf{v_r}\}$ is an orthogonal base for $\operatorname{Col} A$ and $\operatorname{range} A = r$. The decomposition of A implies a matrix Σ of $m \times n$.

$$\Sigma = \begin{bmatrix} D & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 0 \end{bmatrix}$$

In the above matrix Σ, there are r files and columns in D (i.e., D is $r \times r$), and that leaves $m - r$ files and $n - r$ columns comprised of only 0's in Σ. Note that r can be equal to m or n or both, but it can not exceed either dimension.

If A is $m \times n$ with range r, then there exists a matrix Σ that is $m \times n$ in which the diagonal entries in D (which is in Σ) are the singular values of A ($\sigma_1 \geq \cdots \geq \sigma_r > 0$), and there exists an orthogonal matrix U of $m \times m$ and an orthogonal matrix V of $n \times n$ such that $A = U\Sigma V^T$. Note that U and V are orthogonal to each other. Not only are they orthogonal to each other, but they are orthonormal in themselves.

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^T_{n\times n}$$

While D is unique, the forms of U and V are not necessarily.

The columns of U are called the **left singular vectors** of A and the columns of V are called the **right singular vectors** of A.

Note the following **very important properties** of a singular value decomposition:

1. The column space of A is spanned by the columns of U corresponding to non-zero singular values in Σ. These columns form a basis for the column space.
2. The row space of A is spanned by columns of V (i.e., the rows of V^T) corresponding to non-zero singular values in Σ. These rows form a basis for the row space.
3. The null space of A is spanned by the columns of V corresponding to zero singular values in Σ. These columns form a basis for the null space.
4. The null space of A^T is the same as the null space of A. I.e., it is spanned by the columns of V corresponding to zero singular values in Σ.

The final portions of the theorem of the invertible matrix are announced:

- $(\mathrm{Col}\, A)^{\perp} = \{\mathbf{0}\}$
- $(\mathrm{Nul})\, A^{\perp} = \mathbb{R}^n$
- $\mathrm{Fil}\, A = \mathbb{R}^n$
- A has n singular values different from 0.

The following are the algorithmic steps for finding a singular value decomposition for A:

1. Calculate $A^T A$ from A.
2. Calculate the eigenvalues of $A^T A$ by way of $\det(A^T A - \lambda I)$.
3. Calculate the eigenvectors of $A^T A$ by way of reducing $A - \lambda_c I$.
4. Normalize the eigenvectors so that they are unitary.
5. Create V with said eigenvectors and order them from greatest to least according to their respective eigenvalues.
6. If V is does not have n columns, expand V into an orthonormal base for $\mathbb{R}^n$. This involves finding mutually orthogonal vectors and normalizing them. I.e., find $\text{Nul}(\text{Col}(V^T))$ and use the *Gram-Schmidt Process* on the null space's base. Ensure that the new columns are sorted by eigenvalue.
7. Create Σ using the singular values of A (i.e., the square roots of the eigenvalues of $A^T A$) in descending order. Σ should have the same dimensions as A.
8. Normalize $A\mathbf{v_1}, \ldots, A\mathbf{v_r}$. (As A has range r, the first r columns of U are the normalized vectors $A\mathbf{v_1}, \ldots, A\mathbf{v_r}$ not equal to $\mathbf{0}$. Note that $||A\mathbf{v_c}|| = \sigma_c$, so $\mathbf{u_c} = \frac{1}{\sigma_c} A\mathbf{v_c}$.)
9. Create U with the normalized vectors from the previous step.
10. If U is does not have m columns, expand U into an orthonormal base for $\mathbb{R}^m$. This involves finding mutually orthogonal vectors and normalizing them. I.e., find $\text{Nul}(\text{Col}(U^T))$ and use the *Gram-Schmidt Process* on the null space's base. Ensure that the new columns are sorted by eigenvalue.
11. Assemble the decomposition according to $A = U\Sigma V^T$.

The Complete Theorem of the Invertible Matrix

- A has an inverse and A^T has an inverse: $(A^T)^{-1} = (A^{-1})^T$.
- A is row-equivalent to I_n.
- A has n pivot positions.
- The columns of A form a linearly independent set.
- $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- The linear transformation $\mathbf{x} \rightarrowtail A\mathbf{x}$ is one-to-one: A is injective.
- The linear transformation $\mathbf{x} \rightarrowtail A\mathbf{x}$ maps $\mathbb{R}^n$ over $\mathbb{R}^n$: A is surjective.
- A is both injective and surjective, and therefore it is bijective.
- $A\mathbf{x} = \mathbf{b}$ has a unique solution $A^{-1}\mathbf{b} = \mathbf{x}$ for all $\mathbf{b}$ in $\mathbb{R}^n$.
- The columns of A generate $\mathbb{R}^n$.
- The columns of A form a base of $\mathbb{R}^n$.
- The column space is equal to $\mathbb{R}^n$: $\text{Col } A = \mathbb{R}^n$
- The file space is equal to $\mathbb{R}^n$: $\text{Fil}A = \mathbb{R}^n$
- The dimension of the column space is n: $\dim \text{Col } A = n$
- The range of A is n: $\text{range } A = n$
- The nullspace of A is only the zero vector: $\text{Nul } A = \{\mathbf{0}\}$
- The dimension of the null space is zero: $\dim \text{Nul } A = 0$
- There is no eigenvalue of 0 for A: $0 \notin \{\lambda_1, \dots \lambda_n\}$.
- The determinant is not equal to 0: $\det A \neq 0$.
- The column space is perpendicular to the zero vector (and thus the null space): $(\text{Col}A)^\perp = \{\mathbf{0}\}$

- The null space is perpendicular to the column space $\mathbb{R}^n$: $(\text{Nul}A)^{\perp} = \mathbb{R}^n$
- A has n singular values different from 0.