CO145 - Mathematical Methods

Prelude

The content discussed here is part of CO145 - Mathematical Methods (Computing MEng); taught by Michael Huth, and Mario Berta, in Imperial College London during the academic year 2018/19. The notes are written for my personal use, and have no guarantee of being correct (although I hope it is, for my own sake). This should be used in conjunction with the lecture notes. This module differs as there isn't as much new content, but it requires practice - as such, I will likely be including worked examples for my own benefit (which are probably incorrect).

Also, the formatting of this document will be a mess. I'd waste too much time trying to figure out how to adjust line spacing otherwise.

Sequences

Formal Definition of a Limit

A sequence a_n, for $n \geq 1$, converges to some limit $l \in \mathbb{R}$ if, and only if, we can prove $\forall \epsilon > 0[\exists N_\epsilon \in \mathbb{N}[\forall n > N_\epsilon[|a_n - l| < \epsilon]]]$.

To show convergence for the sequence $a_n = \frac{1}{n}$, we need to first make a guess for the limit - suppose $l = 0$. We can now attempt to find some N_ϵ. As $\frac{1}{n} - 0$ is positive for all $n \in \mathbb{N}$, we can drop the absolute, thus it's sufficient to find n such that $\frac{1}{n} < \epsilon$. Since both are positive (hence non-zero), we can take reciprocals on both sides, to get $n > \frac{1}{\epsilon}$. However, we are restricted by the fact that n must be an integer, hence it follows $N_\epsilon = \lceil \frac{1}{\epsilon} \rceil$. For any value of ϵ, we can get some N_ϵ with the function, thus it proves that a limit exists.

Common Converging Sequences

Note that for all of these, we are implicitly saying $\lim_{n \to \infty}$, and that $a_n \to 0$.

a_n	condition	N_ϵ
$\frac{1}{n^c}$	for some $c \in \mathbb{R}^+$	$\lceil \frac{1}{\epsilon^c} \rceil$
$\frac{1}{c^n}$	for some $c \in \mathbb{R}$, such that $|c| > 1$	$\lceil \log_c(\frac{1}{\epsilon}) \rceil$
c^n	for some $c \in \mathbb{R}$, such that $|c| < 1$	$\lceil \log_c(\epsilon) \rceil$
$\frac{1}{n!}$		
$\frac{1}{\ln(n)}$	$n > 1$	$\lceil e^{\frac{1}{\epsilon}} \rceil$

Combining Sequences

Suppose that $a_n \to a$, and $b_n \to b$, as $\lim_{n \to \infty}$;

- $\lim_{n \to \infty} \lambda a_n = \lambda a$ given $\lambda \in \mathbb{R}$
- $\lim_{n \to \infty} (a_n + b_n) = a + b$
- $\lim_{n \to \infty} (a_n b_n) = ab$
- $\lim_{n \to \infty} \frac{a_n}{b_n} = \frac{a}{b}$ given $b \neq 0$

For example, the sequence $a_n = \frac{4n^2 + 3n}{7n^2 + 3n - 2}$, it's trivial to find the limit as $n \to \infty$ by inspection as $\frac{4}{7}$. However, if we divide every term by n^2, we end up with $a_n = \frac{4 + \frac{3}{n}}{7 + \frac{3}{n} - \frac{2}{n^2}}$, which we can break into $a_n = \frac{b_n}{c_n}$, where $b_n = 4 + \frac{3}{n}$, and $c_n = 7 + \frac{3}{n} - \frac{2}{n^2}$. Using the rules from above, we can further break down the sequences (but I really cannot be bothered to do so), to a point where we get $a = \frac{4 + 0}{7 + 0 - 0} = \frac{4}{7}$.

Sandwich Theorem

In the sandwich theorem, where we want to prove that $\lim_{n\to\infty} a_n = l$, we need two sequences that form upper, and lower bounds for a_n, namely u_n, and l_n. If such sequences exist, and satisfy $\exists N \in \mathbb{N}[\forall n \geq N[l_n \leq a_n \leq u_n]]$, and both $\lim_{n\to\infty} u_n = \lim_{n\to\infty} l_n = l$, then we get $\lim_{n\to\infty} a_n = l$.

For example, consider the sequence $a_n = \frac{\cos(n)}{n}$. We know that $-1 \leq \cos(n) \leq 1$, therefore $l_n = -\frac{1}{n} \leq a_n \leq \frac{1}{n} = u_n$. However, as both $u_n \to 0$, and $l_n \to 0$, when $n \to \infty$, it follows that $\lim_{n\to\infty} a_n = 0$.

The sandwich theorem can be proven by finding $N_{\epsilon l}$, and $N_{\epsilon u}$ for l_n, and u_n respectively. As they both converge to the same limit, we can justify that for some $N_\epsilon = \max(N_{\epsilon l}, N_{\epsilon u})$, any $n > N_\epsilon$, we have $|l_n - l| < \epsilon$, and $|u_n - l| < \epsilon$. By removing the modulus signs, we get $-\epsilon < l_n - l < \epsilon$, and $-\epsilon < l_n - l < \epsilon$. By rearranging this, and knowing $l_n < u_n$ by definition, we have $l - \epsilon < l_n < u_n < l + \epsilon$. In order to apply the sandwich theorem, we have to assume $l_n \leq a_n \leq u_n$, hence it follows $l - \epsilon < l_n \leq a_n \leq u_n < l + \epsilon$. This can then be arranged to get $-\epsilon < a_n - l < \epsilon$, thus for all $n > N_\epsilon$, $|a_n - l| < \epsilon$.

Ratio Tests

The follow tests allow us to check if a sequence a_n converges to 0, or diverges. If we want to verify a sequence $\lim_{n\to\infty} b_n = l$, where $l \neq 0$, we can reformulate it as $a_n = b_n - l$, and prove it for a_n.

If $|\frac{a_{n+1}}{a_n}| \leq c < 1$, for $c \in \mathbb{R}$, where n is sufficiently large, then $\lim_{n\to\infty} a_n = 0$. Conversely, if $|\frac{a_{n+1}}{a_n}| \geq c > 1$, then a_n diverges.

Suppose $a_n = 2^{-n}$, in order to show it converges, we want to find some c, such that $|\frac{2^{-(n+1)}}{2^{-n}}| \leq c < 1$. By arithmetic, we can say $|\frac{2^{-(n+1)}}{2^{-n}}| = \frac{1}{2} \leq c < 1$. As there exists real $c = \frac{1}{2} < 1$, we can conclude that a_n converges to 0.

However, we cannot prove that $a_n = \frac{1}{n!}$ converges, despite it being a fairly straightforward proof with just the standard ration tests. For example, we would end up with $|\frac{a_{n+1}}{a_n}| = \frac{1}{n+1}$. While this is fairly conclusive that it's less than 1, we don't get a constant c which sits between it, and 1.

Instead of analysing the behaviour of consecutive terms, we take the limit as $r = \lim_{n\to\infty} |\frac{a_{n+1}}{a_n}|$. If $r < 1$, it converges to 0, but if $r > 1$, it then diverges. Applying this to the previous factorial example, we can clearly see that $\lim_{n\to\infty} \frac{1}{n+1} = 0$, therefore a_n converges to 0.

Manipulating Absolute Values

A useful decomposition with absolute limits is $|x| < a \Leftrightarrow -a < x < a$. Further useful properties are;

- $|xy| = |x| \cdot |y|$
- $|\frac{x}{y}| = \frac{|x|}{|y|}$
- $|x + y| \leq |x| + |y|$ $\hspace{3cm}$ triangle inequality
- $|x - y| \geq ||x| - |y||$

Properties of Real Numbers

Suppose we have some set S of real numbers, then the following properties hold;

- u is an upper bound on S if $\forall s \in S[u \geq s]$
- l is a lower bound on S if $\forall s \in S[l \leq s]$
- S is bounded above if it has an upper bound, bounded below if it has a lower bound, and bounded if it has both
- a set may have no bounds, or many, therefore we say $\sup(S)$ is the least upper bound of S, and $\inf(S)$ to denote the greatest upper bound of S

- if a set has no upper bound, then $\sup(S) = \infty$, and similarly if there is no lower bound, then $\inf(S) = -\infty$

The fundamental axiom of analysis states that if an increasing sequence of reals is bounded above, then it must converge.

Series

A series is a sum of a sequence - we consider an infinite series as a summation in the form $S = \sum\limits_{n=1}^{\infty} a_n$.

We can also take a partial sum, where we sum from 1 to n instead of ∞ as $S_n = \sum\limits_{i=1}^{n} a_i$.

Since we're dealing with finite numbers at this point, we can consider S_n as a sequence - thus whether a series converges, or diverges, depends on whether the sequence of partial sums converge, or diverge. If we know that $\forall i \geq 1[a_i > 0]$, then the partial sum is an increasing sequence where $S_1 < S_2 < S_3 < ...$, if this is bounded above, then it must converge (from the axiom of analysis).

For a series to have a chance of converging, we require $\lim\limits_{n\to\infty} a_n = 0$. However, just proving that is insufficient for concluding a series converges. It's important to note that we don't really care about the first values of a series, only how it behaves on large numbers, as it tends to infinity.

If S converges, or diverges, then $\sum\limits_{n=N}^{\infty} a_n$ also converges, or diverges, as it's equivalent to $S - S_{N-1}$.

Geometric Series

To determine whether a limit G exists for the geometric series, we need to determine that the partial sums converge. Hence we do the proof on the right hand side. It's important to note that the only term that changes based on n is x^{n+1}, therefore we can conclude that G must converge when $|x| < 1$ (from earlier results).

Let us represent the geometric series as $G = \sum\limits_{n=1}^{\infty} x^n$. Should a limit G exist - see the left side;

$$G = \sum_{n=1}^{\infty} x^n$$
$$= x + \sum_{n=2}^{\infty} x^n$$
$$= x + x \sum_{n=1}^{\infty} x^n$$
$$= x + xG$$
$$= \frac{x}{1-x}$$

$$G_n = \sum_{i=1}^{n} x^i$$
$$= x + \sum_{i=2}^{n} x^i$$
$$= x + x \sum_{i=1}^{n-1} x^i$$
$$= x + x(G_n - x^n)$$
$$= \frac{x - x^{n+1}}{1-x}$$

Harmonic Series

The harmonic series is often used in comparison tests to prove that another series diverges.

Let harmonic series be written as $S = \sum\limits_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \underbrace{(\frac{1}{3} + \frac{1}{4})}_{>\frac{1}{4}+\frac{1}{4}} + \underbrace{(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8})}_{>\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}} + ... > 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + ...$

Therefore, with partial sums, we get $S_n = \sum\limits_{i=1}^{2^n} \frac{1}{i} > 1 + \frac{n}{2}$. As S_n clearly diverges, S must also diverge.

Inverse Squares

Suppose we have the sum of squares $S = \sum\limits_{n=1}^{\infty} \frac{1}{n^2}$, and an auxiliary partial sum $T_n = \sum\limits_{i=1}^{n} \frac{1}{i(i+1)}$.

To prove that S converges, we should consider $T_n = 1 - \frac{1}{n+1}$ (write out the first few terms and a pattern should be spotted). Hence $\lim\limits_{n\to\infty} T_n = 1$. Using this, we can argue how the partial sums of the series of inverse squares are bounded. It's trivial to say the following inequality; $\frac{1}{i(i+1)} < \frac{1}{i^2} < \frac{1}{i(i-1)}$.

If we sum this inequality from $i = 2$ to n, we can then justify the following - it's important to note that the sums on the left, and right, differ by only the first, and last term (therefore we can reduce it to a partial sum);

$$\frac{1}{2(3)} + \frac{1}{3(4)} + \frac{1}{4(5)} + ... + \frac{1}{n(n+1)} < \sum_{i=2}^{n} \frac{1}{i^2} < \frac{1}{1(2)} + \frac{1}{2(3)} + \frac{1}{3(4)} + ... + \frac{1}{n(n-1)}$$

$$\Leftrightarrow T_n - \frac{1}{2} < \sum_{i=2}^{n} \frac{1}{i^2} < T_{n-1}$$

$$\Leftrightarrow T_n + \frac{1}{2} < \sum_{i=1}^{n} \frac{1}{i^2} < T_{n-1} + 1$$

$$\Leftrightarrow \frac{3}{2} + \frac{1}{n+1} < S_n < 2 + \frac{1}{n}$$

Therefore, it follows that the partial sums are bounded by $\frac{3}{2}$, and 2. The existence of an upper bound suggests the sequence converges, therefore the series must also converge.

Important Series

These examples will be useful for the comparison tests covered later on. We can use these results to aid us in proving whether some other series converges, or diverges.

name	S	condition	diverges or converges
harmonic series	$\sum\limits_{n=1}^{\infty} \frac{1}{n}$		diverges
harmonic primes	$\sum\limits_{p:\text{prime}} \frac{1}{p}$		diverges
geometric series	$\sum\limits_{n=1}^{\infty} x^n$	$|x| \geq 1$	diverges
geometric series	$\sum\limits_{n=1}^{\infty} x^n$	$|x| < 1$	converges
inverse squares series	$\sum\limits_{n=1}^{\infty} n^2$		converges
???	$\sum\limits_{n=1}^{\infty} \frac{1}{n^c}$	$c > 1$	converges

Convergence Tests

For brevity, suppose we have the following;

- $S = \sum\limits_{i=1}^{\infty} a_i$ a series we want to reason about, and $a_i \geq 0$

- $\sum\limits_{i=1}^{\infty} c_i$ a series we have already established converges to c

- $\sum\limits_{i=1}^{\infty} d_i$ a series we have already established diverges

In the **comparison test**, we have some $\lambda > 0$, and a $N \in \mathbb{N}$. If $\forall i > N[a_i \leq \lambda c_i]$, then S converges. On the other hand, if $\forall i > N[a_i \geq \lambda d_i]$, then S diverges.

To avoid having to find such a λ, we can use the **limit comparison test**, in which we can say if $\lim\limits_{i \to \infty} \frac{a_i}{c_i}$ exists, then S converges, but if $\lim\limits_{i \to \infty} \frac{d_i}{a_i}$ exists, then S diverges.

Another useful test for finding convergence, without needing a pre-established series, is **D'Alembert's ratio test**. Once again, we consider this from some point $N \in \mathbb{N}$, where the cases are;

1. $\forall i \geq N[\frac{a_{i+1}}{a_i} \geq 1]$, then S diverges
2. there exists a k such that $\forall i \geq N[\frac{a_{i+1}}{a_i} \leq k < 1]$, then S converges

This is similar to the comparison test for sequences, and once again the requirement of finding some k between the ratio, and 1 can be avoided by taking limits. This leads to **D'Alembert's limit ratio test**, in which we have the following cases;

1. $\lim\limits_{i \to \infty} \frac{a_{i+1}}{a_i} > 1$, then S diverges
2. $\lim\limits_{i \to \infty} \frac{a_{i+1}}{a_i} = 1$, then it is inconclusive
3. $\lim\limits_{i \to \infty} \frac{a_{i+1}}{a_i} < 1$, then S converges

The second case, in which it is inconclusive, can be illustrated by setting $a_n = 1$, which obviously causes S to diverge, or by setting $a_n = \frac{1}{n^2}$ which causes S to converge. However, both lead to a limit test of 1.

The **integral test** for a sequence, where $a_n = f(n)$ is a decreasing function relies on the following idea;

$$\sum_{n=1}^{\infty} a_{n+1} < \int_1^{\infty} f(x)\,\mathrm{d}x < \sum_{n=1}^{\infty} a_n$$

In practice, if the integral were to diverge once evaluated, the series would also diverge by the right hand inequality, but if it were to converge, then the series would converge with the left hand inequality.

Absolute Convergence

Consider a series with negative a_n values, and another series $S' = \sum\limits_{n=1}^{\infty} |a_n|$. If S' converges, so does S.

However, the same isn't true the other way around, for example, when we have $a_n = (-1)^{n-1}\frac{1}{n}$, S converges to $\ln(2)$, however, doing this with the absolute series, S', it would diverge, as if it were the normal harmonic series. In order to test for **absolute** convergence - we have the following cases;

1. if $\lim\limits_{n \to \infty} |\frac{a_{n+1}}{a_n}| > 1$ then $\sum\limits_{n=1}^{\infty} a_n$ diverges

2. if $\lim\limits_{n \to \infty} |\frac{a_{n+1}}{a_n}| = 1$ then $\sum\limits_{n=1}^{\infty} a_n$ can converge, or diverge

3. if $\lim\limits_{n \to \infty} |\frac{a_{n+1}}{a_n}| < 1$ then $\sum\limits_{n=1}^{\infty} a_n$ converges absolutely, and therefore also converges

Power Series, and Radius of Convergence

We can represent a function as a power series; $f(x) = \sum\limits_{n=0}^{\infty} a_n x^n$.

Consider the case where $f(x) = \frac{x}{1-x}$, which has the expansion $\sum\limits_{n=1}^{\infty} x^n$, which only converges for $|x| < 1$.

Take the power series $S = \sum\limits_{n=1}^{\infty} n^2 x^n$, and apply the **D'Alembert ratio test**, we'd get the following;

$$\lim_{n\to\infty} |\frac{a_{n+1}}{a_n}| = \lim_{n\to\infty} |\frac{(n+1)^2 x^{n+1}}{n^2 x^n}|$$
$$= \lim_{n\to\infty} |x(1 + \frac{1}{n})^2|$$
$$= |x|$$

As we want to establish convergence, we set the result $|x| < 1$, therefore it follows that the ratio of convergence for this particular power series is 1.

Power Series

As previously mentioned, a function can be represented in a series expansion on x; $f(x) = \sum\limits_{n=0}^{\infty} a_n x^n$.

Maclaurin Series

Let some function $f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + ...$, then we can easily work out a_0, by setting $x = 0$, such that $f(0) = a_0$. By differentiating the function, we get $f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + ...$, therefore it follows that $f'(0) = a_1$. By differentiating again, we get $f''(x) = 2a_2 + 2 \cdot 3a_3 x + ...$, therefore it follows that $\frac{f''(0)}{2!} = a_2$. Hence, generally $a_n = \frac{f^{(n)}(0)}{n!}$, where $n \geq 0$.

A **Maclaurin series** is the representation of $f(x) = \sum\limits_{i=0}^{\infty} a_i x^i$.

This function can then be differentiated n times to get $f^{(n)}(x) = \sum\limits_{i=n}^{\infty} a_i (i - 1)...(i - n + 1) x^{i-n}$.

From this, we can see that $f^{(n)}(0) = n! a_n$, hence $f(x) = \sum\limits_{n=0}^{\infty} f^n(0) \frac{x^n}{n!}$.

For example, if we take $f(x) = \ln(1 + x)$, and differentiate repeatedly;

$$
\begin{aligned}
f(x) &= \ln(1 + x) & f(0) &= 0 \\
f'(x) &= (1 + x)^{-1} & f'(0) &= 1 \\
f''(x) &= (-1)(1 + x)^{-2} & f''(0) &= -1 \\
f'''(x) &= (-2)(-1)(1 + x)^{-3} & f''(0) &= 2 \\
f^{(n)}(x) &= \frac{(-1)^{n-1}(n-1)!}{(1+x)^n} & f^{(n)}(0) &= (-1)^{n-1}(n - 1)! \\
a_n &= \frac{(-1)^{n-1}}{n}
\end{aligned}
$$

Therefore, we can justify that $f(x) = \ln(1 + x) = \frac{x}{1} - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ...$ By setting $x = -1$, we'd get $-\infty$, and the power series would become the harmonic series, which we know to diverge. However, by setting $x = 1$, we get $\ln(2)$, is the convergence of the alternating harmonic series (as previously mentioned, in the ratio test example).

Taylor Series

When we're interested in values around a certain point, we can use a more generalised version of a **Maclaurin series**. The general result, where we apply the series around a is;

$$f(x) = f(a) + \frac{f^{(1)}(a)}{1!}(x - a)^1 + \frac{f^{(2)}(a)}{2!}(x - a)^2 + ... = \sum\limits_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x - a)^n$$

For example, if we take $f(x) = \ln(x)$ around 2, and differentiate repeatedly;

$$
\begin{aligned}
f(x) &= \ln(x) & f(2) &= \ln(2)
\end{aligned}
$$

$$
\begin{aligned}
f'(x) &= \tfrac{1}{x} \\
f''(x) &= (-1)\tfrac{1}{x^2} \\
f'''(x) &= (-2)(-1)\tfrac{1}{x^3} \\
f^{(n)}(x) &= \tfrac{(-1)^{n-1}(n-1)!}{x^n} \\
a_n &= \tfrac{(-1)^{n-1}}{n2^n}
\end{aligned}
\qquad\qquad
\begin{aligned}
f'(2) &= \tfrac{1}{2} \\
f''(2) &= -\tfrac{1}{4} \\
f''(2) &= \tfrac{1}{4} \\
f^{(n)}(2) &= \tfrac{(-1)^{n-1}(n-1)!}{2^n}
\end{aligned}
$$

For example, we can approximate $\ln(x)$ around 2 as; $\ln(x) = \ln(2) + \sum\limits_{n=1}^{\infty} \frac{(-1)^{n-1}}{n2^n}(x-2)^2$.

It's important to note the bounds for this, as we start from $n = 1$, and manually state the constant term, since that doesn't fall within the "pattern" we had. We use the same technique to calculate the radius of convergence, as follows;

$$
\begin{aligned}
\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| &= \lim_{n\to\infty}\left|\frac{\frac{(x-2)^{n+1}}{(n+1)2^{n+1}}}{\frac{(x-2)^n}{n2^n}}\right| \\
&= \lim_{n\to\infty}\frac{n}{n+1}\cdot\frac{|x-2|}{2} \\
&= \frac{|x-2|}{2}
\end{aligned}
$$

And then, by setting $\frac{|x-2|}{2} < 1$, we have a radius of convergence of 2, around $x = 2$. Generally, the radius is limited by the closest singularity (where a point is not defined, or isn't well-behaved).

Taylor Series Error Term

In order to ensure termination, instead of summing to ∞, we often truncate at some arbitrary k^{th} term.

$$
f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = \sum_{n=0}^{k} \frac{f^{(n)}(a)}{n!}(x-a)^n + \underbrace{\frac{f^{k+1}(c)}{(k+1)!}(x-a)^{k+1}}_{\text{Lagrange error term}}
$$

The Lagrange error term replaces the tail of the infinite series, from, which would otherwise be a sum from $k+1$ to ∞. The constant c lies between x, and a. By taking the bound $a < c < x$, or $x < c < a$, it can be used to generate the worst-case error.

The mean value theorem states that there is some point c between x, and a, such that $f'(c) = \frac{f(x)-f(a)}{x-a}$, therefore we can get $f(x) = f(a) + (x-a)f'(c)$.

Cauchy Error Term

Define the set of partial sums of the Taylor series as $F_k(t) = \sum\limits_{n=0}^{k} \frac{f^{(n)}(t)}{n!}(x-t)^n$, where t is the offset.

It follows that $F_k(x) = f(x)$ for all k (justified as every term becomes 0, other than the first term, $n = 0$, where it's $\frac{f(x)}{0!}$). Therefore the error term is derived as follows, where $R_k(x)$ is the remainder after k terms;

$$
F_k(x) - F_k(a) = f(x) - \sum_{n=0}^{k} \frac{f^{(n)}(a)}{n!}(x-a)^n = R_k(x)
$$

Power Series Solutions of ODE

When we have a differential equation in the form $\frac{\mathrm{d}y}{\mathrm{d}x} = ky$, for a constant k, we can consider the power series for y. By differentiating term by term, we can compare coefficients;

$$
\frac{\mathrm{d}y}{\mathrm{d}x} \equiv \frac{\mathrm{d}}{\mathrm{d}x}\sum_{i=0}^{\infty} a_i x^i \equiv \sum_{i=1}^{\infty} a_i i x^{i-1} \equiv ky \equiv k\sum_{i=0}^{\infty} a_i x^i \equiv k\sum_{i=1}^{\infty} a_{i-1} x^{i-1} \equiv \sum_{i=1}^{\infty} k a_{i-1} x^{i-1}
$$

By matching coefficients of x^{i-1}, where $i \geq 1$, we have $ka_{i-1} = a_i i$, therefore we end up with the relation $a_i = \frac{k}{i}a_{i-1}$. By "expanding" this, we end up with $a_i = \frac{k}{i}a_{i-1} = \frac{k}{i} \cdot \frac{k}{i-1}a_{i-2} = ... = \frac{k^i}{i!}a_0$.

If we were to have some boundary condition, such as when $x = 0$, $y = 1$, we know that $y(0) = a_0$, hence $a_0 = 1$. Therefore, this can be solved to the following;

$$y = \sum_{i=0}^{\infty} \frac{(kx)^i}{i!} = e^{kx}$$

Linear Algebra

Linear Equation Systems

For brevity, instead of writing;

$$
\begin{array}{ccccccc}
ax_1 & + & bx_2 & + & cx_3 & = & j \\
dx_1 & + & ex_2 & + & fx_3 & = & k \\
gx_1 & + & hx_2 & + & ix_3 & = & l
\end{array}
\qquad
\begin{array}{c}
(1) \\
(2) \\
(3)
\end{array}
$$

We will write $\mathbf{A}\vec{x} = \vec{y}$, where $\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$, $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, and $\vec{y} = \begin{bmatrix} j \\ k \\ l \end{bmatrix}$

In general, for a real linear equation system, we have the following possible outcomes;

- no solution

  $$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 2 & 0 & 3 \end{bmatrix}, \text{ and } \vec{y} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

  By adding equations (1), and (2), we get $2x_1 + 3x_3 = 5$, which contradicts equation (3), hence it is **inconsistent**, and therefore has no solutions.

- unique solution

  $$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 0 & 1 & 1 \end{bmatrix}, \text{ and } \vec{y} = \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix}$$

  By doing equation (1) - equation (3), we get $x_1 = 1$. Adding equations (1), and (2), we end up with $2x_1 + 3x_3 = 5$, however, we know the value of x_1, therefore it follows that $x_3 = 1$. Finally, by using this value in equation (3), we have $x_2 = 1$, therefore the only solution is;

  $$\vec{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- infinite solutions

  $$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 2 & 0 & 3 \end{bmatrix}, \text{ and } \vec{y} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$$

  Note that this is the same $\mathbf{A}$ from the first example. However, this time adding equations (1), and (2), is consistent with equation (3). If we let $x_3 = \lambda$, such that x_3 is a free variable, with $\lambda \in \mathbb{R}$, we can get both x_1, and x_2 in terms of λ. By doing equations (1) - (2), we get $2x_2 - x_3 = 1$, hence $x_2 = \frac{1}{2} + \frac{1}{2}\lambda$. By doing equations (1) + (2), we get $2x_1 + 3x_3 = 5$, which leads to $x_1 = \frac{5}{2} - \frac{3}{2}\lambda$. As such, we have an infinite set of solutions, given $\lambda \in \mathbb{R}$;

  $$\vec{x} = \begin{bmatrix} \frac{5}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} + \lambda \begin{bmatrix} -\frac{3}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix}$$

Groups

With a set G, and some operation $\otimes : G \to G$, the criteria for $(G, \otimes)$ being a **group** are (1 - 4), and (1 - 5) for an **Abelian group**;

1. **closure** $\hspace{6cm} \forall x, y \in G[x \otimes y \in G]$

2. **associativity** $\hspace{4.5cm} \forall x, y, z \in G[(x \otimes y) \otimes z = x \otimes (y \otimes z)]$

3. **neutral element** $\hspace{4cm} \exists e \in G[\forall x \in G[x \otimes e = x \wedge e \otimes x = x]]$

4. **inverse element** $\hspace{3.5cm} \forall x \in G[\exists y \in G[x \otimes y = e \wedge y \otimes x = e]]$

   often written as $y = x^{-1}$, note that this does **not** mean $y = \frac{1}{x}$, defined with respect to $\otimes$

5. **commutative** $\hspace{5.5cm} \forall x, y \in G[x \otimes y = y \otimes x]$

For example, we can say $(\mathbb{Z}, +)$ is a group, but $(\mathbb{N}, +)$ is not, since the latter doesn't have an inverse element; for example $-5 \notin \mathbb{N}$, so 5 doesn't have an inverse. While $(\mathbb{R} \backslash \{0\}, \cdot)$ is an Abelian group, $(\mathbb{R}, \cdot)$ isn't a group as 0 doesn't have an inverse element $(\neg \exists x \in \mathbb{R}[0 \cdot x = 1])$. $(\mathbb{R}^n, +)$ is an Abelian groups under addition, as long as addition is defined component-wise.

Matrices

We can defined the set of all real-valued $m \times n$ matrices $(m, n \in \mathbb{N})$, where m is the number of rows, and n is the number of columns, as $\mathbb{R}^{m \times n}$. We can refer to each element as $a_{i,j}$, where $i \in [1, m]$, and $j \in [1, n]$, and $a_{i,j} \in \mathbb{R}$; thus giving us the following rectangular matrix;

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

A $\mathbb{R}^{1 \times n}$ matrix is referred to as a **row (vector)**, and a $\mathbb{R}^{m \times 1}$ matrix is referred to as a **column (vector)**, with component-wise addition, we are able to say $(\mathbb{R}^{m \times n}, +)$ is a group.

Multiplication on matrices, $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times k}, \mathbf{C} \in \mathbb{R}^{m \times k}$, where $\mathbf{C} = \mathbf{AB}$, is defined as;

$$c_{i,j} = \sum_{l=1}^{n} a_{i,l} b_{l,j}, \text{ with } i \in [1, m], \text{ and } j \in [1, k]$$

In general, matrix multiplication is not commutative, hence $\mathbf{AB} \neq \mathbf{BA}$. However, it is associative, and distributive (given that the dimensions match).

The identity matrix, written as $\mathbf{I}_n \in \mathbb{R}^{n \times n}$, is all 0s, other than $a_{i,i} = 1$, where $i \in [1, n]$. The identity matrix has the following property; $\mathbf{I}_m \mathbf{A} = \mathbf{A} = \mathbf{A} \mathbf{I}_n$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$.

If a square matrix $(\mathbf{A} \in \mathbb{R}^{n \times n})$ is non-singular, we can find an inverse for it, $\mathbf{A}^{-1} \in \mathbb{R}^{n \times n}$, such that $\mathbf{A} \mathbf{A}^{-1} = \mathbf{A}^{-1} \mathbf{A} = \mathbf{I}_n$. However, not every matrix has an inverse. We also have the result $(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}$.

We can define the transpose of $\mathbf{A} \in \mathbb{R}^{n \times m}$ as $\mathbf{B} = \mathbf{A}^\top \in \mathbf{R}^{m \times n}$, where $b_{j,i} = a_{i,j}$. On the rectangular matrix representation, we are essentially mirroring it on the leading diagonal. A **square** matrix is symmetric $\Leftrightarrow \mathbf{A} = \mathbf{A}^\top$. The general results are $(\mathbf{A}^\top)^\top = \mathbf{A}$, $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$, and $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$. Also, if $\mathbf{A}$ is non-singular, then $(\mathbf{A}^{-1})^\top = (\mathbf{A}^\top)^{-1}$.

A multiplication by scalar $\lambda \in \mathbb{R}$, means that for $\mathbf{A} \in \mathbb{R}^{m \times n}$, and $\mathbf{B} = \lambda \mathbf{A}$, it follows that $b_{i,j} = \lambda a_{i,j}$. This operation is distributive, and also associative.

Solving Linear Equation Systems

Given the general form of an equation system as;

$$
\begin{array}{ccccccccc}
a_{1,1}x_1 & + & a_{1,2}x_1 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\
a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots & + & a_{2,n}x_n & = & b_2 \\
\vdots & & \vdots & & & & \vdots & & \vdots \\
a_{m,1}x_1 & + & a_{m,2}x_2 & + & \cdots & + & a_{m,n}x_n & = & b_m
\end{array}
$$

We can easily represent it as $\mathbf{A}\vec{x} = \vec{b}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\vec{x} \in \mathbb{R}^{n(\times 1)}$, and $\vec{b} \in \mathbb{R}^{m(\times 1)}$.

For example, consider a simple linear equation system as follows;

$$
\mathbf{A}\vec{x} = \begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}
$$

Let us also say $\mathbf{A} = \begin{bmatrix} \vec{c_1} \mid \vec{c_2} \mid \vec{c_3} \mid \vec{c_4} \end{bmatrix}$, where $\vec{c_i}$ represents the ith column of $\mathbf{A}$, hence $\vec{b} = \sum_{i=1}^{4} x_i \vec{c_i}$.

In order to find a particular solution, we can easily create $\vec{b}$ as a linear combination of $\vec{c_1}$, and $\vec{c_2}$, such that $\vec{b} = 42\vec{c_1} + 8\vec{c_2} + 0\vec{c_3} + 0\vec{c_4}$, therefore we can get a **particular** solution; $\vec{x} = \begin{bmatrix} 42 & 8 & 0 & 0 \end{bmatrix}^{\top}$.
This doesn't capture all the solutions of this equation system - if there are more unknowns than equations, then there are an infinite number of solutions. Recall that $\vec{b} + \vec{0} = \vec{b}$, hence we can attempt to find non-trivial $\vec{0}$s to generate the set of all solutions.

Using the fact that $\mathbf{A}$ is already in **row echelon form**, we can try say that $\vec{c_3} = 8\vec{c_1} + 2\vec{c_2}$, therefore, it follows that $\vec{0} = 8\vec{c_3} + 2\vec{c_2} - \vec{c_3} + 0\vec{c_4}$, as such we've managed to create a zero solution when $\vec{x} = \begin{bmatrix} 8 & 2 & -1 & 0 \end{bmatrix}^{\top}$. We can do the same with $\vec{c_4}$, such that $\vec{c_4} = -4\vec{c_1} + 12\vec{c_2} \Leftrightarrow \vec{0} = -4\vec{c_1} + 12\vec{c_2} + 0\vec{c_3} - \vec{c_4}$, hence $\vec{x} = \begin{bmatrix} -4 & 12 & 0 & -1 \end{bmatrix}^{\top}$ is also a zero solution. With these two $\vec{0}$s, we can say that the general solution is as follows (where $\lambda_1, \lambda_2 \in \mathbb{R}$);

$$
\vec{x} = \begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 8 \\ 2 \\ -1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} -4 \\ 12 \\ 0 \\ -1 \end{bmatrix}
$$

In this solution, we first find a **particular** solution for $\mathbf{A}\vec{x} = \vec{b}$, and then combine this with **all** solutions to $\mathbf{A}\vec{x} = \vec{0}$. This was fairly trivial to solve, as the matrix was already in a special form. Generally, equation systems won't be given to us like this, however we can transform it into this form by **Gaussian elimination**.

Elementary Transformations

We are allowed to apply the following operations, which let us transform the equation system into a simpler state, but preserve the solution set;

- swap two equations (which means we can swap rows in the matrix)
- multiply an equation (matrix row) by a non-zero constant $\lambda \in \mathbb{R} \backslash \{0\}$
- add (or subtract) an equation (row) to / from another equation (row)

We can say that a matrix is in **row echelon form (REF)**, if all zero rows are below the non-zero rows, and the pivot of each row is **strictly** to the right of the pivot in the row above. The pivot of a row is defined as the first non-zero element in a given row, reading left-to-right.

Consider the following system of equations, and the result of it after Gaussian elimination;

$$\begin{array}{rcrcrcrcrcr}
-2x_1 & + & 4x_2 & - & 2x_3 & - & x_4 & + & 4x_5 & = & -3 \\
4x_1 & - & 8x_2 & + & 3x_3 & - & 3x_4 & + & x_5 & = & 2 \\
x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 \\
x_1 & - & 2x_2 & + & 0x_3 & - & 3x_4 & + & 4x_5 & = & a
\end{array}$$

$$\rightsquigarrow$$

$$\begin{array}{rcrcrcrcrcr}
x_1 & - & 2x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0 \\
 & & & & x_3 & - & x_4 & + & 3x_5 & = & -2 \\
 & & & & & & x_4 & - & 2x_5 & = & 1 \\
 & & & & & & & & 0 & = & a+1
\end{array}$$

We can now work on the augmented matrix to transform it into **REF**;

$$\left[\begin{array}{ccccc|c}
-2 & 4 & -2 & -1 & 4 & -3 \\
4 & -8 & 3 & -3 & 1 & 2 \\
1 & -2 & 1 & -1 & 1 & 0 \\
1 & -2 & 0 & -3 & 4 & a
\end{array}\right]\begin{array}{l} \leftrightarrow R_3 \\ \\ \leftrightarrow R_1 \\ \\ \end{array}
\rightsquigarrow
\left[\begin{array}{ccccc|c}
1 & -2 & 1 & -1 & 1 & 0 \\
4 & -8 & 3 & -3 & 1 & 2 \\
-2 & 4 & -2 & -1 & 4 & -3 \\
1 & -2 & 0 & -3 & 4 & a
\end{array}\right]\begin{array}{l} \\ -4R_1 \\ +2R_1 \\ -R_1 \end{array}
\rightsquigarrow$$

$$\left[\begin{array}{ccccc|c}
1 & -2 & 1 & -1 & 1 & 0 \\
0 & 0 & -1 & 1 & -3 & 2 \\
0 & 0 & 0 & -3 & 6 & -3 \\
0 & 0 & -1 & -2 & 3 & a
\end{array}\right]\begin{array}{l} \\ \\ \\ -R_2 \end{array}
\rightsquigarrow
\left[\begin{array}{ccccc|c}
1 & -2 & 1 & -1 & 1 & 0 \\
0 & 0 & -1 & 1 & -3 & 2 \\
0 & 0 & 0 & -3 & 6 & -3 \\
0 & 0 & 0 & -3 & 6 & a-2
\end{array}\right]\begin{array}{l} \\ \\ \\ -R_3 \end{array}
\rightsquigarrow$$

$$\left[\begin{array}{ccccc|c}
1 & -2 & 1 & -1 & 1 & 0 \\
0 & 0 & -1 & 1 & -3 & 2 \\
0 & 0 & 0 & -3 & 6 & -3 \\
0 & 0 & 0 & 0 & 0 & a+1
\end{array}\right]\begin{array}{l} \\ \cdot -1 \\ \cdot -\frac{1}{3} \\ \end{array}
\rightsquigarrow
\left[\begin{array}{ccccc|c}
1 & -2 & 1 & -1 & 1 & 0 \\
0 & 0 & 1 & -1 & 3 & -2 \\
0 & 0 & 0 & 1 & -2 & 1 \\
0 & 0 & 0 & 0 & 0 & a+1
\end{array}\right]$$

It follows that there is only a solution when $a = -1$. To clarify some definitions, we refer to a variable as being **basic**, if it corresponds to the pivot element of any given row, hence in this case, the basic variables are x_1, x_3, x_4, and the rest of the variables are **free** (x_2, and x_5 in our case). The pivot columns in our case (written as $\vec{p_i}$ corresponding to x_i) are;

$$\vec{p_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \vec{p_3} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \text{ and } \vec{p_4} = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix}$$

We implicitly set the **free** variables to 0 when we find our particular solution using this method.

$$x_1\vec{p_1} + x_3\vec{p_3} + x_4\vec{p_4} = \vec{b} \Leftrightarrow x_1\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4\begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} \Leftrightarrow \vec{x} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}$$

We can get $x_4 = 1$ straight away, and from that we can get $x_3 = -1$, and also $x_1 = 2$. By using the same method as before, where we try create the non-pivot columns as a linear combination of the pivot columns, we can get the following general solution ($\forall \lambda_1, \lambda_2 \in \mathbb{R}$) - note that we could just as easily negate all the values in the column vectors, since our λ_1, λ_2 can take any real value;

$$\vec{x} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} -2 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} -2 \\ 0 \\ 1 \\ -2 \\ -1 \end{bmatrix}$$

Minus-1 Trick

We say that some equation system is in **reduced row echelon form (RREF)** (this may also be referred to as row-reduced echelon form, or row canonical form in other texts), if it is already in row echelon form, every pivot item is 1, and the pivot is the only non-zero entry in the column. Consider the following matrix, with the pivots coloured purple, and the matrix after adding rows of the form $\begin{bmatrix} 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \end{bmatrix}$, to create a square matrix. By reading the columns containing the new -1s, we can get all solutions for $\mathbf{A}\vec{x} = \vec{0}$, once again $\forall \lambda_1, \lambda_2 \in \mathbb{R}$;

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \end{bmatrix} \qquad \mathbf{A}' = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \qquad \vec{x} = \lambda_1 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}$$

Gaussian Elimination to Calculate the Inverse

This is a fairly simple technique for calculating the inverse matrix, where we essentially augment the original matrix with the identity, and apply Gaussian elimination to get the identity on the left hand side; $\begin{bmatrix} \mathbf{A} \mid \mathbf{I}_n \end{bmatrix} \rightsquigarrow \cdots \rightsquigarrow \begin{bmatrix} \mathbf{I}_n \mid \mathbf{A}^{-1} \end{bmatrix}$.

This is one of the parts in this module that are fairly easy to understand, but just requires practice for the exams. This technique is applied to a simple $\mathbf{A} \in \mathbb{R}^{2 \times 2}$;

$$\begin{bmatrix} 1 & 2 & \vert & 1 & 0 \\ 3 & 4 & \vert & 0 & 1 \end{bmatrix} \overset{-3R_1}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & \vert & 1 & 0 \\ 0 & -2 & \vert & -3 & 1 \end{bmatrix} \overset{\cdot -\frac{1}{2}}{\rightsquigarrow} \begin{bmatrix} 1 & 2 & \vert & 1 & 0 \\ 0 & 1 & \vert & \frac{3}{2} & -\frac{1}{2} \end{bmatrix} \overset{-2R_2}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & \vert & -2 & 1 \\ 0 & 1 & \vert & \frac{3}{2} & -\frac{1}{2} \end{bmatrix}$$

Vector Spaces

We define a real-valued **vector space** ($\mathbb{R}$-vector space) is a set V with two operations;

- $+ : V \times V \to V$ inner operation
- $\cdot : \mathbb{R} \times V \to V$ outer operation

The following properties hold;

- $(V, +)$ is an Abelian group
- the outer operation is distributive

$$\lambda \cdot (\vec{x} + \vec{y}) = \lambda \cdot \vec{x} + \lambda \cdot \vec{y} \qquad\qquad \forall \lambda \in \mathbb{R}, \vec{x}, \vec{y} \in V$$
$$(\lambda + \psi) \cdot \vec{x} = \lambda \cdot \vec{x} + \psi \cdot \vec{x} \qquad\qquad \forall \lambda, \psi \in \mathbb{R}, \vec{x} \in V$$

- the outer operation is associative

$$\lambda \cdot (\psi \cdot \vec{x}) = (\lambda \cdot \psi) \cdot \vec{x} \qquad\qquad \forall \lambda, \psi \in \mathbb{R}, \vec{x} \in V$$

- the outer operation has a neutral element of 1

$$1 \cdot \vec{x} = \vec{x} \qquad\qquad \forall \vec{x} \in V$$

Examples of vector spaces are $V = \mathbb{R}^n$, $V = \mathbb{R}^{m \times n}$, and $V = \mathbb{C}$ (the standard representation of the complex numbers).

Vector Subspaces

Suppose we have $U \subset V$, where V is a vector space, then U is a vector / linear subspace if it is a vector space with the same outer, and inner operations, but restricted to $U \times U$, and $\mathbb{R} \times U$, respectively.

For us to show that U is a vector subspace, we must show the following properties;

- $U \neq \varnothing$, especially have to show $\vec{0} \in U$

- closure of U

  outer operation: $\lambda \vec{x} \in U$ $\hspace{6cm} \forall \lambda \in \mathbb{R}, \vec{x} \in U$

  inner operation: $\vec{x} + \vec{y} \in U$ $\hspace{6cm} \forall \vec{x}, \vec{y} \in U$

We have trivial vector subspaces of the vector space V, namely V itself, and $\{\vec{0}\}$ (the subspace containing only the zero vector). The intersection of an arbitrary number of subspaces is also a subspace, and the intersection of all subspaces of V is referred to as the **linear hull** of V.

The solution space for a homogeneous linear equation system $\mathbf{A}\vec{x} = \vec{0}$, with n unknowns, is a subspace of $\mathbb{R}^n$. Note that a homogeneous equation means that it has the zero vector on the right hand side. The same also applies in reverse; such that for a vector subspace $U \subset V$, there is a homogeneous system which characterises U. In contrast, an inhomogeneous linear equation system $\mathbf{A}\vec{x} = \vec{b}$, where $\vec{b} \neq \vec{0}$ does not have a vector subspace for its solutions. Assume that it does have some subspace U, therefore, $\vec{0} \in U$ by definition. However, $\vec{0}$ doesn't satisfy the equation, since we know $\vec{b} \neq \vec{0}$.

Linear Independence

We can define a vector space V, with $\vec{x_1}, \vec{x_2}, ..., \vec{x_k}$, as being linearly dependent if, and only if, there exists a non-trivial solution (such that there exists at least one $\lambda_i \neq 0$) in the following equation;

$$\vec{0} = \sum_{i=1}^{k} \lambda_i \vec{x_i}$$

By the closure of vector spaces, a set of linearly dependent vectors contains some redundancy.

We can check if vectors $\vec{x_1}, \vec{x_2}, ..., \vec{x_k} \in V$ are linearly independent by constructing an augmented matrix; $\mathbf{A} = \begin{bmatrix} \vec{x_1} \mid \vec{x_2} \mid \cdots \mid \vec{x_k} \end{bmatrix}$. Once put into row echelon form, if all the columns are pivot columns, then it follows that it is linearly independent, otherwise the non-pivot columns can be made from the pivot columns to the left of it (therefore it is linearly dependent).

Consider the following vectors in $\mathbb{R}^4$;

$$\vec{x_1} = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \ \vec{x_2} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \text{ and } \vec{x_3} = \begin{bmatrix} -1 \\ -2 \\ 1 \\ 1 \end{bmatrix}$$

We essentially want to solve $\lambda_1 \vec{x_1} + \lambda_2 \vec{x_2} + \lambda_3 \vec{x_3} = \vec{0}$, which means we can set this up as follows;

$$\begin{bmatrix} 1 & 1 & -1 & | & 0 \\ 2 & 1 & -2 & | & 0 \\ -3 & 0 & 1 & | & 0 \\ 4 & 2 & 1 & | & 0 \end{bmatrix} \begin{matrix} \\ -2R_1 \\ +3R_1 \\ -4R_1 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 & | & 0 \\ 0 & -1 & 0 & | & 0 \\ 0 & 3 & -2 & | & 0 \\ 0 & -2 & 5 & | & 0 \end{bmatrix} \begin{matrix} \\ \\ +3R_2 \\ -2R_2 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 & | & 0 \\ 0 & -1 & 0 & | & 0 \\ 0 & 0 & -2 & | & 0 \\ 0 & 0 & 5 & | & 0 \end{bmatrix} \begin{matrix} \\ \cdot -1 \\ \cdot -\frac{1}{2} \\ \end{matrix} \rightsquigarrow$$

$$\begin{bmatrix} 1 & 1 & -1 & | & 0 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 1 & | & 0 \\ 0 & 0 & 5 & | & 0 \end{bmatrix} \begin{matrix} \\ \\ \\ -5R_3 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & -1 & | & 0 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 1 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}$$

As every column here is a pivot column, such that every column is linearly independent of the ones to its left, it follows that there isn't a non-trivial solution, as the only solution is $\lambda_1 = \lambda_2 = \lambda_3 = 0$, therefore the vectors are linearly independent.

Linear Combinations (or Something?)

Suppose there is some vector space V, with k linearly independent vectors $\vec{b_1}, \vec{b_2}, ..., \vec{b_k}$, and m linear combinations, expressed as follows (with $j \in [1, m]$);

$$\vec{x_j} = \sum_{i=1}^{k} \lambda_{i,j} \vec{b_i}$$

We can then apply the same approach to test for linear independence;

$$\vec{0} = \sum_{j=1}^{m} \mu_j \vec{x_j} = \sum_{j=1}^{m} \mu_j \sum_{i=1}^{k} \lambda_{i,j} \vec{b_i} = \sum_{j=1}^{m} \sum_{i=1}^{k} \mu_j \lambda_{i,j} \vec{b_i} = \sum_{i=1}^{k} \left(\sum_{j=1}^{m} \mu_j \lambda_{i,j} \right) \vec{b_i}$$

This means that $\vec{x_1}, \vec{x_2}, ..., \vec{x_m}$ only has linear independence if, and only if the following column vectors are linearly independent;

$$\vec{\lambda_1} = \begin{bmatrix} \lambda_{1,1} \\ \vdots \\ \lambda_{k,1} \end{bmatrix} \in \mathbb{R}^k, ..., \vec{\lambda_m} = \begin{bmatrix} \lambda_{1,m} \\ \vdots \\ \lambda_{k,m} \end{bmatrix} \in \mathbb{R}^k$$

Consider the following example, with a set of linearly independent vectors $\vec{b_1}, \vec{b_2}, \vec{b_3}, \vec{b_4} \in \mathbb{R}^3$, and;

$$\begin{array}{rrrrrrrrr}
\vec{b_1} & - & 2\vec{b_2} & + & \vec{b_3} & - & \vec{b_4} & = & \vec{x_1} \\
-4\vec{b_1} & - & 2\vec{b_2} & + & 0\vec{b_3} & + & 4\vec{b_4} & = & \vec{x_2} \\
2\vec{b_1} & + & 3\vec{b_2} & - & \vec{b_3} & - & 3\vec{b_4} & = & \vec{x_3} \\
17\vec{b_1} & - & 10\vec{b_2} & + & 11\vec{b_3} & + & \vec{b_4} & = & \vec{x_4}
\end{array}$$

Now, to check whether $\vec{x_1}, \vec{x_2}, \vec{x_3}, \vec{x_4}$ are linearly independent, we need to find the reduced row echelon form of the coefficient matrix;

$$\begin{bmatrix} 1 & -4 & 2 & 17 \\ -2 & -2 & 3 & -10 \\ 1 & 0 & -1 & 11 \\ -1 & 4 & -3 & 1 \end{bmatrix} \begin{matrix} \\ +2R_1 \\ -R_1 \\ +R_1 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & -4 & 2 & 17 \\ 0 & -10 & 7 & 24 \\ 0 & 4 & -3 & -6 \\ 0 & 0 & -1 & 18 \end{bmatrix} \begin{matrix} \\ \cdot -\frac{1}{10} \\ \\ \cdot -1 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & -4 & 2 & 17 \\ 0 & 1 & -\frac{7}{10} & -\frac{24}{10} \\ 0 & 4 & -3 & -6 \\ 0 & 0 & 1 & -18 \end{bmatrix} \begin{matrix} \\ \\ -4R_2 \\ \end{matrix} \rightsquigarrow$$

$$\begin{bmatrix} 1 & -4 & 2 & 17 \\ 0 & 1 & -\frac{7}{10} & -\frac{24}{10} \\ 0 & 0 & -\frac{2}{10} & \frac{36}{10} \\ 0 & 0 & 1 & -18 \end{bmatrix} \begin{matrix} \\ \\ +\frac{2}{10}R_4 \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & -4 & 2 & 17 \\ 0 & 1 & -\frac{7}{10} & -\frac{24}{10} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -18 \end{bmatrix} \begin{matrix} \\ \\ \leftrightarrow R_4 \\ \leftrightarrow R_3 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & -4 & 2 & 17 \\ 0 & 1 & -\frac{7}{10} & -\frac{24}{10} \\ 0 & 0 & 1 & -18 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} -2R_3 \\ +\frac{7}{10}R_3 \\ \\ \end{matrix} \rightsquigarrow$$

$$\begin{bmatrix} 1 & -4 & 0 & 53 \\ 0 & 1 & 0 & -15 \\ 0 & 0 & 1 & -18 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} +4R_2 \\ \\ \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & -15 \\ 0 & 0 & 1 & -18 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since we have a non-pivot column, it follows that it's solvable non-trivially, and therefore it is linearly dependent. Therefore, the $\vec{x}$ vectors are linear dependent since $\vec{x_4}$ lies in the span of $\vec{x_1}, \vec{x_2}, \vec{x_3}$.

Basis

Given a finite number of vectors $\vec{x_1}, \vec{x_2}, ..., \vec{x_k} \in V$, we can say that any vector of the following form, is a linear combination of the vectors $\vec{x_1}, \vec{x_2}, ..., \vec{x_k}$ (given $\lambda_1, \lambda_2, ..., \lambda_k \in \mathbb{R}$);

$$\vec{v} = \sum_{i=1}^{k} \lambda_i \vec{x_i} \in V$$

Suppose we have $A \subset V$, and $A = \{\vec{x_1}, \vec{x_2}, ..., \vec{x_k}\}$, if we can generate every vector $\vec{v} \in V$, as a linear combination of vectors in A, then A is a **generating set / span** of the vector space V. This is denoted as $V = [A] = [\vec{x_1}, \vec{x_2}, ..., \vec{x_k}]$.

We can define a generating set of V as a **basis** of V, if is linearly independent, and minimal. We say a generating set of V is minimal if there isn't a smaller set which spans V. With a real vector space

V, and $B \subset V$, $B \neq \varnothing$, B is a basis of $V \Leftrightarrow B$ is a minimal generating set $\Leftrightarrow B$ is a maximal linearly independent subset of V.

It is also equivalent to say that every $\vec{v} \in V$ is a **unique** linear combination of vectors from B;

$$\vec{v} = \sum_{i=1}^{k} \lambda_i \vec{b}_i = \sum_{i=1}^{k} \mu_i \vec{b}_i \Rightarrow \forall i \in [1, k][\mu_i = \lambda_i]$$

While bases aren't unique, the number of elements (or $\dim(V)$) is constant for a vector space V. In $\mathbb{R}^3$, we refer to the **standard** basis as the following set of vectors, thus $\dim(\mathbb{R}^3) = 3$, obviously;

$$B = [\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}]$$

With a vector subspace $U \subset V$, we can set an upper bound on the dimensions as $\dim(U) \leq \dim(V)$, and $\dim(U) = \dim(V) \Leftrightarrow U = V$.

Consider the following subspace $U \subset \mathbb{R}^5$, spanned by;

$$\vec{x_1} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \vec{x_2} = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \\ -2 \end{bmatrix}, \vec{x_3} = \begin{bmatrix} 3 \\ -4 \\ 3 \\ 5 \\ -3 \end{bmatrix}, \text{ and } \vec{x_4} = \begin{bmatrix} -1 \\ 8 \\ -5 \\ -6 \\ 1 \end{bmatrix}$$

As we're once again trying to determine linear independence, we end up with another homogeneous equation system, that can be represented by taking the augmented matrix of the vectors, and then apply Gaussian elimination on it, down to reduced row echelon form (still doing all of it, for practice);

$$\begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & -4 & 8 \\ -1 & 1 & 3 & -5 \\ -1 & 2 & 5 & -6 \\ -1 & -2 & -3 & 1 \end{bmatrix} \begin{matrix} \\ -2R_1 \\ +R_1 \\ +R_1 \\ +R_1 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & -5 & -10 & 10 \\ 0 & 3 & 6 & -6 \\ 0 & 4 & 8 & -7 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} \\ \\ \cdot\frac{1}{3} \\ \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & -5 & -10 & 10 \\ 0 & 1 & 2 & -2 \\ 0 & 4 & 8 & -7 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} \\ +5R_3 \\ \\ -4R_3 \\ \end{matrix} \rightsquigarrow$$

$$\begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} \\ \rightarrow R_4 \\ \rightarrow R_2 \\ \rightarrow R_3 \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} +R_3 \\ +2R_3 \\ \\ \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} -2R_2 \\ \\ \\ \\ \end{matrix} \rightsquigarrow$$

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

From this, we can see that the vectors corresponding to the pivot columns are $\vec{x_1}$, $\vec{x_2}$, and $\vec{x_4}$, hence we can say that $[\vec{x_1}, \vec{x_2}, \vec{x_4}]$ forms a basis of U.

On the other hand, we can try to find a simple basis of U, which means we want as many coordinates equal to 0 as possible. We can do this by writing the vectors as **rows** in the matrix, and then apply the same Gaussian elimination;

$$\begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 2 & -1 & 1 & 2 & -2 \\ 3 & -4 & 3 & 5 & -3 \\ -1 & 8 & -5 & -6 & 1 \end{bmatrix} \begin{matrix} \\ -2R_1 \\ -3R_1 \\ +R_1 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 0 & -5 & 3 & 4 & 0 \\ 0 & -10 & 6 & 8 & 0 \\ 0 & 10 & -6 & -7 & 0 \end{bmatrix} \begin{matrix} \\ \\ \\ +R_3 \end{matrix} \rightsquigarrow$$

$$\begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 0 & -5 & 3 & 4 & 0 \\ 0 & -10 & 6 & 8 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{matrix} \\ \\ +2R_2 \\ \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 0 & -5 & 3 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{matrix} \\ \cdot -\frac{1}{5} \\ \leftrightarrow R_4 \\ \leftrightarrow R_3 \end{matrix} \rightsquigarrow$$

$$\begin{bmatrix} 1 & 2 & -1 & -1 & -1 \\ 0 & 1 & -\frac{3}{5} & -\frac{4}{5} & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} +R_3 \\ +\frac{4}{5}R_3 \\ \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & -1 & 0 & -1 \\ 0 & 1 & -\frac{3}{5} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} -2R_2 \\ \\ \\ \end{matrix} \rightsquigarrow$$

$$\begin{bmatrix} 1 & 0 & \frac{1}{5} & 0 & -1 \\ 0 & 1 & -\frac{3}{5} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad U = [\begin{bmatrix} 1 \\ 0 \\ \frac{1}{5} \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -\frac{3}{5} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}]$$

By reading the rows with the leading 1s, we have a simple basis; $U = [\vec{b_1}, \vec{b_2}, \vec{b_3}]$

Rank

The rank of a matrix is the number of linearly independent columns, denoted $\text{rk}(\mathbf{A})$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$. We are also given that the column rank is the same as the row rank, hence $\text{rk}(\mathbf{A}) = \text{rk}(\mathbf{A}^\top)$.

The columns of $\mathbf{A}$ span a subspace $U \in \mathbb{R}^m$, with $\dim(U) = \text{rk}(\mathbf{A})$, and the rows span a subspace $W \in \mathbb{R}^n$, with $\dim(W) = \text{rk}(\mathbf{A})$. A basis of U, and W, can be found by applying Gaussian elimination to $\mathbf{A}$, and $\mathbf{A}^\top$ respectively.

$\mathbf{A} \in \mathbb{R}^{n \times n}$ is invertible (regular) if, and only if, $\text{rk}(\mathbf{A}) = n$. For all $\mathbf{A} \in \mathbb{R}^{m \times n}$, and all $\vec{b} \in \mathbb{R}^m$, the equation system $\mathbf{A}\vec{x} = \vec{b}$ can only be solved if $\text{rk}(\mathbf{A}) = \text{rk}(\mathbf{A} \mid \vec{b})$. The subspace of solutions for $\mathbf{A}\vec{x} = \vec{0}$ (**kernel / nullspace**) has dimension $n - \text{rk}(\mathbf{A})$. A matrix is said to have **full rank** if $\text{rk}(\mathbf{A}) = \min(m, n)$, otherwise it is said to be **rank deficient**.

For example, to calculate the rank of $\mathbf{A}$, as below, we apply Gaussian elimination to row echelon form, and look at the number of pivot columns (which corresponds to the linearly independent columns);

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \begin{matrix} \\ +2R_2 \\ -3R_2 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & -1 & -3 \end{bmatrix} \begin{matrix} \\ \\ +R_2 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

As we have two pivot columns, it follows that $\text{rk}(\mathbf{A}) = 2$.

Subspace Intersection

Suppose we have two subspaces $U_1 = [\vec{b_1}, \vec{b_2}, ..., \vec{b_k}] \subset V$, and $U_2 = [\vec{c_1}, \vec{c_2}, ..., \vec{c_l}] \subset V$. We're then interested in finding the basis of $U_1 \cap U_2$, which is all the $\vec{x} \in V$, such that $\vec{x} \in U_1 \wedge \vec{x} \in U_2$. For something to be in U_1, or U_2, it must be a linear combination of the basis vectors of the respective subspaces; hence we can get the following;

$$\sum_{i=1}^k \lambda_i \vec{b_i} = \vec{x} = \sum_{j=1}^l \mu_j \vec{c_j} \Leftrightarrow \sum_{i=1}^k \lambda_i \vec{b_i} - \sum_{j=1}^l \mu_j \vec{c_j} = \vec{0}$$

This gives a homogeneous equation system, where we want to solve;

$$\begin{bmatrix} \vec{b_1} \mid \vec{b_2} \mid \cdots \mid \vec{b_k} \mid -\vec{c_1} \mid -\vec{c_2} \mid \cdots \mid -\vec{c_l} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \mu_1 \\ \vdots \\ \mu_l \end{bmatrix} = \vec{0}$$

However, we only need to determine $\lambda_1, \lambda_2, ..., \lambda_k$, or $\mu_1, \mu_2, ..., \mu_l$, to find the intersection of U_1, and U_2.

Consider the example, with the following vector subspaces;

$$U_1 = [\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}] \subset \mathbb{R}^4, \text{ and also } U_2 = [\begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}] \subset \mathbb{R}^4$$

By generating the matrix shown above, we can now apply Gaussian elimination to reduced row echelon form;

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} \\ -R_1 \\ \\ \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & -1 \\ 0 & 1 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} \\ \\ -R_2 \\ \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & -1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} \\ \\ \\ -R_4 \end{matrix} \rightsquigarrow$$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{matrix} \\ \\ \leftrightarrow R_4 \\ \leftrightarrow R_3 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

From this, we can determine that $\mu_2 = 0$, and that μ_1 is a free variable, since it doesn't have a corresponding pivot column. Therefore, it follows that $\vec{x} = \mu_1 \vec{c_1}$, hence we've determined that the basis of the intersection is as follows;

$$U_1 \cap U_2 = [\begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}]$$

Linear Mappings

A linear mapping preserves the structure of vector spaces. Therefore, for two $\mathbb{R}$-vector spaces V, W, and a mapping Φ (note that it can also be called a **vector space homomorphism**);

$$\Phi : V \to W \text{ is } \textbf{linear} \Leftrightarrow \forall \lambda, \mu \in \mathbb{R} \forall \vec{x}, \vec{y} \in V [\Phi(\lambda \vec{x} + \mu \vec{y}) = \lambda \Phi(\vec{x}) + \mu \Phi(\vec{y})]$$

Special cases of this are;

- **isomorphism** $\hfill \Phi : V \to W$ is linear, and bijective

    finite dimensional $\mathbb{R}$-vector spaces V, and W are isomorph $\Leftrightarrow \dim(V) = \dim(W)$

- **endomorphism** $\hfill \Phi : V \to V$ is linear
- **automorphism** $\hfill \Phi : V \to V$ is linear, and bijective
- **identity mapping** $\hfill \text{id}_V : V \to V$ such that $\vec{x} \mapsto \vec{x}$

For example, consider the mapping $\Phi : \mathbb{R}^2 \to \mathbb{C}$, where $\Phi(\vec{x}) = x_1 + ix_2$. We can show that this is a linear mapping, by taking arbitrary $\vec{x}, \vec{y} \in \mathbb{R}^2$, and also arbitrary $\lambda \in \mathbb{R}$;

$$\begin{aligned} \Phi(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}) &= \Phi(\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}) \\ &= (x_1 + y_1) + i(x_2 + y_2) \\ &= x_1 + ix_2 + y_1 + iy_2 \\ &= \Phi(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) + \Phi(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}) \end{aligned} \qquad \begin{aligned} \Phi(\lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) &= \Phi(\begin{bmatrix} \lambda x_1 \\ \lambda x_2 \end{bmatrix}) \\ &= (\lambda x_1) + i(\lambda x_2) \\ &= \lambda(x_1 + ix_2) \\ &= \lambda \Phi(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) \end{aligned}$$

Image, and Kernel / Null Space

For the mapping $\Phi : V \to W$, we define the **kernel / null space**, and **image / range** respectively;

- $\ker(\Phi) \triangleq \{\vec{v} \in V \mid \Phi(\vec{v}) = \vec{0}_W\} = \Phi^{-1}(\{\vec{0}_W\})$

    this is the set of vectors in V that maps to the neutral element $\vec{0}_W$ in W

- $\mathrm{Im}(\Phi) \triangleq \{\vec{w} \in W \mid \exists \vec{v} \in V[\Phi(\vec{v}) = \vec{w}]\} = \Phi(V)$

    this is the set of vectors in W that can be reached by Φ from a vector in V

Note that $\Phi(\vec{0}_V) = \vec{0}_W$, hence the **unique** $\vec{0}_V \in \ker(\Phi)$, therefore the null space $\neq \varnothing$. By the definitions above, we can easily see that $\mathrm{Im}(\Phi) \subset W$, and also $\ker(\Phi) \subset V$, hence they are subspaces of their respective vector spaces. A mapping, Φ, is injective $\Leftrightarrow \ker(\Phi) = \{\vec{0}\}$.

A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ represents the mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, where $\Phi : \vec{x} \mapsto \mathbf{A}\vec{x}$. If we consider the column vectors of $\mathbf{A}$, such that we have $\vec{a_1}, \vec{a_2}, ..., \vec{a_n}$, we can apply the following;

$$\mathrm{Im}(\Phi) = \{\mathbf{A}\vec{x} \mid \vec{x} \in \mathbb{R}^n\} = \{\sum_{i=1}^{n} x_i \vec{a_i} \mid x_1, ..., x_n \in \mathbb{R}\} = [\vec{a_1}, ..., \vec{a_n}] \subset \mathbb{R}^m$$

Which tells us that the image is the span of the columns of $\mathbf{A}$.

Consider the following example, of a mapping represented by a matrix;

$$\Phi : \mathbb{R}^4 \mapsto \mathbb{R}^2, \text{ where } \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

By what we've stated above, we can simply get the image / range, by taking the column vectors;

$$\mathrm{Im}(\Phi) = [\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}]$$

To compute the kernel of Φ, we solve $\mathbf{A}\vec{x} = \vec{0}$, which can be done fairly easily with Gaussian elimination into reduced row echelon form, and then applying the Minus-1 Trick;

$$\begin{bmatrix} 1 & 2 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \leftrightarrow R_2 \\ \leftrightarrow R_1 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 2 & -1 & 0 \end{bmatrix} -R_1 \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & -1 & -1 \end{bmatrix} \cdot \tfrac{1}{2} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -\tfrac{1}{2} & -\tfrac{1}{2} \end{bmatrix}$$

Now that the matrix is in RREF, we can apply the Minus-1 Trick to compute the basis of the kernel;

$$\ker(\Phi) = [\begin{bmatrix} 0 \\ -\tfrac{1}{2} \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -\tfrac{1}{2} \\ 0 \\ -1 \end{bmatrix}]$$

For a linear mapping between two vector spaces; $\Phi : V \to W$, it holds $\dim(\ker(\Phi)) + \mathrm{Im}(\ker(\Phi)) = \dim(V)$. This is referred to as the **Rank-Nullity Theorem**.

The following properties also hold for $\mathbb{R}$-vector spaces V, W, X, and the linear mappings $\Phi : V \to W$, $\Psi : W \to X$, and $\Lambda : V \to W$;

- $\Psi \circ \Phi : V \to X$ is also a linear mapping
- if Φ is an isomorphism (bijective), then it follows that the inverse, $\Phi^{-1} : W \to V$, is also an isomorphism
- $\Phi + \Lambda$, and $\lambda\Phi$ are also linear, for $\lambda \in \mathbb{R}$

Matrix Representations

Any finite n-dimensional $\mathbb{R}$-vector space V, is isomorphic to another n-dimensional $\mathbb{R}$-vector space W, especially $\mathbb{R}^n$, if the dimensions match, such that $\dim(V) = \dim(W)$. This allows us to use $\mathbb{R}^n$ without loss of generality. If we define an **ordered** basis $[\vec{b_1}, \vec{b_2}, ..., \vec{b_n}]$ of V, we can construct such an isomorphism.

Suppose there is some $\mathbb{R}$-vector space V, with the ordered basis B, as denoted above. For $\vec{x} \in V$, we can obtain a unique linear combination, same as before;

$$\vec{x} = \sum_{i=1}^{n} \alpha_i \vec{b_i}$$

Therefore, we take the n-tuple given by α_i, to be the **coordinates** of $\vec{x}$ with respect to B, and have it written as a vector.

Consider $\mathbb{R}^2$, where the **standard** basis is $[\vec{e_1}, \vec{e_2}] = [\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}]$, and $\vec{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ with respect to $[\vec{e_1}, \vec{e_2}]$.

However, this same vector doesn't have to be represented with respect to the standard basis. If we were to define another ordered basis;

$$B = [\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}], \text{ then the same vector is } \begin{bmatrix} -\frac{1}{2} \\ \frac{5}{2} \end{bmatrix} \text{ with respect to the new basis}$$

For some n-dimensional vector space V, with ordered basis B, the mapping $\Phi : \mathbb{R}^n \to V$, with $\Phi(\vec{e_i}) = \vec{b_i}$ is linear, and also an isomorphism.

Suppose we have vector spaces V, and W with ordered bases $B = [\vec{b_1}, \vec{b_2}, ..., \vec{b_n}]$, and $C = [\vec{c_1}, \vec{c_2}, ..., \vec{c_m}]$ respectively. Then we can consider the representation of each of the basis vectors in B with respect to C; for $j \in [1, n]$;

$$\Phi(\vec{b_j}) = \sum_{i=1}^{m} \alpha_{i,j} \vec{c_i}, \text{ where we have } \mathbf{A}_\Phi \in \mathbb{R}^{m \times n} \triangleq (\alpha_{i,j})$$

The coordinates of $\Phi(\vec{b_j})$ with respect to the ordered basis C of W are the jth column of $\mathbf{A}_\Phi$. The rank of the transformation matrix, is the same as the dimension of the image of the transformation, such that $\mathrm{rk}(\mathbf{A}_\Phi) = \dim(\mathrm{Im}(\Phi))$.

If we have $\vec{\tilde{x}}$ as the coordinate vector of $\vec{x} \in V$, which is defined with respect to B, and $\vec{\tilde{y}}$ being the coordinate vector of $\vec{y} = \Phi(\vec{x}) \in W$ with respect to C, where $\Phi : V \to W$, then;

$$\vec{\tilde{y}} = \mathbf{A}_\Phi \vec{\tilde{x}}$$

Thus, we're able to map coordinates with respect to some ordered basis in V, to coordinates with respect to some other ordered basis in W, by using the transformation matrix we found.

Consider a homomorphism $\Phi : V \to W$, with ordered bases $B = [\vec{b_1}, \vec{b_2}, \vec{b_3}]$, and $C = [\vec{c_1}, \vec{c_2}, \vec{c_3}, \vec{c_4}]$ respectively.

$$\Phi(\vec{b_1}) = \vec{c_1} - \vec{c_2} + 3\vec{c_3} - \vec{c_4}$$
$$\Phi(\vec{b_2}) = 2\vec{c_1} + \vec{c_2} + 7\vec{c_3} + 2\vec{c_4}$$
$$\Phi(\vec{b_3}) = 3\vec{c_2} + \vec{c_3} + 4\vec{c_4}$$

We can then represent that as the transformation matrix $\mathbf{A}_\Phi$, which also gives us the coordinate vectors α_j of $\Phi(\vec{b_j})$, with respect to C.

$$\mathbf{A}_\Phi = [\vec{\alpha_1} \mid \vec{\alpha_2} \mid \vec{\alpha_3}] = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \\ 3 & 7 & 1 \\ -1 & 2 & 4 \end{bmatrix}$$

Suppose the coordinate vector $\vec{\hat{x}}$ representing $\vec{x} \in V$ with respect to B is;

$$\vec{\hat{x}} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1\vec{b_1} + x_2\vec{b_2} + x_3\vec{b_3}$$

Therefore, the coordinate vector $\vec{\hat{y}}$ representing $\Phi(\vec{x}) \in W$ with respect to C is;

$$
\begin{aligned}
\vec{\hat{y}} &= \Phi(\vec{\hat{x}}) \\
&= x_1\Phi(\vec{b_1}) + x_2\Phi(\vec{b_2}) + x_3\Phi(\vec{b_3}) \\
&= x_1(\vec{c_1} - \vec{c_2} + 3\vec{c_3} - \vec{c_4}) + x_2(2\vec{c_1} + \vec{c_2} + 7\vec{c_3} + 2\vec{c_4}) + x_3(3\vec{c_2} + \vec{c_3} + 4\vec{c_4}) \\
&= \mathbf{A}_\Phi\vec{\hat{x}}
\end{aligned}
$$

The above can be done since we know that Φ is a linear map, and that $\vec{\hat{x}}$ is a real valued vector.

Basis Change

Continuing on from the previous spaces V, and W, and the mapping $\Phi : V \to W$, suppose we have another ordered basis of V, and W, respectively;

$$\bar{B} = [\vec{\bar{b_1}}, \vec{\bar{b_2}}, ..., \vec{\bar{b_n}}], \text{ and also } \bar{C} = [\vec{\bar{c_1}}, \vec{\bar{c_2}}, ..., \vec{\bar{c_m}}]$$

Like before, because we know that B, and C are both bases, we can construct anything in $\bar{B}$, and $\bar{C}$ as a linear combination of vectors in their respective original basis;

$$\vec{\bar{b_j}} = \sum_{i=1}^{n} s_{i,j}\vec{b_i}, \text{ and also } \vec{\bar{c_k}} = \sum_{l=1}^{m} t_{l,k}\vec{c_l}, \text{ where } j \in [1, n], \text{ and } k \in [1, m]$$

Therefore, we can define regular matrices $\mathbf{S} \in \mathbb{R}^{n\times n} \triangleq (s_{i,j})$, and $\mathbf{T} \in \mathbb{R}^{m\times m} \triangleq (t_{l,k})$, which maps coordinates with respect to $\bar{B}$, onto coordinates with respect to B, and similar for $\bar{C}$ to C. Refer to **3Blue1Brown**'s *Change of basis | Essence of linear algebra, chapter 13*, for a more intuitive perspective on this.

Like before, the jth column of $\mathbf{S}$, and $\mathbf{T}$, are the coordinate representations of $\vec{\bar{b_j}}$, and $\vec{\bar{c_j}}$ with respect to B, and C, respectively.

Suppose we have some $\bar{\mathbf{A}}_\Phi$, which represents the mapping $\Phi : V \to W$, with bases $\bar{B}$, and $\bar{C}$ respectively. The elements of $\bar{\mathbf{A}}_\Phi$ are represented with $\bar{\alpha}$.

$$
\begin{aligned}
\Phi(\vec{\bar{b_j}}) &= \sum_{k=1}^{m} \bar{\alpha}_{k,j}\vec{\bar{c_k}} && \text{the matrix which we're trying to find} \\
&= \sum_{k=1}^{m} \bar{\alpha}_{k,j} \sum_{i=1}^{m} t_{i,k}\vec{c_i} && \text{by our definition of } \vec{\bar{c_k}} \\
&= \sum_{i=1}^{m}(\sum_{k=1}^{m} t_{i,k}\bar{\alpha}_{k,j})\vec{c_i} && \text{rearranging } \sum
\end{aligned}
$$

To solve for $\bar{\mathbf{A}}_\Phi$, we consider how we represent $\vec{\bar{b_j}}$ with respect to B, and how we represent each $\Phi(\vec{b_k})$;

$$
\begin{aligned}
\Phi(\vec{\bar{b_j}}) &= \Phi(\sum_{k=1}^{n} s_{k,j}\vec{b_k}) && \text{by our definition of } \vec{\bar{b_j}} \\
&= \sum_{k=1}^{n} s_{k,j}\Phi(\vec{b_k}) && \text{by linearity of } \Phi \\
&= \sum_{k=1}^{n} s_{k,j} \sum_{i=1}^{m} \alpha_{i,k}\vec{c_i} && \text{by the mapping of } \Phi \\
&= \sum_{i=1}^{m}(\sum_{k=1}^{n} \alpha_{i,k}s_{k,j})\vec{c_i} && \text{rearranging } \sum
\end{aligned}
$$

Looking at the final lines of the above equations, it follows that;

$$\Phi(\vec{b}_j) = \sum_{i=1}^{m}(\sum_{k=1}^{m} t_{i,k}\bar{\alpha}_{k,j})\vec{c}_i = \sum_{i=1}^{m}(\sum_{k=1}^{n} \alpha_{i,k}s_{k,j})\vec{c}_i \Leftrightarrow \sum_{k=1}^{m} t_{i,k}\bar{\alpha}_{k,j} = \sum_{k=1}^{n} \alpha_{i,k}s_{k,j}$$

Therefore, it follows that we have $\mathbf{T}\bar{\mathbf{A}}_\Phi = \mathbf{A}_\Phi \mathbf{S} \Leftrightarrow \bar{\mathbf{A}}_\Phi = \mathbf{T}^{-1}\mathbf{A}_\Phi \mathbf{S}$. Note that we can represent this as mapping composition; $\Phi_{\bar{C}\bar{B}} = \Phi_{\bar{C}C} \circ \Phi_{CB} \circ \Phi_{B\bar{B}} = \Phi_{C\bar{C}}^{-1} \circ \Phi_{CB} \circ \Phi_{B\bar{B}}$. The diagram below shows the mappings, and the respective transformation matrices in violet;

$$
\begin{array}{ccc}
\text{vector spaces} & V \xrightarrow{\quad\Phi\quad} & W \\[2mm]
\text{ordered bases} & B \xrightarrow{\quad\Phi_{CB}\quad} & C \\
& \Phi_{B\bar{B}}\Big\uparrow \mathbf{S} \quad \mathbf{A}_\Phi \qquad \mathbf{T}^{-1}\Big\downarrow \Phi_{\bar{C}C} = \Phi_{C\bar{C}}^{-1} \\
& \bar{B} \xrightarrow{\quad\Phi_{\bar{C}\bar{B}}\quad} & \bar{C}
\end{array}
$$

We say that two matrices, $\mathbf{A}, \bar{\mathbf{A}} \in \mathbb{R}^{m \times n}$ are equivalent if there exists regular matrices $\mathbf{S} \in \mathbb{R}^{n \times n}$, and $\mathbf{T} \in \mathbb{R}^{m \times m}$, such that $\bar{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{S}$. We say they are similar if $\mathbf{A}, \bar{\mathbf{A}} \in \mathbb{R}^{n \times n}$, has a corresponding regular matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$, such that $\bar{\mathbf{A}} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$. Therefore, it follows that similarity is a stronger property than equivalence, since all similar matrices are equivalent, but not vice versa.

With $\mathbb{R}$-vector spaces V, W, X, mappings $\Phi : V \to W$, $\Psi : W \to X$; we define the composed mapping $\Psi \circ \Phi : V \to X$, to have the transformation matrix $\mathbf{A}_{\Psi \circ \Phi} = \mathbf{A}_\Psi \mathbf{A}_\Phi$

Determinants

Determinants are only defined for square matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$, and will be denoted as $\det(\mathbf{A})$, or $|\mathbf{A}|$. There are also some properties about determinants of smaller matrices, but I'm skipping that.

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ 0 & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{n,n} \end{bmatrix} \text{ is an upper triangular matrix, therefore } \det(\mathbf{A}) = \prod_{i=1}^{n} a_{i,i}$$

We can also apply the following properties of determinants, without proof;

- $\det(\mathbf{AB}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$
- $\det(\mathbf{A}) = 0 \Leftrightarrow \mathbf{A}$ is singular (hence $\det(\mathbf{A}) \neq 0 \Leftrightarrow$ if it is regular)
- $\det(\mathbf{A}) = \det(\mathbf{A}^\top)$
- if $\mathbf{A}$ is regular; $\det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}$
- similar matrices have the same determinant; therefore all $\mathbf{A}_\Phi$ of $\Phi : V \to V$ have the same determinant

In order to get the determinant of a matrix by using the triangular form, we can use Gaussian elimination, with the following row operation rules;

- adding a multiple of a column/row to another doesn't change $\det(\mathbf{A})$
- multiplying a column/row by $\lambda \in \mathbb{R}$ scales $\det(\mathbf{A})$ by λ, hence $\det(\lambda \mathbf{A}) = \lambda^n \det(\mathbf{A})$
- swapping two columns/rows changes the sign of $\det(\mathbf{A})$

For example, consider the following sequence of row operations;

$$
\begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 2 & -1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ -2 & 0 & 2 & -1 & 2 \\ 2 & 0 & 0 & 1 & 1 \end{vmatrix} \begin{matrix} \\ -R_1 \\ \\ +R_1 \\ -R_1 \end{matrix} \rightsquigarrow \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{vmatrix} \begin{matrix} \\ \\ +R_1 \\ \\ \end{matrix} \rightsquigarrow \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & -1 & 1 \end{vmatrix} \begin{matrix} \\ \\ \\ -3R_3 \\ +R_3 \end{matrix} \rightsquigarrow
$$

$$
\begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & -1 & 4 \end{vmatrix} \begin{matrix} \\ \\ \\ \\ +R_4 \end{matrix} \rightsquigarrow \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 & -3 \end{vmatrix} = \quad 2 \cdot -1 \cdot 1 \cdot 1 \cdot -3 = 6
$$

Another method for calculating the determinant of a matrix is via **Laplace Expansion**. We define a matrix $\mathbf{A} \in \mathbb{R}^{n \times n} = (a_{i,j})$, and another matrix $\mathbf{A}_{i,j} \in \mathbb{R}^{(n-1) \times (n-1)}$ as $\mathbf{A}$ without its ith row, nor its jth column.

- $\det(\mathbf{A}) = \sum\limits_{k=1}^{n} (-1)^{k+j} a_{i,k} \det(\mathbf{A}_{i,k})$ \hspace{2cm} expansion along row i

- $\det(\mathbf{A}) = \sum\limits_{k=1}^{n} (-1)^{k+j} a_{k,j} \det(\mathbf{A}_{k,j})$ \hspace{2cm} expansion along column j

The chequerboard pattern formed by the signs are as follows, for $\mathbf{A} \in \mathbb{R}^{n \times n}$;

$$
n \text{ is odd}; \begin{bmatrix} + & - & \cdots & + \\ - & + & \cdots & - \\ \vdots & \vdots & \ddots & \vdots \\ + & - & \cdots & + \end{bmatrix}, \text{ or when } n \text{ is even}; \begin{bmatrix} + & - & \cdots & - \\ - & + & \cdots & + \\ \vdots & \vdots & \ddots & \vdots \\ - & + & \cdots & + \end{bmatrix}
$$

It's better to expand along rows/columns with as many 0s as possible, as that would reduce the amount of work that needs to be done.

Eigenvalues, and Eigenvectors

We define λ as an eigenvalue for an endomorphism $\Phi : V \to V$, $\Leftrightarrow \exists \vec{v} \in V \vec{0}[\vec{v} \neq \vec{0} \wedge \Phi(\vec{v}) = \lambda \vec{v}]$. The set of all eigenvectors corresponding to some λ in $\mathbf{A}_\Phi$ is defined as the **eigenspace** of $\mathbf{A}_\Phi$, with respect to λ, and is written as E_λ. The set of all eigenvalues of $\mathbf{A}_\Phi$ is called the **spectrum** of $\mathbf{A}_\Phi$.

We can also say that λ is an eigenvalue of $\mathbf{A}$, when $\text{rk}(\mathbf{A} - \lambda \mathbf{I}_n) < n$, hence $\det(\mathbf{A} - \lambda \mathbf{I}_n) = 0$.

From the latter definition, we can define the characteristic polynomial $p(\lambda) \triangleq \det(\mathbf{A} - \lambda \mathbf{I}_n)$. By setting $p(\lambda) = 0$, the roots of our equations are the eigenvalues of $\mathbf{A}$.

To determine the eigenspace of $\mathbf{A}_\Phi$, with eigenvalue λ, we take $\ker(\mathbf{A}_\Phi - \lambda \mathbf{I}_n)$. Since we defined an eigenvalue to satisfy $\mathbf{A}_\Phi \vec{x} = \lambda \vec{x} \Leftrightarrow (\mathbf{A}_\Phi - \lambda \mathbf{I}_n)\vec{x} = \vec{0}$, so we're simply solving a homogeneous system of equations.

We can say that two similar matrices $\mathbf{A}, \bar{\mathbf{A}} \in \mathbb{R}^{n \times n}$, have the same characteristic polynomial;

$$
\begin{aligned}
\det(\bar{\mathbf{A}} - \lambda \mathbf{I}_n) &= \det(\mathbf{S}^{-1}\mathbf{A}\mathbf{S} - \lambda \mathbf{I}_n) && \text{by definition of similar matrices} \\
&= \det(\mathbf{S}^{-1}\mathbf{A}\mathbf{S} - \lambda \mathbf{S}^{-1}\mathbf{I}_n\mathbf{S}) && \mathbf{S}^{-1}\mathbf{S} = \mathbf{I}_n \\
&= \det(\mathbf{S}^{-1}(\mathbf{A} - \lambda \mathbf{I}_n)\mathbf{S}) && \text{distributivity of matrix multiplication} \\
&= \det(\mathbf{S}^{-1}) \cdot \det(\mathbf{A} - \lambda \mathbf{I}_n) \cdot \det(\mathbf{S}) && \text{property of determinants} \\
&= \det(\mathbf{S})^{-1} \cdot \det(\mathbf{A} - \lambda \mathbf{I}_n) \cdot \det(\mathbf{S}) && \text{property of determinants} \\
&= \det(\mathbf{A} - \lambda \mathbf{I}_n)
\end{aligned}
$$

There are some worked examples in the notes, but this is something that **needs** to be practised. The

concept itself isn't too difficult, but being able to apply it quickly needs some practice.

Diagonalization

The Panopto recording for this lecture wasn't very legible, so it is mostly based from the notes, and some intuition from **3Blue1Brown**.

As an "exercise"; consider taking powers of $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$, diagonalize it, and find the nth power.

$\mathbf{A} \in \mathbb{R}^{n \times n}$ is diagonalizable when it is similar to a **diagonal** matrix. We say a matrix is diagonal if it is in the following form, such that anything **not** on the leading diagonal is 0;

$$\begin{bmatrix} c_1 & 0 & \cdots & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & c_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & c_n \end{bmatrix}$$

When we have an endomorphism Φ of an n-dimensional $\mathbb{R}$-vector space V, if Φ is diagonalizable, it follows that the transformation matrix $\mathbf{A}_\Phi$ is also diagonalizable, and there is a basis in V consisting of the eigenvectors (hence there must be n eigenvectors; the sum of the dimensions of the eigenspaces is n). Φ is only diagonalizable; if the following conditions hold;

- it's characteristic polynomial is in the form $p(\lambda) = (-1)^n (\lambda - c_1)^{r_1} \cdots (\lambda - c_k)^{r_k}$; $r_i \in \mathbb{N}$, $c_i \in \mathbb{R}$

    the characteristic equation decomposes into linear factors

- $i \in [1, k]$ - $\dim(\text{Im}(\Phi - c_i \text{id}_V)) = n - r_i \Leftrightarrow \dim(\ker(\Phi - c_i \text{id}_V)) = r_i$ rank-nullity

    the dimension of the eigenspace E_{c_i} must correspond to the algebraic multiplicity (how many times an eigenvalue appears in the characteristic polynomial) r_i of the eigenvalue in $p(\lambda)$

    each eigenvalue c_i will appear r_i times on the diagonal

Suppose we have a diagonalizable matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, and the ordered eigenbasis $[\vec{b_1}, ..., \vec{b_n}]$ of $\mathbf{A}$, with $\mathbf{A}\vec{b_i} = c_i \vec{b_i}$, with $i \in [1, n]$ then;

$$\mathbf{S} = \begin{bmatrix} \vec{b_1} & | & \cdots & | & \vec{b_n} \end{bmatrix}, \text{ and } \mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{D} = \begin{bmatrix} c_1 & 0 & \cdots & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & c_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & c_n \end{bmatrix}$$

Therefore, $\mathbf{D}$ is the transformation matrix of $\vec{x} \mapsto \mathbf{A}\vec{x}$, with respect to the ordered eigenbasis.

Consider the following example;

$$\mathbf{A} = \begin{bmatrix} 3 & 2 & -1 \\ 2 & 6 & -2 \\ 0 & 0 & 2 \end{bmatrix}$$

To check if it's diagonalizable, we first find the characteristic polynomial $p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}_3)$;

$$
\begin{aligned}
p(\lambda) &= \det(\mathbf{A} - \lambda \mathbf{I}_3) \\
&= \begin{vmatrix} 3 - \lambda & 2 & -1 \\ 2 & 6 - \lambda & -2 \\ 0 & 0 & 2 - \lambda \end{vmatrix} \\
&= (2 - \lambda) \begin{vmatrix} 3 - \lambda & 2 \\ 2 & 6 - \lambda \end{vmatrix} \\
&= (2 - \lambda)[(6 - \lambda)(3 - \lambda) - 4] \\
&= (2 - \lambda)(\lambda^2 - 9\lambda + 14) \\
&= (2 - \lambda)(\lambda - 7)(\lambda - 2) \\
&= (-1)^3 (\lambda - 2)^2 (\lambda - 7)
\end{aligned}
$$

The eigenspace E_2 can be computed by doing Gaussian elimination on $\mathbf{A} - 2\mathbf{I}_3$;

$$
\begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} \\ -2R_1 \\ \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
$$

By the Minus-1 Trick, we get $E_2 = [\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}]$

Likewise, the eigenspace E_7 can be computed by doing another Gaussian elimination on $\mathbf{A} - 7\mathbf{I}_3$;

$$
\begin{bmatrix} -4 & 2 & -1 \\ 2 & -1 & -2 \\ 0 & 0 & -5 \end{bmatrix} \begin{matrix} \\ +\frac{1}{2}R_1 \\ \cdot -\frac{1}{5} \end{matrix} \rightsquigarrow \begin{bmatrix} -4 & 2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{matrix} \cdot -\frac{1}{4} \\ \leftrightarrow R_3 \\ \leftrightarrow R_2 \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & -\frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} -\frac{1}{4}R_2 \\ \\ \end{matrix} \rightsquigarrow \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}
$$

By the Minus-1 Trick (and also multiplying to get rid of the fraction), we get $E_7 = [\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}]$

Looking at the characteristic polynomial, it's clear that the algebraic multiplicities matches the dimensions of the eigenspaces, therefore we can combine the spectrum as an ordered eigenbasis;

$$
\mathbf{S} = \begin{bmatrix} -2 & 1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}, \text{ and } \mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{bmatrix}
$$

Putting a matrix into this form has a great benefit when computing powers; as we are able to say $\mathbf{A}^k = (\mathbf{SDS}^{-1})^k = \mathbf{SD}^k\mathbf{S}^{-1}$.

Cayley-Hamilton Theorem

The **Cayley-Hamilton Theorem** states that the endomorphism $\Phi : V \to V$, with a transformation matrix $\mathbf{A}_\Phi$, on n-dimensional $\mathbb{R}$-vector space V, and the characteristic polynomial p, then $p(\Phi) = 0$ (the zero mapping), and $p(\mathbf{A}_\Phi) = \mathbf{0}$ (the zero matrix). The proof for this is non-trivial (pretty much the only time you'll hear that in this degree).

Consider the matrix $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$, which has the characteristic polynomial $p(\lambda) = \lambda^2 - 2\lambda + 3$ therefore;

$$
\begin{aligned}
& \mathbf{A}^2 - 2\mathbf{A} + 3\mathbf{I}_2 = \mathbf{0} \\
\Leftrightarrow{} & -\mathbf{A}^2 + 2\mathbf{A} = 3\mathbf{I}_2 \\
\Leftrightarrow{} & \mathbf{A}\tfrac{1}{3}(-\mathbf{A} + 2\mathbf{I}_2) = \mathbf{I}_2 \\
\Leftrightarrow{} & \mathbf{A}^{-1} = \tfrac{1}{3}(2\mathbf{I}_2 - \mathbf{A})
\end{aligned}
$$

Inner Products

Inner products introduce rigour into intuitive geometric properties.

For example, the n-dimensional scalar product is defined for $\vec{x}, \vec{y} \in \mathbb{R}^n$, as $\vec{x}^\top \vec{y} = \sum_{i=1}^n x_i y_i$.

However, generally, suppose we have a vector space V, and a bilinear mapping $\beta : V \times V \to \mathbb{R}$ (both arguments are linear);

- β is **symmetric** $\hfill \forall \vec{x}, \vec{y} \in V [\beta(\vec{x}, \vec{y}) = \beta(\vec{y}, \vec{x})]$
- β is **positive definite** $\hfill \forall \vec{x} \in V[\vec{x} \neq \vec{0} \Rightarrow \beta(\vec{x}, \vec{x}) > 0], \text{ and } \beta(\vec{0}, \vec{0}) = 0$
- if it is both, then it is defined as an **inner product** on V, and written as $\langle \vec{x}, \vec{y} \rangle$
- $(V, \langle \cdot, \cdot \rangle)$ is called an **inner product space**, amongst other things - if it's the dot / scalar product, then we call it a **Euclidean vector space**

Suppose we have some inner product space $(V, \langle \cdot, \cdot \rangle)$, then $||\vec{x}|| \triangleq \sqrt{\langle \vec{x}, \vec{x} \rangle}$ is the **length / norm** of $\vec{x} \in V$. The mapping is $|| \cdot || : V \to \mathbb{R}$, such that $|| \cdot || : \vec{x} \mapsto ||\vec{x}||$. When this is in an Euclidean vector space, the norm is the geometric length of the vector. The norm has the following properties;

- $\forall \vec{x} \in V [||\vec{x}|| \geq 0]$, and $||\vec{x}|| = 0 \Leftrightarrow \vec{x} = \vec{0}$
- $\forall \vec{x} \in V \forall \lambda \in \mathbb{R} [||\lambda \vec{x}|| = |\lambda| \cdot ||\vec{x}||]$
- $\forall \vec{x}, \vec{y} \in V [||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||]$ $\hfill$ **Minkowski inequality**

With the same inner product space, we define $d(\vec{x}, \vec{y}) \triangleq ||\vec{x} - \vec{y}||$, as the **distance** of $\vec{x}, \vec{y} \in V$, with the mapping $d : V \times V \to \mathbb{R}$, and $d : (\vec{x}, \vec{y}) \mapsto d(\vec{x}, \vec{y})$. The mapping is called **metric**. A metric d is positive definite, symmetric, and satisfies the triangular inequality, such that $d(\vec{x}, \vec{z}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$.

We define two vectors $\vec{x}, \vec{y}$ as **orthogonal** if $\langle \vec{x}, \vec{y} \rangle = 0$, and is often denoted as $\vec{x} \perp \vec{y}$.

With an inner product space $(V, \langle \cdot, \cdot \rangle)$, and $\vec{x}, \vec{y}, \vec{z} \in V$, we can use the following properties;

1. $|\langle \vec{x}, \vec{y} \rangle| \leq ||\vec{x}|| \cdot ||\vec{y}||$ $\hfill$ Cauchy-Schwarz inequality
2. $||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||$ $\hfill$ Minkowski inequality
3. $d(\vec{x}, \vec{z}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$ $\hfill$ triangular inequality
4. $||\vec{x} + \vec{y}|| + ||\vec{x} - \vec{y}|| = 2||\vec{x}||^2 + 2||\vec{y}||^2$ $\hfill$ parallelogram law
5. $4\langle \vec{x}, \vec{y} \rangle = ||\vec{x} + \vec{y}||^2 - ||\vec{x} - \vec{y}||^2$
6. $\vec{x} \perp \vec{y} \Leftrightarrow ||\vec{x} + \vec{y}||^2 = ||\vec{x}||^2 + ||\vec{y}||^2$

By property 1, we can define $\theta \in [0, \pi)$, as long as $\vec{x} \neq \vec{0}$, and $\vec{y} \neq \vec{0}$;

$$-1 \leq \frac{\langle \vec{x}, \vec{y} \rangle}{||\vec{x}|| \cdot ||\vec{y}||} \leq 1, \text{ and therefore } \cos(\theta) = \frac{\langle \vec{x}, \vec{y} \rangle}{||\vec{x}|| \cdot ||\vec{y}||}$$

Orthogonal Projections

Suppose we have an $\mathbb{R}$-vector space V, and $W \subset V$, we can define a linear mapping $\pi : V \to W$ as a **projection** if $\pi^2 = \pi \circ \pi = \pi$. As π is a linear map, we can define a **projection matrix**, $\mathbf{P}_\pi$, which has the property $\mathbf{P}_\pi = \mathbf{P}_\pi^2$. This causes all eigenvalues of $\mathbf{P}_\pi$ to be 0, or 1, and therefore $E_0 = \ker(\mathbf{P}_\pi)$, and $E_1 = \text{Im}(\mathbf{P}_\pi)$. To derive projections in $(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$, we assume the scalar product $\langle \vec{x}, \vec{y} \rangle = \vec{x}^\top \vec{y}$

For the 1-dimensional case (projection to a line), let there be a subspace $W \subset \mathbb{R}^n$, spanned by only $\vec{b}$. When we project some n-dimensional vector $\vec{x}$ to W, we write $\vec{p} = \pi_W(\vec{x})$. Geometrically, the projection point $\vec{p}$ is closest to $\vec{x}$, such that we minimise $||\vec{x} - \vec{p}||$ - therefore $\vec{p} - \vec{x}$ is orthogonal to W, and therefore to the basis of W. In one dimension, this is fairly easy to solve, as we simply have $\langle \vec{p} - \vec{x}, \vec{b} \rangle = (\vec{p} - \vec{x})^\top \vec{b} = 0$.

However, we know that $\vec{p}$ must lie on the span of W, and with a single basis vector, we have $\vec{p} = \lambda \vec{b}$, where $\lambda \in \mathbb{R}$.

$$\vec{x} - \lambda \vec{b} \perp \vec{b}$$
$$\Leftrightarrow \langle \vec{x} - \lambda \vec{b}, \vec{b} \rangle = 0$$
$$\Leftrightarrow \langle \vec{x}, \vec{b} \rangle - \lambda \langle \vec{b}, \vec{b} \rangle = 0$$
$$\Leftrightarrow \lambda = \frac{\langle \vec{x}, \vec{b} \rangle}{\langle \vec{b}, \vec{b} \rangle} = \frac{\langle \vec{x}, \vec{b} \rangle}{||\vec{b}||^2}$$

By using the dot product, we get $\vec{p} = \frac{\vec{b}^\top \vec{x}}{||\vec{b}||^2} \vec{b}$, however, knowing that $\vec{p} = \mathbf{P}_\pi \vec{x}$, we get $\mathbf{P}_\pi = \frac{\vec{b} \vec{b}^\top}{||\vec{b}||^2}$

However, we want to generalise this to more than one dimension. We're still working with $\vec{x} \in \mathbb{R}^n$, but instead we're mapping to a general subspace $W \subset \mathbb{R}^n$, which is spanned by the ordered basis;

$$\left[\vec{b_1}, \vec{b_2}, ..., \vec{b_m} \right], \text{ which can be expressed as a matrix } \mathbf{B} = \left[\vec{b_1} \mid \cdots \mid \vec{b_m} \right]$$

We can then express the projection in the following form;

$$\pi_W(\vec{x}) = \vec{p} \in W \Rightarrow \vec{p} = \sum_{i=1}^m \lambda_i \vec{b_i} = \mathbf{B} \vec{\lambda}$$

Since we want the minimal distance, the line between $\vec{p} \in W$, and $\vec{x} \in \mathbb{R}^n$, must be orthogonal to all basis vectors; and we get the following homogeneous system of equations (assuming we're in an Euclidean vector space);

$$\forall i \in [1, m][\langle \vec{b_i}, \vec{x} - \mathbf{B} \vec{\lambda} \rangle = \vec{b_i}^\top (\vec{x} - \mathbf{B} \vec{\lambda}) = 0]$$
$$\leadsto$$
$$\begin{bmatrix} \vec{b_1}^\top \\ \vdots \\ \vec{b_1}^\top \end{bmatrix} (\vec{x} - \mathbf{A} \vec{\lambda}) = \vec{0} \Leftrightarrow \mathbf{B}^\top (\vec{x} - \mathbf{A} \vec{\lambda}) = \vec{0} \Leftrightarrow \mathbf{B}^\top \mathbf{B} \vec{\lambda} = \mathbf{B}^\top \vec{x}$$

However, since we know that the rows / columns of $\mathbf{B}$ are linearly independent (since they were constructed from basis vectors), we can invert the matrix, such that $\vec{\lambda} = \underbrace{(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top}_{\text{pseudo-inverse of } \mathbf{B}} \vec{x}$.

Since we've established $\vec{p} = \mathbf{B} \vec{\lambda}$, it follows that $\pi_W(\vec{x}) = \vec{p} = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \vec{x}$. We also get the projection matrix $\mathbf{P}_\pi = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top$.