QUICK NOTES ON *CALCULUS* BY TOM APOSTOL.

ERNEST YEUNG - PRAHA 10, ČESKÀ REPUBLIKA

1. One-dimensional Continuity and Differentiation.

These theorems form the foundation for continuity and will be valuable for differentiation later.

Theorem 1 (Bolzano's Theorem)**.**
Let f be cont. at $\forall x \in [a,b]$.
Assume $f(a),\ f(b)$ have opposite signs.
Then $\exists$ at least one $c \in (a,b)$ s.t. $f(c) = 0$.

Proof. Let $f(a) < 0,\ f(b) > 0$.
Want: Find one value $c \in (a,b)$ s.t. $f(c) = 0$
Strategy: find the largest c.
Let $S = \{$ all $x \in [a,b]$ s.t. $f(x) \leq 0$ $\}$.
S is nonempty since $f(a) < 0$. S is bounded since all $S \subseteq [a,b]$.
$\Longrightarrow S$ has a suprenum.
Let $c = supS$.
If $f(c) > 0$, $\exists (c-\delta, c+\delta)$ s.t. $f > 0$
$c - \delta$ is an upper bound on S
but c is a least upper bound on S. Contradiction.
If $f(c) < 0$, $\exists (c-\delta, c+\delta)$ s.t. $f < 0$
$c + \delta$ is an upper bound on S
but c is an upper bound on S. Contradiction. □

Theorem 2 (Sign-preserving Property of Continuous functions)**.**
Let f be cont. at c and suppose that $f(c) \neq 0$.
then $\exists (c-\delta, c+\delta)$ s.t. f be on $(c-\delta, c+\delta)$ has the same sign as $f(c)$.

Proof. Suppose $f(c) > 0$.
$\forall \epsilon > 0$, $\exists \delta > 0$ s.t. $f(c) - \epsilon < f(x) < f(c) + \epsilon$ if $c - \delta < x < c + \delta$ (by continuity).
Choose δ for $\epsilon = \frac{f(c)}{2}$. Then

$$\frac{f(c)}{2} < f(x) < \frac{3f(c)}{2} \quad \forall x \in (c-\delta, c+\delta)$$

Then f has the same sign as $f(c)$. □

Theorem 3 (Intermediate value theorem)**.**
Let f be cont. at each pt. on $[a,b]$.
Choose any $x_1, x_2 \in [a,b]$ s.t. $x_1 < x_2$. s.t. $f(x_1) \neq f(x_2)$.
Then f takes on every value between $f(x_1)$ and $f(x_2)$ somewhere in (x_1, x_2).

Proof. Suppose $f(x_1) < f(x_2)$
Let k be any value between $f(x_1)$ and $f(x_2)$
Let $g = f - k$

$$g(x_1) = f(x_1) - k < 0$$
$$g(x_2) = f(x_2) - k > 0$$

By Bolzano, $\exists c \in (x_1, x_2)$ s.t. $g(c) = 0 \quad \Longrightarrow f(c) = k$ □

Theorem 4 (Mean Value Theorem for Integrals)**.** *Let f be cont. on $[a,b]$. Then $\exists\, c \in (a,b)$ s.t.*

$$f(c) = \frac{1}{b-a}\int_a^b f(x)dx$$

Proof. Let $\begin{matrix} M = \max f \\ m = \min f \end{matrix}$ on $[a, b]$.

Then $m \leq f(x) \leq M$ or $m \leq \frac{1}{b-a}\int_a^b f(x) = A(f) \leq M$.
By intermediate value theorem, $A(f) = f(c)$, for some c in $[a, b]$. □

Theorem 5 (Second Mean Value Theorem)**.** *If f, g cont. on $[a, b]$, g never changes sign in $[a, b]$,then for some $c \in [a, b]$,*

$$\int_a^b f(x)g(x)dx = f(c)\int_a^b g(x)dx$$

Proof. Assume $g \geq 0$ on $[a, b]$. Suppose $m \leq f \leq M$ on $[a, b]$.

$$mg \leq fg \leq M \text{ or } m\int_a^b g \leq \int_a^b fg \leq M\int_a^b g$$

If $\int_a^b g = 0$, equality, done. Else if $\int_a^b g > 0$,

$$m \leq \int_a^b fg / \int_a^b g \leq M$$

Apply intermediate value theorem $\implies f(c) = \int_a^b fg / \int_a^b g$. □

Theorem 6 (Theorem 4.3)**.**
Let f be defined on I.
Assume f has a rel. extrema at an int. pt. $c \in I$.
If $\exists\, f'(c),\ f'(c) = 0$; the converse is not true.

Proof. $Q(x) = \frac{f(x)-f(c)}{x-c}$ if $x \neq c$, $\quad Q(c) = f'(c)$
$\exists f'(c)$, so $Q(x) \to Q(c)$ as $x \to c$ so Q is continuous at c.
If $Q(c) > 0$, $\quad \frac{f(x)-f(c)}{x-c} > 0$. For $x - c \gtreqless 0$, $\quad f(x) \gtreqless f(c)$, thus contradicting the rel. max or rel. min. (no neighborhood about c exists for one!)
If $Q(c) < 0$, $\quad \frac{f(x)-f(c)}{x-c} < 0$. For $x - c \gtreqless 0$, $\quad f(x) \lesseqgtr f(c)$, thus contradicting the rel. max or rel. min. (no neighborhood about c exists for one!)
Converse is not true: e.g. saddle points. □

Theorem 7 (Rolle's Theorem)**.**
Let f be cont. on $[a, b]$, $\quad \exists f'(x) \quad \forall x \in (a, b)$ and let

$$f(a) = f(b)$$

then $\exists$ at least one $c \in (a, b)$, such that $f'(c) = 0$.

Proof. Suppose $f'(x) \neq 0 \quad \forall x \in (a, b)$.
By extreme value theorem, $\exists$ abs. max (min) M, m somewhere on $[a, b]$.
M, m on endpoints a, b (Thm 4.3).
$F(a) = f(b)$, so $m = M$. f constant on $[a, b]$. Contradict $f'(x) \neq 0$ □

Theorem 8 (Mean-value theorem for Derivatives)**.** *Assume f is cont. everywhere on $[a, b]$, $\exists f'(x) \quad \forall x \in (a, b)$.*
$\exists$ at least one $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a) \tag{1}$$

Proof.

$$\begin{aligned} h(x) &= f(x)(b-a) - x(f(b) - f(a)) \\ h(a) &= f(a)b - f(a)a - af(b) + af(a) \\ h(b) &= f(b)(b-a) - b(f(b) - f(a)) = bf(a) - af(b) = h(a) \\ \implies \exists c &\in (a, b), \text{ such that } h'(c) = 0 = f'(c)(b-a) - (f(b) - f(a)) \end{aligned}$$

□

Theorem 9 (Cauchy's Mean-Value Formula)**.** *Let f, g cont. on $[a, b]$, $\exists f'$, g' $\quad \forall x \in (a, b)$*
Then $\exists\, c \in (a, b)$. x

$$f'(c)(g(b) - g(a)) = g'(c)(f(b) - f(a)) \quad \textit{(note how it's symmetrical)} \tag{2}$$

Proof.

$$\begin{aligned} h(x) &= f(x)(g(b) - g(a)) - g(x)(f(b) - f(a)) \\ h(a) &= f(a)(g(b) - g(a)) - g(a)(f(b) - f(a)) = f(a)g(b) - g(a)f(b) \\ h(b) &= f(b)(g(b) - g(a)) - g(b)(f(b) - f(a)) \\ \implies h'(c) &= f'(c)(g(b) - g(a)) - g'(c)(f(b) - f(a)) = 0 \quad \text{(by Rolle's Thm.)} \end{aligned}$$

□

2. Polynomial Approximations to functions

2.1. The behavior of $\log x$ and e^x for large x.

Theorem 10. *If $a, b > 0$,*

$$\lim_{x \to +\infty} \frac{(\log x)^b}{x^a} = 0 \tag{3}$$

$$\lim_{x \to +\infty} \frac{x^b}{e^{ax}} = 0 \tag{4}$$

Proof. If $c > 0$, $t \geq 1$, then $t^{-1} \leq t^{c-1}$.

Hence, if $x > 1$,

$$0 < \log x = \int_1^x \frac{1}{t} dt \leq \int_1^x t^{c-1} dt = \frac{x^c - 1}{c} < \frac{x^c}{c}$$

$$\frac{(\log x)^b}{x^a} < \frac{x^{bc-a}}{c^b}$$

Let $c = \frac{a}{2b}$.

Since $x^{bc-a} = x^{-a/2} \to 0$ as $x \to \infty$,

$$\frac{(\log x)^b}{x^a} \to 0$$

For $\frac{x^b}{e^{ax}}$,

Let $t = e^x$, $x = \ln t$,

Then $\frac{x^b}{e^{ax}} = \frac{(\ln t)^b}{t^a}$

but $x \to \infty$ as $t \to \infty$, so $\lim_{x \to \infty} \frac{x^b}{e^{ax}} = 0$ □

3. Series

It's important to make the following distinction:

$$\text{If } C \text{ is satisfied, then } \sum a_n \text{ converges} \qquad \text{(sufficient)}$$

$$\text{If } \sum a_n \text{ converges, then } C \text{ is satisfied}$$

$$\sum a_n \text{ converges if and only if } C \text{ is satisfied}$$

Note that the statements "If P, then Q" and "If $(\text{not}Q)$, then $(\text{not}P)$" are logically equivalent.

Theorem 11 (Divergence Test by nth term, or nth term, necessary condition for convergence)**.** *If $\sum a_n$ converges,* $\lim_{n \to \infty} a_n = 0$.

Note the contrapositive: If not $\lim_{n \to \infty} a_n = 0$*, then $\sum a_n$ diverges.*

This theorem is useful for testing divergence through its contrapositive.

Proof. $s_n = \sum_{j=1}^n a_j$, $\qquad a_n = s_n - s_{n-1}$,
since $\lim_{n \to \infty} s_n = \lim_{n \to \infty} s_{n-1}$, $\quad \lim_{n \to \infty} a_n = 0$ □

Theorem 12 (Integral Test)**.**
Let f be a positive decreasing function, defined for all real $x \geq 1$.
For $\forall n \geq 1$, let $s_n = \sum_{k=1}^{n} f(k)$ and $t_n = \int_1^n f(x)dx$.
Then both sequences $\{s_n\}$ and $\{t_n\}$ converge or both diverge.

Using the geometric series $\sum x^n$ as a comparison series, Cauchy developed two useful tests known as the root test and the ratio test.

Theorem 13 (Ratio test)**.** *Let $\sum a_n$ be a series of positive terms such that*

$$\frac{a_{n+1}}{a_n} \to L \text{ as } n \to \infty$$

(1) *If $L < 1$, the series converges.*
(2) *If $L > 1$, the series diverges.*
(3) *If $L = 1$, the test is inconclusive.*

Proof.
If $L < 1$

Consider $L < x < 1$

$x - L > \frac{a_{j+1}}{a_j} - L$ since $\lim_{j\to\infty} \frac{a_{j+1}}{a_j} = L$

$\forall j \geq N$ for some $N = N(x - L) > 0$

$$x > \frac{a_{j+1}}{a_j} \to \frac{a_{j+1}}{x^{j+1}} < \frac{a_j}{x^j}$$

So then $\frac{a_j}{x^j}$ is decreasing with j. Then $\frac{a_j}{x_j} \leq \frac{a_N}{x^N} = C$ or $a_j \leq Cx^j$ for $j \geq N$. By comparison test, $\sum a_j$ converges. □

Theorem 14 (Leibniz's Rule)**.** *If a_j is a monotonically decreasing sequence with limit 0, $\sum_{j=1}^{\infty}(-1)^{j-1}a_j$ converges.*

If $S = \sum_{j=1}^{\infty} a_j$, $\quad s_n = \sum_{j=1}^{n}(-1)^{j-1}a_j$,

$$0 < (-1)^j(S - s_j) < a_{j+1}$$

Proof. s_{2n} are monotonically increasing, s_{2n+1} are monotonically decreasing since

$$s_{2n+2} - s_{2n} = a_{2n+1} - a_{2n+2} > 0 \quad s_{2n+1} - s_{2n-1} = a_{2n+1} - a_{2n} < 0$$
$$s_2 < s_{2n} \qquad\qquad s_1 > s_{2n+1}$$

Since any monotonically increasing or decreasing sequence that's bounded converges, s_{2n}, s_{2n+1} converges.

Now

$$\lim_{n\to\infty} s_{2n} - \lim_{n\to\infty} s_{2n+1} = \lim_{n\to\infty}(-a_{2n}) = 0 \text{ so } \lim_{n\to\infty} s_{2n} = \lim_{n\to\infty} s_{2n-1}$$

Also, we have

$$S - s_{2n} > s_{2n+1} - s_{2n} = a_{2n+1} \qquad \begin{gathered} S - s_{2n-1} > s_{2n} - s_{2n-1} = -a_{2n} \\ (-1)(S - s_{2n-1}) < a_{2n} \end{gathered}$$

□

Theorem 15.
Assume $\sum |a_n|$ converges.

Then $\sum a_n$ converges and $|\sum a_j| \leq \sum |a_j|$.

Proof. Consider $b_j = a_j + |a_j|$.

$$0 \leq b_j \leq 2|a_j| \text{ so } \sum b_j \leq 2\sum |a_j|$$

By comparison test $\sum b_j = \sum a_j + \sum |a_j|$ converges. Then so does $\sum a_j$

Suppose $a_j = u_j + iv_j$.

Since $|u_j| \leq |a_j|$ and $\sum |a_j|$ converges, by comparison test $\sum |u_j|$ converges.

$\sum u_j$ converges.

Same with $\sum |v_j|$ converging; $\sum v_j$ converges.

Then $\sum u_j + iv_j = \sum a_j$ converges.

Use triangle inequality to show $|\sum a_j| \leq \sum |a_j|$.

$\square$

Theorem 16 (Apostol Vol.1, Theorem 7.5)**.** *Assume f has a cont. f'' in some neighborhood of a. Then $\forall\, x \in$ neighborhood,*

$$f(x) = f(a) + f'(a)(x-a) + E_1(x), \text{ where } E_1(x) = \int_a^x (x-t)f''(t)dt$$

Proof.

$$E_1 = f(x) - f(a) - f'(a)(x-a) = \int_a^x f' - f'(a)\int_a^x dt = \int_a^x (f'(t) - f'(a))dt =$$

$$= \int_a^x udv, \text{ where } \begin{array}{l} u = f'(t) - f'(a) \\ v = (t-x) \end{array}$$

$$\Longrightarrow E_1 = -\int_a^x f''(t)(t-x) \text{ since } u(a) = 0,\ v(x) = 0$$

$\square$

Theorem 17 (Apostol Vol.2, Theorem 7.6)**.** *Assume f has cont. $f^{(n+1)}$ on some interval of a. Then $\forall\, x \in$ interval,*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + E_n(x), \text{ where } E_n = \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)dt$$

Proof. Proof by induction. $E_{n+1}(x) = E_n(x) - \frac{f^{(n+1)}(a)}{(n+1)!}(x-a)^{n+1}$.

Note $\frac{(x-a)^{n+1}}{n+1} = \int_a^x (x-t)^n dt$ so

$$E_{n+1}(x) = \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)dt - f^{(n+1)}(a)\int_a^x (x-t)^n dt = \frac{1}{n!}\int_a^x (x-t)^n (f^{(n+1)}(t) - f^{(n+1)}(a))dt =$$

$$= \int_a^x udv, \text{ where } \begin{array}{l} u = f^{(n+1)}(t) - f^{(n+1)}(a) \\ v = \dfrac{-(x-t)^{n+1}}{n+1} \end{array}$$

Now $u(a) = 0$, $v(x) = 0$.

$$\Longrightarrow E_{n+1}(x) = -\int_a^x -\frac{(x-t)^{n+1}}{(n+1)!} f^{(n+2)}(t)$$

$\square$

Note, by second mean-value theorem for integration,

$$E_n = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \text{ for } a < c < x$$

Theorem 18 (Abel's Partial Summation Formula)**.**
Let a_n, b_n be two sequences of complex numbers.
Let $A_n = \sum_{j=1}^n a_j$
Then $\sum_{j=1}^n a_j b_j = A_n b_{n+1} + \sum_{j=1}^n A_j(b_j - b_{j+1})$

Proof.

$$A_0 = 0$$

$$a_j = A_j - A_{j-1} \qquad j = 1, 2, \ldots n$$

$$\sum_{j=1}^n a_j b_j = \sum_{j=1}^n (A_j - A_{j-1})b_j = \sum_{j=1}^n A_j b_j - \sum_{j=1}^n A_{j-1}b_j =$$

$$= \sum_{j=1}^n A_j b_j - \sum_{j=2}^{n-1} A_j b_{j+1} = A_n b_{n+1} + \sum_{j=1}^n A_j (b_j - b_{j+1})$$

$\square$

Theorem 19 (Dirichlet's Test)**.** *Consider* $\sum a_n$*,* $a_n \in \mathbb{Z}$ *such that* $\left|\sum_{j=1}^{n} a_j\right| = |A_n| \leq M$
Let b_n *be a decreasing sequence converging to zero, i.e.* $b_n > b_{n+1}, \quad \lim_{n\to\infty} b_n = 0$
Then $\sum a_n b_n$ *converges.*

Proof. Since there is an $M > 0$ such that $|A_j| \leq <, \quad \forall j \in n$

$$A_n b_{n+1} \to 0 \text{ as } n \to \infty$$

$$|A_j(b_j - b_{j+1})| \leq M(b_j - b_{j+1}) \text{ but } \sum b_j - b_{j+1} = b_1 \text{ converges}$$

$$\implies \sum |A_j(b_j - b_{j+1})| \leq \sum M(b_j - b_{j+1}) = Mb_1$$

$$\text{since } \sum A_j(b_j - b_{j+1}) \text{ converges, then} \sum a_j b_j \text{ converges}$$

□

Theorem 20 (Abel's test)**.**
Let $\sum a_j$ *be a convergent series with* $a_j \in \mathbb{C}$*.*
b_n *monotonically convergent sequence,* $b_n \in \mathbb{R}$
Then $\sum a_n b_n$ *converges.*

Proof. $\sum a_j$ converges, so A_j converges, so $A_j b_{j+1}$ converges.
A_j is bounded (by the limit).
$A_j b_{j+1}$ converges so $\sum A_j(b_j - b_{j+1})$ converges.
$\sum a_j b_j$ converges. □

Theorem 21 (Existence of a circle of convergence (Thm. 11.7))**.**
Assume $\sum a_j z^j$ *converges for at least one* $z_1 \neq 0$
diverges for at least one $z_2 \neq 0$

$\exists\, r > 0$ *s.t.* $\sum a_j z^j$ *absolutely* *converges for* $|z| < r$
diverges for $|z| > r$

Proof.
Let $A =$ set of all $|z|$ s.t. $\sum a_j z^j$ converges
(A is not empty since we've given at least one z_1)

Given $\sum a_j z^j$ diverges for z_2, A^c (where $A \bigcup A^c = \mathbb{Z}^+$) is not empty,

Suppose $|z_3| \in A, \quad |z_3| > |z_2|$
Then by Thm. 11.6, $\sum a_j z^j$ converges for z_2, but $z_2 \in A^c$. Contradiction.
$\forall\, z \in A, |z| < |z_2|$

Since A is a nonempty set of positive numbers that's bounded above,
there exists a least upper bound (limsup) r

Thus $\sum a_j z^j$ diverges if $|z| > r$ (by definition of limsup of A)

If $|z| < r$, $\exists\, x \in A$ s.t. $|z| < x < r$
Then by Thm. 11.6, $\sum a_j z^j$ converges absolutely for z. □

Let's review some of the ideas from this section.
Sufficient Condition for convergence.

Theorem 22 (Bernstein's Theorem)**.** *Assume* $\forall x \in [0, r],\ f(x),\ f^{(j)}(x) \geq 0 \quad \forall j \in \mathbb{N}$*.*
Then if $0 \leq x < r$

$$\sum^{\infty} \frac{f^{(k)}(0)}{k!} x^k \quad \textit{converges}$$

Proof. If $x = 0$, we're done. Assume $0 < x < r$.

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k + E_n(x)$$
$$E_n(x) = \frac{x^{n+1}}{n!} \int_0^1 u^n f^{(n+1)}(x - xu) du$$
$$F_n(x) = \frac{E_n(x)}{x^{n+1}} = \frac{1}{n!} \int_0^1 u^n f^{(n+1)}(x - xu) du$$

Since $f^{(n+1)} > 0, \quad f^{(n+1)}(x(1-u)) \leq f^{(n+1)}(r(1-u))$
$$\implies F_n(x) \leq F_n(r) \implies \frac{E_n(x)}{x^{n+1}} \leq \frac{E_n(r)}{r^{n+1}}$$

For $f(x) = \sum_{j=0}^{n} \frac{f^{(j)}(0)}{j!} x^j + E_n(x) \implies E_n(x) \leq \left(\frac{x}{r}\right)^{n+1} E_n(r)$

$$f(r) = \sum_{j=0}^{n} \frac{f^{(j)}(0)}{j!} r^j + E_n(r) \geq E_n(r) \text{ since } f^{(j)}(0) \geq 0 \quad \forall j$$

So then $0 \leq E_n(x) \leq \left(\frac{x}{r}\right)^{n+1} f(r)$
$n \to \infty$ and $f(t)$ will be some non-infinite value, so $E_n(x) \xrightarrow{n\to\infty} 0$. □

Theorem 23. *Let f be represented by $f(x) = \sum_{j=0}^{\infty} a_j (x-a)^j$ in the $(a - r, a + r)$ interval of convergence*

(1) $\sum_{j=1}^{\infty} j a_j (x-a)^{j-1}$ *also has radius of convergence r.*
(2) $f'(x)$ *exists* $\forall x \in (a - r, a + r)$ *and*

$$f'(x) = \sum_{j=1}^{\infty} j a_j (x-a)^{j-1} \tag{5}$$

We can integrate power series (this is significant).

Theorem 24 (Integrability of Power Series)**.**
Assume f is represented by

$$f(x) = \sum_{j=0}^{\infty} a_j (x-a)^j$$

in an open interval $(a - r, a + r)$. Then f is continuous on this interval, and $\forall x \in (a - r, a + r)$

$$\int_a^x f(t)dt = \sum_{j=0}^{\infty} a_j \int_a^x (t-a)^j dt = \sum_{j=0}^{\infty} \frac{a_j}{j+1} (x-a)^{j+1} \tag{6}$$

4. Ordinary Differentiation Equations.

4.1. Second-Order ODEs.

4.1.1. *Complete solution to $L(y) = y'' + ay' + by = 0$.*

Theorem 25 (Theorem 8.7, Complete solution to the Second-Order homogeneous ODE)**.**
Let $d = a^2 - 4b$

Then $y = e\left(\frac{-ax}{2}\right)(c_1 u_1 + c_2 u_2), \quad \forall x \in (-\infty, \infty)$

if $d = 0, \quad u_1 = 1, \quad u_2 = x$

if $d > 0, \quad u_1 = e^{kx}, \quad u_2 = e^{-kx}; \quad k = \frac{1}{2}\sqrt{d}$

if $d < 0, \quad u_1 = \cos(kx), \quad u_2 = \sin(kx); \quad k = \frac{1}{2}\sqrt{-d}$

4.2. **Nonhomogeneous linear equations of second-order with constant coefficients.** (Apostol, Vol. 1, Sec. 8.15, pp. 329)

Consider $y'' + ay' + by = R = L(y)$

If $y_1,\, y_2$ are solutions to $L(y) = R$, i.e. $L(y_1) = L(y_2) = R$,
then $L(y_1 - y_2) = 0 \qquad \Longrightarrow y_1 - y_2 = y_h = cv_1 + c_2v_2$

Now $v_1 = e\left(\frac{-ax}{2}\right)u_1; \quad v_2 = e\left(\frac{-ax}{2}\right)u_2$

$W = v_1v_2' - v_2v_1'$

Theorem 26 (Theorem 8.9)**.** *Let* $v_1,\, v_2$ *be solns. to* $L(y) = 0$

Let $W =$ *Wronskian of* v_1, v_2

Then $y_1 =$ *particular soln. s.t.* $L(y_1) = R$

(7)
$$y_1 = t_1v_1 + t_2v_2$$

where
$$t_1 = \int v_2\frac{R}{W}; \qquad t_2 = \int v_1\frac{R}{W}$$

Proof.
$$y_1' = t_1v_1' + t_2v_2' + t_1'v_1 + t_2'v_2$$
$$y_1'' = t_1v_1'' + t_2v_2'' + t_1'v_1' + t_2'v_2' + (t_1'v_1 + t_2'v_2)'$$
$$L(y_1) = t_1'v_1' + t_2'v_2' + (t_1'v_1 + t_2'v_2)' + a(t_1'v_1 + t_2'v_2)$$

Cleverly, let
$$t_1'v_1' + t_2'v_2' = R \quad t_1'v_1 + t_2'v_2 = 0$$

Now since $W = v_1v_2' - v_2v_1' \neq 0 \quad \forall\, x,$
$$\Longrightarrow t_1' = \frac{-v_2R}{W}; \quad t_2' = \frac{v_1R}{W}$$

□

Introduction to homogeneous (of degree zero) ODEs and geometry. f homogeneous (of degree zero) if $f(tx, ty) = f(x, y)$,
$$y' = f(1, \frac{y}{x})$$

e.g. $v = \frac{y}{x}$
$$y' = v'x + v = f(1, v)$$

(8)
$$\int \frac{dv}{f(1, v) - v} = \frac{1}{2}x^2$$

$y' = f(x, y)$

Consider *isoclines* of the equation, curve where y' is constant. e.g. $y' = \frac{-2y}{x}$; Consider $y = mx \qquad (1, m) \in L$

Suppose there's an integral curve through each pt. of $y = mx \quad \Longrightarrow f(a, b) = f(a, ma) = a^0f(1, m)$

i.e. the integral curve through (a, b) has the same slope as the integral curve through $(1, m)$

Similarity transformation carries set S into a new set kS,
by multiplying the coordinates of each point of S by a constant factor $k > 0$

every line through the origin remains fixed under a similarity transformation

therefore the isoclines of a homogeneous equation don't change under a similarity transformation;

hence the appearance of the direction field doesn't change either.

this suggests similarity transformations carry integral curves into integral curves

$$y = F(x)$$
$$F'(x) = f(x, F(x))$$
$$\text{Let } (x,y) \in kS. \qquad \text{Then } \left(\frac{x}{k}, \frac{y}{k}\right) \in S$$
$$F\left(\frac{x}{k}\right) = \frac{y}{k}$$
$$kS \text{ described by } g(x) = kF\left(\frac{x}{k}\right); \quad g' = F'\left(\frac{x}{k}\right)$$
$$F'\left(\frac{x}{k}\right) = f\left(\frac{x}{k}, F\left(\frac{x}{k}\right)\right) = \frac{1}{k^0} f(x, kF\left(\frac{x}{k}\right)) = f(x,g)$$

So g, integral curve in kS is also in S.

5. Basic Tricks for Algebra

In factoring a polynomial consisting of x^n and y^n, consider that for n even, the two factors should be "symmetric" and for n odd, the two factors should have a "cubic" form.

6. Linear Algebra.

6.1. Linear Spaces.

Definition 1 (Vector Space)**.** A vector space consists of a set V of objects (called vectors), equipped with 2 operations:

(1) vector addition if $x, y \in V, \quad x + y \in V$
(2) scalar multiplication $\alpha \in \mathbb{R}, x \in V; \quad \alpha x \in V$

such that 10 axioms are satisfied:

(9) $x + y = y + x$ (commutative law)

(10) $(x+y)+z = x+(y+z)$

(11) $\exists\, 0 \in V$ such that $0 + x = x + 0 = x$

(12) $\exists -x \in V, \quad \forall x \in V$ such that $x + (-x) = (-x) + x = 0$

(13) $\exists\, 1 \in V$ such that $1 \cdot x = x$

(14) $(\alpha + \beta)x = \alpha x + \beta y$

(15) $\alpha(x+y) = \alpha x + \alpha y$

(16) $(\alpha\beta)x = \alpha(\beta x)$

(17) $\forall x, y \in V, \quad x + y \in V$ (closure under vector addition)

(18) $\forall x \in V, \quad \alpha x \in V$ (closure under scalar multiplication)

Definition 2 (Subspace)**.**
A subspace of a vector space V is a nonempty subset H of V such that
H is closed under addition; if $x, y \in H; \quad x + y \in H$
H is closed under scalar multiplication if $\alpha \in \mathbb{R}, \quad x \in H; \quad \alpha x \in H$

An example of a subspace for $\mathbb{R}^3$.
if $L = \{\lambda\vec{u} | \vec{u} \in L, \lambda \in \mathbb{R}\}$
$\vec{x}, \vec{y} \in L; \quad \alpha \in \mathbb{R}$
$\vec{x} + \vec{y} = x_1\vec{u} + y_1\vec{u} = (x_1 + y_1)\vec{u} \in L$
$\alpha\vec{x} = \alpha(x_1\vec{u}) = (\alpha x_1)\vec{u} \in L$

Definition 3 (Definition of a matrix)**.**

$$A = [a_{ij}]_{(mn)} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \tag{19}$$

with A being an $m \times n$ matrix, mn denotes the size of the matrix.

We also can define matrix addition easily with this notation:

$$A + B = [a_{ij} + b_{ij}]_{(mn)} \tag{20}$$

Possibly useful notation: $(AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$

Definition 4 (Matrix Multiplication definition)**.**

Let $A = [a_{ij}]_{(mn)}$ be an $m \times n$ matrix
$B = [b_{ij}]_{(np)}$ be an $n \times p$ matrix
AB is a $m \times p$ matrix such that

$$\boxed{AB = \left[\sum_{j=1}^{n} a_{ij}b_{jk}\right]_{(mp)} \quad ; \quad (AB)_{ik} = \sum_{j=1}^{n} (A)_{ij}(B)_{jk}} \tag{21}$$

Note that

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \dots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

$$Ax = \sum_{j=1}^{n} x_j [a_{ij}]_{m1}$$

This is a useful expression for a column space expansion.

Notice that the solution set of the system of homogeneous equations is exactly the nullspace N of the coefficient matrix A.

Definition 5. The solution set for the equation $Ax = 0$ is called the **nullspace** of A.
The number of pivot entries of a matrix A is the rank of A.

rank of A is the number of pivot entries, i.e. number of variables that are dependent on other arbitrary variables.

Definition 6. Matrices that do not possess inverses are said to be **noninvertible** or **singular** (In my opinion these words are not useful, they are simply more names upon names, but other people will use these terms).

6.1.1. *Theorems for Linear Spaces.*

Theorem 27. *S spans $\forall\, x \in L(S)$ uniquely iff S spans 0 uniquely.*

Proof.
If S spans $\forall\, x \in L(S)$ uniquely,
S spans $0 \in L(S)$ uniquely.

If S spans 0 uniquely,
Consider $X \in L(S)$.
Suppose S spans X in two ways, $x = \sum c_j A_j = \sum d_j A_j$

$$\implies \sum (c_j - d_j)A_j = 0$$

but S spans 0 in a unique way, s.t. $c_j - d_j = 0 \quad \forall\, j \implies c_j = d_j$
S spans X in a unique way. □

Theorem 28 (Apostol's Thm. 12.8 for Vol.1, Thm. 1.5 for Vol.2)**.** *Let $S = \{A_1, \dots, A_q\} =$ linear independent set of q elements, $S \subseteq V_N$*

Consider $L(S)$, linear span of S.

Then $\forall$ set of $q+1$ elements in $L(S)$, the set is dependent.

Proof. Consider $q = 1$.
$\{A_1\} = S,\ \ A_1 \neq 0$, since S linear independent.

Consider $B_1,\ B_2 \in L(S)$; $B_1 \neq B_2$
then $B_1 = c_1 A_1,\ \ B_2 = c_2 A_2; \quad c_1,\ c_2 \neq 0$

Then $c_2 B_1 - c_1 B_2 = 0$, which is a nontrivial representation of 0, $\{B_1, B_2\}$ dependent.

Assume $q-1$ case is true.

Consider any set of $q+1$ elements in $L(S)$, $T = \{B_1, \ldots, B_{q+1}\}$

$$B_j = \sum_{k=1}^{q} a_{jk} A_k \qquad \forall\, j = 1, \ldots, q+1$$

Case 1. $a_{j1} = 0, \forall\, j = 1, \ldots, q+1$
Then $\forall\, j$, $B_j = \sum_{k=2}^{q} a_{jk} A_k, \quad B_j \in L(S'); S' = \{A_2, \ldots, A_q\}$
But S' linear independent and consists of $q-1$ linear independent elements. By $q-1$ case, T is dependent.
Case 2. Not all a_{j1} are zero.

Assume $a_{11} \neq 0$ (if necessary, we can renumber the B's to achieve this)

$$c_j = a_{j1}/a_{11}$$

$$c_j B_1 = a_{j1} A_1 + \sum_{k=2}^{q} c_j a_{1k} A_k$$

$$B_j = a_{j1} A_1 + \sum_{k=2}^{q} a_{jk} A_k$$

$$c_j B_1 - B_j = \sum_{k=2}^{q} (c_j a_{1k} - a_{jk}) A_k$$

Then $c_j B_1 - B_j \in L(S')$. But by $q-1$ case, q elements in $L(S')$ are linearly dependent.

$\sum_{j=2}^{q+1} t_j (a_j B_1 - B_j) = 0$ and t_j not all $t_j = 0$

$\Longrightarrow$ T linearly dependent since there's a nontrivial representation of 0 with B_j, $j = 1, \ldots, q+1$ $\square$

Theorem 29 (Basis Properties, Apostol's Thm. 12.10 of Vol. 1 or Thm. 1.7 of Vol. 2)**.** *In V_n, basis properties include:*

(1) *$\forall$ basis $\mathcal{B}_{V_n}$, $\mathcal{B}_{V_n}$ contains exactly n elements.*
(2) *Any set of linear independent elements is a subset of some basis.*
(3) *Any set of n linear independent elements is a basis.*

Proof. (1) $(e_1, \ldots, e_n)$ forms one basis.

Want : If for any 2 basis, they contain the same number of elements, we're done.
Let S, T be 2 bases, S has q elements, T has r elements.
If $r > q$, T contains at least $q+1$ elements in $L(S)$, since $L(S) = V_n$
Then T linearly dependent by Thm. 12.8 ($\forall$ set of $q+1$ elements in $L(S)$ is dependent, if $|S| = q$ and S linearly independent)
contradicts T is a basis. Then $r \leq q$

reverse S, T to get $q \leq r \Longrightarrow r = q$

(2) Let $S = \{A_1, \ldots, A_q\}$ be any linearly independent set, $S \subseteq V_n$
If $L(S) = V_n$, done.
else, then $\exists\, X \in V_n$ s.t. $X \notin L(S)$
Let $S' = \{A_1, \ldots, A_q, X\}$
If S' dependent, then for

$$\sum_{j=1}^{q} c_j A_j + c_{q+1} X = 0, \qquad c_1, \ldots, c_{q+1} \text{ not all zero}$$

If $c_{q+1} = 0$, then $c_j = 0$ for $j = 1, \ldots, q$, (since $A_1, \ldots, A_q$ are independent) and we said they're not all zero.

Then $c_{q+1} \neq 0 \Longrightarrow X = \frac{-1}{c_{q+1}} \sum_{j=1}^{q} c_j A_j$

But we said that $X \notin L(S)$
$\Longrightarrow S'$ linearly independent.

$|S'| = q + 1$.
If $L(S') = V_n$, done.
Otherwise, repeat until $S^{(n-q)}$ is obtained; otherwise, $S^{(n+1-q)}$ contains $n+1$ elements, contradicting Thm.12.8 ($\forall$ set of $n+1$ elements in $L(S)$ is dependent, if $|S| = n$ and S linearly independent).

(3) Let S be any linearly independent set of n elements.
By (b) or part (2), S is a subset of some basis $\mathcal{B}_{V_n}$
But by (a) or part(1), $\mathcal{B}_{V_n}$ contains n elements, $S = \mathcal{B}_{V_n}$

$\square$

6.1.2. *Lines, Planes, Conic Sections: Basic Analytical Geometry.*

Theorem 30 (Apostol's Thm. 13.7)**.** *For* $\begin{aligned} M &= \{P + sA + tB\} \\ M' &= \{P + sC + tD\} \end{aligned}$; $M = M'$ *iff* $L(\{A, B\}) = L(\{C, D\})$

Proof.
If $L(\{A, B\}) = L(\{C, D\})$, then $M = M'$
If $M = M'$
$\forall\, X = P + sA + tB \in M$, $X \in M'$
so that $P + sA + tB = P + s_2C + t_2D \Longrightarrow sA + tB = s_2C + t_2D \Longrightarrow L(\{A, B\}) \subseteq L(\{C, D\})$

$\forall\, X' = P + sC + tD \in M'$, $X' \in M$
so that $P + sC + tD = P + s_2A + t_2B \Longrightarrow sC + tD = s_2A + t_2B \Longrightarrow L(\{C, D\}) \subseteq L(\{A, B\})$ $\square$

Theorem 31 (Apostol's Thm. 13.8)**.** *For* $\begin{aligned} M &= \{P + sA + tB\} \\ M' &= \{Q + sA + tB\} \end{aligned}$, $M = M'$ *iff* $Q \in M$

Proof.
If $M = M'$, $\quad Q \in M$
If $Q \in M$, $Q = P + s_1A + t_1B$
$\forall\, X = P + sA + tB \in M$, $X = Q - s_1A - t_1B + sA + tB = Q + (s - s_1)A + (t - t_1)B$
$X \in M'$

$\forall\, X' = Q + sA + tB \in M'$, $X' = P + s_1A + t_1B + sA + tB = P + (s_1 + s)A + (t_1 + t)B$
$X' \in M$

$\Longrightarrow M = M'$ $\square$

Theorem 32 (Apostol's Thm. 13.9)**.** *Given plane M, pt. $Q \notin M$, $\exists$ only one plane M', $M \parallel M'$ and $Q \in M'$*

Proof. Consider $M = \{P + sA + tB\}$
Surely $\exists\, M' = \{Q + sA + tB\}$
Supppose $\exists\, M'' = \{Q + sC + tD\}$, $M'' \parallel M \quad M'' \parallel M$ so M'' must have $L(C, D) = L(A, B)$

Then by Thm. 13.7, $M'' = M'$ $\square$

Theorem 33 (Apostol's Thm. 13.10)**.** *If P, Q, R are noncollinear pts. $\exists$ only one plane M containing P, Q, R, namely*

$$M = \{P + s(Q - P) + t(R - P)\}$$

Proof.
Assume $P = 0$, Q, R are linearly independent.
$M' = \{sQ + tR\}$

$M'' = \{sA + tB\}$ (any other plane through the origin)

If $Q, R \in M''$, $\quad \begin{aligned} Q &= aA + bB \\ R &= cA + dB \end{aligned} \quad$ so $M' \subseteq M''$

For $M'' \subseteq M'$, $(ad-bc)A = dQ - bR$

$ad - bc \neq 0$ otherwise Q, R are linearly dependent.

$$\implies A = \frac{dQ - bR}{ad - bc}$$
$$B = \frac{-cQ + aR}{ad - bc} \text{ then } M'' \subseteq M'$$

Let $M = \{P + s(Q-P) + t(R-P)\}$

Let $\begin{array}{l} C = Q - P \\ D = R - P \end{array}$ C, D linearly independent; otherwise $C = tD = Q - P = tR - tP$ or $Q = tR + (1-t)P$, contradicting that P, Q, R are linearly independent.

So M is a plane containing P, Q, R.

Suppose $\exists\, M' = \{P + sA + tB\}$ s.t. $P, Q, R \in M'$

Consider $M_0' = \{sA + tB\}$

Then $X \in M'$ only if $X - P \in M_0''$

$Q, R \in M'$ so $C, D \in M_0''$

But we showed above that $\exists$ only one M_0' containing linearly independent C, D

$M_0' = \{sC + tD\} \implies M' = \{P + sC + tD\} = M$ $\square$

Theorem 34 (Apostol's Thm. 13.11)**.** *$A, B, C \in V_n$ are linearly dependent iff A, B, C lie on the same plane through the origin.*

Proof. Assume A, B, C are dependent.
Then $C = sA + tB$
If A, B independent, $C, A, B \in \{C + sA + tB\}$
If A, B dependent, C, A, B all lie on the same line.

Assume A, B, C lie on the same plane, M, through the origin,
If A, B dependent, A, B, C are dependent (whether or not C is pairwise dependent with A, B or not)
If A, B independent, A, B span plane M' through the origin. By Thm. 13.10, $M' = M$, and $C \in M$,
so $C = aA + bB$, A, B, C linearly independent. $\square$

Theorem 35 (Scalar Multiple as check on linear dependent in V_3; Apostol's Thm. 13.14)**.** *$A, B, C \in V_3$ linearly dependent iff $A \cdot B \times C = 0$*

Proof.
If A, B, C dependent,
If B, C dependent, done.
If B, C independent,
but A, B, C dependent, so $A = bB + cC$

$$A \cdot (B \times C) = (bB + cC) \times (B \times C) = 0$$

If $A \cdot (B \times C) = 0$
If B, C dependent, A, B, C are dependent, done.
If B, C independent, $A = c_1 B + c_2 C + c_3 (B \times C)$, since $B, C, B \times C$ form a basis, by theorem.

$$A \cdot (B \times C) = c_3 (B \times C)^2 = 0 \implies c_3 = 0$$
$$A = c_1 B + c_2 C \implies A, B, C \text{ dependent}$$

$\square$

Theorem 36 (Cramer's Rule)**.** *If $xA + yB + zC = D$, and $A \cdot (B \times C) \neq 0$, then to find x, y, z separately, just do this trick:*

$$x = \frac{D \cdot (B \times C)}{A \cdot (B \times C)} \qquad y = \frac{D \cdot (C \times A)}{A \cdot (B \times C)} \qquad z = \frac{D \cdot (A \times B)}{A \cdot (B \times C)}$$

6.2. **Applications for Linear Spaces.**

Definition 7 (Plane)**.** Consider $M \subseteq V_n$, V_n a real Euclidean space.
$M \equiv$ plane
if $\exists P \in V_n$ and $A, B \in V_n$ linearly independent, such that

$$M = \{P + sA + tB | s, t \in \mathbb{R}\} \tag{22}$$

If you are given an orthogonal vector to the plane, this form for the equation of the plane may prove useful.

Proposition 1 (Equation of a Plane)**.** *If $N \perp A, B$ for plane $M = \{P+sA+tB|s, t \in \mathbb{R}\}$ then $N \cdot (X - P) = 0, \quad \forall \in M, P \in M$.*

You can obtain an equation for the plane this way, using N, the orthogonal vector to the plane.

Definition 8 (General definition of conic sections (i.e. vector-formulated definition))**.** Given line L, directrix,
pt. F, focus, $F \notin L$
$e > 0$, eccentricity,
$d =$ distance of L from F
conic section $C = \{X\}$ s.t.

$$\|X - F\| = ed(X, L) \tag{23}$$

Now
$e < 1$ ellipse
$e = 1$ parabola
$e > 1$ hyperbola
Further, if N unit normal to L, then $\forall\, P \in L$,

$$d(X, L) = \frac{|(X - P) \cdot N|}{\|N\|} = |(X - P) \cdot N|$$

If $(X - P) \cdot N \gtreqless 0$, $X \in$ positive half-plane (negative half-plane)

If $F \in$ negative half-plane determined by N ,

$$F + dN \in L$$

This is the key, to recognize that $F + dN = P \in L$

$$\|X - F\| = e|(X - (F + dN)) \cdot N| \tag{24}$$

Consider symmetry about the origin (i.e. if $X, \exists - X$).

$$\|X - F\| = ed(X, L) = |eX \cdot N - e(F \cdot N + d)|$$

Let $\boxed{a = e(F \cdot N + d)}$

$$\|X - F\|^2 = \|X\|^2 - 2x \cdot F + \|F\|^2 = e^2(X \cdot N)^2 - 2aeX \cdot N + a^2$$

$X \to -X$ (and now add the 2 equations together, so that we're only left with)

$$X \cdot F = aeX \cdot N \text{ or } X \cdot (F - aeN) = 0$$

$$\text{then } F = aeN$$

Note that if $e = 1$, $d = 0$, impossible. So no symmetry for parabolas

$$\text{If } e \neq 1, \boxed{a = \frac{ed}{1 - e^2}}$$

$$\Longrightarrow \boxed{\|X\|^2 + (ae)^2 = e^2(X \cdot N)^2 + a^2}$$

major axis	if $X = \pm aN$ (vertices)	$\|X\|^2 + (ae)^2 = e^2(X \cdot N)^2 + a^2$ is satisfied
minor axis	if $X = \pm bN'$	$b^2 + (ae)^2 = e^2(0) + a^2; \quad b^2 = a^2(1 - e^2)$

Theorem 37 (Quick Review of Parabolas). *F on positive half plane to N*

$$\|X - F\| = e|(X - (F - dN)) \cdot N|$$

Let $N = \vec{e}_x,\ d = 2c,\ F = (c, 0);\ e = 1$

$$(x-c)^2 + y^2 = e^2((x-c) + 2c)^2 = (x-c)^2 + 4c(x-c) + 4c^2$$

$$y^2 = 4cx$$

Thus, for ellipses, the vertex is equidistant to the focus and directric (confirming the other definition)

Let $N = \vec{e}_y;\ d = 2c,\ F = (0, c),\ e = 1$

$$x^2 + (y-c)^2 = ((y-c) + 2c)^2 = (y-c)^2 + 4c(y-c) + 4c^2$$

$$x^2 = 4cy$$

The conic sections could also be defined in such a way:

An ellipse is the set of all points in a plane s.t. the sum of whose distance is d_1 and d_2 from 2 fixed points F_1 and F_2 (the foci) is constant.

$$d_1 + d_2 = \text{ constant}$$

Hyperbola is the set of all points for which the difference $|d_1 - d_2|$ is constant.
Parabola is the set of all points in a plane for which the distance to a fixed pt. F (focus) is equal to the distance to a given time (directrix).
(This is just a restatement of $\|X - F\| = ed(X, L) \xrightarrow{e=1} \|X - F\| = d(X, L)$).

My proofs:

Proof. $\|X - F\| = ed(X, L)$
If we assume symmetry about the origin, i.e. $X \to -X$, and about the foci.

$$\|X - F\| = e|(X - (F + dN)) \cdot N| = e((F \cdot N + d) - X \cdot N)$$

$$\|X - (-F)\| = ed(X, L) = e|(X + F + dN) \cdot N| = e(X \cdot N + F \cdot N + d)$$

$$\|X + F\| + \|X - F\| = 2e(F \cdot N + d) = 2a = \text{ const.}$$

□

Example 4. *Reflection properties of conic sections.*

For the parabola and hyperbola,
$F_1 \to$ origin.

$$\begin{array}{llll} u_1 \parallel X & d_1 = \|X\| & X = d_1 u_1 & X' = d_1' u_1 + d_1 u_1' \\ u_2 \parallel X - F_2 & d_2 = \|X - F_2\| & X = F_2 + d_2 u_2 & \quad = d_2' u_2 + d_2 u_2' \end{array}$$

$$\left.\begin{array}{l} u_1 \cdot u_1' = 0 \\ u_2 \cdot u_2' = 0 \end{array}\right. \text{ since } u_1, u_2 \text{ have constant length}$$

$$\begin{array}{ll} X' \cdot u_1 = d_1' & X' \cdot (u_1 + u_2) = d_1' + d_2' \\ X' \cdot u_2 = d_2' & X' \cdot (u_1 - u_2) = d_1' - d_2' \end{array}$$

$$\begin{array}{ll} \text{ellipse} & \to d_1' + d_2' = 0 \\ \text{hyperbola} & \to d_1' - d_2' = 0 \end{array}$$

$$\begin{array}{ll} X' \cdot (u_1 + u_2) = 0 & \text{on the ellipse} \\ X' \cdot (u_1 - u_2) = 0 & \text{on the hyperbola} \end{array}$$

$$\text{Let } T = \frac{X'}{\|X'\|}$$

$$\begin{array}{llcll} T \cdot u_2 = -T \cdot u_1 & \text{on the ellipse} & \multirow{2}{*}{$\rightarrow$} & \cos\theta_2 = -\cos\theta_1 & \text{on the ellipse} \\ T \cdot u_2 = T \cdot u_1 & \text{on the hyperbola} & & \cos\theta_2 = \cos\theta_1 & \text{on the hyperbola} \end{array}$$

Parabola.

$$\|X - F\| = d(X, L) = |(X - (-2a)e_x) \cdot e_x| = X \cdot e_x + 2a$$

Let $F = 0$, $\quad \|X - F\| = \|X\|$
Let $X = d_1 u_1$

$$\begin{array}{ll} X' = d_1' u_1 + d_1 u_1' & \|X\| = d_1 = X \cdot e_x + 2a \\ X' \cdot u_1 = d_1' = X' \cdot e_x & d_1' = X' \cdot e_x \end{array}$$

$$\xrightarrow{\frac{X'}{\|X'\|} = T} \begin{array}{l} T \cdot u_1 = T \cdot e_x \\ \cos\theta_1 = \cos\theta_2 \end{array}$$

Theorem 38. *If A is a matrix with fewer rows than columns, then the equation $Ax = 0$ has a nontrivial (that is, nonzero) solution for x.*

Proof. $[A|0]$ is put into reduced row-echelon form $[C|0]$ by elementary row operations. Suppose C has j pivot entries and k columsn without pivot entries.
(so C has $j + k$ columns altogether and x has $j + k$ components).

C has j rows beginning with pivot entries; its other rows if any are all zero.

$$Cx = 0 \Longrightarrow \begin{array}{l} x_{i1} = \dots \\ \vdots \\ x_{ij} = \dots \end{array}$$

where the right side of each equation contains only nonpivot variables.

Set all k nonpivot variables equal to a nonzero number. Then resulting x is nonzero and satisfies $Cx = 0$, thus satisfying $Ax = 0$.
C can have at most one pivot entry per row

$$\begin{array}{c} \leftarrow j\text{columns} \rightarrow \quad \leftarrow k \text{ columns } \rightarrow \\ \left[\begin{array}{cccccc|c} 1 & & & & & & b_1 \\ & 1 & & & & & b_2 \\ & & \ddots & & & & \vdots \\ & & & 0 & 0 & \dots \; 0 & b_m \end{array}\right] \end{array}$$

Since $j \leq m$ and $m < n = j + k$, $\quad k > 0$ so then there are k non-pivot variables.
k variables that the j pivot variables can be dependent upon. □

Proposition 2. *If H is a subspace of a vector space V, then 0 belongs to H.*

Proof. Let $x \in H$. Since $\alpha x \in H \quad \forall \alpha \in \mathbb{R}, \quad 0x = 0 \in H$ □

Proposition 3. *If H is a subspace of vector space V, and $x \in H$, then $-x \in H$.*

Proof. Let $x \in H$. Since $\alpha x \in H \quad \forall \alpha \in \mathbb{R}, \quad (-1)x = -x \in H$. □

Theorem 39. *If H is a subspace of a vector space V, H is itself a vector space under the operations of addition and scalar multiplication defined on V.*

Proof.

Axioms (9), (10) hold in H because they hold in V.
By Proposition (2), $0 \in H$ then Axiom (11) holds in H.
By Proposition (3), $-x \in H, \quad \forall x \in H$ so Axiom (12) holds.
Axioms (13), (14), (15), (16) are satisfied in H, since they are satisfied in V.

□

Note: This theorem gives us a powerful method for constructing many new examples of vector spaces (Consider a vector space we know, like $\mathbb{R}^n$ or V. Then construct a subspace from $\mathbb{R}^n$ or V.

Usually, we set up row-echelon form to calculate the inverse of a matrix. A good check is to set up the inverse matrix to multiply the original matrix to see if identity is obtained. Try to see the solution.

6.3. **Metric Spaces.**

6.3.1. *Euclidean Spaces.*

Definition 9 (Euclidean Spaces)**.** A set V is a Euclidean space if V is a linear space and has an inner product.

A real linear space V has an inner product
if $\forall x, y \in V$, $\exists$ unique $(x, y) \in \mathbb{R}$ such that $\forall z \in V; c \in \mathbb{R}$

$$(x, y) = (y, x) \text{ (symmetry)} \tag{25}$$
$$(x, y + z) = (x, y) + (x, z) \text{ (linearity)} \tag{26}$$
$$c(x, y) = c(x, y) \text{ (homogeneity)} \tag{27}$$
$$(x, x) > 0 \text{ if } x \neq 0 \text{ (positivity)} \tag{28}$$

A complex linear space V has an inner product
if $\forall x, y \in V$, $\exists$ unique $(x, y) \in \mathbb{C}$ such that $\forall z \in V; c \in \mathbb{C}$

$$(x, y) = \overline{(y, x)} \text{ (symmetry)} \tag{29}$$
$$(x, y + z) = (x, y) + (x, z) \text{ (linearity)} \tag{30}$$
$$c(x, y) = c(x, y) \text{ (homogeneity)} \tag{31}$$
$$(x, cy) = \overline{c}(x, y) \text{ (homogeneity)} \tag{32}$$
$$(x, x) > 0 \text{ if } x \neq 0 \text{ (positivity)} \tag{33}$$

$(x, cy) = \overline{(cy, x)} = \overline{c}\overline{(y, x)} = \overline{c}(x, y)$

Theorem 40 (Cauchy-Schwarz Inequality)**.** *Every inner product for a Euclidean space V satisfies Cauchy-Schwarz inequality, which is*

$$|(x, y)|^2 \leq (x, x)(y, y) \quad \forall x, y \in V \tag{34}$$

Proof for real Euclidean space. We simply make a change of notation in the following.

If A or $B = 0$, then we're done. Thus assume both A, B nonzero.

Let

$$C = xA - yB = (B \cdot B)A - (A \cdot B)B$$

Now $C \cdot C \geq 0$ by positivity of the inner product.

$$\begin{aligned} C \cdot C &= (xA - yB) \cdot (xA - yB) = x^2 A \cdot A - 2xyA \cdot B + y^2 B \cdot B = \\ &= (B \cdot B)^2 (A \cdot A) = 2(B \cdot B)(A \cdot B)^2 + (A \cdot B)^2 B \cdot B = (B \cdot B)^2 (A \cdot A) - (B \cdot B)(A \cdot B)^2 \\ & B \cdot B > 0 \text{ since } B \neq 0, \text{ so divide the right hand side by } B \cdot B \text{ to get} \\ & (B \cdot B)(A \cdot B) - (A \cdot B)^2 \geq 0 \\ & \implies (A \cdot B)^2 \leq (B \cdot B)(A \cdot A) \end{aligned}$$

$\square$

Proof for complex Euclidean space. $A, B \in V$
if A or B = 0, then we're done; $0 = 0$. Thus, assume $A, B \neq 0$.

Let

$$C = xA - yB = (B \cdot B)A - (A \cdot B)B$$

Now $C \cdot C \geq 0$ by positivity of the inner product.

$$\begin{aligned}
(C,C) &= (xA - yB, xA - yB) = \\
&= (xA - yB, xA) + (xA - yB, -yB) = \overline{(xA, xA - yB)} + \overline{(-yB, xA - yB)} = \\
&= \overline{(xA, xA) + (xA, -yB)} + \overline{(yB, xA) + (-yB, -yB)} \text{ (by linearity and symmetry for previous) } = \\
&= |x|^2\|A\|^2 + \overline{(xA, -yB)} + (xA, -yB) + |y|^2\|B\|^2 = |x|^2\|A\|^2 + -\overline{x}y\overline{(A,B)} - x\overline{y}(A,B) + \|y\|^2\|B\|^2 \\
&= \|B\|^4\|A\|^2 + -\|B\|^2\|(A,B)\|^2 - \|B\|^2\|(A,B)\|^2 + \|(A,B)\|^2\|B\|^2 = \|B\|^4\|A\|^2 - \|(A,B)\|^2\|B\|^2 \\
&\quad \|B\|^2 > 0 \text{ since } B \neq 0, \text{ so divide the right hand side by } B \cdot B \text{ to get}
\end{aligned}$$

$\|B\|^2\|A\|^2 - \|(A,B)\|^2 \geq 0. \quad \Longrightarrow \|(A,B)\|^2 \leq \|B\|^2\|A\|^2 (B \cdot B)(A \cdot B) - (A \cdot B)^2 \geq 0$

□

Definition 10 (Norm)**.** if V is a Euclidean space,

$$\|x\| = (x,x)^1/2 = \text{ norm of } x \tag{35}$$

Theorem 41 (Properties of a Norm)**.** *if V is a Euclidean space, $\forall x, y \in V$; $\quad$ c scalar*

(1) (a) $\|x\| = 0$ *if* $x = 0$
 (b) $\|x\| > 0$ *if* $x \neq 0$
 (positivity)
(2) $\|cx\| = |c|\|x\|$
(3) (a) $\|x + y\| \leq \|x\| + \|y\|$
 (b) $\|x + y\| = \|x\| + \|y\|$ *if $x = 0$ and $y = 0$, or if $y = cx$, for some $c > 0$*
 triangle inequality

Proof. (1)

$$(x,x) \neq 0 \text{ if } x = 0$$
$$(x,x) > 0 \text{ if } x \neq 0$$

(2)

$$(cx, cx) = c\overline{c}(x,x) = |c|^2(x,x) \quad \text{(by axioms for inner products)}$$

(3)

$$\begin{aligned}
\|x+y\|^2 = (x+y, x+y) = (x+y, x) + (x+y, y) &= \overline{(x, x+y)} + \overline{(y, x+y)} = \\
= \overline{(x,x)} + \overline{(x,y)} + \overline{(y,x)} + \overline{(y,y)} &= \|x\|^2 + \|y\|^2 + (x,y) + \overline{(x,y)}
\end{aligned}$$

By the Cauchy-Schwarz inequality, $|(x,y)| \leq \|x\|\|y\|$,

$$\Longrightarrow \|x+y\|^2 \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2$$

When $y = cx$,

$$\|x+y\|^2 = \|x + cx\| = (1+c)\|x\| = \|x\| + \|cx\| = \|x\| + \|y\|$$

□

6.3.2. *Orthogonality.*

Definition 11 (Angle between two elements)**.** In a real Euclidean space V,

$$\cos\theta = \frac{(x,y)}{\|x\|\|y\|} \tag{36}$$

Definition 12 (Orthogonality)**.** In a real Euclidean space V,

$$x, y \in V \text{ orthogonal if } (x,y) = 0 \tag{37}$$

For $S \subseteq V$, S is an orthogonal set if $(x,y) = 0, \forall x, y \in S, x \neq y$.
S is an orthonormal set if S is an orthogonal set and $\forall x \in S, \|x\| = 1$

Definition 13 (Orthogonal complement)**.** Let $S \subseteq V, V$ being a Euclidean space.

$$S^\perp = \{x | x \in V \text{ and } (x,y) = 0 \forall y \in S\} \tag{38}$$

Theorem 42 (Orthogonal Decomposition)**.** *Let V be a Euclidean space. Let S be a finite-dimensional subspace of V.*

then $\forall x \in V; x = s + s^{\perp}$, where $s \in S, s^{\perp} \in S^{\perp}$ and

$$\|x\|^2 = \|s\|^2 + \|s^{\perp}\|^2 \tag{39}$$

Note that $S \cup S^{\perp} = \{0\}$ and $S \oplus S^{\perp} = V$.

Proof. Since S is finite-dimensional, it has a finite orthonormal basis $\{e_1, e_2, \dots, e_n\}$. Let $s = \sum_{j=1}^{n} (x_j, e_j) e_j; s^{\perp} = x - s$.

$s \in S$ since $s \in sp(\{e_1, e_2, \dots, e_n\})$.

$$(s^{\perp}, e_j) = (x - s, e_j) = (x, e_j) - (s, e_j)$$

$$\text{but } (s, e_j) = (\sum_{k=1}^{n} (x, e_k) e_k, e_j) = ((x, e_j) e_j, e_j) = (x, e_j)$$

$$\text{so } (s^{\perp}, e_j) = 0 \implies s^{\perp} \in S^{\perp}$$

Then $x = s + s^{\perp}$ exists and is well-defined.

(uniqueness). Suppose $x = s + s^{\perp}$ and $x = t + t^{\perp}; s, t \in S; s^{\perp}, t^{\perp} \in S^{\perp}$

$$s - t = (x - s^{\perp}) - (x - t^{\perp}) = t^{\perp} = s^{\perp}$$

so then $s - t \in S, t^{\perp} - s^{\perp} \in S^{\perp}$
so $(s - t, t^{\perp} - s^{\perp}) = 0$ and $s - t = t^{\perp} - s^{\perp}$.
only 0 is such that $(0, 0) = 0$ and $0 = 0$

$$s - t = 0; t^{\perp} - s^{\perp} = 0 \implies t = s; t^{\perp} = s^{\perp}$$

□

Definition 14 (Projection)**.**
Let S be a finite-dimensional subspace of Euclidean space V.
Let $\{e_1, e_2, \dots, e_n\}$ be an orthonormal basis for S.

$$s = \sum_{j=1}^{n} (x, e_j) e_j \text{ if } x \in V \tag{40}$$

then $s \equiv$ projection of x on the subspace S.

6.3.3. *Gram-Schmidt process.*

Theorem 43 (Orthogonal sets are independent sets)**.** *Let V be a Euclidean space. Let S be an orthogonal set. $S \subseteq V; \forall x \in S, x \neq 0$. S is independent. (i.e. orthogonal sets are independent sets)*

if $dim V = n$, and S contains n distinct, non-zero elements, then S is a basis for V.

Proof. Consider

$$\sum_{j=1}^{n} \alpha_j x_j = 0 \quad \forall x_j \in V$$

Consider $x_i \in S$

$$\left(\sum_{j=1}^{n} \alpha_j x_j, x_i \right) = \sum_{j=1}^{n} \alpha_j (x_j, x_i) = 0$$

$$(x_j, x_i) = \delta_{ij}$$

$$\text{Considering the } i = j \text{ term }, \alpha_i (x_i, x_i) = \alpha_i \|x_i\|^2 = 0$$

$$\alpha_i = 0 \text{ since } \|x_i\|^2 > 0$$

this must be true for all $i = 1, 2, \dots, n$ or $i = 1, 2 \dots$.

If $dim V = n$ and $S = \{x_1, x_2, \dots x_n\}$, then by above, $x_1, x_2, \dots x_n$ independent. then by theorem, S is a basis for V. □

This theorem shows a finite-dimensional Euclidean space with an orthonormal basis, the inner product can be computed in terms of their components.

Theorem 44. *Let $V \equiv$ finite-dimensional Euclidean space . $dimV = n$.*
Let $\{e_1, e_2, \ldots, e_n\} \equiv$ orthogonal basis for V. then $\forall x, y \in V$,

$$(x, y) = \sum_{j=1}^{n} (x, e_j)\overline{(y, e_j)}$$

if $x = y$,

$$\|x\|^2 = \sum_{j=1}^{n} |(x, e_j)|^2$$

Proof.

$$(x, y) = \left(\sum_{j=1}^{n} (x, e_j)e_j, \sum_{j=1}^{n} (y, e_j)e_j \right) = \sum_{j=1}^{n} \sum_{k=1}^{n} (x, e_j)\overline{(y, e_k)}(e_j, e_k) = \sum_{j=1}^{n} (x, e_j)\overline{(y, e_j)}$$

Note that if we specified a real Euclidean space, then

$$(x, y) = \sum_{j=1}^{n} (x, e_j)(y, e_j)$$

□

Theorem 45 (Orthogonalization theorem)**.**
Let $x_1, x_2, \ldots$ be a finite or infinite sequence of elements in a Euclidean space V.
Consider $sp(\{x_1, x_2, \ldots, x_k\})$,
then there exists a corresponding sequence of elements $y_1, y_2, \dots \in V$
such that $\forall k$ integer.

(1) *y_k orthogonal to $\forall y \in sp(\{y_1, y_2, \ldots, y_{k-1})$*
(2) *$sp(\{y_1, y_2, \ldots, y_k\}) = sp(\{x_1, x_2, \ldots, x_k\})$*
(3) *$y_1, y_2, \ldots$ unique, modulo scalar factors, i.e. if $y_1', y_2' \ldots$ also satisfy the above statements, then $y_k' = c_k y_k$, where c_k is a scalar, $\forall k$.*

Proof. Use induction. The $n = 1$ case is easy.

$$y_1 = x_1$$

Consider $y_{n+1} = x_{n+1} = \sum_{j=1}^{n} a_j y_j$.
$a_j = 1$?

$$(y_{n+1}, y_j) = (x_{n+1}, y_j) - \sum_{k=1}^{n} a_k (y_k, y_j) = (x_{n+1}, y_j) - a_j (y_i, y_j), \quad (\text{ for } j \leq n, (y_k, y_j) = \delta_{kj})$$

$$\text{we can make } (y_{n+1}, y_j) = 0 \text{ if } a_j = \frac{(x_{n+1}, y_j)}{(y_j, y_j)}$$

if $y_j = 0$, then y_{n+1} is orthogonal to y_j for any a_j and in this case we choose $a_j = 0$.
so y_{n+1} is well-defined.
$\Longrightarrow (y_{n+1}, y_j) = 0 \quad \forall j = 1, 2, \ldots, n \Longrightarrow (y_{n+1}, y) = 0 \, \forall y \in sp(y_1, y_2, \ldots, y_n)$
We prove (2). Use induction.
$n = 1$.
$x_1 = y_1$, so $sp(x_1) = sp(y_1)$.

Assume nth case is true.
Recall $y_{n+1} = x_{n+1} - \sum_{j=1}^{n} \frac{(x_{n+1}, y_j)}{(y_j, y_j)} y_j$.
Since $y_1, y_2, \ldots, y_n \in sp(x_1, x_2, \ldots x_n)$ (given)

$$y_1, y_2, \ldots, y_n \subseteq sp(x_1, x_2, \ldots x_n)$$

and $x_{n+1} \in sp(x_1, x_2, \ldots x_n)$.
then $y_{n+1} \in sp(x_1, x_2, \ldots x_{n+1})$

$\Longrightarrow sp(y_1, y_2, \ldots y_{n+1}) \subseteq sp(x_1, x_2, \ldots x_{n+1})$.

Similarly, since $x_{n+1} = y_{n+1} + \sum_{j=1}^{n} \frac{(x_{n+1}, y_j)}{(y_j, y_j)} y_j \in sp(y_1, y_2, \ldots y_{n+1})$,
$\Longrightarrow sp(x_1, x_2, \ldots x_{n+1}) \subseteq sp(y_1, y_2, \ldots y_{n+1})$.

We prove (3). Again, use induction.
$n = 1$.
For $x = \alpha x_1 \in sp(x_1); x = \beta y_1' \in sp(y_1')$.

$$y_1' = \left(\frac{\alpha}{\beta}\right) x_1 = cy_!$$

Assume nth case.

$y_{n+1}' \in sp(x_1, x_2, \ldots x_{n+1}) \subseteq sp(y_1, y_2, \ldots y_{n+1})$ (by assumption and property (2) above, respectively)

$$y_{n+1}' = \sum_{j=1}^{n+1} c_j y_j = z_n + c_{n+1} y_{n+1}; z_n \in sp(y_1, y_2, \ldots, y_n)$$

By property (1), $(y_{n+1}', z_n) = (y_{n+1}, z_n) = 0$.

$$(y_{n+1}', z_n) = (z_n, z_n) + c_{n+1}(y_{n+1}, z_n) = (z_n, z_n) + 0 = 0$$
$$\Longrightarrow z_n = 0$$

so $y_{n+1}' = c_{n+1} y_{n+1}$. □

Theorem 46 (Gram-Schmidt Process)**.**
Let $x_1, x_2, \ldots, x_n \in V$, V being a Euclidean space.
Let $y_1, y_2, \ldots, y_n \in V$.
Suppose for

$$y_{n+1} = x_{n+1} - \sum_{j=1}^{n} \frac{(x_{n+1}, y_j)}{(y_j, y_j)} y_j;\ y_{n+1} = 0 \textit{ for some } n$$

then we can construct $y_1, y_2, \ldots, y_n$ an orthogonal set, and $y_1, y_2, \ldots, y_n \neq 0$ such that

$$y_1 = x_1; y_{j+1} = x_{j+1} - \sum_{k=1}^{j} \frac{(x_{j+1}, y_k)}{(y_k, y_k)} y_k; \quad \forall j = 1, 2, \ldots, n-1 \tag{41}$$

Proof. Suppose

$$y_{n+1} = 0 \text{ for some } n$$

then since

$$y_{n+1} = x_{n+1} - \sum_{j=1}^{n} \frac{(x_{n+1}, y_j)}{(y_j, y_j)} y_j = 0 \Longrightarrow x_{n+1} = \sum_{j=1}^{n} \frac{(x_{n+1}, y_j)}{(y_j, y_j)} y_j$$

and $y_j \in sp(x_1, x_2, \ldots, x_n)$ for $j = 1, 2, \ldots, n$. so $x_1, x_2, \ldots x_{n+1}$ are dependent (since x_{n+1} is a linear combination of $x_1, x_2, \ldots, x_n$).
then if $x_1, x_2, \ldots, x_n$ are independent, corresponding $y_1, y_2, \ldots, y_n$ are nonzero. □

Theorem 47 (Approximation theorem)**.**
Let S be a finite dimensional subspace of a Euclidean space V.
Let $x \in V$.
Let S be the projection of x on S.

$$\|x - s\| \leq \|x - t\|$$

$\forall t \in S$

Proof. By Theorem (42), $x = s + s^{\perp}$; $s \in S; s^{\perp} \in S^{\perp}$.

$$x - t = (x - s) + (s - t)$$
$$x - s \in S^{\perp}; s - t \in S$$
$$\|x - t\|^2 = \|x - s\|^2 + \|s - t\|^2$$
$$\|x - t\|^2 \geq \|x - s\|^2$$
$$\text{if } \|x - t\| = \|x - s\|, \quad \|s - t\| = 0 \Longrightarrow s = t$$
$$\text{if }, \|x - t\| = \|x - s\|$$

□

6.4. **Determinants.**

Definition 15 (Axiomatic Definition of a Determinant function)**.**
Let d be a real or complex valued function defined for each ordered n-tuple of vectors $A_1, A_2, \ldots, A_n$ in n-space.
$d \equiv$ determinant of order n if it satisfies the following axioms $\forall A_1, A_2, \ldots, A_n, C \in$ n=space:

(42)
If kth row A_k is multiplied by scalar t
$d(\ldots, tA_k, \ldots) = td(\ldots, A_k, \ldots)$
(Homogeneity in each row)

(43)
$\forall k = 1, 2, \ldots n,$
$d(A_1, A_2, \ldots, A_k + C, \ldots A_n) = d(A_1, A_2, \ldots, A_k, \ldots A_n) + d(A_1, A_2, \ldots, C, \ldots A_n)$
(Additivity in each row)

(44)
$d(A_1, A_2, \ldots, A_n) = 0$ if $A_i = A_j$ for some i, j with $i \neq j$
(The Determinant Vanishes if any 2 rows are equal)

(45)
$d(I_1, I_2, \ldots I_n) = 1$ where $I_k \equiv k$th unit coordinate vector
(The Determinant of the identity matrix is equal to 1)

Note that by induction, Axioms (42) and (43) can be combined to state linearity:

(46) $$d(\sum_{k=1}^{p} t_k C_k, A_2, \ldots, A_n) = \sum_{k=1}^{p} t_k d(C_k, A_2, \ldots A_n)$$

Also note that sometimes a weaker version of Axiom (44) is used:

(44') $(A_1, A_2, \ldots, A_n) = 0$ if $A_k = A_{k+1}$ for some $k = 1, 2, \ldots n - 1$
(The Determinant Vanishes if Two Adjacent Rows are Equal)

Theorem 48 (Properties of Determinants)**.** *A determinant function satisfying linearity (Axioms (42) and (43)) and Axiom (**??**') has the following properties:*

(47)
$d(A_1, A_2, \ldots, A_n) = 0$ *if* $A_k = 0$ *for some* k

(48)
$d(\ldots, A_k, A_{k+1}, \ldots) = -d(\ldots, A_{k+1}, A_k, \ldots)$

(49)
The determinant sign changes if any 2 rows A_i and A_j, with $i \neq j$, are interchanged

(50)
$d(A_1, A_2, \ldots, A_n) = 0$ *if* $A_i = A_j$ *for some i and j with $i \neq j$*

(51)
$d = 0$ *if its rows are dependent*

Proof.

- (47) By linearity of the determinant,

$$d(\dots 0, \dots) = d(\dots 0A_k, \dots) = 0d(\dots, A_k, \dots) = 0$$

- (48)

Let B be a matrix having the same rows as A except for row k and row $k+1$

Let $B_k = B_{k+1} = A_k + A_{k+!}$

Then $detB = 0$ (by Axiom (**??**'))

$$d(\dots, A_k + A_{k+1}, A_k + A_{k+1}, \dots) = 0$$

$$d(\dots, A_k, A_k, \dots) + d(\dots, A_{k+1}, A_k, \dots) + d(\dots, A_k, A_{k+1}, \dots) + d(\dots, A_{k+1}, A_{k+1}, \dots) = 0$$

$$\implies d(\dots, A_{k+1}, A_k, \dots) = -d(\dots A_k, A_{k+1} \dots)$$

$$(\text{ since } d(\dots, A_k, A_k, \dots) = d(\dots, A_{k+1}, A_{k+1}, \dots) = 0$$

- (49) Assume $i < j$ (without loss of generality)
  interchange row A_j successively with earlier adjacent rows.

$$\implies A_{j-1}, A_{j-2}, \dots, A_i \quad (j-1 \text{ interchanges })$$

interchange row A_i successively with the later adjacent rows.

$$A_{i+1}, A_{i+2}, \dots, A_{j-1} \quad (j-1-i \text{ interchanges })$$

So then there are $(j-1) + (j-i-1) = 2(j-i) - 1$ total interchanges.
There are always an odd number of interchanges,
so the determinant changes sign an odd number of times.

- (50) Let B be a matrxi obtained from A by interchanging rows A_i and A_j

$$detB = -detAA$$

But $A_i = A_j$, so that $detA = detB$. Then $detA = 0$.

- (51) Suppose $\exists c_1, c_2, \dots c_n$ scalars such that $\sum_{k=1}^{n} c_k A_k = 0$

$$\text{Let } A_1 = \sum_{k=2}^{n} t_k A_k \quad \text{(without loss of generality)}$$

$$d(A_1, A_2, \dots A_n) = \sum_{k=2}^{n} t_k(A_k, A_2, \dots, A_n) = 0 \quad (\text{ since } d(A_k, A_2, \dots, A_n) = 0 \, \forall k = 2, 3, \dots, n)$$

Since $i = 1$ for A_i was arbitrarily chosen, this must be true for any row that is dependent.

□

6.5. **Linear Transformations.**

Definition 16 (Linear Transformation)**.** Let V, W be linear spaces.
$T : V \mapsto W$
T is a linear transformation if

1. $T(x+y) = T(x) + T(y) \quad \forall x, y \in V$
2. $T(\alpha x) = \alpha T(x) \quad \forall x \in V$; α scalar

Theorem 49 (Basic Properties of Linear Transformations)**.**
Let V, W be linear spaces.
Let $T : V \mapsto W$ be a linear transformation. Then

1. *if $x = 0, T(x) = 0$ (Note that the converse is not necessarily true)*
2. $T(\sum_{j=1}^{n} \alpha_j x_j) = \sum_{j=1}^{n} \alpha_j T(x_j)$

Proof.

1. $T(0) = T((0)x) = (0)T(x) = 0$ (we pulled out the zero scalar, allowed by definition)

(2) Assume nth case: $T(\sum_{j=1}^{n} \alpha_j x_j) = \sum_{j=1}^{n} \alpha_j T(x_j)$

$$T(\sum_{j=1}^{n+1} \alpha_j x_j) = T(\sum_{j=1}^{n} \alpha_j x_j + \alpha_{n+1} x_{n+1}) = T(\sum_{j=1}^{n} \alpha_j x_j) + T(\alpha_{n+1} x_{n+1}) =$$
$$= \sum_{j=1}^{n} \alpha_j T(x_j) + \alpha_{n+1} T(x_{n+1}) = \sum_{j=1}^{n+1} \alpha_j T(x_j)$$

□

Definition 17 (Range)**.**
If $T : V \mapsto W$, T linear transformation

$$rangeT \equiv \{y|y = T(x); \forall x \in V\} \subseteq W$$

Definition 18 (Nullspace or Kernal)**.** For $T : V \mapsto W$ linear transformation

(52) $$kerT = \{x|x\ inV \text{ and } T(x) = 0\} \equiv \text{ nullspace or kernal of } T$$

6.5.1. *Nullity-Rank theorem.*

Theorem 50 (Nullity-Rank Theorem)**.** *Consider finite-dimensional V, i.e. $dimV = n < \infty$. Then $T(V)$ finite-dimensional and*

(53) $$dimkerT + dimrangeT = dimV$$

Proof.
Let $n = dimV$.
Let $\{e_1, e_2, \dots, e_k\} = \mathcal{B}_{kerT} \equiv$ basis for $kerT$.

By Theorem (29), $\mathcal{B}_{kerT}$ form part of some basis for V, say

$$\{e_1, e_2, \dots e_k, e_{k+1}, \dots, e_{k+r}\} = \mathcal{B}_V \quad \text{where } k + r = n$$

Consider $rangeT$. $\forall y \in rangeT, y = T(x)$. $x \in V$, so $x = \sum_{j=1}^{k+r} c_j e_j$.

$$T(x) = T\left(\sum_{j=1}^{k+r} c_j e_j\right) = \sum_{j=1}^{k+r} c_j T(e_j) = 0 + \sum_{j=k+1}^{k+r} c_j T(e_j)$$

Reversing the steps, we get

$$sp(\{T(e_{k+1}), T(e_{k+2}), \dots, T(e_{k+r})) = rangeT$$

Consider $\sum_{j=k+1}^{k+r} c_j T(e_j) = 0$. Then

$$T\left(\sum_{j=k+1}^{k+r} c_j e_j\right) = 0 \Longrightarrow \sum_{j=k+1}^{k+r} c_j e_j \in kerT$$

So then $\sum_{j=k+1}^{k+r} c_j e_j = \sum_{j=1}^{k} c_j e_j$
$\Longrightarrow \sum_{j=1}^{k+r} c_j e_j = 0$.
Now the e_j's are independent, so $c_1 = c_2 = \dots = c_k = c_{k+1} = \dots = c_{k+r} = 0$.
Thus $T(e_{k+1}), T(e_{k+2}), \dots, T(e_{k+r})$ are linearly independent.
$\Longrightarrow \{T(e_{k+1}), T(e_{k+2}), \dots, T(e_{k+r})\} \equiv$ basis for $rangeT$. □

Theorem 51 (Nullity-Rank theorem for Matrix Theory)**.** *Consider finite-dimensional V, i.e. $dimV = n < \infty$. Then $T(V)$ finite-dimensional*

6.5.2. *Algebra of Linear Transformation.*

Definition 19.
Let $S : V \mapsto W$ and
$T : V \mapsto W$
Let c be any scalar.
sum $S + T$ and product of cT are defined as

(54) $$(S + T)(x) = S(x) + T(x); \quad (cT)(x) = cT(x) \forall \in V$$

Theorem 52. *$\mathcal{L}(V,W) \equiv$ the set of all linear transformations on V into W is a linear space with operations of addition and multiplication by scalars defined through Theorem (19).*

Proof. Prove all ten axioms hold true.

(1) (uniqueness) Let $S, T \in \mathcal{L}(V,W)$.
Consider $S+T$

$$(S+T)(x) = S(x) + T(x) = y_1 + y_2 \in W, \quad y_1, y_2 \in W$$

Suppose $U'(x) = y_1 + y_2$.

$$U'(x) = (S+T)(x) \quad \forall x \in V \Longrightarrow U' = S+T$$

(2) (uniqueness) Consider cT

$$(cT)(x) = cT(x) = cy_1 \in W$$

Suppose $T'(x) = cy_1$

$$cy_1 = cT(x) = T'(x) \quad \forall x \in V; \text{ so } cT = T'$$

(3) $(S+T)(x) = S(x) + T(x) = T(x) + S(x) = (T+S)(x)$

(4) $((S+T)+U)(x) = (S+T)(x) + U(x) = S(x) + T(x) + U(x) = S(x) + (T+U)(x) = (S+(T+U))(x)$

(5) define $\mathbf{0}(x) = 0 \quad \forall x \in V$

(6) define $\forall T \in \mathcal{L}(V,W), (-T)$ such that

$$(T + (-1)T)(x) = T(x) + (1)T(x) = 0 \quad \forall x \in V$$

$\Longrightarrow (T + (-1)T) = \mathbf{0}$

(7) Consider a, b scalars.

$a(bT)(x) = a(bT(x)) = (ab)T(x) = ((ab)T)(x)$ as long as scalars obey closure under multiplication

(8)

$$a(S+T)(x) = a(S(x) + T(x)) = (aS)(x) + (aT)(x) = (aS + aT)(x) \quad \forall x \in V$$
$$a(S+T) = aS + aT$$

(9) $(aT + bT)(x) = aT(x) + bT(x) = (a+b)T(x) = ((a+b)T)(x) \quad \forall x \in V$
$\Longrightarrow (aT + bT) = (a+b)T$

(10) define $T = \mathbf{1}$ such that $\forall x \in V, \mathbf{1}x = x$.

□

Definition 20 (Composition of linear transformations)**.** Let U, V, W be sets.
Let $T : U \mapsto V$
Let $S : V \mapsto W$
then define the composition $ST : U \mapsto W$ by

$$(ST)(x) = S(T(x)) \quad \forall\, x \in U \tag{55}$$

Theorem 53. *if* $\quad T : U \mapsto V$
$S : V \mapsto W$
$R : W \mapsto X$

$$R(ST) = (RS)T \tag{56}$$

Proof. Both $R(ST)$ and $(RS)T$ have U as a domain.

$$\left.\begin{aligned} (R(ST))(x) &= R((ST)(x)) = R(S(T(x))) \\ (RS)T(x) &= (RS)(T(x)) = R(S(T(x))) \quad \forall x \in V \end{aligned}\right. \Longrightarrow R(ST) = (RS)T$$

□

Definition 21.
Let $T : V \mapsto V$. We define integer powers of T inductively as follows:

$$T^0 = \mathbf{1};\ T^n = TT^{n-1} \text{ for } n \geq 1 \tag{57}$$

Theorem 54.
Let U, V, W be linear spaces.
Let $T : U \mapsto V$
Let $S : V \mapsto W$ be linear transformation.
$ST : U \mapsto W$ is linear

Proof. $\forall x, y \in U$; a, b scalars

$$(ST)(ax + by) = S(T(ax + by)) = S(aT(x) + bT(y)) = aS(T(x)) + bS(T(y)) = \\ = a(ST)(x) + b(ST)(y)$$

□

Theorem 55.
Let U, V, W be linear spaces with the same field of scalars.
Let $S, T \in \mathcal{L}(V, W)$ and let c be any scalar.
(1) *For any function R with values in V*

$$(S + T)R = SR + TR \text{ and } (cS)R = c(SR) \tag{58}$$

(2) *For any linear transformation $R : W \mapsto U$,*

$$R(S + T) = RS + RT \text{ and } R(cS) = c(RS) \tag{59}$$

Proof. Use definition of composition.
(1)

$$(S + T)R(x) = S(R(x)) + T(R(x)) = (SR)(x) + (TR)(x) = (SR + TR)(x)$$

$$(cS)R(x) = c(S(R(x))) = c(SR)(x) \implies (cS)R = c(SR)$$

(2)

$$R(S+T)(x) = R(S(x)+T(x)) = R(S(x))+R(T(x)) = (RS)(x)+(RT)(x) = (RS+RT)(x)$$

$R(cS)(x) = R(cS(x)) = c(R(S(x)))$ (note that we needed linearity of R)

$$\implies = c(RS)(x)$$

□

6.5.3. *Inverses.*

Definition 22 (Left Inverses)**.**
Let V, W be sets.
Let $T : V \mapsto W$ be a function.
function $S : T(V) \mapsto V$ is a left inverse of T.
if $S(T(x)) = x \, \forall x \in V$, i.e. if

$$ST = I_V$$

where I_V is the identity transformation on V.
function $R : T(V) \mapsto V$ is called a right inverse of T.
if $T(R(y)) = y \, \forall y \in T(V)$

$$TR = I_{T(V)}$$

where $I_{T(V)}$ is the identity transformation on $T(V)$.

Left inverses need not exist. Right inverses need not be unique. For example, consider a function with no left inverse but with two right inverses.
Let $V = \{1, 2\}$
Let $W = \{0\}$
Define $T : V \mapsto W$ such that

$$T(1) = T(2) = 0$$

define $R : W \mapsto V$; $R' : W \mapsto V$ such that

$$R(0) = 1 \quad R'(0) = 2$$

T cannot have a left inverse S since

$$1 = S(T(1)) = S(0) \qquad 2 = S(T(2)) = S(0)$$

Theorem 56. *Every function $T : V \mapsto W$ has at least one right inverse.*

Proof. $\forall \in T(v), \exists x \in V$ such that $T(x) = y$
R is well-defined and R is defined by

$$R(y) = x \text{ then } T(R(y)) = T(x) = y$$

$\square$

Theorem 57.
A function $T : V \mapsto W$ can have at most one left inverse.
If T has a left inverse S, then S is also a right inverse.

Proof. Assume T has 2 left inverses, $S : T(V) \mapsto V$ and $S' : T(V) \mapsto V$.
Choose any $y \in T(x)$. $y = T(x)$ for some $x \in V$, so

$$S(T(x)) = x \quad S'(T(x)) = x$$
$$S(y) = x = S'(y) \quad \forall y \in T(v). \quad \text{so } S = S'; S \text{ is unique}$$

We now want to show its a right inverse as well.
Choose any $y \in T(V)$. $y = T(x)$ for some $x \in V$.

$$S(y) = S(T(x)) = x$$
$$(\text{ apply } T) \implies T(S(y)) = T(x) = y = (TS)(y) \implies (TS) = I_{T(V)}$$

$\square$

Definition 23 (onto)**.**
For $T : V \mapsto W$, T linear transformation,
If T onto W then

$$\forall y \in W, \exists x \in V \text{ such that } T(x) = y$$

Definition 24 (one-to-one)**.**
T is one-to-one
  if $v_1 \neq v_2; \quad v_1, v_2 \in V$
    then $T(v_1) \neq T(v_2); \quad T(v_1), T(v_2) \in W$
or
  if $T(v_1) = T(v_2)$;
    then $v_1 = v_2$

Theorem 58. *A function $T : V \mapsto W$ has a left inverse iff T maps distinct elements of V onto distinct elements of W (one-to-one) i.e.*

$$\exists S \text{ left inverse iff } \forall x, y \in V, \text{ if } x \neq y, T(x) \neq T(y) \tag{60}$$

Proof.

Assume T has S left inverse.
  Assume $T(x) = T(y)$. Apply S.
    $S(T(x)) = S(T(y)) = x = y$.
  Then T is one-to-one.
Assume T is one-to-one.
  $\forall y \in T(V), \exists x \in V$ such that $y = T(x)$.
    Since T is one-to-one, only one $x \in V$ exists such that $y = T(x)$.
  Let $S : T(V) \mapsto V$.
    $S(y) = x$ where $y = T(x)$, $\forall y \in T(V)$.
  S is well-defined since $\forall y \in T(V)$, only one $x \in V$ exists such that $y = T(x)$.

$$S(y) = S(T(x)) = (ST)(x) = x$$
$$ST = I_V \quad \text{so } T \text{ has a left inverse}$$

□

Definition 25 (Invertibility and Notation for Unique left inverse)**.** The unique left inverse of a function T is denoted by T^{-1}.
We say that T is invertible and T^{-1} is the inverse of T.

Theorem 59. *Let $T : V \mapsto W$ be a linear transformation in $\mathcal{L}(V, W)$. Then the following statements are equivalent:*

(1) *T is one-to-one on V*
(2) *T invertible and $T^{-1} : T(V) \mapsto V$ is linear*
(3) *$\forall x \in V; T(x) = 0$ implies $x = 0$, i.e. $nullspace(T) = \{0\}$*

Proof. Assume T is one-to-one.
By Theorem (58), T^{-1} exists.
(Show linearity). Choose $u, v \in T(V)$ such that $u = T(x), v = T(y)$ for any a, b scalars.

$$au + bv = aT(x) + bT(y) = T(ax + by)$$

apply T^{-1}

$$T^{-1}(au + bv) = T^{-1}T(ax + by) = ax + by = aT^{-1}(u) + bT^{-1}(v)$$

so T^{-1} linear. (we used the linearity of T and definition of T^{-1}. Assume T invertible and T^{-1} linear.
Suppose $T(x) = 0$.

$$T^{-1}(T(x)) = T^{-1}(0) = 0 \Longrightarrow x = 0$$

Assume $nullspace(T) = \{0\}$. Consider $T(x) = T(y)$.

$$T(x) - T(y) = T(x - y) = 0 \Longrightarrow x - y = 0 \Longrightarrow x = y$$

$\Longrightarrow T$ is one-to-one. □

Theorem 60. *Let $T : V \mapsto W$ be a linear transformation in $\mathcal{L}(V, W)$.*
Assume V is finite-dimensional, i.e. $dimV = n < \infty$
The following are equivalent.

(1) *T is one-to-one on V*
(2) *if $e_1, e_2, \ldots, e_p$ are independent in V,*
then $T(e_1), T(e_2), \ldots, T(e_p)$ are independent in $T(V)$.
(3) *$dimT(V) = n$*
(4) *if $\{e_1, e_2, \ldots, e_n\}$ = basis for V*
then $\{T(e_1), T(e_2), \ldots, T(e_n)\}$ = basis for $T(V)$.

Proof. Assume T is one-to-one.
Consider $\sum_{j=1}^{p} c_j T(e_j) = 0$.

$$\sum_{j=1}^{p} c_j T(e_j) = T\left(\sum_{j=1}^{p} c_j e_j\right) = 0 \Longrightarrow \text{ then } \sum_{j=1}^{p} c_j e_j = 0$$

e_j's are independent, so $c_j = 0$
$\Longrightarrow T(e_1), T(e_2), \ldots, T(e_p)$ are independent.
Assume if $e_1, e_2, \ldots, e_p$ are independent in V, then $T(e_1), T(e_2), \ldots, T(e_p)$ are independent in $T(V)$.
Suppose $\{e_1, e_2, \ldots, e_n\} = \mathcal{B}_V \equiv$ basis for V.
then $T(e_1), T(e_2), \ldots, T(e_n)$ are independent and form part of a basis (by theorem).
$\Longrightarrow dimT(V) \geq n$.
But by nullity-rank theorem, Theorem (50), $dimT(V) \leq n$. $\Longrightarrow$ $dimT(V) = n$.

Assume $dimT(V) = n$. Suppose $\{e_1, e_2, \ldots, e_n\} = \mathcal{B}_V \equiv$ basis for V.

$$\forall y \in T(V), \exists x \in V \text{ such that } y = T(x)$$
$$\forall x \in V; x = \sum_{j=1}^{n} c_j e_j$$
$$T(x) = \sum_{j=1}^{n} c_j e_j = \sum_{j=1}^{n} c_j T(e_j)$$
$$\implies T(e_j)\text{'s span } T(V)$$

Now we assumed $dimT(V) = n$ $\implies \{T(e_1), T(e_2), \ldots, T(e_n)\}$ form a basis for $T(V)$. Assume if $\{e_1, e_2, \ldots, e_n\} =$ basis for V, then $\{T(e_1), T(e_2), \ldots, T(e_n)\} =$ basis for $T(V)$.

$$x \in V \implies T(x) = T\left(\sum_{j=1}^{n} c_j e_j\right) = \sum_{j=1}^{n} c_j T(e_j) = 0$$

but $T(e_j)$'s are independent $\implies c_j = 0$.
so $T(x) = 0$ implies $x = 0$. T is one-to-one on V. □

6.5.4. *Matrix Representations for Linear Transformations.*

Theorem 61 (Basis elements under Linear Transformation)**.** *Let $\{e_1, e_2, \ldots e_n\} = \mathcal{B}_V \equiv$ basis for linear space V, $dimV = n$*
Let $w_1, w_2, \ldots, w_n \in W$.
then $\exists$ one and only one linear transformation $T : V \mapsto W$ such that
if $T(e_j) = w_j$; $\forall j = 1, 2, \ldots n$
then $\forall x \in V$, $x = \sum_{j=1}^{n} x_j e_j$
$T(x) = \sum_{j=1}^{n} x_j w_j$
(this theorem says that how T acts on the basis vectors of V completely determines how T acts on any $x \in V$)

Proof. $\forall x \in V$, $x = \sum_{j=1}^{n} x_j e_j$
Suppose $T(x) = \sum_{j=1}^{n} x_j w_j$ $\forall x \in V$
$T(e_j) = w_j$ since we let $x_k = \delta_{jk}$ (So T is well-defined)

$$T(\alpha_1 x_1 + \alpha_2 x_2) = T(\alpha_1 \sum_{j=1}^{n} x_{1j} e_j + \alpha_2 \sum_{j=1}^{n} x_{2j} e_j) = T(\sum_{j=1}^{n} (\alpha_1 x_{1j} + \alpha_2 x_{2j}) e_j) =$$
$$= \sum_{j=1}^{n} (\alpha_1 x_{1j} + \alpha_2 x_{2j}) w_j =$$
$$= \alpha_1 \sum_{j=1}^{n} x_{1j} w_j + \alpha_2 \sum_{j=1}^{n} x_{2j} w_j = \alpha_1 T(x_1) + \alpha_2 T(x_2)$$

$\implies T$ is linear
Now we prove there is only one linear transformation.

$$S(x) = S(\sum_{j=1}^{n} x_j e_j) = \sum_{j=1}^{n} x_j S(e_j) = \sum_{j=1}^{n} x_j w_j = T(x)$$

□

Theorem 62 (Matrix Representation of Linear Transformations)**.**
Let $T \in \mathcal{L}(V, W)$, $dimV = n, dimW = m$
Let $\{e_1, e_2, \ldots e_n\} = \mathcal{B}_V \equiv$ basis for linear space V,
Let $\{w_1, w_2, \ldots w_m\} = \mathcal{B}_W \equiv$ basis for linear space W,
Let $(t_{ik}) = m \times n$ matrix such that

if $T(e_k) = \sum_{k=1}^{m} t_{ik} w_i \quad (\mathcal{B}_V \to \mathcal{B}_W)$
then $\forall x \in V; \quad x = \sum_{k=1}^{n} x_k e_k$;

$$T(x) = \sum_{j=1}^{m} y_j w_j \in W$$

$$y_i = \sum_{k=1}^{m} t_{ik} x_k \quad \forall i = 1, 2, \dots m$$

(61)

Proof.

$$T(x) = \sum_{j=1}^{n} x_j T(e_j) = \sum_{j=1}^{n} x_j \sum_{i=1}^{m} t_{ij} w_i =$$

$$= \sum_{j=1}^{m} \left(\sum_{k=1}^{n} t_{ik} x_k \right) w_j = \sum_{j=1}^{m} y_j w_j$$

□

Theorem 63 (Isomorphism theorem, Apostol's Thm. 2.15)**.** *$\forall S, T \in \mathcal{L}(V, W)$, $\forall c$,*

$$m(S+T) = m(S) + m(T) \text{ and } m(cT) = cm(T)$$

and $m(S) = m(T)$ implies $S = T$ so m is one-to-one on $\mathcal{L}(V, W)$

Proof. For $T(e_j) = \sum_{k=1}^{m} t_{kj} w_k$, and similarly for $S(e_j)$

$$(S+T)(e_j) = \sum_{k=1}^{m} (s_{kj} + t_{kj}) w_k \qquad (cT)(e_j) = \sum_{k=1}^{m} c t_{kj} w_k$$

$m(S+T) = (s_{ik} + t_{ik}) = m(S) + m(T)$ and $m(cT) = (ct_{ik}) = cm(T)$. So m is linear.
To prove m one-to-one,
Suppose $m(S) = m(T)$, $S = (s_{ik})$, $T = (t_{ik})$
$m(S) = m(T) \to s_{kj} = t_{kj} \to S(e_j) = T(e_j) \, \forall \, e_j$, so $S(x) = T(x) \quad \forall \, x \in V$.
Thus $S = T$ □

6.5.5. *Matrix Theory Applications.*

Theorem 64. *In complex Euclidean space,*
$\forall m \times n$ complex matrix A,

$$(\text{ column space of } A)^{\perp} = \text{ nullspace of } A^*$$

Proof. Consider A^* such that $(A^*)_{ij} = \overline{a}_{ji}$; $i = 1, 2, \dots, n; j = 1, 2, \dots, m$.
if $y \in$ nullspace A^*,

$$(A^* y)_{ij} = \sum_{k=1}^{m} a^*_{ik} y_k = \sum_{k=1}^{m} \overline{a}_{ki} y_k = 0$$

$$\Longrightarrow \overline{\left(\sum_{k=1}^{n} \overline{a}_{ki} y_k \right)} = \sum_{k=1}^{m} \overline{y}_k a_{ki} =$$

$$= \overline{y}^T \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix} = y^* \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix} = 0$$

$y \perp i$th column of A.
Likewise, reverse the steps to prove the converse.
$\Longrightarrow$ (column space of $A)^{\perp} =$ nullspace of A^* □

Theorem 65. *In a real Euclidean space, $\forall m \times n$ matrix A,*

(62)
$$(\text{ row space })^{\perp} = \text{ nullspace } A$$

Proof. Consider $x \in$ (nullspace A); consider $Ax = 0$.

$$(Ax)_{ij} = \sum_{k=1}^{n} A_{ik}x_k = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = 0$$

$$\implies i\text{th row } \perp x$$

Consider $y \in$ (row space $)^{\perp}$.

$$\sum_{i=1}^{m} \alpha_i \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = 0 = \sum_{i=1}^{m} \alpha_i \sum_{k=1}^{n} a_{ik}y_k$$

Choice of α_i is arbitrary.
so $\sum_{k=1}^{n} a_{ik}y_k = 0$
$y \in$ nullspace of A
$\implies$ (row space $)^{\perp} =$ nullspace A. □

6.6. **Minors, cofactors, cofactor matrix, Cramer's rule, revisited.** Example (lead-in):
Consider $n \times n$ matrix A.
Every row of A can be expressed as a linear combination of n unit coordinate vectors $I_1, \dots, I_n$
first row of $A = A_1 = \sum_{j=1}^{n} a_{1j}I_j$

$$d(A_1, A_2, \dots, A_n) = d(\sum_{j=1}^{n} a_{1j}I_j, A_2, \dots, A_n) = \sum_{j=1}^{n} a_{1j}d(I_j, A_2, \dots, A_n) = detA = \sum_{j=1}^{n} a_{1j}detA'_{1j}$$

Theorem 66 (Expansion by cofactors, Apostol's Thm. 3.8)**.**
Let $A'_{kj} = A$ replacing kth row by unit coordinate vector I_j.

$$detA = \sum_{j=1}^{n} a_{kj}detA'_{kj} \tag{63}$$

cofactor of $a_{kj} = detA'_{kj}$

Proof. Let $A'_{kj} = A$, replacing kth row by unit coordinate vector I_j: $(A'_{kj})_{lm} = \begin{cases} a_{lm} & \text{if } l \neq k \\ \delta_{mj} & \text{if } l = k \end{cases}$

$$\begin{aligned} detA = d(A_1, A_2, \dots, A_n) = d(A_1, A_2, \dots, \sum_{j=1}^{n} a_{kj}I_j, \dots, A_n) = \\ = \sum_{j=1}^{n} a_{kj}d(A_1, A_2, \dots, I_j, \dots, A_n) = \sum_{j=1}^{n} a_{kj}det(A'_{kj}) \end{aligned}$$

□

Definition 26. Given $n \times n$ matrix A,

$A_{kj} \equiv k, j$ minor of A $= n-1$ order matrix obtained by deleting kth row and j column

Theorem 67 (Expansion by kth-row minors, Apostol's Thm. 3.9)**.**
For any $n \times n$ matrix A, $n \geq 2$,

$$\textit{cofactor } a_{kj} = detA'_{kj} = (-1)^{k+j}detA_{kj}; \quad A_{kj} = k, j \textit{ minor of } A$$

then $detA = \sum_{j=1}^{n} (-1)^{k+j} a_{kj}detA_{kj}$

Proof. Special case of $k = j = 1$

$$A'_{11} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \qquad A^0_{11} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & a_{22} & \dots & a_{2n} \\ 0 & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

$detA'_{11} = detA^0_{11}$ (by elementary row operations)

$detA^0_{11} = detA_{11}$ (by Thm. 3.7, we have det of block diagonal matrices)

$k = 1, j$ arbitrary.
Want: $detA'_{1j} = (-1)^{j-1} detA_{1j}$. If true,
For A'_{kj}, A'_{kj} transformed into B'_{1j} by $k-1$ successive interchanges of adjacent rows.

$$detA'_{kj} = (-1)^{k-1} detB'_{ij}$$

where B'_{1j} is an $n \times n$ matrix whose first row is I_j and $1, j$ minor $B_{1,j}$ is A_{kj}.
then $detB'_{1j} = (-1)^{j-1} detB_{1j} = (-1)^{j-1} detA_{kj}$, so $detA'_{kj} = (-1)^{k+j} detA_{kj}$

$$A'_{1j} = \begin{bmatrix} 0 & \dots 1 & \dots 0 & & \\ a_{21} & \dots & a_{2j} & \dots a_{2n} & \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{bmatrix} \qquad A^0_{1j} = \begin{bmatrix} 0 & \dots & 1 & \dots 0 & \\ a_{21} & \dots & 0 & \dots & a_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & 0 & \dots & a_{nn} \end{bmatrix}$$

$detA'_{1j} = detA^0_{1j}$ (det is unchanged by row operations)

$$A_{1j} = \begin{bmatrix} a_{21} & \dots & a_{2,j-1} & a_{2,j+1} & \dots a_{2n} & \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{n,j-1} & a_{n,j+1} & \dots & a_{nn} \end{bmatrix}$$

$detA^0_{1j} = f(A_{1j})$ (function of the $n-1$ rows of A_{1j}).
f satisfies the first 3 axioms for a determinant function of order $n-1$.

$$f(A_{1j}) = f(I) detA_{1j}; \quad I = \text{ identity matrix of } n-1 \text{ ,} \quad \text{(by uniqueness thm.)}$$

$$C = \begin{array}{c} \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \underbrace{0}_{j\text{th column}} & 0 & \dots 1 & \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \\ \leftarrow j\text{th row} \\ \\ \\ \\ \end{array}$$

∎

6.6.1. *The cofactor matrix.*

Definition 27 (Definition of the cofactor matrix)**.**

(64) $$\text{cofactor matrix of } A = cofA = cofa_{ij} = (-1)^{i+j} detA_{ij}$$

(65) $$cofA = (cofa_{ij})^n_{i,j=1} = ((-1)^{i+j} detA_{ij})^n_{i,j=1}$$

Theorem 68 (Apostol's Thm. 3.12)**.** *For any $n \times n$ matrix A with $n \geq 2$,*

(66) $$A(cofA)^T = (detA)I$$

If $detA \neq 0$, $A^{-1} = \frac{1}{detA}(cofA)^T$

Proof. $detA = \sum_{j=1}^{n} a_{kj} cofa_{kj}$
Keep k fixed.
Matrix B, whose ith row is equal to the kth row of A for some $i \neq k$, and whose remaining rows are the same as those of A.
$detB = 0$ because ith and kth rows are equal.
$detB = \sum_{j=1}^{n} b_{ij} cofb_{ij} = 0$
Since ith row of B is equal to kth row of A, $b_{ij} = a_{kj}$ and $cofb_{ij} = cofa_{ij}, \forall\, j$
$detB = \sum_{j=1}^{n} a_{kj} cofa_{ij} = 0$ if $k \neq i$

$$\sum_{j=1}^{n} a_{kj} cofa_{ij} = \begin{cases} detA & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases} = \sum_{j=1}^{n} a_{kj} cofa_{ji}^{T} = (A(cofa)^{T})_{ki}$$
$$\implies A(cofa)^{T} = (detA)I$$

□

Theorem 69 (Cramer's Rule, Apostol's Thm. 3.14)**.** *If* $\sum_{j=1}^{n} a_{ij} x_j = b_i$ *(*$i = 1, 2, \ldots, n$*) and* A *nonsingular,*
then $\exists$ *unique solution* $x_j = \frac{1}{detA} \sum_{k=1}^{n} b_k cofa_{kj}$ *for* $j = 1, 2, \ldots, n$

Proof.
$$Ax = B$$
$$\implies X = A^{-1}B = \frac{1}{detA}(cofA)^{T}B$$
Note that $x_j = \frac{detC_j}{detA}$,
where C_j is the matrix obtained from A by replacing jth column of A by column matrix B. □

6.6.2. *Change of Basis.* Armed with isomorphism, we can make a change of basis.

Theorem 70 (Change of Basis for Matrix Representation; same domain and range)**.** *If* $2n \times n$ *matrices* A, B *represent the same linear transformation* T*,*
then $\exists$ *invertible* C *such that*

(67) $$B = C^{-1}AC$$

if A *is the matrix of* T *relative to basis* $E = [e_1, e_2, \ldots e_n]$ *and*
B *is the matrix of* T *relative to basis* $U = [u_1, u_2, \ldots u_n]$

(68) $$\text{then } U = EC;\ u_j = \sum_{k=1}^{n} e_k c_{kj}; \quad \forall j = 1, 2, \ldots n$$

Note that the converse is true.
Also note that C^{-1} *changes coordinates relative to* $\mathcal{B}_V$ *into coordinates relative to* $\mathcal{C}_V$*.*

Proof. Let $T : V \mapsto W$
Assume $V = W$ so $T : V \mapsto V$, $dimV = dimW = n$,
Let $\{e_1, e_2, \ldots e_n\} = \mathcal{B}_V \equiv$ basis for V
Let $\{u_1, u_2, \ldots u_n\} = \mathcal{C}_V \equiv$ another basis for V
Theorems (61), (62) already tell us about matrix representations and uniqueness of the matrxi representation and how
$$T(e_j) = \sum_{i=1}^{n} t_{ij} u_j \quad (\mathcal{B}_V \mapsto \mathcal{C}_V)$$
Consider
$$T(e_j) = \sum_{k=1}^{n} a_{kj} e_k = \sum_{k=1}^{n} e_k a_{kj} \quad \forall j = 1, 2, \ldots n\, (\mathcal{B}_V \to \mathcal{B}_V)$$
$$T(u_j) = \sum_{k=1}^{n} b_{kj} u_k = \sum_{k=1}^{n} u_k b_{kj} \quad \forall j = 1, 2, \ldots n\, (\mathcal{C}_V \to \mathcal{C}_V)$$

since $u_j \in V; \quad \forall u_j$

$$u_j = \sum_{k=1}^{n} c_{kj} e_k = \sum_{k=1}^{n} e_k c_{kj}; \quad \forall j = 1, 2, \ldots n \Longrightarrow U = EC$$
$$T(u_j) = \sum_{k=1}^{n} c_{kj} T(e_k) = \sum_{k=1}^{n} T(e_k) c_{kj}$$

Note how C^{-1} changes coordinates from $\mathcal{B}_V$ basis to coordinates in $\mathcal{C}_V$ basis, because we can express something like $x = \sum_{j=1}^{n} x_j e_j$ into $\sum_{j=1}^{n} y_j u_j$ using the equation $e_j = \sum_{k=1}^{n} u_k c_{kj}^{-1}$.

Let $E = [e_1, e_2, \ldots e_n]$; $U = [u_1, u_2, \ldots u_n]$; $C = (c_{kj})$
$E' = [T(e_1), T(e_2), \ldots T(e_n)]$; $U' = [T(u_1), T(u_2), \ldots T(u_n)]$
then $E' = EA$; $U' = UB$; $U = EC$, $U' = E'C$

Note that $C : V \mapsto V$ is an isomorphism, so C is invertible.
Finding a relationship between A and B tells us how to make a change of basis for the matrix representation of T.

$$U' = E'C = EAC = UC^{-1}AC = UB$$
$$\Longrightarrow C^{-1}AC = B$$

□

When $T : V \mapsto W$, $dimV \neq dimW$, what can we do?

Theorem 71 (Change of basis for matrix representations in general)**.**
Let

$$\{e_1, e_2, \ldots e_n\} = \mathcal{B}_V \equiv \text{ basis for } V$$
$$\{v_1, v_2, \ldots v_n\} = \mathcal{C}_V \equiv \text{ basis for } V$$
$$\{u_1, u_2, \ldots u_n\} = \mathcal{B}_W \equiv \text{ basis for } W$$
$$\{w_1, w_2, \ldots w_n\} = \mathcal{C}_V \equiv \text{ basis for } W$$

Consider the linear transformation $T : V \mapsto W$, $dimV = n$; $dimW = m$
Let $A = (a_{ij})$ *be a matrix representation for* T *such that*
$T(e_j) = \sum_{i=1}^{m} a_{ij} u_i = \sum_{k=1}^{m} u_k a_{kj}; \quad \forall j = 1, 2, \ldots n$
Let $B = (b_{ij})$ *be a matrix representation for* T *such that*
$T(v_j) = \sum_{i=1}^{m} b_{ij} w_i = \sum_{k=1}^{m} w_k b_{kj}; \quad \forall j = 1, 2, \ldots n$
Then $\exists C$*, change of basis matrix for* V *from* $\mathcal{B}_V$ *to* $\mathcal{C}_V$ *and*

$\exists D$*, change of basis matrix for* W *from* $\mathcal{B}_W$ *to* $\mathcal{C}_W$ *i.e.*

$$C = (c_{kj}) \quad n \times n \text{ matrix and}$$
$$D = (d_{kj}) \quad m \times m \text{ matrix such that}$$
$$v_j = \sum_{k=1}^{n} e_k c_{kj} \quad \forall j = 1, 2, \ldots m$$
$$w_j = \sum_{k=1}^{n} u_k d_{kj} \quad \forall j = 1, 2, \ldots m$$

Then $D^{-1}AC = B$

Proof. By Theorem (**??**), $\exists$ unique isomorphisms $C : E \mapsto V$
$D : U \mapsto W$ So we're given that

$$v_j = \sum_{k=1}^{n} e_k c_{kj}; \quad \forall j = 1, 2, \ldots n; \qquad V = EC$$
$$w_j = \sum_{k=1}^{m} e_k d_{kj}; \quad \forall j = 1, 2, \ldots m; \qquad W = UD$$

Notice how matrix C consists of columns that are basis elements of $\mathcal{C}_V$ in terms of $\mathcal{B}_V$ basis elements,

matrix D consists of columns that are basis elements of $\mathcal{C}_W$ in terms of $\mathcal{B}_W$ basis elements.

(Just set E, U to be the identity matrix)

$$\begin{aligned} E' &= UA \\ V' &= WB \end{aligned} \quad \text{are matrix representations for } T.$$

$$\begin{aligned} V &= EC \\ W &= UD \end{aligned} \quad \text{are definitions of } C, D \text{ change of basis matrices.}$$

$$V' = E'C$$

So

$$E'C = UAC = WB = (UD)B \implies AC = DB$$

$$\boxed{D^{-1}AC = B}$$

The following commutator diagram was useful in discovering this theorem.

$$\begin{tikzcd} E \arrow[r, "T"] \arrow[d, "C"'] & E' \arrow[d, "C"'] & rangeT \text{ in } U's \arrow[l, "U"'] \\ V \arrow[r, "T"] & V' & rangeT \text{ in } W's \arrow[l, "W"'] \arrow[u, "D"'] \end{tikzcd}$$

□

We also get a procedure for the change of basis coordinates for the vectors themselves from $U = CE$ in the proof of Eqn. (68).

Theorem 72 (Change of coordinates for vectors)**.**
Consider linear space V; $dimV = n$.
Consider 2 bases, one possibly containing at least one vector not contained in the other.

$$\begin{aligned} \mathcal{B}_V &= \{e_1, e_2, \dots e_n\} \\ \mathcal{C}_V &= \{u_1, u_2, \dots u_n\} \end{aligned}$$

Consider $x \in V$; $x = \sum_{j=1}^n x_j e_j = \sum_{j=1}^n y_j u_j$ (i.e. x having different sets of coordinates relatives to $\mathcal{B}_V$ or $\mathcal{C}_V$ basis)

then $\exists A = (a_{ij})$ $\quad n \times n$ matrix that is a matrix representation of an isomorphism from $\mathcal{B}_V$ to $\mathcal{C}_V$.

Note that $A = C^{-1}$ of the previous theorem, Thm. (70), for $U = EC$.

Proof. Consider $x \in V; x = \sum_{j=1}^n x_j e_j = \sum_{j=1}^n y_j u_j$
and suppose $\sum_{k=1}^n a_{kj} u_k = e_j$

$$x = \sum_{j=1}^n x_j e_j = \sum_{j=1}^n x_j \sum_{k=1}^n a_{kj} u_k = \sum_{j=1}^n \left(\sum_{k=1}^n a_{jk} x_k \right) u_j = \sum_{i=1}^n \left(\sum_{k=1}^n a_{ik} x_k \right) u_i$$

$$\implies y_i = i\text{th coordinate of } x \text{ in } \mathcal{C}_V \text{ basis } = \sum_{k=1}^n a_{ik} x_k = (Ax)_{i1}$$

So "$A[x]_{\mathcal{B}_V} = [x]_{\mathcal{C}_V}$" □

6.7. **Eigenvalues and Eigenvectors.**

Definition 28. Let $T : S \mapsto V$.

$\lambda \equiv$ eigenvalue of T if $\exists x \neq 0, x \in S$ such that

$$T(x) = \lambda x$$

where $x \equiv$ eigenvector of T belonging to λ

Note that λ can be $\lambda = 0$, while $x \neq 0$ by definition.

Definition 29 (Characteristic Polynomial)**.** if A is an $n \times n$ matrix,

$$f(\lambda) = det(\lambda I - A) \equiv \text{ characteristic polynomial of } A \tag{69}$$

Theorem 73 (Properties of Characteristic Polynomials)**.** *For a characteristic polynomial $f(\lambda)$ which is in general a nth order polynomial, so that*

$$f(\lambda) = \sum_{j=0}^{n} c_j \lambda^j$$

then

(70)

$c_0 = (-1)^n detA = (-1)^n \lambda_1 \dots \lambda_n$ *(i.e. zero order constant coefficient is the $(-1)^n$ the determinant)*

(71)

$c_{n-1} = -1(\lambda_1 + \dots + \lambda_n) = -trA$ *(i.e. the $n-1$th order coefficient is the (-1) the trace)*

Proof. In general, the characteristic polynomial $f(\lambda)$ of a $n \times n$ matrix A can be factorized using its n roots:

$$f(\lambda) = (\lambda - \lambda_1) \dots (\lambda - \lambda_n)$$

$$\implies f(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + \dots + c_1 \lambda + c_0 = \sum_{j=0}^{n} c_j \lambda^j$$

Just be comparing this with the factored form and comparing powers of λ, then

$$c_0 = (-1)^n \lambda_1 \dots \lambda_n = (-1)^n detA$$

$$c_{n-1} = -(\lambda_1 + \dots + \lambda_n) = -trA = -\sum_{j=1}^{n} \lambda_j$$

□

6.8. **All about Trace.** Traces of matrices are called characters in group representation theory and provide a description for equivalent representations for elements of a group.

Definition 30 (Trace of a matrix A)**.** If A is an $n \times n$ matrix, $A = [a_{ij}]_{(nn)}$,

$$trA = a_{11} + a_{22} + \dots + a_{nn} = \sum_{i=1}^{n} a_{ii} \tag{72}$$

The trace is the sum of the diagonal entries of A.

Theorem 74 (Properties of the Trace)**.** *We have the following properties from the trace:*

$$tr\alpha A = \alpha trA \tag{73}$$

$$tr(A+B) = trA + trB \tag{74}$$

$$trA = trA^T \tag{75}$$

$$tr(AB) = tr(BA) \tag{76}$$

Note that we can define an inner product for $n \times n$ matrices.

Theorem 75 (Square Matrix inner product)**.** *If A, B are $n \times n$ matrices,*

$$(A, B) = trAB^T \tag{77}$$

is a possible inner product. Also from the following proof, we get this fact:

$$tr(AA^\dagger) = \sum_{i,j=1}^{n} |a_{ij}|^2 \tag{78}$$

$tr(AA^\dagger)$ is the sum of the square magnitudes of the entries of A and is a possible candidate for a "norm" of a matrix A.

Proof. Recall the definition of a complex-valued inner product.

$$(x, x) \geq 0$$
$$(x, y) = \overline{(y, x)}$$
$$c(x, y) = (cx, y)$$
$$(x, x) \geq 0$$

We have

$$(A, A) = trAA^T = \sum_{i,j=1}^{n} |a_{ij}|^2 \implies (A, A) = 0 \text{ if } a_{ij} = 0 \forall i, j = 1, \dots, n$$

$$(A, B) = tr(AB^\dagger) = \sum_{i=1}^{n} (AB^\dagger) = \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik}(b^\dagger)_{ki} = \sum_{i,k=1}^{n} a_{ik}\overline{b}_{ik}$$

$$(B, A) = \sum_{i,k=1}^{n} b_{ik}\overline{a}_{ik} = \overline{(A, B)}$$

$$(A, B + C) = tr(A(B + C)) = tr(AB + AC) = tr(AB) + tr(AC) = (A, B) + (A, C)$$

$$(cA, B) = tr(cAB^T) = ctr(AB^T) = c(A, B)$$

□

6.9. Eigenvalues, Eigenvectors, Similar Matrices, continued.

Definition 31 (Similar Matrices)**.** Two $n \times n$ matrices A and B are similar if $\exists$ invertible C such that

$$B = C^{-1}AC \tag{79}$$

Theorem 76 (Existence of diagonalization.)**.** *Let $T : V \mapsto V$; $dimV = n$.*
T has a diagonal matrix representation, iff $\exists\{u_1, u_2, \dots u_n\} \subseteq V, u_1, u_2, \dots u_n$ independent and $\lambda_1, \lambda_2, \dots, \lambda_n$ scalars such that

$$T(u_k) = \lambda_k u_k;\ k = 1, 2, \dots n$$

Note: having $u_1, u_2, \dots u_n$ independent eigenvectors is a necessary and sufficient condition to having a diagonal matrix representation.

Proof. Assume T has a diagonal matrix representation $A = (a_{ik})$ relative to some basis $(e_1, e_2, \dots e_n)$.

$$T(e_k) = \sum_{j=1}^{n} a_{jk}e_j = a_{kk}e_k \text{ since } a_{ik} = a_{kk}\delta_{ik}$$

so with $u_k = e_k$, $\lambda_k = a_{kk}$, then by a change of notation $\implies T(u_k) = \lambda_k u_k$

Suppose $\exists\{u_1, u_2, \dots u_n\}$ $\lambda_1, \lambda_2, \dots \lambda_n$ scalars such that

$$T(u_k) = \lambda_k u_k$$

Since $\{u_1, u_2, \dots u_n\|$ independent and $dimV = n$; $\{u_1, u_2, \dots u_n\}$ form a basis by theorem. Let $\lambda_k = a_{kk}$; $a_{ik} = a_{kk}\delta_{ik}$ then $A = (a_{ik})$ is a diagonal matrix representation of T relative to $\{u_1, u_2, \dots u_n\}$ basis. □

Theorem 77 (Independence of Eigenvectors)**.**
Let $u_1, u_2, \dots u_k$ be eigenvectors of linear transformation $T : S \mapsto V$.
Assume corresponding eigenvalues $\lambda_1, \lambda_2, \dots \lambda_k$ distinct.
Then $u_1, u_2, \dots u_k$ are independent.

Proof. Use induction.
$k = 1$. u_1 is independent by itself.
Assume kth case is true.
Suppose $T(u_{k+1}) = \lambda_{k+1}u_{k+1}$.
Consider

$$\sum_{j=1}^{k+1} c_j u_j = 0$$

Then

$$T\left(\sum_{j=1}^{k+1} c_j u_j\right) = \sum_{j=1}^{k+1} c_j T(u_j) = \sum_{j=1}^{k+1} c_j \lambda_j u_j$$

$$\sum_{j=1}^{k+1} \lambda_j c_j u_j - \sum_{j=1}^{k+1} \lambda_{k+1} c_j u_j = \sum_{j=1}^{k+1} (\lambda_j - \lambda_{k+1}) c_j u_j$$

Now $\lambda_j \neq \lambda_{k+1}$ for $j \neq k+1$

$c_j = 0$ for $j = 1, 2, \ldots k$ because of independence of the first kth elements

$$\Longrightarrow \sum_{j=1}^{k+1} c_j u_j = 0 + c_{k+1} u_{k+1} = 0 \Longrightarrow c_{k+1} = 0$$

□

Theorem 78 (Distinct Eigenvalues)**.**
If $dimV = n$, every $T : V \mapsto V$ linear transformation has at most n distinct eigenvalues.
If T has n distinct eigenvalues,
then the corresponding eigenvectors form a basis for V and the matrix of T relative to this basis is diagonal with the eigenvalues as diagonal entries.
Note that Theorem (78) tells us that the existence of n distinct eigenvalues is a sufficient condition for T to have a diagonal matrix representation, but is not necessary (consider the identity transformation).
i.e. the converse of this theorem is not necessarily true.

Proof.
If T had $n+1$ distinct eigenvalues by Theorem (77), V would have $n+1$ independent elements. This is not possible since $dimV = n$.

If T has n distinct eigenvalues, the corresponding eigenvectors are independent since $dimV = n$, the n corresponding eigenvectors form a basis. By Theorem (76), since there are n independent eignevectors, T can be diagonalized. □

Theorem 79. *Let $T : V \mapsto V$ be a linear transformation. V has scalars in F, $dimV = n$. Then, the eigenvalues of T consists of the roots of $f(\lambda) = det(\lambda I - A)$*

Proof. Suppose λ is an eigenvalue of T.
Then

$$T(x) = \lambda x \Longrightarrow (\lambda I - T)(x) = 0$$

since λ is an eigenvalue, then x is an eigenvector and $x \neq 0$.
then $\lambda I = T$ must be noninvertible.

if $\lambda I - T$ is non-invertible,
then $det(\lambda I - T) = 0$

Thus, if λ is an eigenvalue for T, then $det(\lambda I - T) = 0$
Note that if $det(\lambda I - T) = 0$, $(\lambda I - T)(x) = 0$ for $x = 0$ and/but not $x \neq 0$. □

6.9.1. *Similar matrices and Eigenvalues.*

Theorem 80.
Similar matrices have the same characteristic polynomial and therefore the same eigenvalues.
Converse appears to work if A and B are diagonalizable.

Proof. **Here is the trick:**

$$\lambda I - B = \lambda I - C^{-1}AC = \lambda C^{-1}IC - C^{-1}AC = C^{-1}(\lambda I - A)C$$

$$det(\lambda I - B) = det(C^{-1}(\lambda I - A)C) = det(\lambda I - A)$$

What about the converse?

$$det(\lambda I - A) = det(\lambda I - B)$$
$$C^{-1}AC = diagA$$
$$D^{-1}BD = diagB$$
$$det(\lambda I - A) = det(\lambda I - C(diagA)C^{-1}) = det(C(\lambda I - diagA)C^{-1}) = det(\lambda I - diagA) =$$
$$= det(\lambda I - diagB) = det(\lambda I - B)$$
$$\Longrightarrow (\lambda - a_{11})(\lambda - a_{22})\dots(\lambda - a_{nn}) = (\lambda - b_{11})(\lambda - b_{22})\dots(\lambda - b_{nn})$$
By comparing terms
$$\Longrightarrow a_{11} = b_{11}, a_{22} = b_{22}, \dots, a_{nn} = b_{nn}$$
$$diagA = diagB$$
$$\Longrightarrow \quad C^{-1}AC = D^{-1}BD$$
$$(DC^{-1})A(CD^{-1}) = B$$

□

Definition 32 (Hermitian and skew-Hermitian transformations (symmetric and skew-symmetric transformations))**.**
Let $\begin{array}{l} E \equiv \text{ Euclidean space} \\ V \equiv \text{ subspace of } E \end{array}$. Transformation $T : V \to E$ Hermitian (skew-Hermitian) if

(80) $$(T(x), y) = \pm(x, T(y)) \quad \forall\, x, y \in V$$

If E is real, then T is a *symmetric (skew-symmetric)* transformation.

6.9.2. *Existence of an orthonormal set of eigenvectors for Hermitian and skew-Hermitian operators acting on finite-dimensional spaces.* Eigenvalues need not exist, but if T acts on a finite-dimensional complex space, the eigenvalues always exist since they are roots of the characteristic polynomial.

Theorem 81 (Existence of orthonormal eigenvectors for Hermitian (skew-Hermitian) operators)**.** *Assume $dimV = n$.*
Let $T : V \mapsto V$ be a Hermitian or skew-Hermitian.
$\exists n$ eigenvectors $u_1, u_2, \dots, u_n$ of T which form an orthonormal basis for V.

Proof. Use induction.
If $n = 1$, T has exactly one eigenvalue. Any eigenvectors u_1 of norm 1 is an orthonormal basis for V.
Assume $n - 1$ case.
For V, $dimV = n$, choose eigenvalue λ_1 and eigenvector u_1, $\|u_1\| = 1$
$$T(u_1) = \lambda u_1$$
Let $S = sp(u_1)$. Consider $S^\perp = \{x | x \in V, (x, u_1) = 0\}$.
We want $dimS^\perp = n - 1$ and $T : S^\perp \mapsto S^\perp$.
By theorem, u_1 is part of a basis for V, say $(u_1, v_2, \dots, v_n)$.
without loss of generality, assume $(u_1, v_2, \dots, v_n)$ to be orthonormal (apply Gram-Schmidt process if necessary).
$$x \in S^\perp \subset V \text{ so } x = x_1 u_1 + x_2 v_2 + \dots + x_n v_n$$
$$(x, u_1) = x_1 = 0. \quad x = \sum_{j=2}^{n} x_j v_j \text{ so } dimS^\perp = n - 1$$
Assume T is Hermitian (skew-Hermitian).
$$(T(x), u_1) = \pm(x, T(u_1)) = \pm(x, \lambda_1 u_1) = \pm\overline{\lambda}_1 (x, u_1) = 0$$
so $T(x) \in S^\perp$.
Use induction hypothesis ($n - 1$ case) for T on $S^\perp$ so that
$\exists n - 1$ eigenvectors $u_2, \dots, u_n$ that form an orthonormal basis for $S^\perp$.
$u_1 \perp S^\perp$ by definition so
$u_1, u_2, \dots, u_n$ eigenvectors form an orthonormal basis for V. □

Theorem 82.
Let $(e_1, e_2, \dots, e_n)$ *be a basis for* V *and*
let $T : V \mapsto V$ *be a linear transformation.*
T Hermitian (skew-Hermitian) iff $(T(e_j), e_i) = \pm(e_j, T(e_i))\, \forall i, j.$

Proof.

$$x = \sum x_j e_j \quad y = \sum y_j e_j$$
$$(T(x), y) = \sum_{j=1}^{n} x_j (T(e_j), y) = \sum_{j=1}^{n}\sum_{i=1}^{n} x_j \overline{y}_i (T(e_j), e_i)$$
$$(T(y), x) = \sum_{j=1}^{n}\sum_{i=1}^{n} x_j \overline{y}_i (e_j, T(e_i))$$
$$\text{now } x_j, \overline{y}_i \text{ arbitrary, so } \forall i, y;\ (T(e_j), e_i) = (e_j, T(e_i))$$

□

Theorem 83.
Let $(e_1, e_2, \dots, e_n)$ *be an orthonormal basis for* V.
Let $A = (a_{ij})$ *be a matrix representation of linear transformation* $T : V \mapsto V$ *relative to* $\mathcal{B}_V$.
then

$$T \text{ is Hermitian (skew-Hermitian) } iff\, a_{ij} = \pm \overline{a}_{ji} \quad \forall i, j$$

Note: every real Hermitian matrix is symmetric.

Proof.

$$T(e_j) = \sum_{k=1}^{n} a_{kj} e_k \quad (\text{ matrix representation relative to } V)$$
$$(T(e_j), e_i) = \sum_{k=1}^{n} a_{kj} (e_k, e_i) = a_{ij}$$
$$(e_j, T(e_i)) = (e_j, \sum_{k=1}^{n} a_{ki} e_k) = \sum_{k=1}^{n} \overline{a}_{ki} (e_j, e_k) = \overline{a}_{ji}$$
$$(T(e_j), e_i) = \pm(e_j, T(e_i)) \Longrightarrow a_{ij} = \pm \overline{a}_{ji} \quad \text{ if } T \text{ is Hermitian (skew-Hermitian)}$$

□

Definition 33 (Definition of a Hermitian *matrix*)**.**
A square matrix $A = (a_{ij})$ is Hermitian if

$$a_{ij} = \overline{a}_{ji}\, \forall\, i, j \text{ or } A = \overline{A}^T$$

A square matrix A is skew-Hermitian if

$$a_{ij} = -\overline{a}_{ji}\, \forall\, i, j \text{ or } A = -\overline{A}^T$$

Definition 34.

$$\overline{A}^T = A^* \equiv \text{ adjoint of } A$$

Theorem 84 (Apostol Vol. 2, Thm. 5.7, pp. 122)**.**
Every $n \times n$ *Hermitian or skew-Hermitian matrix* A *is similar to the diagonal matrix,*
$\Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_n)$ *of its eigenvalues.*
Also $\Lambda = C^{-1} A C$
where C *is invertible, i.e.* $C^{-1} = C^*$.
To say the whole theorem in another way,

If you have a Hermitian or skew-Hermitian matrix, it can be diagonalized with its eigenvalues

Proof.
Let $V =$ space of n-tuples of complex numbers.
Let $(e_1, e_2, \dots, e_n)$ be the orthonormal basis of unit coordinate vectors.
if $x = \sum x_i e_i$; $y = \sum y_i e_i$ let $(x, y) = \sum x_i \overline{y}_i$.

By the existence of eigenvectors for Hermitian operators, Theorem (81), V has an orthogonal basis of eigenvectors $(u_1, u_2, \dots, u_n)$ relative to T with diagonal matrix representation $\Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_n)$.

Since A and Λ represent T, A and Λ are similar, so

$$\Lambda = C^{-1}AC$$

where

$$[u_1, u_2, \dots u_n] = [e_1, e_2, \dots, e_n]C$$

This equation shows that the jth column of C consists of the components of u_j relative to $(e_1, e_2, \dots e_n)$.

$$(u_j, u_i) = \sum_{k=1}^{n} c_{kj}\bar{c}_{ki} = \delta_{ij} \Longrightarrow CC^* = I \Longrightarrow C^{-1} = C^*$$

□

6.9.3. *Special kinds of linear transformations and matrices.* First, we'll talk about Unitary and Orthogonal matrices. Now unitary and orthogonal matrices are useful because they are the "change of basis matrices" for Hermitian (skew-Hermitian) and real symmetric (skew-symmetric) matrices, respectively.

Definition 35 (Unitary and Orthogonal Matrices)**.**
A square matrix A is unitary if $AA^* = I$.
A square matrix A is orthogonal if $AA^T = I$

Note: every real unitary matrix is orthogonal since $A^* = A^T$

Definition 36.
A square matrix A with real or complex entries is symmetric if $A = A^T$. A square matrix A with real or complex entries is skew-symmetric if $A = -A^T$.

Theorem 85. *Every square matrix A can be expressed as a unique decomposition:*

$A = B+C$ $\quad$ *(B and C are Hermitian and skew-Hermitian or are symmetric and skew-symmetric)*

Proof.

$$\text{Let } B = \frac{A + A^*}{2},\ C = \frac{A - A^*}{2}$$
$$\Longrightarrow A = B + C \text{ and}$$
$$B^* = \left(\frac{A + A^*}{2}\right)^* = B,\ C^* = \frac{A - A^*}{2} = \frac{A^* - A}{2} = -C$$

□

Theorem 86. *If A is orthogonal,* $1 = det(AA^T) = (detA)(detA^T) = (detA)^2 \Longrightarrow detA = \pm 1$.

The following straightforward theorem tells us why we call orthogonal matrices orthogonal.

Theorem 87. *An $N \times N$ real matrix is orthogonal iff its columns form an orthonormal basis for $\mathbb{R}^n$.*

Proof. If A is orthogonal, $(A^T A)_{ij} = (\infty)_{ij} = \delta_{ij}$.

$$(A^T A)_{ij} = \sum_{k=1}^{N} (A^T)_{ik} a_{kj} = \sum_{k=1}^{N} a_{ki} a_{kj} = \delta_{ij}$$

Now $((A)_i, (A)_j) = \sum_{k=1}^{N} a_{ki} a_{kj}$.
So if A is orthogonal, its columns are orthonormal. If its columns are orthonormal, A is orthogonal (by definition). □

This theorem says that any length-preserving and angle-preserving transformation on $\mathbb{R}^N$ must be orthonormal.

Theorem 88.
Let $T : V \mapsto V$ be a linear transformation on Euclidean space V.
Let $\mathcal{B}_V = \{b_1, b_2, \dots b_n\}$ = orthonormal basis for V.
Let A_T = matrix representation of T relative to $\mathcal{B}_V$.
Then T is an isometry of V iff A_T is an orthogonal matrix.

Proof.
If T is an isometry of V, $\forall x_1, x_2 \in V$; $(T(x_1), T(x_2)) = (x_1, x_2)$.
$\forall b_i, b_j \in \mathcal{B}_V$; $(T(b_i), T(b_j)) = (b_i, b_j) = \delta_{ij}$
So then, using the definition of matrix representation in the following first step,

$$(T(b_i), T(b_j)) = (\sum_{k=1}^n a_{ki} b_k, \sum_{k=1}^n a_{kj} b_j) = \sum_{k=1}^n \sum_{l=1}^n a_{ki} a_{lj} (b_k, b_l) = \sum_{k=1}^n a_{ki} a_{kj} =$$
$$= (i\text{th column of } A_T, j\text{th column of } A_T) = \delta_{ij}$$

Then columns of A_T form an orthonormal basis.
$\Longrightarrow$ Then by theorem (Thm. (87)), A_T is an orthogonal matrix.
If A_T is an orthogonal matrix, reverse steps so that,

$$\delta_{ij} = (i\text{th column of } A_T, j\text{th column of } A_T) = (T(b_i), T(b_j))$$

Now $(b_i, b_j) = \delta_{ij}$ (definition of orthonormal basis), so

$\Longrightarrow (T(b_i), T(b_j)) = \delta_{ij} = (b_i, b_j)$. Inner space is preserved.
Now we have to show that T is an isomorphism because an isometry is angle preserving and an isomorphism.
$T(b_1), T(b_2), \dots, T(b_n)$ are an orthonormal set, so $\Longrightarrow$ by theorem, $T(b_1), T(b_2), \dots, T(b_n)$ form an orthonormal basis for V.

$$\forall y \in V, y = \sum_{j=1}^n \alpha_j T(b_j) = T\left(\sum_{j=1}^n \alpha_j b_j\right)$$

where $\sum_{j=1}^n \alpha_j b_j \in V$. So T is onto.
By nullity-rank theorem,

$$nullT = dimV - rankT = n - n = 0$$

So $T(0) = 0$. T is one-to-one. $\Longrightarrow T$ is an isomorphism $\Longrightarrow T$ is an isometry. □

6.9.4. *Arbitrary Rotation Matrices in $\mathbb{R}^N$.* .
Particularly with 'advanced' Classical Mechanics and Quantum Mechanics, it's useful to start formulating the theory in terms of arbitrary rotations. Cartesian coordinates still remain useful because they don't change direction (once fixed) with "active rotations."
Consider an arbitrary rotation of a set of orthonormal Cartesian coordinates, $R_{\theta\phi}$, with no inversion (so we expect $detR_{\theta\phi} = 1$ as opposed to being -1).

$$R_{\theta\phi} = \begin{bmatrix} \sin\theta\cos\phi & -\sin\phi & -\cos\theta\cos\phi \\ \sin\theta\sin\phi & \cos\phi & -\cos\theta\sin\phi \\ \cos\theta & 0 & \sin\theta \end{bmatrix} \tag{81}$$

Notice how the columns of $R_{\theta\phi}$ are orthonormal and that $detR_{\theta\phi} = 1$.

Definition 37 (nilpotent)**.** matrix N is a nilpotent if for some $k \in \mathbb{Z}^+$; $N^k = 0$.

Definition 38 (idempotent)**.** operator or matrix P is idempotent if $P^2 = P$.

Definition 39 (2-dim rotation matrix)**.**

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

rotates the vector by θ in the counter-clockwise (positive) direction or
rotates the coordinate axes by θ in the clockwise direction.
We could also think of it as rotation of vectors in the xy plane in the positive, right-hand orientation.

6.9.5. *Quadratic forms (Apostol Vol. 2, Ch. 5).*

Definition 40 (Quadratic forms)**.** Let $T : V \to E$ be Hermitian transformation. Recall $(T(x), y) = (x, T(y))\ \forall\, x, y \in V$. Define Q s.t. $Q(x) = (T(x), x)$.

Theorem 89 (Theorem 5.8, Apostol Vol. 2)**.**
Let $(e_1, \dots, e_n) \equiv$ orthonormal basis for Euclidean space $V = \mathcal{B}_V$
Let T be Hermitian transformation; let $A = (a_{ij})$ be matrix of T relative to $\mathcal{B}_V$.
Then $Q(x) = \sum_{i,j=1}^n a_{ij} x_i x_j$, if $x = \sum_{i=1}^n x_i e_i$

Proof.

$$T(x) = \sum x_i T(e_i) \quad \text{(linearity)}$$
$$Q(x) = (T(x), x) = \sum_i \overline{x}_i (T(e_i), \sum_j x_j e_j) = \sum_{i,j} \overline{x}_i x_j (e_i, T(e_j)) = \sum_{i,j} a_{ij} \overline{x}_i x_j$$

since $(e_i, T(e_j)) = a_{ij}$. □

Even if A is not symmetric,

Definition 41 (Quadratic form (any matrix))**.**
Let $V \equiv$ Euclidean space
Let $(e_1, \dots, e_n)$ orthonormal basis for $V \equiv \mathcal{B}_V$
Let $(A = (a_i j)$ be any $n \times n$ matrix of scalars.

$$Q(x) \equiv \text{ quadratic form } = \sum_{i,j=1}^n a_{ij} \overline{x}_i x_j \quad (Q = \sum a_{ij} x_i x_j, \text{ if } V \text{ real})\, \forall\, x = \sum_i x_i e_i \in V$$

Note that in matrix form, $Q = XAX^\dagger$ (XAX^T if V real).

Theorem 90 (Thm. 5.10, Apostol Vol. 2)**.** $\forall\, n \times nA$, $\forall$ *row matrix X*

$$XAX^T = XBX^T \text{ where } B = \frac{1}{2}(A + A^T), \quad B \text{ symmetric (!!!)}$$

Proof.

$$X(\frac{1}{2}(A+A^T))X^T = \frac{1}{2}(XAX^T + (XAX^T)^T) = XAX^T \quad \text{(transpose of a number is a number)}$$

□

6.9.6. *Notes on Sec. 5.13 Reduction of a real quadratic form to a diagonal form.*

Theorem 91 (Thm. 5.11)**.** *Consider XAX^T, A Hermitian. A Hermitian, so $\exists\, C$ unitary s.t. $C^\dagger AC = \Lambda$ diagonal. Then $XAX^\dagger = Y\Lambda Y^\dagger = \sum_{i=1}^n \lambda_i y_i^2$, where $Y = [y_1, \dots, y_n]$ is the row matrix $Y = XC$, and $\lambda_1, \dots, \lambda_n$ are eigenvalues of A.*

Proof. Since C unitary, $C^\dagger = C^{-1}$. Then $Y = XC$ implies $X = YC^\dagger$.
Then $XAX^\dagger = YC^\dagger A(YC^\dagger)^\dagger = Y\Lambda Y^\dagger$ □

6.9.7. *Unitary Triangularization.*

Theorem 92. *Let A be an $n \times n$ matrix. Then there's a unitary matrix U such that*

$$U^* AU = T = (t_{ij}) \quad (\text{ with } t_{ij} \text{ for } i > j\ ) \tag{82}$$

The diagonal elements t_{ii} are eigenvalues of A.

Proof. Use induction.
For $n = 1$, $(1)a_{11}(1) = \lambda_1 = T$
Assume $n - 1$ case is true.
Let u_1 be a unit eigenvector of A.

$$Au_1 = \lambda_1 u_1 \quad \|u_1\| = 1$$

u_1 exists because $det(\lambda I - A) = 0$ for at least one complex number $\lambda = \lambda_1$.
Also note that we're guaranteed n complex eigenvalues since complex numbers are algebraically closed.

Since $u_1 \neq 0$, u_1 has at least one nonzero component, say the rth component. Consider u_1 along with "standard" basis elements, without e_r.

$$u_1, e_1, e_2, \ldots, e_{r-1}, e_{r+1}, \ldots, e_n$$

Use Gram-Schmidt process to orthogonalize the n vectors.

$$v_1, v_2, \ldots, v_n \quad (v_i, v_j) = \delta_{ij} \text{ and } u_1 = v_1$$

Let V be the unitary matrix with columns v_j.

AV has columns

$$Av_1 = \lambda_1 v_1; \quad Av_2, \ldots, Av_n$$

V^*AV has the first column

$$V^*(Av_1) = V^*\lambda_1 v_1 = [\lambda_1 (v_i)^* v_1] \quad (i = 1, 2, \ldots n)$$

which equals $\lambda_1 col(1, 0, 0, \ldots, 0)$ since $v_i^* v_1 = (v_i, v_1) = \delta_{i1}$

Thus V^*AV has the form

$$V^*AV = \begin{bmatrix} \lambda_1 & * & \ldots\ldots & * \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{bmatrix}$$

where the numbers $*$ are irrelevant and where B is an $(n-1) \times (n-1)$ matrix.

By induction, the $n-1$ case, we have an $(n-1) \times (n-1)$ unitary matrix W for which

$$W^*BW = T_1, \quad T_1 \equiv \text{ a triangular matrix with all zeros below the main diagonal}$$

Consider

$$Y = \begin{bmatrix} 1 & 0 & \ldots\ldots & 0 \\ 0 & & & \\ \vdots & & W & \\ 0 & & & \end{bmatrix}$$

Y is unitary because its columns are mutually orthogonal and of norm 1.

$$Y^*(V^*AV)Y = \begin{bmatrix} \lambda_1 & * & \ldots\ldots\ldots\ldots & * \\ 0 & & & \\ \vdots & & W^*BW & \\ 0 & & & \end{bmatrix} = T$$

But $W^*BW = T_1$ which has all zeros below the main diagonal.

Therefore T has all zeros below the main diagonal.

And $U = VY$ is unitary.

Thus $Y^*(V^*AV)Y = U^*AU = T$ triangulizes A.

Also note that

$$\begin{aligned} det(\lambda I - T) &= (\lambda - t_{11})(\lambda - t_{22}) \ldots (\lambda - t_{nn}) = \\ &= det(U(\lambda I - A)U^*) = det(\lambda I - A) = (\lambda - \lambda_1)(\lambda - \lambda_2) \ldots (\lambda - \lambda_n) \end{aligned}$$

since the characteristic polynomial is "invariant under a similarity transformation." □

6.10. **The Jordan Canonical Form.**

Definition 42. Zero or nonzero vector p is a principal vector of grade $g \geq 0$ belonging to eigenvalue λ_i if

$$(\lambda_i I - A)^g p = 0$$

and if there is no smaller non-negative integer $\gamma < g$ for which

$$(\lambda_i I - A)^\gamma p = 0$$

Lemma 1 (Cayley-Hamilton Theorem)**.**
If $\phi(\lambda) = det(\lambda I - A)$*, then* $\phi(A) = 0$

Proof. Define $n \times n$ matrix of signed cofactors.

$$C(\lambda) = cof(\lambda I - A)$$

For any square matrix M, $M(cofM)^T = (detM)I$ (by Theorem)

$$\Longrightarrow (\lambda I - A)(C(A))^T = (det(\lambda I - A))I = \phi(\lambda)I$$

If λ is an $n \times n$ matrix,
every component of $C(\lambda)$ is a polynomial of degree $\leq n-1$ in λ.

$$\Longrightarrow (C(\lambda))^T = \lambda^{n-1}C_0 + \lambda^{n-2}C_1 + \cdots + \lambda C_{n-2} + C_{n-1}$$

where C_i is an $n \times n$ matrix of constants.

$$\begin{aligned}
(C(\lambda))^T &= \sum_{j=0}^{n-1} \lambda^{n-1-j} C_j \\
(\lambda I - A)(C(\lambda))^T &= \sum_{j=0}^{n-1} \lambda^{n-1-j}(\lambda I - A)C_j \\
&= \sum_{j=0}^{n-1} \lambda^{n-j}C_j - \sum_{j=0}^{n-1} \lambda^{n-1-j}AC_j = \\
&= \lambda^n C_0 + \sum_{j=1}^{n-1} \lambda^{n-j}C_j - \sum_{j=1}^{n-1} \lambda^{n-j}AC_{j-1} - AC_{n-1} = \\
&= \lambda^n C_0 + \sum_{j=1}^{n-1} \lambda^{n-j}(C_j - AC_{j-1}) - AC_{n-1}
\end{aligned}$$

On the other side, we have

$$\phi(\lambda)I = \lambda^n I + \alpha_1 \lambda^{n-1} I + \cdots + \alpha_{n-1}\lambda I + \alpha_n I$$

then comparing powers of λ,

$$\begin{aligned}
\lambda^n &: I = C_0 \\
\lambda^{n-1} &: C_1 - AC_0 = \alpha_1 I \\
&\vdots \\
\lambda &: C_{n-1} - AC_{n-2} = \alpha_{n-1} I \\
1 &: -AC_{n-1} = \alpha_n I
\end{aligned} \tag{83}$$

Multiply the first equation of (83) by A^n on the left,
the second by A^{n-1}, etc., and add

$$A^n C_0 + A^{n-1}(C_1 - AC_0) + \cdots + A(C_{n-1} - AC_{n-2}) - AC_{n-1} = \phi(A)I$$

$$\begin{aligned}
C_0 &= I \\
C_j - AC_{j-1} &= \alpha_j I \\
-AC_{n-1} &= \alpha_n I
\end{aligned}$$

$$\begin{aligned}
\phi(A) &= A^n C_0 + \sum_{j=1}^{n-1} A^{n-j}(C_j - AC_{j-1}) - AC_{n-1} \\
&= A^n C_0 + \sum_{j=1}^{n-2} A^{n-j}C_j - \sum_{j=1}^{n-2} A^{n-j-1}AC_j - A^n C_0 - AC_{n-1} = 0
\end{aligned}$$

All terms on the left hand side cancel, so

$$\mathbf{0} = \phi(A)I$$

□

Lemma 2.
Let $s \geq 2$
Let $\phi_1(\lambda), \ldots, \phi_s(\lambda)$ be polynomials.
Suppose there's no number λ_0 which is a root of all those polynomials.
Then there are polynomials $\psi_1(\lambda), \ldots, \psi_s(\lambda)$ such that

$$\psi_1(\lambda)\phi_1(\lambda) + \psi_2(\lambda)\phi_2(\lambda) + \cdots + \psi_s(\lambda)\phi_s(\lambda) = 1$$

Note that if for some λ_0, $\phi_0(\lambda) = 0\,(i = 1, \ldots, s)$ *then* $0 + 0 + \cdots + 0 \neq 1$.

Proof.
Let $N = ($ degree of $\phi_1) + \cdots +$ degree of $\phi_s)$
if $N = 0$, all $\phi_i(\lambda)$ are constants.
At least one ϕ_k is nonzero. So let $\psi_k(\lambda) = \frac{1}{\phi_k}$ and let all others $psi_i(\lambda) = 0$.
$\Longrightarrow \psi_k\phi_k = 1$.
If $N \geq 1$, and assume without loss of generality,

$$\text{degree of } \phi_1(\lambda) \leq \cdots \leq \text{ degree of } \phi_s(\lambda)$$

Let $\phi_k(\lambda)$ be the first polynomial not identically zero.
Then $k < s$; otherwise, if $k = s$, every root of ϕ_s would be a root of all ϕ_i.

if $\phi_k(\lambda) =$ constant, let $\psi_k = \dfrac{1}{\phi_k}$ and $\phi_i = 0$ so $\psi_k\phi_k = 1$.

if $\phi_k(\lambda)$ is not a constant, divide all other $\phi_i(\lambda)$ by $\phi_k(\lambda)$

$$\phi_k(\lambda) = q_i(\lambda)\phi_k(\lambda) + r_i(\lambda)\,(i \neq k)OB \tag{84}$$

Note that if $i < k$, $\phi_i(\lambda) + r_i(\lambda) = 0$.

if $i > k$

degree of $r_i(\lambda)$ $<$ degree of $\phi_k(\lambda)$ and so also

degree of $r_i(\lambda)$ $<$ degree of $\phi_i(\lambda)$

Consider $\phi_k(\lambda), r_i(\lambda),\ (i \neq k)$.
$\phi_k(\lambda), r_i(\lambda)$ cannot have a common root λ_0, otherwise, by Eqn. (84), λ_0 would be a root of all polynomials $\phi_1, \ldots, \phi_s$.
Use induction on the sum of the degrees, so that we can assume this as our $N - 1$ case,

$$\psi_k^*(\lambda)\phi_k(\lambda) + \sum_{i\neq k}\psi_i^*(\lambda)r_i(\lambda) = 1 \Longrightarrow \boxed{\psi_k^*(\lambda)\phi_k(\lambda) + \sum_{i\neq k}\psi_i^*(\lambda)(\phi_i(\lambda) - q_i(\lambda)\phi_k(\lambda)) = 1}$$

where we had defined

$$\psi_k(\lambda) = \psi_k^*(\lambda) - \sum_{i\neq k}\psi_i^*(\lambda)q_i(\lambda)$$

$$\psi_i(\lambda) = \psi_i^* \text{ for } i \neq k$$

To further explain how we used induction for our "$N - 1$" case, consider

$N = ($ degree of $\phi_1)+\cdots+($ degree of $\phi_s) > 0+\cdots+($ degree of $\phi_k)+($ degree of $r_{k+1})+\cdots+($ degree of $r_s)$

Simply take the $N - 1$ case to be polynomials of $0, 0, \ldots, \phi_k, r_{k+1}, \ldots, r_s$. □

Theorem 93 (Principal Vector representation)**.**
Let A *by an* $n\times n$ *matrix with different eigenvalues* $\lambda_1, \ldots, \lambda_s$ *with multiplicities* $m_1, \ldots, m_s$.
Then every n*-component column vector* x *has a representation*

$$x = p^{(1)} + p^{(2)} + \cdots + p^{(s)} \tag{85}$$

where p^l *is a uniquely defined principal vector belonging to* λ_l *of grade* $\leq m_l$.

Proof. If $s = 1$, $m_1 = n$, then $det(\lambda I - A) = (\lambda - \lambda_1)^n$.
By Cayley-Hamilton Theorem, Lemma (1), $(A - \lambda_1 I)^n = 0$.
Then every x itself is a principal vector of grade $\leq m_1 = n$ belonging to λ_1.
If $s \geq 2$,
define

$$\phi_i(\lambda) = \prod_{j=1, j\neq i}^{s}(\lambda - \lambda_j)^{m_j}$$

Note that $(\lambda - \lambda_i)^{m_i}\phi_i(\lambda) = det(\lambda I - A) = \phi(\lambda)$
and by Cayley-Hamilton Theorem, $(A - \lambda_i I)^{m_i}\phi_i(A) = \phi(A)$

$\phi_i(\lambda)$ polynomials have no common root (you must consider each and every polynomial; for instance, while $\phi_i(\lambda), \phi_n(\lambda)$ may share many factors, there'll be a polynomial that won't contain a particular factor)

$$\psi_1(\lambda)\phi_1(\lambda) + \cdots + \psi_s(\lambda)\phi_s(\lambda) = 1$$

If $\omega(\lambda) =$ a polynomial $\equiv 1$, we must have $\omega(A) = I$ for any square matrix A.

since permissible manipulations of addition, subtraction, and (scalar) multiplication are the same for the scalar variable λ, as for a single square matrix A (helped by the fact that A commutes with I and itself).

(note in the derivation of Lemma (2), we divided by polynomials). So for

$$det(\lambda I - A) = \phi(\lambda) = (A - \lambda_1 I)(A - \lambda_2 I) \dots (A - \lambda_n I)$$

the factors commute with each other.

$$\Longrightarrow \phi_1(A)\psi_1(A) + \dots + \phi_s(A)\psi_s(A) = I$$

$$\text{multiply by } x$$

$$\Longrightarrow [\phi_1(A)\psi_1(A)x] + \dots + [\phi_s(A)\psi_s(A)x] = x$$

But each [ ] is a principal vector. To see, this consider

$(A - \lambda_i I)^{m_i}(\phi_i(A)\psi_i(A)x) = ((A - \lambda_i I)^{m_i}\phi_i(A))\psi_i(A)x = 0$ (since $\phi(A) = 0$)

To show uniqueness;

Suppose $x = q^{(1)} + \dots + q^{(s)}$,

where $q^1 \neq p^1$ and where each q^i is a principal vector of grade $\leq m_i$ belonging to λ_i

$$x = p^{(1)} + p^{(2)} + \dots + p^{(s)}$$

$$x = q^{(1)} + \dots + q^{(2)}$$

$$\Longrightarrow \begin{array}{c} 0 = r^{(1)} + \dots + r^{(s)} \\ \text{where } r^{(i)} = p^{(i)} = q^{(i)} \end{array} \tag{86}$$

Now $r^{(1)}$ is also a nonzero principal vector belonging to λ_1.

Let $r^{(1)}$ have a grade $g_1 \geq 1$.

Let $c^1 = (A - \lambda_1 I)^{g_1 - 1} r^{(1)}$; $r^{(1)} \neq 0$.

Then

c^1 is an eigenvector belonging to λ_1 since

$$(A - \lambda_1 I)c^{(1)} = 0$$

Now multiply Eqn. (86) by

$$(A - \lambda_1 I)^{g_1 - 1} \prod_{j=2}^{s} (A - \lambda_j I)^{m_j}$$

then

$$\Longrightarrow 0 = \sum_{j=2}^{s} (\lambda_1 - \lambda_j)^{m_j} c^{(1)} + 0 + \dots + 0$$

Thus, a contradiction; the right hand side is nonzero.

For further clarification, consider that $c^{(1)}$ is an eigenvector.

$$(A - \lambda_1 I)c^{(1)} = 0$$

$$Ac^{(1)} = \lambda_1 c^{(1)}$$

$$\text{so then } (A - \lambda_j I)c^{(1)} = (\lambda_1 - \lambda_j)c^{(1)}$$

$$\text{so then} \phi_1(A)c^{(1)} = \phi_1(\lambda_1)c^{(1)}$$

Also note the commutability of $(A - \lambda_1 I)(A - \lambda_j I)$

$$(A - \lambda_1 I)(A - \lambda_j I) = A^2 - \lambda_j A - \lambda_1 A + \lambda_1 \lambda_j I$$

$$(A - \lambda_j I)(A - \lambda_1 I) = A^2 - \lambda_1 A - \lambda_j A + \lambda_1 \lambda_j I$$

□

Definition 43. For any A $n \times n$ matrix with some eigenvalue λ_i, define

$$M = A - \lambda_i I$$

Then define the linear space P_g of principal vectors of grade g belonging to the eigenvalue λ_i of A, i.e.

$$P_g = \{x | M^g x = 0\} \quad (g = 0, 1, \dots, m) \tag{87}$$

where $m = m_i$ the multiplicity of λ_i

Lemma 3.
For $i = 1, \dots, s$, let B_i be any basis for the linear space of principal vectors of grade $\leq m_i$ belonging to λ_i.
Then the collection of vectors $B_1, B_2, \dots, B_n$ is a basis for the n-dimensional space E^n.

Proof. By the Principal Vector Representation Theorem, Theorem (93) above, $\forall x \in E^n$ has a unique representation in principal vectors. So

$$x = p^{(1)} + p^{(2)} + \cdots + p^{(s)} \tag{88}$$

where $p^{(i)}$ is a principal vector of grade $\leq m_i$ belonging to λ_i.

Since B_i is a basis, $p^{(i)}$ has a unique representation as a linear combination of the vectors comprising B_i.

So Eqn. (88) implies $\forall \in E^n$ has a unique representation of vectors comprising $B_1, B_2, \dots,$ and B_s.

Note that the vectors in $B_1, \dots, B_s$ are linearly independent also because $x = 0$ has the unique representation (88) with all $p^i = 0$ (i.e. if $x = 0$ then $x = 0 + 0 + \cdots + 0$ is a representation; since it's unique, it must be the only one). $\square$

We can, with Lemma (3), consider separately the spaces of principal vectors belonging to $\lambda_i, \dots, \lambda_s$.

Let's define Jordan basis.

Definition 44 (Jordan basis)**.** A **Jordan basis for** P_g , J, where

$$J = (v_1^1, v_2^1, \dots, v_{l_1}^1, v_1^2, \dots, v_{l_2}^2, \dots, v_1^m, \dots, v_{l_m}^m) \tag{89}$$

is a basis for P_g and obeying the following:

$$\begin{array}{cccc} v_1^1 & v_1^2 & \dots & v_1^m \\ v_2^1 = Mv_1^1 & v_2^2 = Mv_1^2 & \dots & v_2^m = Mv_1^m \\ v_3^1 = Mv_2^1 & v_3^2 = Mv_2^2 & \dots & v_3^m = Mv_2^m \\ \vdots & \vdots & \vdots & \vdots \\ v_{l_1}^1 = Mv_{l_1-1}^1 & v_{l_2}^2 = Mv_{l_2-1}^2 & \dots & v_{l_m}^m = Mv_{l_m-1}^m \\ Mv_{l_1}^1 = 0 & Mv_{l_2}^2 = 0 & \dots & Mv_{l_m}^m = 0 \end{array} \tag{90}$$

where each column or **chain** has length $\leq g$

Lemma 4 (Jordan basis existence)**.** *The space P_m defined in Eqn. (87) has a Jordan basis.*

Proof.
If $m = 1$, $P_1 = \{x | Mx = 0\}$

If $x_1^1, \dots, x_m^1$ is a basis for P_1, each x_j^1 is an eigenvector of M.

Thus any basis for P_1 is a Jordan basis (by Jordan basis definition).

If $m > 1$,

Let $y^1, \dots, y^\beta$ be a basis for P_{g-1}, where $g \geq 2$. Let

$$x^1, \dots, x^\alpha, y^1, \dots, y^\beta \tag{91}$$

be a basis for P_g formed by appending any necessary additional vectors $x^1, \dots, x^\alpha$.

If $\alpha = 0$, $P_g = P_{g-1}$ and there are no x's to consider.

Suppose $\alpha > 1$.

Suppose $g = 2$.

Form $Mx^1, \dots, Mx^\alpha$.

Suppose

$$a_1 x^1 + \cdots + a_\alpha x^\alpha + b_1 M x^1 + \cdots + b_\alpha M x^\alpha = \sum_{j=1}^{\alpha} (a_j x^j + b_j M x^j) = 0$$

$$\text{(Multiply both sides by } M \text{)} \implies M \left(\sum_{j=1}^{\alpha} (a_j x^j + b_j M x^j) \right) = M(a_1 x^1 + \cdots + a_\alpha x^\alpha) + 0 + \cdots + 0 = 0$$

since $M^2 x^i = 0$ for $x^i \in P_2$

So then $\sum_{j=1}^{\alpha} a_j x^j \in P_1$, so $\sum_{j=1}^{\alpha} a_j x^j = \sum_{j=1}^{\beta} \gamma_j y_j$.
$\implies a_j = 0$ since $x^1, \ldots, x^\alpha, y^1, \ldots, y^\alpha$ are given as being independent.

$\implies M(\sum b_i x^i) = 0$ since $a_j = 0$.

$$\sum_{j=1}^{\alpha} b_j x^j = \sum_{i=1}^{\beta} (\gamma_j) y_i$$

$$\sum_{j=1}^{\alpha} b_j x^j + \sum_{i=1}^{\beta} (-\gamma_j) y_i = 0$$

$b_j = 0$ by the independence of x^j, y^i's

Thus, the 2α vectors $x^1, \ldots, x^\alpha, Mx^1, \ldots, Mx^\alpha$ are independent.
$Mx_1, \ldots, Mx^\alpha$ are independent vectors in P_1.
If $Mx^1, \ldots Mx^\alpha$ do not span P_1, adjoin $z^{\alpha+1}, \ldots, z^\beta$ so that $Mx^1, \ldots, Mx_\alpha, z^{\alpha+1}, \ldots, z^\beta$ are a basis for P^1.
So we've shown P_2 to have a Jordan basis.

$$\begin{array}{ccccc} x^1 & x^2 & \ldots & x^\alpha & \\ Mx^1 & Mx^2 & \ldots & Mx^\alpha & z^{\alpha+1}, \ldots, z^\beta \end{array}$$

(Note that $z^{\alpha+1}, \ldots, z^\beta$ are in P_1, and so are already eigenvectors).
We've proven the $g = 2$ case of the following assertion:
Assertion: The vectors $y^1, \ldots, y^\beta$ in Eqn. (91) may be replaced by vectors $z^1, \ldots, z^\beta$ such that

$$x^1, \ldots, x^\alpha, z^1, \ldots, z^\beta$$

is a Jordan basis for P_g.
So for $g = 2$ case, in this notation, $Mx^1 \equiv z^1, \ldots Mz^\alpha \equiv z^\alpha$.
Consider $g > 2$.
If $\alpha = 0$, $P_g = P_{g-1}$, and the assertion follows by induction, since P_{g-1} has a Jordan basis.
Suppose $\alpha \geq 1$. Form $Mx^1, \ldots, Mx^\alpha$
Consider

$$\sum a_i x^i + \sum b_i M x^i = 0$$

$$\text{(Multiply both sides by } M^{g-1}\text{)} \implies M^{(g-1)} \left(\sum_{j=1}^{\alpha} (a_j x^j \right) + \sum 0 = 0$$

since $M^g x^i = 0$ for $x^i \in P_g$

$x^1, \ldots, x^\alpha$ are independent vectors in P_g that are also independent of $y^1, \ldots, y^\beta$, which form a basis for P_{g-1}.

So then $M^{g-1}(\sum a_i x^i) = 0$ implies $\sum_{j=1}^{\alpha} a_j x^j \in P_{g-1}$, so $\sum_{j=1}^{\alpha} a_j x^j = \sum_{i=1}^{\beta} \gamma_i y_i$.
$\implies a_j = 0$ since $x^1, \ldots, x^\alpha, y^1, \ldots, y^\alpha$ are given as being independent.

$$\implies \sum b_i M x^i = M(\sum b_i x^i) = 0 \text{ since } a_j = 0. \text{ Thus } \sum b_i x^i \in P_1 \subset P_{g-1}$$

$$\implies \sum b_i x^i = \sum_{j=1}^{\beta} \gamma_j y^j \implies b_i = 0 \text{ since } x^i, y^j \text{ are independent}$$

Thus x^i, Mx^i are 2α independent vectors.
Let $w^1, \ldots, w^\gamma$ be any basis for P_{g-2}.

Consider $Mx^1, \ldots, Mx^\alpha; w^1, \ldots, w^\gamma \in P_{g-1}$
These vectors are independent since if

$$\sum a_i M x^i + \sum b_j w^j = 0$$

$$\text{then } M^{g-2}(\sum a_i M x^i) + M^{g-2} \sum b_j w^j = M^{g-1}(\sum a_i x^i) + 0 = 0$$

So then $M^{g-1}(\sum a_i x^i) = 0$ implies $\sum_{j=1}^{\alpha} a_j x^j \in P_{g-1}$, so $\sum_{j=1}^{\alpha} a_j x^j = \sum_{i=1}^{\beta} \gamma_i y_i$.
$\implies a_j = 0$ since $x^1, \ldots, x^\alpha, y^1, \ldots, y^\alpha$ are given as being independent.

$$\Longrightarrow \sum b_j w^j = 0 \text{ since } a_j = 0.$$
$$\Longrightarrow b_i = 0 \text{ since } w^j \text{ are independent}$$

If $Mx^1, \dots, Mx^\alpha, w^1, \dots, w^\gamma$ are not a basis for P_{g-1}, adjoin $q^1, \dots, q^\delta$, so that

$$Mx^1, \dots, Mx^\alpha, q^1, \dots, q^\delta, w^1, \dots, w^\gamma \tag{92}$$

form a basis for P_{g-1}.

By induction case, we can apply the assertion to the $g-1$ case. We can replace $w^1, \dots, w^\gamma$ in Eqn. (**??**) by a new basis $z^1, \dots, z^\gamma$ for P_{g-2} so that

$$Mx^1, \dots, Mx^\alpha, q^1, \dots, q^\delta, z^1, \dots, z^\gamma \tag{93}$$

is a Jordan basis for P_{g-1}.

Recall that we had to add $x^1, \dots, x^\alpha$ to make a P_g basis from a P_{g-1} basis. So append the x^i's to the Jordan basis for P_{g-1} given in Eqn. (93) to form a basis for P_g. But already the x^i's satisfy the condition to be a Jordan basis (Eqn. (90)). Then

$$\begin{matrix} x^1 & x^2 & \dots & x^\alpha \\ Mx^1 & Mx^2 & \dots & Mx^\alpha, \end{matrix} \quad q^1, \dots q^{\delta} x^1, \dots, z^\gamma$$

□

Lemma 5.
Let λ_i be an eigenvalue of multiplicity m_i belonging to $n \times n$ matrix A.
Let B_i be a Jordan basis for x such that

$$(A - \lambda_i I)^{m_i} x = 0$$

Then, if B_i is a matrix whose columns are basis vectors

$$AB_i = B_i \Lambda_i \tag{94}$$

where Λ_i has the form

$$\Lambda_i = \begin{bmatrix} \lambda_i & 0 & 0 & \dots & 0 & 0 \\ * & \lambda_i & 0 & \dots & 0 & 0 \\ 0 & * & \lambda_i & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_i & 0 \\ 0 & 0 & 0 & \dots & * & \lambda_i \end{bmatrix} \tag{95}$$

Proof.
Let $M = A - \lambda_i I$.
Let Jordan basis B_i be a matrix with elements in B_i as columns:

$$B_i = [v_1^1, \dots, v_{l_i}^1, v_1^2, \dots, v_{l_2}^2, \dots, v_1^\alpha, \dots, v_{l_\alpha}^\alpha]$$
$$\text{(multiply by M)} \Longrightarrow MB_i = [Mv_1^1, \dots, Mv_{l_1}^1, Mv_1^2, \dots, Mv_{l_2}^2, \dots, Mv_1^\alpha, \dots, Mv_{l_\alpha}^\alpha]$$
$$= [v_2^1, \dots, 0, \dots, v_2^2, \dots, 0, \dots, v_2^\alpha, \dots, 0]$$
$$MB_i = AB_i - \lambda_i B_i; \; AB_i = \lambda_i B_i + MB_i$$
$$AB_i = [\lambda_i v_1^1 + v_2^1, \dots, \lambda_i v_{l_1}^1 + 0, \lambda_i v_1^2 + v_2^2, \dots, \lambda_i v_{l_2}^2, \dots, \lambda_i v_1^\alpha + v_2^\alpha, \dots, \lambda_i v_{l_\alpha}^\alpha]$$
$$= [v_1^1, \dots, v_{l_1}^1, v_{l_2}^2, \dots, v_1^\alpha, \dots, v_{l_\alpha}^\alpha]\Lambda_i$$

provided that

$$** \cdots * = \underbrace{1 \dots 1}_{l_1 - 1} 0 \underbrace{1 \dots 1}_{l_2 - 1}, 0, \dots, \underbrace{1 \dots 1}_{l_{\alpha-1} - 1} 0; \underbrace{1 \dots 1}_{l_\alpha - 1} 0$$

□

Theorem 94 (Jordan Canonical Form Theorem)**.** *Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \dots, \lambda_s$ with multiplicities $m_1, \dots, m_s$, i.e.*

$$det(\lambda I - A) = \prod_{j=1}^{s} (\lambda - \lambda_j)^{m_j}$$

Then A is similar to J of the form

$$J = \begin{bmatrix} \Lambda_1 & 0 & \dots & 0 \\ 0 & \Lambda_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \Lambda_s \end{bmatrix} \tag{96}$$

where Λ_i is an $m_i \times m_i$ matrix of the form

$$\Lambda_i = \begin{bmatrix} \lambda_i & 0 & 0 & \dots & 0 & 0 & 0 \\ * & \lambda_i & 0 & \dots & 0 & 0 & 0 \\ 0 & * & \lambda_i & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & * & \lambda_i & 0 \\ 0 & 0 & 0 & \dots & 0 & * & \lambda_i \end{bmatrix} \tag{97}$$

or i.e.

$$(\Lambda_k)_{ij} = \lambda_k \delta_{ij} + (*)\delta_{i-1,j}$$

with each $*$ equal to 0 *or* 1.

Proof. Form $B = [B_1, B_2, \dots, B_s]$.
By Lemma (3), the columns of B are a basis for E^n.
Thus B is an $n \times n$ matrix with inverse B^{-1}.

$$\begin{aligned} AB &= [AB_1, AB_2, \dots, AB_s] \\ &= [B_1\Lambda_1, B_2\Lambda_2, \dots, B_s\Lambda_s] \\ [B_1, B_2, \dots, B_s] &\begin{bmatrix} \Lambda_1 & 0 & \dots & 0 \\ 0 & \Lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Lambda_s \end{bmatrix} = BJ \\ &\implies B^{-1}AB = J \end{aligned}$$

Now, we want to show that Λ_i is an $m_i \times m_i$ matrix.
Since A is similar to J,

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

$$\implies \prod_{i=1}^{s} (\lambda - \lambda_i)^{m_i} = \prod_{i=1}^{s} det(\lambda I - \Lambda_i) \tag{98}$$

Now suppose Λ_i is an $o_i \times o_i$ matrix.
Then triangular Λ_i has characteristic determinant

$$det(\lambda I - \Lambda_i) = (\lambda - \lambda_i)^{o_i}$$

since $\lambda_1, \dots, \lambda_s$ are distinct eigenvalues of A, matching term by term $(\lambda - \lambda_i)$ in Eqn. (98) implies $m_i = 0$.

In Eqn. (94), $AB_i = B_i\Lambda_i$, so the number of columns of B_i equals the order of Λ_i. Thus, the order of Λ_i is m_i.

Thus, **the multiplicity, m_i, of λ_i equals the dimension of the space of principal vectors of grade $\leq m_i$ belonging to λ_i** □

7. Differential Calculus of Scalar and Vector Fields (Apostol, Vol. 2, Ch. 8)

7.1. **Open balls, open sets.**

$T : \mathbb{R}^n \to \mathbb{R}^m$
Let $a \in \mathbb{R}^n$, $r > 0$

open ball $= \{x, \, x \in \mathbb{R}^n \mid \|x - a\| < r\} = B(a) = B(a; r)$

Definition 45 (Definition of an Interior Point)**.**
Let $S \subseteq \mathbb{R}^n$, $a \in S$
if $\exists\, B(a;r) \subseteq S$

$$a = \text{ interior point of } S$$

$$\{ \text{ int. pt. of } S \} = \text{ interior of } S = intS = \{a | \exists \text{ at least one } B(a) \subseteq S\}$$

Definition 46 (Definition of an Open Set)**.** $S \subseteq \mathbb{R}^n$ is open if all its pts. S are interior pts., i.e.

S open iff $S = intS$

$$\text{neighborhood of } a = \text{ open set containing } a$$

Definition 47 (Definition of exterior and boundary)**.**
Exterior pt. x to set $S \subseteq \mathbb{R}^n$ if $\exists$ n-ball $B(x)$ containing no pts. of S i.e.

$$\text{If } \exists\, B(x) \text{ s.t. } \forall\, x_1 \in B(x),\ x_1 \notin S, \text{ then } x \text{ is exterior to } S$$

$\{$ ext. pt. of S $\} = extS$.

A pt. that's neither ext. to S nor int. to S is a boundary pt. of S
$\{$ boundary pts. of S $\} = \partial S$, i.e. the set of all boundary pts. of S is called the boundary of S and is denoted by ∂S.

7.2. **Limits and continuity (Apostol's 8.4, Vol. 2).**

f is continuous at a if $\exists\, f(a)$ and if

$$\lim_{x \to a} f(x) = f(a) \Longleftrightarrow \forall\, \epsilon > 0,\ \exists\, \delta > 0 \text{ s.t. } |f(x) - f(a)| < \epsilon \text{ if } \|x - a\| < S$$

Important example: Example 7. f is not continuous on all paths for $f(x,y) = \frac{xy}{x^2+y^2}$ if $(x,y) \neq (0,0)$; $f(0,0) = 0$.

$$\left.\begin{array}{l} \text{if } x \neq 0 \text{ constant}, \lim\limits_{y \to 0} f = 0 \\ \text{if } y \neq 0 \text{ constant}, \lim\limits_{x \to 0} f = 0 \end{array}\right. \text{but if } y = x, \text{ along this line, } f = 1/2 \text{ (!!!)}$$

Definition 48 (Definition of the derivative of a scalar field with respect to a vector)**.** Given $f : S \to \mathbb{R}$; $S \subseteq \mathbb{R}^n$
Let $a \in intS$, $y \in \mathbb{R}^n$

$$f'(a;y) = \lim_{h \to 0} \frac{f(a+hy) - f(a)}{h} \tag{99}$$

Definition 49 (Definition of directional and partial derivatives)**.** If y s.t. $|y| = 1$, $f'(a;y) =$ directional derivative of f at a in y direction.
$\partial_k f(a) = f'(a;e_k) =$ partial derivative.

7.3. **Total derivative.** Consider this illustrative example.
Let $f(x,y) = \frac{xy^2}{x^2+y^4}$ if $x \neq 0$, $f(0,y) = 0$.
Let $a = (0,0)$, $t = (a,b)$

$$\frac{f(a+ht) - f(a)}{h} = \frac{f(ht) - f(0)}{h} = \frac{f(ha,hb)}{h} = \frac{ab^2}{a^2 + h^2 b^4} \xrightarrow{h \to 0} b^2/a$$

$$f'(0;t) = b^2/a$$

For $t = (a,b)$, $f'(0;t) = 0$.
Thus, $f'(0;t)$ exists for all directions of t.
For $y = kx$ and $x = 0$, $f(x,y) = \frac{k^2x^3}{x^2+k^4+x^4} \xrightarrow{x \to 0} 0$,
but for $x = y^2$, $f(x,y) = 1/2$

The existence of all directional derivatives at a point fails to imply continuity at that point.

Let's review a number of important concepts with R^n fields. Differentiability must be redefined through a $n - dim$ Taylor expansion.

Definition 50 (Definition of a Differentiable Scalar Field)**.**
Let $f : S \to \mathbb{R}$
Let a be an int. pt. of S.
Let $B(a;r)$ s.t. $B(a;r) \subseteq S$
Let v s.t. $\|v\| < r$, so $a + v \in B(a;r)$ Then

f diff. at a
if $\exists\, T_a,\ E$ s.t.
linear $T_a : \mathbb{R}^n \to \mathbb{R}$
scalar $E(a,v),\ E(a,v) \to 0$ as $\|v\| \to 0$ and

$$f(a+v) = f(a) + T_a(v) + \|v\|E(a,v) \tag{100}$$

The next theorem shows that if the total derivative exists, it is unique. It also tells us how to compute $T_a(y)$, $\forall\, y \in \mathbb{R}^n$.

Theorem 95 (Uniqueness of total derivative)**.** *Assume f diff. at a with total derivative T_a*
Then $\exists\, f'(a;y) \quad \forall y \in \mathbb{R}^n$ and

$$T_a(y) = f'(a;y)$$

Also,

$$f'(a;y) = \sum_{j=1}^n D_j f(a) y_j \text{ for}$$
$$y = (y_1, \dots, y_j, \dots, y_n)$$

Proof.
If $y = 0$, $\quad T_a(0) = 0$ and $f'(a;0) = 0$. Done.

Suppose $y \neq 0$

$$f(a+v) = f(a) + T_a(v) + \|v\|E(a,v) \qquad \text{(since we assume } f \text{ diff.)}$$

$$v = hy$$

$$\implies \frac{f(a+hy) - f(a)}{h} = \frac{1}{h}T_a(hy) + \frac{\|hy\|}{h}E(a,hy) \xrightarrow{h\to 0} f'(a,y) = T_a(y) + 0$$

Now use linearity of T_a:

$$T_a(y) = \sum T_a(y_j e_j) = \sum y_j T_a(e_j) = \sum y_j f'(a;e_j) = \sum y_j D_j f(a)$$

□

Then the gradient was introduced, $\nabla f(a) = (\partial_1 f(a), \dots, \partial_n f(a))$ so that
$f'(a;y) = \sum_{j=1}^n \partial_j f(a) y_j = \nabla f(a) \cdot y$ so then also

$$\implies f(a+v) = f(a) + \nabla f(a) \cdot v + \|v\|E(a;v)$$

Theorem 96 (Differentiability implies Continuity)**.**
If a scalar field f is differentiable at a, then f is cont. at a

Proof. Since f is diff.
$|f(a+v) - f(a)| = |\nabla f(a) \cdot v + \|v\|E(a,v)|$
By Cauchy-Schwarz inequality,

$0 \leq |f(a+v) - f(a)| \leq \|\nabla f(a)\|\|v\| + \|v\||E(a;v)|$

As $v \to 0$, $|f(a+v) - f(a)| \to 0$ so f cont. at a. □

If f is diff. at a, then all its partials exist (but the converse isn't true).
existence of partials doesn't necessarily imply f is diff.
e.g. $f(x,y) = \frac{xy^2}{x^2+y^4}$

Theorem 97 (Sufficient Condition for Differentiability)**.**
Assume $\exists\, \partial_1 f, \dots, \partial_n f$ in some n-ball $B(a)$ and are cont. at a. Then f diff. at a.

Proof.
Let $\lambda = \|v\|$; then $v = \lambda u, \quad \|u\| = 1$
Express $f(a+v) - f(a)$ as a telescoping sum.

$$f(a+v) - f(a) = f(a+\lambda u) - f(a) = \sum_{k=1}^{n} (f(a+\lambda v_k) - f(a+\lambda v_{k-1}))$$

where $\{v_k\}$ s.t. $\begin{array}{c} v_0 = 0 \\ v_n = u \end{array}$. Then choose v_k's s.t.

$$\begin{aligned} v_k &= v_{k-1} + u_k e_k \\ v_1 = u_1 e_1; \quad v_2 &= u_1 e_1 + u_2 e_2, \dots v_n = u_1 e_1 + \dots + u_n e_n \\ f(a+\lambda v_k) - f(a+\lambda v_{k-1}) &= f(a+\lambda v_{k-1} + \lambda u_k e_k) - f(a+\lambda v_{k-1}) = \\ &= f(b_k + \lambda u_k e_k) - f(b_k) \end{aligned}$$

b_k, $b_k + \lambda u_k e_k$ differ only by their kth component so apply the mean value theorem

$$\implies f(b_k + \lambda u_k e_k) - f(b_k) = (\lambda u_k)\partial_k f(c_k)$$

$$\text{as } b_k \to a, \text{ as } \lambda \to 0, \text{ so } c_k \to a$$
$$\implies f(a+v) - f(a) = \lambda \sum_{k=1}^{n} u_k \partial_k f(c_k)$$

Now $\nabla f(a) \cdot v = \lambda \sum u_k \partial_k f(a)$.
$\implies f(a+v) - f(a) - \nabla f(a) \cdot v = \lambda \sum u_k(\partial_k f(c_k) - \partial_k f(a)) = E(a,v)$

$c_k \to a$ as $\|v\| \to 0$, and given $\partial_k f$ are cont., $E(a,v) \to 0$ as $\|v\| \to 0$.
By def. of diff., f is diff. $\square$

Definition 51 (Vector field derivative)**.** Let $f : S \to \mathbb{R}^m; \;\; S \subseteq \mathbb{R}^n$
if $a \in intS, \;\; y \in \mathbb{R}^n$

$$f'(a;y) = \lim_{h\to 0} \frac{f(a+hy) - f(a)}{h} \tag{101}$$

Definition 52 (f differentiable)**.** f diff. at int. pt. a if

$$T_a : \mathbb{R}^n \to \mathbb{R}^m$$
$$f(a+v) = f(a) + T_a(v) + \|v\|E(a;v) \text{ where } E(a;v) \to 0 \text{ as } v \to 0 \tag{102}$$

Theorem 98 (Apostol's Thm. 8.9)**.** *Assume f diff. at a with T_a,*
Then $\exists\, f'(a;y) \quad \forall\, y \in \mathbb{R}^n$ and

$$T_a(y) = f'(a;y) \tag{103}$$

Moreover, if $f = (f_1, \dots, f_m)$; $y = (y_1, \dots, y_n)$

$$T_a(y) = \sum_{j=1}^{m} \nabla f_j(a) \cdot y e_i = (\nabla f_1(a) \cdot y, \dots, \nabla f_m(a) \cdot y) \tag{104}$$

Proof. Let $v = hy$

$$\begin{aligned} f(a+hy) - f(a) = T_a(hy) + \|hy\|E(a,v) &= hT_a(y) + |h|\|y\|E(a;v) \\ \frac{f(a+hy) - f(a)}{h} &\xrightarrow{h\to 0} f'(a;y) = T_a(y) \\ f'(a;y) = \sum_{j=1}^{m} f_i'(a;y)e_i &= \sum_{j=1}^{m} \nabla f_i(a) \cdot y e_i \end{aligned}$$

$\square$

7.3.1. *Matrix formation of the total derivative.*

$$T_a(y) = Df(a)y$$

$$Df(a) = \begin{bmatrix} D_1f_1(a) & D_2f_1(a) & \dots & D_nf_1(a) \\ D_1f_2(a) & D_2f_2(a) & \dots & D_nf_2(a) \\ \vdots & \vdots & \ddots & \vdots \\ D_1f_m(a) & D_2f_m(a) & \dots & D_nf_m(a) \end{bmatrix}$$

$$(T_a(y))_j = \sum_{k=1}^{n} (Df(a))_{jk}y_k = \nabla f_j(a) \cdot y = (\partial_k f_j)(a)y_k$$

Interesting to note that $f\nabla = Df$; $(Df)_{jk} = \partial_k f_j$

Theorem 99. *If f diff. at a, f cont. at a.*

Proof.

$$f(a+v) = f(a) + f'(a)(v) + \|v\|E(a;v)$$

$v \to 0$, $\|v\|E(a,v) \to 0$; $f'(a)(v) \to 0$ since $f'(a)(v)$ is a linear transformation continuous at 0. ☐

$\|f'(a)(v)\| \leq M_f(a)\|v\|$; where $M_f(a) = \sum_{j=1}^{m} \|\nabla f_j(a)\|$
since

$$\|f'(a)(v)\| = \| \sum_{j=1}^{m} (\nabla f_j(a) \cdot v)e_j \| \leq \sum_{j=1}^{m} |\nabla f_j(a) \cdot v| \leq \sum_{j=1}^{m} \|\nabla f_j(a)\|\|v\|$$

7.4. **The chain rule for derivatives of vector fields (Apostol's 8.20, Vol. 2).**

Theorem 100 (Chain Rule)**.**
Let f and g be vector fields s.t. $h = f \circ g$ defined in a neighborhood of pt. a
Assume g diff. at a, total derivative $g'(a)$
Let $b = g(a)$
Assume f diff. at b, total derivative $f'(b)$

$$h'(a) = f'(b) \circ g'(a) \tag{105}$$

Proof.

$$h(a+y) - h(a) = f(g(a+y)) - f(g(a)) = f(b+v) - f(b)$$

Let $v = g(a+y) - g(a)$

$$v = g'(a)y + \|y\|E_g(a,y) \text{ where } E_g(a,y) \to 0 \text{ as } y \to 0$$

$$f(b+v) - f(b) = f'(b)v + \|v\|E_f(b,v) \text{ where } E_f(b,v) \to 0 \text{ as } v \to 0$$

$$f(b+v) - f(b) = f'(b)g'(a)y + f'(b)\|y\|E_g(a,y) + \|v\|E_f(b,v) = f'(b)g'(a)y + \|y\|E(a;y)$$

where $E(a,0) = 0$ and $E(a,y) = f'(b)E_g(a,y) + \frac{\|v\|}{\|y\|}E_f(b,v)$

Want: $E(a,y) \to 0$ as $y \to 0$

$$f'(b)E_g(a,y) \to 0 \text{ since } E_g(a,y) \to 0 \text{ as } y \to 0$$

$$\|v\| \leq M_g(a)\|y\| + \|y\||E_g(a,y)|$$

$$\text{then } \frac{\|v\|}{\|y\|}E_f(b,v) \to 0 \text{ as } y \to 0$$

☐

$$h'(a) = f'(b) \circ g'(a)$$

Since composition of linear transformations corresponds to multiplication of matrices.

$$Dh(a) = Df(b)Dg(a); \; b = g(a)$$

$$a \subseteq \mathbb{R}^p; \quad g(a) \in \mathbb{R}^n; \quad f(b) \in \mathbb{R}^m$$

$$(Dh(a))_{jk} = (\partial_k h_j)(a) = \sum_{l=1}^{n} (Df(b))_{jl}(Dg(a))_{lk} = \sum_{l=1}^{n} (\partial_l f_j(b))(\partial_k g_l(a))$$

Theorem 101 (Sufficient condition for equality of mixed partial derivatives, Apostol's Thm. 8.12)**.** *Assume $D_1 f, D_2 f, D_{1,2} f, D_{2,1} f$ exist on open S.*
If $D_{1,2}f, D_{2,1}f$ are cont.,

$$D_{1,2}f(a,b) = D_{2,1}f(a,b)$$

Proof. Consider a rectangle $R(h,k)$ with vertices

$$\begin{matrix} (a,b+k) & (a+h,b+k) \\ (a,b) & (a+h,b) \end{matrix}$$

Consider $\Delta(h,k) = f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b)$
Let $G(x) = f(x,b+k) - f(x,b) \quad \forall\, x \in (a,a+h)$.
(geometrically, we are considering the values of f at those points at which an arbitrary vertical line cuts the horizontal edges of $R(h,k)$.

$$\Delta(h,k) = G(a+h) - G(a) \xrightarrow{\text{1 dim. mean-value thm.}} G(a+h) - G(a) = hG'(x_1) \quad x_1 \in (a,a+h)$$

$$G'(x) = D_1 f(x,b+k) - D_1 f(x,b)$$

$$\Longrightarrow \Delta(h,k) = h[D_1 f(x_1,b+k) - D_1 f(x_1,b)] \xrightarrow{\text{1 dim. mean-value thm.}}$$

$$\xrightarrow{\text{1 dim. mean-value thm.}} \Delta(h,k) = hkD_{2,1}f(x_1,y_1) \quad y_1 \in (b,b+k)$$

(x_1,y_1) lies somewhere in rectangle $R(h,k)$.
Apply the same procedure to $H(y) = f(a+h,y) - f(a,y)$

$$\Delta(h,k) = hkD_{1,2}f(x_2,y_2) \text{ where } (x_2,y_2) \in R(h,k)$$

$$\Longrightarrow D_{1,2}f(x_1,y_1) = D_{2,1}f(x_2,y_2)$$

Let $(h,k) \to (0,0)$ and use the continuity of $D_{1,2}f$, $D_{2,1}f$ $\square$

7.5. **A first-order partial differential equation with constant coefficients (Apostol's Sec. 9.2, Vol. 2).** Consider

$$a\partial_x f + b\partial_y f = (a,b)\cdot(\nabla f) = A \cdot \nabla f = 0$$

Now $r' \cdot (\nabla f)(r) = 0$ when $f(r) = c$ (level curves),
So in this case, when $y = \frac{b}{a}x + c$ or $bx - ay = c$, then f is constant. Then

$$f = g(bx - ay)$$

(because f only changes by the value of $bx - ay$)

Check: $\begin{aligned} \partial_x f &= bg'(bx-ay) \\ \partial_y f &= -ag'(bx-ay) \end{aligned} \qquad a\partial_x f + b\partial_y f = abg' - abg' = 0$

Suppose $a\partial_x f + b\partial_y f = 0$

$x = Au + By$

$y = Cu + Dv$

$h(u,v) = f(Au + Bv, Cu + Dv)$

Choose A, B, C, D, arbitrarily, s.t. $\partial_u h = 0$. $\partial_u h = A\partial_x f + C\partial_y f = (A + C\left(\frac{-a}{b}\right))\partial_x f$

Let $A = \frac{a}{b}C \Longrightarrow \begin{aligned} x &= \frac{a}{b}Cu + Bv \\ y &= Cu + Dv \end{aligned}$

Then for $A = \frac{a}{b}C$, $h(u,v)$, by $\partial_u h = 0$, is a function of v alone.

$$\Longrightarrow h(u,v) = g(v)$$

Then

$$g\left((x - \frac{a}{b}y)/(B - \frac{a}{b}D)\right) = g\left(\frac{bx - ay}{bB - aD}\right) = g_1(bx - ay)$$

Theorem 102 (Apostol's Thm. 9.3)**.** *Given*

$$F(x_1,\dots,x_n) = 0$$

$$x_n = f(x_1,\dots,x_{n-1})$$

then $\partial_k f = -\frac{D_k f}{D_n f}$
where $D_k F$, $D_n F$ are evaluated at $(x_1,x_2,\dots,x_{n-1},f(x_1,\dots,x_{n-1}))$

Proof.

$$F(x_1,\dots,x_n) = F(x_1,\dots,x_{n-1}, f(x_1,\dots,x_{n-1}))$$

$$\frac{d}{dx_k}F = \partial_k F + (\partial_k f)\partial_n F = 0 \Longrightarrow \partial_k f = \frac{-\partial_k F}{\partial_n F}$$

□

Alternative view; suppose we have 2 surfaces on an n-dim. hyperspace.

$$F(x_1,\dots,x_n) = 0 \qquad G(x_1,\dots,x_n) = 0$$

If the surfaces intersect along a curve in the n-dim. hyperspace,

then we could solve $\begin{matrix} F = 0 \\ G = 0 \end{matrix}$ to obtain a parametrized C.

Suppose it's possible to solve for x_j; $j = 1,\dots,n-1$

$$x_j = x_j(x_n)$$

Then $\begin{matrix} F = 0 \\ G = 0 \end{matrix}$ for $X = (x_1(x_n),\dots,x_{n-1}(x_n), x_n)$

Let $\begin{matrix} f(x_n) = F(x_1(x_n),\dots,x_{n-1}(x_n),x_n) \\ g(x_n) = G(x_1(x_n),\dots,x_{n-1}(x_n),x_n) \end{matrix}$

$$f'(x_n) = \sum_{j=1}^{n-1} \partial_{x_n} x_j \partial_{x_j} F + \partial_{x_n} F = 0 \qquad \sum_{j=1}^{n-1} \partial_{x_n} x_j \partial_{x_j} F = -\partial_{x_n} F$$

$$g'(x_n) = \sum_{j=1}^{n-1} \partial_{x_n} x_j \partial_{x_j} G + \partial_{x_j} G = 0 \qquad \sum_{j=1}^{n-1} \partial_{x_n} x_j \partial_{x_j} G = -\partial_{x_n} G$$

e.g. $n = 3$

$$\begin{matrix} x'F_x + y'F_y = -F_z \\ x'G_x + y'G_y = -G_z \end{matrix} \quad \Longrightarrow \quad \begin{matrix} X' = -\dfrac{\begin{vmatrix} F_z & F_y \\ G_z & G_y \end{vmatrix}}{\begin{vmatrix} F_x & F_y \\ G_x & G_y \end{vmatrix}} \\ \\ Y' = -\dfrac{\begin{vmatrix} F_x & F_z \\ G_x & G_z \end{vmatrix}}{\begin{vmatrix} F_x & F_y \\ G_x & G_y \end{vmatrix}} \end{matrix}$$

$$\frac{\partial(f_1,\dots,f_n)}{\partial(x_1,\dots,x_n)} = \begin{vmatrix} \partial_{x_1} f_1 & \partial_{x_2} f_1 & \dots & \partial_{x_n} f_1 \\ \vdots & \vdots & \ddots & \vdots \\ \partial_{x_1} f_n & \partial_{x_2} f_n & \dots & \partial_{x_n} f_n \end{vmatrix} \quad \text{(Jacobian notation)}$$

$$\Longrightarrow \begin{matrix} X' = \dfrac{\partial(F,G)/\partial(y,z)}{\partial(F,G)/\partial(x,y)} \\ Y' = \dfrac{\partial(F,G)/\partial(z,x)}{\partial(F,G)/\partial(x,y)} \end{matrix}$$

Definition 53. a is a relative maximum (relative minimum) iff $\exists$ open ball $B_a(r) \subseteq D$ s.t. $\forall\, x \in B_a(r),\ \ f(x) \le f(a)\ \ (f(x) \ge f(a))$.

Lemma 6. *If a is a relative extrema of diff. scalar f, then $\nabla f(a) = 0$.*

Proof. $\forall\, j = 1,\dots,n$, let $g_j(t) = f(a_1,\dots,a_j + t,\dots,a_n)$, so $g_j(0) = f(a)$.
Since f has a relative extremum at a, g_j has a local extremum at 0.
Since f diff., g_j diff, so $g_j'(0) = 0\ \ \forall\, j$. But $g_j'(0) = \frac{\partial f}{\partial x_j}(a)$. Then $\nabla f(a) = 0$. □

Definition 54. $\nabla f(a) = 0$ at stationary pt. a. Hence, $f(a+y) - f(a) = \|y\| E(a,y)$.

Definition 55. Consider $\frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n D_{ij} f(a) y_i y_j$.

$$H(x) = [D_{ij} f(x)]_{i,j=1}^n = \text{ Hessian matrix, and } \sum_{i=1}^n \sum_{j=1}^n D_{ij} f(a) y_i y_j = yH(a)y^T$$

Theorem 103 (Apostol's Thm. 9.4, 2nd. order Taylor Formula for Scalar Fields)**.** *Let f be a scalar field with cont., 2nd. order partial derivatives $D_{ij}f$ in an n-ball $B(a)$. Then $\forall\, y \in \mathbb{R}^n$ s.t. $a + y \in B(a)$,*

$$f(a+y) - f(a) = \nabla f(a) \cdot y + \frac{1}{2!} y H(a+cy) y^T \text{ where } 0 < c < 1 \text{ or} \tag{106}$$

$$f(a+y) - f(a) = \nabla f(a) \cdot y + \frac{1}{2!} y H(a) y^T + \|y\|^2 E_2(a,y) \text{ where } E_2(a,y) \to 0 \text{ as } y \to 0 \tag{107}$$

Proof. Keep y fixed. Define $g(u)$ s.t. $g(u) = f(a + uy)$ for $-1 \le u \le 1$
Then $f(a+y) - f(a) = g(1) - g(0)$.
2nd. order Taylor formula on $[0,1] \implies g(1) - g(0) = g'(0) + \frac{1}{2!} g''(0)$ where $0 < c < 1$.

$$g = g(u) = f(r(u)),\ r(u) = a + uy$$

$$\implies g'(u) = \nabla f(r(u)) \cdot r'(u) = \nabla f \cdot y = \sum_{j=1}^n (\partial_j f)(r(u)) y_j \text{ if } r(u) \in B(u)$$

$$g'(0) = \nabla f(a) \cdot y$$

$$g''(u) = \sum_{i=1}^n \partial_i \left(\sum_{j=1} 6n (\partial_j f)(r(u)) y_j \right) y_i = \sum_{j=1}^n \partial_{ij} f(r(u)) y_i y_j = yH((u)) y^T$$

Hence $g''(c) = yH(a+cy)y^T$.

Define $E_2(a,y)$ by $\|y\|^2 E_2(a,y) = \frac{1}{2!} y (H(a+cy) - H(a)) y^T$ if $y \neq 0$ and let $E_2(a,0) = 0$.

$$\implies f(a+y) - f(a) = \nabla f(a) \cdot y + \frac{1}{2!} y H(a) y^T + \|y\|^2 E_2(a,y)$$

Want: $E_2(a,y) \to 0$ as $y \to 0$.

$$\|y\|^2 |E_2(a,y)| = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (D_{ij} f(a+cy) - D_{ij} f(a)) y_i y_j \le \frac{1}{2} \sum_{i,j=1}^n |D_{ij} f(a+cy) - D_{ij} f(a)| \|y\|^2$$

$$\implies |E_2(a,y)| \le \frac{1}{2} \sum_{i,j=1}^n |D_{ij} f(a+cy) - D_{ij} f(a)|, \text{ for } y \neq 0$$

$D_{ij} f$ cont. at a, so $D_{ij} f(a+cy) \to D_{ij} f(a)$ as $y \to 0$; $E_2(a,y) \to 0$ as $y \to 0$ ☐

Theorem 104 (Apostol's Thm. 9.5, Vol. 2)**.** *Let $A = [a_{ij}]$ be an $n \times n$ real symmetric matrix and*

$$Q(y) = yAy^T = \sum_{i,j}^n a_{ij} y_i y_j \tag{108}$$

then $Q(y) \gtrless 0 \ \ \forall\, y \neq 0$, iff eigenvalues of A are positive (negative); Q is positive definite (negative definite).

Proof. By Thm. 5.11, $\exists$ orthogonal matrix C s.t. $yAy^T = \sum_{i=1}^n \lambda_i x_i^2$, where $x = (x_1, \dots, x_n) = yC$, $\lambda_1, \dots, \lambda_n$ are eigenvalues of A. (i.e. Hermitian matrices are diagonalizable).
Eigenvalues of A are real since A is symmetric.

If all eigenvalues are positive, $Q(y) > 0$, if $x \neq 0$. Since $x = yC$, $y = xC^{-1}$, so $x \neq 0$, iff $y \neq 0$. Then $Q(y) > 0 \ \forall\, y \neq 0$.
Converse, if $Q(y) > 0 \ \forall\, y \neq 0$, choose y s.t. $x = yC$ is e_k.
For this y, $Q(y) = \lambda_k$, so $\lambda_k > 0$.

☐

Theorem 105 (Apostol Vol.2, Thm. 9.6)**.** *Let f be a scalar field, cont. $D_{ij}f$ in n-ball $B(a)$; $H(a)$ Hessian matrix of stationary pt. a*
(a) If eigenvalues of $H(a)$ are positive (negative), f has a rel. min. (rel. max.) at a.
(c) If $H(a)$ has positive and negative eigenvalues, then f has a saddle pt. at a.

Proof. Let $Q(y) = yH(a)y^T$. Then $f(a+y)-f(a) = \frac{1}{2}Q(y)+\|y\|^2E_2(a;y)$. $(\nabla f(a) = 0)$ where $E_2(a,y) \to 0$ as $y \to 0$.

Want: $\exists\, r > 0$ s.t. if $0 < \|y\| < r,\ \ f(a+y) - f(a)$ has same sign as $Q(y)$.
Assume eigenvalues $\lambda_1, \dots, \lambda_n$ of $H(a)$ are positive. Let h be the smallest.
If $u < h$, $\lambda_1 - u, \dots, \lambda_n - u$ are also positive, and are eigenvalues of $H(a) - uI$.
By Thm. 9.5, $y(H(a) - uI)y^T$ positive definitive. $y(H(a) - uI)y^T > 0, \forall\, y \neq 0$.

$$\implies yH(a)y^T > y(uI)y^T = u\|y\|^2 \quad \forall\, u < h$$

Let $u = \frac{1}{2}h$, so $Q(y) > \frac{1}{2}h\|y\|^2\, \forall\, y \neq 0$. Since $E_2(a,y) \to 0$ as $y \to 0$, $\exists\, r > 0$ s.t.

$$|E_2(a,y)| < \frac{h}{2} \text{ whenever } 0 < \|y\| < r$$

For such y, we have $0 \leq \|y\|^2|E_2(a,y)| < \frac{1}{2}h\|y\|^2 < \frac{1}{2}Q(y)$

So

$$f(a+y) - f(a) \geq \frac{1}{2}Q(y) - \|y\|^2|E_2(a,y)| > 0$$

Then f has a rel. min. at a.

Prove statement (c). Let λ_1, λ_2 be 2 eigenvalues of $H(a)$ of opposite signs.
Let $h = \min\{|\lambda_1|, |\lambda_2|\}$. Then $\forall\, u$ s.t. $-h < u < h$. $\lambda_1 - u$, $\lambda_2 - u$ are eigenvalues of opposite sign of $H(a) - uI$.
Thus $u \in (-h, h)$. $y(H(a) - uI)y^T$ takes both positive and negative values in every neighborhood of $y = 0$.
Choose $r > 0$, so $|E_2(a,y)| < \frac{h}{2}$, if $0 < \|y\| < r$.
then $f(a+y) - f(a)$ has same sign as $Q(y)$. Since positive and negatives occur as $y \to 0$, f has a saddle pt. at a.

□