MAT224 Linear Algebra

Definitions, Lemmas, Theorems, Corollaries
and their related proofs

Tingfeng Xia

Winter 2019

by Tingfeng Xia

Materials in this booklet are based heavily on Prof. Nicholas Hoell's lectures as well as A COURSE IN LINEAR ALGEBRA by David B. Damiano and John B. Little.

Items in this booklet are, in fact, very similar to those in the book and are intended to be used as a bullet point guide to (almost) all the knowledge points and should *certainly not* be used as a substitute for the actual learning material. Please notice that I make *no promise* about the accuracy of statements appearing in these notes.

The course website could be found here:
`http://www.math.toronto.edu/nhoell/MAT224/`

Contents

Chapter 1

Vector Spaces

1.1 (Real) Vector Space

1.1.1. **Definition of real vector space:** A real vector space is a set V together with

(a) **Closure under vector addition:** an operation called vector addition, which for each pair of vectors $\mathbf{x}, \mathbf{y} \in V$ produces another vector in V denoted $\mathbf{x} + \mathbf{y}$, (i.e. $\forall \mathbf{x}, \mathbf{y} \in V, \mathbf{x} + \mathbf{y} \in V$) and

(b) **Closure under scalar multiplication:** an operation called multiplication by a scalar (a real number), which for each vector $\mathbf{x} \in V$, an each scalar $c \in \mathbb{R}$ produces another vector in V denoted $c\mathbf{x}$. (i.e. $\forall \mathbf{x} \in V, \forall c \in \mathbb{R}, c\mathbf{x} \in V$)

Furthermore, the two operations must satisfy the following axioms:

(a) $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in V, (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$

(b) $\forall \mathbf{v}, \mathbf{y} \in V, \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$

(c) $\exists \mathbf{0} \in V \, s.t. \forall \mathbf{x} \in V, \mathbf{x} + \mathbf{0} = \mathbf{x}$ (Note that this property is a.k.a existence of additive identity)

(d) $\forall \mathbf{x} \in V, \exists (-\mathbf{x}) \in V \; s.t. \; \mathbf{x} + (-\mathbf{x}) = \mathbf{0}$ (Note that this property is a.k.a existence of additive inverse)

(e) $\forall \mathbf{x}, \mathbf{y} \in V, c \in \mathbb{R}, c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$

(f) $\forall \mathbf{x} \in V, c, d \in \mathbb{R}, (c + d)\mathbf{x} = c\mathbf{x} + d\mathbf{x}$

(g) $\forall \mathbf{x} \in V, c, d \in \mathbb{R}, (cd)\mathbf{x} = c(d\mathbf{x})$

(h) $\forall \mathbf{x} \in V, 1\mathbf{x} = \mathbf{x}$

Remark. Note that here we are not explicitly defining a vector space to be non-empty, however, if a set ever fails to have a larger than zero cardinality, it must not be a vector space. This is a consequence of axiom (c), the existence of the one and only zero vector in the space.

1.1.6. **Propositions for a R-v.s.** Let V be a vector space. Then

(a) The zero vector is unique. Note that it might not necessarily be actually the zero vector in $\mathbb{R}^n$ that we are somewhat used to use.
Proof:
Suppose, for the sake of contradiction, that $\mathbf{a}, \mathbf{b}$ are two *different* zero vectors of the vector space V. Then, by the definition of zero vector, we have

$$\forall \mathbf{x} \in V, \mathbf{x} + \mathbf{a} = \mathbf{x} \wedge \mathbf{x} + \mathbf{b} = \mathbf{x}$$

Simple algebraic manipulation yields us $\mathbf{x}+\mathbf{a} = \mathbf{a}+\mathbf{x} \implies \mathbf{a} = \mathbf{b}$
$\longrightarrow\!\!\!\!\!\!\times\!\!\!\!\!\!\longleftarrow$ Contradiction! $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

(b) $\forall \mathbf{x} \in V, 0\mathbf{x} = 0$
Proof:
We have $0\mathbf{x} = (0+0)\mathbf{x} = 0\mathbf{x} + 0\mathbf{x}$, by axiom 6. By axiom (d), I know there exists a additive inverse of $0\mathbf{x}$, so I subtract on both sides of the equation such additive inverse. This yields us $\mathbf{0} = 0\mathbf{x}$ as wanted. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

(c) $\forall \mathbf{x} \in V$, the additive inverse is unique. Note that it might not necessarily be actually just (-1) times the vector in $\mathbb{R}^n$ that we are somewhat used to use.
Proof:
Let $\mathbf{x} \in V$, and let $(-\mathbf{x}), (-\mathbf{x})'$ be two additive inverse of $\mathbf{x}$. Then, on one hand, by axioms 1, 4 and 3 we have

$$\begin{aligned} \mathbf{x} + (-\mathbf{x}) + (-\mathbf{x})' &= (\mathbf{x} + (-\mathbf{x})) + (-\mathbf{x})' \\ &= \mathbf{0} + (-\mathbf{x})' \\ &= (-\mathbf{x})' \end{aligned}$$

On the other hand, by axiom 2, we have

$$\begin{aligned}\mathbf{x} + (-\mathbf{x}) + (-\mathbf{x})' &= \mathbf{x} + (-\mathbf{x})' + (-\mathbf{x}) \\ &= (\mathbf{x} + (-\mathbf{x})') + (-\mathbf{x}) \\ &= \mathbf{0} + (-\mathbf{x}) \\ &= -\mathbf{x}\end{aligned}$$

Hence we conclude that $(-\mathbf{x}) = (-\mathbf{x})'$, and this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

(d) $\forall \mathbf{x} \in V,\ \forall c \in \mathbb{R},\ (-c)\mathbf{x} = -(c\mathbf{x})$
Proof:
We have

$$\begin{aligned}c\mathbf{x} + (-c)\mathbf{x} &= (c + -c)\mathbf{x} \\ &= \mathbf{0}\mathbf{x} \\ &= \mathbf{0}\end{aligned}$$

Then we notice that equating the fist and the last of the equations above completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

1.2 Sub-spaces

1.2.4. **Lemma on functions in $C^0(\mathbb{R})$.** Note that by $C^n(\cdot)$ we mean the function in this set are all of $Class - n$. Let $f, g \in C^0(\mathbb{R}), \text{let } c \in \mathbb{R}$. Then,

(a) $f + g \in C^0(\mathbb{R})$, and
(b) $cf \in C^0(\mathbb{R})$

The proof of this lemma relies on limit theorems of calculus.

1.2.6. **(Intuitive) definition of (vector) subspace.** Let V be a vector space and let $W \subseteq V$ be a subset. Then W is a (vector) subspace if W is a vector subspace itself under the operations of vector sum and scalar multiplication from V.

1.2.8. **Quick check rule for a subspace.** Let V be a vector subspace, and let W be a **non empty** subset of V. Then W is a subspace of V *if and only if*

$$\forall \mathbf{x}, \mathbf{y} \in W,\ \forall c \in \mathbb{R},\ \text{we have } c\mathbf{x} + \mathbf{y} \in W$$

Notice that the *if and only if* makes this theorem extremely powerful in the sense that it will save you a lot of time to come up with a disproofing counter example. However, do keep in mind that there are cases that this theorem is very hard to check, for example, Exercise 1.2, Question 3(c).

1.2.9. **Remark on the necessary condition of non-emptiness of subspace.** According to the definition of vector space that we gave in 1.1.1, a vector space must contain an additive identity element, hence it is necessary that we ensure $W \subseteq V$(from 1.2.6) is not an empty set.

1.2.13. **Theorem: Intersection of sub-spaces is a subspace.** Let V be a vector space. Then the intersection of any collection of sub-spaces of V is a subspace of V.

1.2.14. **Corollary: Hyper planes in $\mathbb{R}^n$ are sub-spaces of $\mathbb{R}^n$.** Let $a_{ij}(1 \leq i \leq m)$, let $W_i = \{(x_1, \ldots, x_n) \in \mathbb{R}^n | a_{i1}x_1 + \ldots + a_{in}x_n = 0,\ \forall 1 \leq i \leq m\}$. Then W is a subspace of $\mathbb{R}^n$.

1.3 Linear Combinations

1.3.1. **Definitions regarding L.C. and derived spans.** Let S be a subset of a vector space V, that is $S \subseteq V$.

(a) a *linear combination* of vectors in S is any sum $a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n$, where the $a_i \in \mathbb{R}$, and the $x_i \in S$.

(b) we define the *Span* of a set of vectors as follows to consider the special case of $S \stackrel{?}{=} \emptyset \in V$.
Case1: $S \neq \emptyset$: In this case, we define $Span(S)$ to be all possible linear combinations using vectors in S.
Case2: $S = \emptyset$: In this case, we define $Span(S = \emptyset) = \{\mathbf{0}\}$. We call this the zero-space.

(c) If $W = Span(S)$, we say S *spans(or generates)* W.

1.3.4. **Span of a subset of a vector space is a subspace.** Let V be a vector space and let S be any subset of V. Then $Span(S)$ is a subspace of V.

1.3.5. **Sum of sets(with application to subs-paces).** Let W_1 and W_2 be sub-spaces of a vector space V. The sum of W_1 and W_2 is the set

$$W_1 + W_2 := \{\mathbf{x} \in V | \mathbf{x} = \mathbf{x_1} + \mathbf{x_2}, \text{for some } \mathbf{x_1} \in W_1, \mathbf{x_2} \in W_2\}$$

We think of the sum of the two sub-spaces(the two sets) as the set of vectors that can be built up from the vectors in W_1 and W_2 by linear combinations. Conversely, the vectors in the set $W_1 + W_2$ are precisely the vectors that can be broken down into the sum of a vector in W_1 and a vector in W_2. One may find it helpful to view this as an analogue to a Cartesian product of the two set with a new constraint on the result.

1.3.6. **Example.** If $W_1 = \{(a_1, a_2) \in \mathbb{R}^2 | a_2 = 0\}$ and $W_2\{(a_1, a_2) \in \mathbb{R}^2 | a_1 = 0\}$, then $W_1 + W_2 = \mathbb{R}^2$, since every vector in $\mathbb{R}^2$ can be written as the sum of vector in W_1 and a vector in W_2. For instance, we have $(5, -6) = (0, 5) + (0, -6)$, and $(5, 0) \in W_1 \wedge (0, -6) \in W_2$

1.3.8. **Proposition: The sum of spans of sets is the span of the union of the sets.** Let $W_1 = Span(S_1)$ and $W_2 = Span(S_2)$ be sub-spaces of a*(the same)* vector space V. Then $W_1 + W_2 = Span(S_1 \cup S_2)$. Notice that the proof of this gave the important idea of mutual inclusion in proving sets are equal to each other.

1.3.9. **The sum of sub-spaces is also a subspace.** Let W_1 and W_2 be sub-spaces of a vector space V. Then $W_1 + W_2$ is also a subspace of V.
Proof:
It is clear that $W_1 + W_2$ is non-empty, since neither W_1 nor W_2 is empty. Let $\mathbf{x}, \mathbf{y}$ be two vectors in $W_1 + W_2$, let $c \in \mathbb{R}$. By our choice of $\mathbf{x}$and $\mathbf{y}$, we have

$$\begin{aligned} c\mathbf{x} + \mathbf{y} &= c(\mathbf{x_1} + \mathbf{x}_2) + (\mathbf{y_1} + \mathbf{y_2}) \\ &= (c\mathbf{x}_1 + \mathbf{y}_1) + (c\mathbf{x}_2 + \mathbf{y}_2) \\ &\in W_1 + W_2 \end{aligned}$$

Since W_1 and W_2 are sub-spaces of V, we have $(c\mathbf{x}_1 + \mathbf{y}_1) \in W_1 \wedge (c\mathbf{x}_2 + \mathbf{y}_2) \in W_2$. Then by (1.2.8), we see that indeed $W_1 + W_2$ is a subspace of V. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

1.3.10. **Remark.** In general, if W_1 and W_2 are sub-spaces of V, then $W_1 \cup W_2$ will not be a subspace of V. For example, consider the two sub-spaces of $\mathbb{R}^2$ given in example (1.3.6). In that case $W_1 \cup W_2$ is the union of two lines through the origin in $\mathbb{R}^2$.

1.3.11. **Proposition.** Let W_1 and W_2 be sub-spaces of vector space V and let W be a subspace of V such that $W \supseteq W_1 \cup W_2$, then $W \supseteq W_1 + W_2$. Informally speaking, this proposition saying: "$W_1 + W_2$ is the smallest subspace containing $W_1 \cup W_2$", i.e., Any subspace that contains $W_1 \cup W_2$ must be a super set of $W_1 + W_2$.
Proof:
We want to show: $W \supseteq W_1 \cup W_2 \implies W \supseteq W_1 + W_2$

Assume that $W \supseteq W_1 \cup W_2$. Let $w_1 \in W_1,\ w_2 \in W_2$.

$$\text{We notice that } w_1,\ w_2 \in W_1 \cup W_2 \subseteq W$$
$$\implies w_1, w_2 \in W$$

(Since W is a subspace, so it is closed under addition)

$$\implies w_1 + w_2 \in W$$
$$\implies W_1 + W_2 \subseteq W \iff W \supseteq W_1 + W_2$$

$\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

1.4 Linear (In)dependence

1.4.2. **Algebraic definition of linear dependence.** Let V be a vector space, and let $S \subseteq V$.

(a) A *linear dependence* among the vectors of S is an equation $a_1\mathbf{x} + \ldots + a_n\mathbf{x}_n = \mathbf{0}$ where the $x_i \in S$, and the $a_i \in \mathbb{R}$ are not all zero(i.e., at least one of the $a_i \neq 0$). In familiar[1] words, there exists a non-trivial solution to the equation mentioned above.

(b) the set S is said to be *linearly dependent* if there exists a linear dependence among the vectors in S.

[1]Familiar from MAT223, Prof. Jason Siefken's IBL(Inquiry Based Learning) notes.

Remark: It can be shown that the geometric[2] definition and this are, in-fact, equivalent to each other. I will now produce the proof.
Proof of equivalence of definitions:
Let V be a vector space, and let $S \subseteq V$. Consider the following equation:

$$\begin{aligned} & a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \ldots + a_n\mathbf{x}_n = 0, \text{ where } \exists a_i \neq 0 \\ & \text{(WLOG, assume that } a_n \neq \mathbf{0}) \\ \implies & \frac{a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \ldots + a_n\mathbf{x}_n}{a_n} = \mathbf{0} \\ & \implies \mathbf{x}_n = -\sum_{i=1}^{n-1} a_i\mathbf{x}_i \end{aligned}$$

Notice that the result $\mathbf{x}_n$ is in terms of all the other $(n-1)$ vectors in the set, hence a linear combination of those vectors, and this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$
Re-Remark: We can also use this proof as an argument towards the following problem: Show that at least one of the vectors in a linearly dependent set is redundant. We could take take a similar proof and argue that the linear combination could be written without at least one of the vectors.

1.4.4. **Algebraic definition of linear independence.** Let V be a vector space, and $S \subseteq V$. Then S is *linearly independent* if whenever we have $a_i \in \mathbb{R}$ and $x_i \in S$ such that $a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n = \mathbf{0}$, then $a_i = 0,\ \forall i$. A more conceivable way to understand this is if the aforementioned equation exists and only exists a set of trivial solution then the vectors involved in the equation are *linearly independent*.
Remark: A set of vector is linearly independent *if and only if* it is not linearly dependent.

1.4.7. **Propositions regarding linear (in)dependency.**

(a) Let S be a linearly dependent subset of a vector space V, and let

[2] A set of vectors is said to be dependent of each other there exists a vector in this set, that it is in the Span of all other vectors in the set. I.e., There is some vectors in this set that are "redundant", it's position can be taken by some linear combination of the other vectors in the set.

S' be another subset of V that contains S. Then S' is also linearly dependent.

(b) Let S be a linearly independent subset of vector space V and let S' be another subset of V that is contained in S. Then S' is also linearly independent.

Proof of (a): Since S is linearly dependent, there exists a linear dependence among the vectors in S, say, $a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n = \mathbf{0}$. Since S is contained in S', this is also a linear dependence among the vectors in S'. Hence S' is linear dependent. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

Proof of (b): Consider any equation $a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n = \mathbf{0}$, where the $a_i \in \mathbb{R},\ \mathbf{x}_i \in S'$. Since S' is contained in S, we can also view this as a potential linear dependence among vectors in S. However, S is linearly independent, so it follows that all the $a_i = 0 \in \mathbb{R}$. Hence S' is also linearly independent. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

1.4.*. **Example of showing linear independence.(1)** Show that the set $\{1, x, x^2, x^3, \ldots, x^n\}$ is linearly independent. We consider the following equation:

$$0 = a_0 + a_1x + \ldots + a_{n-1}x^{n-1} + a_nx^n \qquad (*)$$

Then we take the derivative:

$$\frac{d^n}{dx^n}(*) = 0 = n! \cdot a_n$$

Then since $n! \neq 0$, we know $a_n = 0$. We repeat the same process, taking derivatives $(n-1)$ times w.r.t. x we get $a_{n-1} = 0$ and so on. As a last step we have:

$$\frac{d}{dx}(*) = 0 = 1! \cdot a_1 \implies a_1 = 0$$

Then we have $0 = 0+0+\ldots+a_0 \implies a_0 = 0$. So the equation has and only has a trivial solution, thus the set $\{1, x, x^2, x^3, \ldots, x^n\}$ is linearly independent. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

1.4.*. **Example of showing linear independence.(2)** Show that the set $\{e^x, e^{2x}\}$ is linearly independent. We consider the following equation:

$$0 = ae^x + be^{2x} \qquad (1)$$

We take derivative on both sides of (1) w.r.t. x and we have:

$$0 = ae^x + 2be^{2x} \quad (2)$$

We subtract (2) from (1) to get:

$$0 = be^{2x} \implies b = 0 \implies 0 = ae^x + 0 \implies a = 0$$

Since $a = b = 0$ is the one and only solution to (1), we claim they are linearly independent. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

1.5 Interlude on Solving SLEs

1.5.*. **Note Aside:** This section of the book is covered, although not rigorously but completely, in MAT223. Hence the vast majority of definitions and corollary in this section were omitted. Consult the book for more detail on this.

1.5.1. **Definition of (homogeneous) SLEs**[3] A system of m equations in n unknowns $x_1, \ldots, x_n$ of the form:

$$\begin{aligned} a_{11}\mathbf{x_1} + \ldots + a_{1n}\mathbf{x_n} &= b_1 \\ a_{21}\mathbf{x_1} + \ldots + a_{2n}\mathbf{x_n} &= b_2 \\ &\ldots \\ a_{m_1}\mathbf{x_1} + \ldots + a_{mn}\mathbf{x_n} &= b_m \end{aligned}$$

where the $a_{ij},\ b_i \in \mathbb{R}$, is called a *system of linear equations.* The scalars a_{ij} are called coefficients or *weights* of the equations. We call this system **homogeneous** if and only if all the b_i are 0's.

1.5.2. **Definition of equivalent SLEs** Two systems of linear equations are said to be equivalent if their sets of solutions are the same(i.e., mutual inclusion of the two solution sets)

1.5.3. **Propositions on operations on SLEs**[4]

[3] System of Linear Equations

[4] This should be familiar from MAT223, operations involved in row reducing an augmented matrix for a system of linear equations

(a) The system obtained buy adding any multiple of any one equation to any second equation, while leaving the other other equations unchanged, is an equivalent system.

(b) The system obtained by multiplying any one equation by a non-zero scalar and leaving the other equations unchanged is an equivalent system.

(c) The system obtained by interchanging any two equations is an equivalent system.

1.5.13. **Corollary** If $m < n$, every homogeneous system of m linear equations in n unknowns has a non-trivial solution.

1.6 Bases and Dimension

1.6.1. **Definition of basis.** A subset S of a vector space V is called a basis if $V = Span(S)$ and the set S is linearly independent.

1.6.3. **Theorem.** Let V be a vector space, and let S be a non-empty subset of V. Then S is a basis of V if and only if $\forall \mathbf{x} \in V, \mathbf{x}$ can be written *uniquely* as a linear combination of the vectors in S.

1.6.6. **Theorem.** Let V be a vector space that has a finite spanning set, and let S be a linearly independent subset of V. Then there exists a basis S' of V, such that $S \subseteq S'$. Note that this theorem is can be summarized as follows: Every linearly independent set of vectors could be extended to a basis. We do so by adding yet another vector that is linearly independent to all the vectors already in the set, but notice that this process should possibly be repeated but *not* infinite.

1.6.8. **Lemma on linear independence towards a set and a vector.** Let S be a linearly independent subset of V and let $\mathbf{x} \in V$, but $\mathbf{x} \notin S$. Then $S \cup \{\mathbf{x}\}$ is linearly independent if and only if $\mathbf{x} \notin Span(S)$.

1.6.10. **Theorem on cardinality of linearly independent set.** Let V be a vector space and let S be a spanning set for V, which has m elements. Then no linearly independent set in V can have more than m elements.
Proof:
To show that there are no linearly independent set in V that can have

more than m elements in it, it suffices to show that every set in V that has more than m elements in it is linearly dependent. Let $S = \{y_1, \ldots, y_m\}$ and $S' = \{y_1, \ldots, y_n\} \subset V$ where $n > m$. Now we consider the following equation:

$$a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n = \mathbf{0} \qquad \text{where } \mathbf{x}_i \in S', a_i \in \mathbb{R} \tag{1.1}$$

Since S is a spanning set for V, then $\exists b_{ij} \in \mathbb{R},\ s.t.\ \forall 1 \leq i \leq m$:

$$\mathbf{x}_i = b_{i1}\mathbf{y}_1 + \ldots + b_{im}\mathbf{y}_m = \sum_{j=1}^{m} b_{ij}\mathbf{y}_j$$

Then we substitute the $\mathbf{x}_i$'s into equation (1.1):

$$a_1\left(\sum_{j=1}^{m} b_{1j}\mathbf{y}_j\right) + \ldots + a_n\left(\sum_{j=1}^{m} b_{nj}\mathbf{y}_j\right) = \mathbf{0} \tag{1.2}$$

Rearranging (1.2), we have:

$$\left(\sum_{j=1}^{n} b_{1j}a_j\right)\mathbf{y}_1 + \ldots + \left(\sum_{j=1}^{n} b_{mj}a_j\right)\mathbf{y}_m = \mathbf{0} \tag{1.3}$$

Now we consider the following SLE(The coefficients of (1.3)):

$$\begin{cases} b_{11}a_1 + \ldots + b_{1n}a_n & = 0 \\ b_{21}a_1 + \ldots + b_{2n}a_n & = 0 \\ & \vdots \\ b_{m1}a_1 + \ldots + b_{mn}a_n & = 0 \end{cases}$$

If we can find a set of $a_1, \ldots, a_n$ that are not all zero that solves the above SLE, then those scalars will give us a linear dependence among the vectors in S' in (1.1). But the above SLE is a system of homogeneous linear equations in the n unknowns a_i's, hence since $m < n$, by Corollary (1.5.13), there exists a non-trivial solution. Then S' is linearly dependent and this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

1.6.11. **Bases of a vector space shall have same cardinality.** Let V be a vector space and let S and S' be two bases of with m and m' elements, respectively. Then $m = m$;.
Proof: (Euclid's style)
Let S and S' with properties given as above. Since they are bases, we know: (By Theorem 1.6.10)

(a) S is a spanning set and S' is a linearly independent set $\implies m' = |S'| \leq |S| = m$

(b) S' is a spanning set and S is a linearly independent set $\implies m = |S| \leq |S'| = m'$

Then, $m' = m$. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

1.6.12. **Definitions of (in)finite-dimension**

(a) If V is a vector space with some finite basis (possibly empty), we sat V is *finite-dimensional*. We say the vector space is *infinite-dimensional* otherwise.

(b) Let V be a *finite-dimensional* vector space. The dimension of V, denoted as $\dim(V)$, is the number of elements in a (hence any) basis of V.

(c) If $V = \{\mathbf{0}\}$, we define $\dim(V) = 0$. **Remark:** Please note that is consistent with our definition of span of a empty set. We defined the span of empty set to be $\{\mathbf{0}\}$, and hence $\{\mathbf{0}\}$ is the resulting space of all the possibly linear combination of vectors in the empty set(where there is none). Then the cardinality of the empty set(which is zero) is here defined as the dimension of the zero vector space.

1.6.13. **Examples on dimensions of vector spaces.** If we have a basis of V, then computing $\dim(V)$ is simply a matter of counting the number of vectors in a (hence any) basis for the vectors space. We explore this by looking at the following examples

(a) For each n, $\dim(\mathbb{R}^n) = n$. This is a consequence of the standard basis for $\mathbb{R}^n$ is of the form: $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ has n elements in it.

(b) $\dim(P_n(\mathbb{R})) = n + 1$, since $|\{1, x^1, x^2, \ldots, x^n\}| = n + 1$.[5]

[5]The absolute value marks around a set returns the cardinality of the set.

(c) The vector spaces $P(\mathbb{R})$[6], $C^n(\mathbb{R})$, where $n \in \mathbb{N}^{\geq 0}$, are not of finite dimension, and hence are called *infinite dimensional*.

1.6.14. **Corollary.** Let W be a subspace of a finite-dimensional vector space V. Then $\dim(W) \leq \dim(V)$. Furthermore, $\dim(W) = \dim(V)$ if and only if $W = V$.

1.6.15. **Corollary.** Let W be a subspace of $\mathbb{R}^n$ defined by a system of homogeneous linear equations. The $\dim(W)$ is equal to the number of free variables in the corresponding echelon form of the equations.
Generalization. It is also worth pointing out that by setting the free variables equal to one in turn[7], we can always generate a basis for the subspace. Referring to example (1.6.16) in the book, we see that we have two free variables, and the two (linearly independent) vectors obtained from the method described above that form a 2-dimensional subspace of $\mathbb{R}^5$.

1.6.18. **Inclusion-Exclusion principle of dimensions.** Let W_1 and W_2 be finite dimensional sub-spaces of a vector space V. Then,

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$$

Remark: This theorem could not be generalized to higher dimensions as does the Inclusion-Exclusion Principle in set theory. (Consequence of the challenge problem of Tutorial 2), more information can be found here, on a paper of generalizing this formula.

[6]Here $P(\mathbb{R})$ means the vector space of all polynomials that are $\mathbb{R} \to \mathbb{R}$.
[7]By in turn we mean set one of them to one and all others zero, one at a time

Chapter 2

Linear Transformations

2.1 Linear Transformations

2.1.1. **Definition of linear transformation.**[1] A function $T : V \to W$ is called a *linear mapping* or a *linear transformation* if it satisfies:

(a) $\forall \mathbf{u}, \mathbf{v} \in V,\ T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{u})$

(b) $\forall \alpha \in \mathbb{R}, \mathbf{v} \in V,\ T(\alpha \mathbf{v}) = \alpha T(\mathbf{v})$

2.1.2. **Proposition: alternative definition of L.T.s** A function $T : V \to W$ is a linear transformation if and only if

$$\forall \alpha, \beta \in \mathbb{R}, \forall \mathbf{u}, \mathbf{v} \in V, T(\alpha \mathbf{u} + \beta \mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v}).$$

Proof($\Longrightarrow$):
Assuming that T is a linear transformation, then the definition in (2.1.1) must satisfy. Let $\alpha, \beta \in \mathbb{R}, \mathbf{u}, \mathbf{v} \in V$, then:

$$\begin{aligned} T(\alpha \mathbf{u} + \beta \mathbf{v}) &= T(\alpha \mathbf{u}) + T(\beta \mathbf{u}) \quad \text{// by (a)} \\ &= \alpha T(\mathbf{u}) + \beta T(\mathbf{v}) \quad \text{// by (b)} \end{aligned}$$

Proof($\Longleftarrow$):
Assuming the alternative definition, we want to show the definition given in (2.1.1). Since the quantifier is $\forall$, we can take $\alpha = \beta = 1$, (arbitrary) $\mathbf{u}, \mathbf{v} \in V$ in the alternative definition which directly yields

[1] Familiar from MAT223

us (a) of (2.1.1). Then we take $\alpha \in \mathbb{R}, \beta = 0, \mathbf{u}, \mathbf{v} \in V$. In this case, we want to show that $T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$. We proceed as follows:

$$\begin{aligned} T(\alpha\mathbf{u} + \beta\mathbf{v}) &= T(\alpha\mathbf{u} + 0\mathbf{v}) = T(\alpha\mathbf{u}) \\ &= \alpha T(\mathbf{u}) + 0T(\mathbf{u}) \\ &= \alpha T(\mathbf{u}) \end{aligned}$$

Since $\alpha \in \mathbb{R}, \mathbf{u} \in V$ are arbitrary, this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

2.1.3. **Corollary.** A function $T : V \to W$ is a linear transformation if and only if

$$\forall a_1, \ldots, a_k \in \mathbb{R}, \forall \mathbf{v}_1, \ldots, \mathbf{v}_k \in V : \; T\left(\sum_{i=1}^{k} a_i \mathbf{v}_i\right) = \sum_{i=1}^{k} a_i T(\mathbf{v}_i) \quad (2.1)$$

*Proof:*To show this if and only if relationship, we have to consider the implication of both directions. Since (2.1.2) is just a special case of (2.1.3) then we are done in proving (2.1.3) $\implies$ (2.1.2). We will prove the other direction by mathematical induction (on k) to generalize the case of $k = 2$ to arbitrary $k \in \mathbb{N}^{\geq 2}$.
Define predicate $P(k) : \mathbb{N}^{\geq 2} \to \{0, 1\}$ as "(2.1) holds".
Claim that $\forall k \in \mathbb{N}^{\geq 2}, P(k)$.
Basis Case: $k = 2$. In this case the wanted equality is the same as the one proved in (2.1.2), so $P(2)$.
Inductive Step: Let $k \in \mathbb{N}^{\geq 2}$, assume $P(k-1)$, we want to show that $P(k)$ follows.

$$\begin{aligned} T\left(\sum_{i=1}^{k} a_i \mathbf{v}_i\right) &= T\left(a_k \mathbf{v}_k + \sum_{i=1}^{k-1} a_i \mathbf{v}_i\right) \\ &= a_k T(\mathbf{v}_k) + T\left(\sum_{i=1}^{k-1} a_i \mathbf{v}_i\right) \qquad \text{// By (2.1.1)} \\ &= a_k T(\mathbf{v}_k) + \sum_{i=1}^{k-1} a_i T(\mathbf{v}_i) \qquad \text{// By } P(k-1) \\ &= \sum_{i=1}^{k} a_i T(\mathbf{v}_i) \end{aligned}$$

So $P(k)$ follows in this case. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

2.1.9. **Angle between two vectors.** If $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^2$ and $\mathbf{0} \neq \mathbf{b} \in \mathbb{R}^2$, then the angle θ between them must be[2]

$$\theta = \arccos\left(\frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|}\right)$$

Remark. Notice that this definition could also be extended to $\mathbb{R}^3$.

2.1.10. **Corollary on orthogonality of vectors.** If $\mathbb{R}^2 \ni \mathbf{a} \neq \mathbf{0}$ and $\mathbb{R}^2 \ni \mathbf{b} \neq \mathbf{0}$, then the angle θ between them is a right angle if and only if $\langle \mathbf{a}, \mathbf{b} \rangle = 0$.

Remark. Note this definition of orthogonality could be extended to all euclidean vector spaces. The orthogonal complement of a subspace is the space of all vectors that are orthogonal to every vector in the subspace. In a three-dimensional Euclidean vector space, the orthogonal complement of a line through the origin is the plane through the origin perpendicular to it, and vice versa. In four-dimensional Euclidean space, for example, the orthogonal complement of a line is a hyper-plane and vice versa, and that of a plane is a plane.

2.1.14. **Proposition.** If $T : V \to W$ is a linear transformation and V is finite-dimensional, then T is uniquely determined by its values on the members of a basis of V. To make this proposition more clear, we present the proof below. We will show that if S and T are linear transformations that take the same values on each member of a basis for V, then in fact $S = T$.

Proof:

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for V, and let S and T be two linear transformations that satisfy $T(\mathbf{v}_1) = S(\mathbf{v}_1), \forall i \in \{1, \ldots, k\}$. If $\mathbf{v} \in V$, $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_k\mathbf{k}$, then

$$\begin{aligned} T(\mathbf{v}) &= T(a_1\mathbf{v}_1 + \ldots + a_k\mathbf{k}) \\ &= a_1T(\mathbf{v}_1) + \ldots + a_kT(\mathbf{v}_k) && // \text{ Since } T \text{ is linear} \\ &= a_1S(\mathbf{v}_1) + \ldots + a_kS(\mathbf{v}_k) \\ &= S(a_1\mathbf{v}_1 + \ldots + a_k\mathbf{k}) && // \text{ Since } S \text{ is linear} \\ &= S(\mathbf{v}) \end{aligned}$$

Hence S and T are equal as mappings from V to W. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

[2]The angle brackets here denotes the inner product of vectors

2.2 Linear Transformations between finite dimensional vector spaces

2.2.1. **Proposition.** Let $T : V \to W$ be a linear transformation between the finite dimensional vector spaces V and W. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis for V and $\{\mathbf{w}_1, \ldots, \mathbf{w}_l\}$ is a basis for W, then $T : V \to W$ is uniquely determined by the $l \times k$ scalars used to express $T(\mathbf{v}_j)$, where $j \in \{1, \ldots, k\}$, in terms of $\mathbf{w}_1, \ldots, \mathbf{w}_l$.

2.2.6. **Matrix of a transformation w.r.t.α, β.** Let $T : V \to W$ be a linear transformation between finite-dimensional vector spaces V and W, and let $\alpha = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ and $\beta = \{\mathbf{w}_1, \ldots, \mathbf{w}_l\}$, respectively, be any basis for V and W. Let a_{ij}, where $1 \leq i \leq l$ and $1 \leq j \leq k$ be the $l \cdot k$ scalars that determined T with respect to the bases α and β. The matrix whole entries are the scalars a_{ij}, given above, is called the *matrix of the linear transformation T with respect to the bases α for V and β for W*. We denote such matrix with following notation: $[T]_{\alpha}^{\beta}$.
Notice that, again, α is a basis for V, the domain of the transformation and β is a basis for W, the co-domain of the transformation.
Remark. This should be familiar from MAT223, where we perform transformation on each and every element in a basis for a space, and transcribe them, in terms of another basis, into a standard matrix for the transformation. In terms of standardized formulaic calculations, we have

$$[T]_{\alpha}^{\beta} = \begin{bmatrix} [T(\alpha_1)]_{\beta} & \ldots & [T(\alpha_n)]_{\beta} \end{bmatrix}$$

where the α_i's denotes the i-th element of the α basis, and is itself a vector.

2.2.15. **Proposition: Linear Transformation on alternative basis.** Let $T : V \to W$ be a linear transformation between vector spaces V of dimension k and W of dimension l. Let $\alpha = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for V, and let $\beta = \{\mathbf{w}_1, \ldots, \mathbf{w}_l\}$ be a basis for W. Then for each $\mathbf{v} \in V$, we have the following:

$$[T(\mathbf{v})]_{\beta} = [T]_{\alpha}^{\beta} [\mathbf{v}]_{\alpha}$$

One can think about this in the sense that the same transformation could have been accomplished under another basis. We just have to

convert the original problem into some easy to solve basis, perform the transformation under that basis, and then convert the result back. We now present the proof for this proposition.
*Proof:*Let $\mathbf{v} = x_1\mathbf{v}_1 + \ldots + x_k\mathbf{v}_k \in V$. Then if $T(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + \ldots + a_{lj}\mathbf{w}_l$,

$$\begin{aligned} T(\mathbf{v}) &= \sum_{j=1}^{k} x_j T(\mathbf{v}_j) \\ &= \sum_{j=1}^{k} x_j \left(\sum_{i=1}^{l} a_{ij}\mathbf{w}_i \right) \\ &= \sum_{i=1}^{l} \left(\sum_{j=1}^{k} x_j a_{ij} \right) \mathbf{w}_i \end{aligned}$$

Thus the i-th coefficient of $T(\mathbf{v})$ in terms of β is $\sum_{j=1}^{k} x_j a_{ij}$, and

$$[T(\mathbf{v})]_\beta = \begin{bmatrix} \sum_{j=1}^{k} x_j a_{1j} \\ \cdots \\ \cdots \\ \cdots \\ \sum_{j=1}^{k} x_j a_{lj} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{l1} & \cdots & a_{lk} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix} = [T]_\alpha^\beta [\mathbf{v}]_\alpha$$

$\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

2.2.18. **Proposition: Property of matrix *times* L.C. of vectors.** Note that in this proposition, we assume the vectors and the matrix are compatible[3]. Let A be an $l \times k$ matrix and $\mathbf{u}$ and $\mathbf{v}$ be column vectors with k entries. Then,

$$\forall \text{ pairs of } a \in \mathbb{R}, b \in \mathbb{R}, A(a\mathbf{u} + b\mathbf{v}) = aA\mathbf{u} + bA\mathbf{v}$$

*Proof:*Since $\mathbf{u}, \mathbf{v} \in \mathbb{R}^k$, let them be $\mathbf{u} = (u_1, \ldots, u_k)^T$ and $\mathbf{v} = (v_1, \ldots, v_k)^T$. Fix $a, b \in \mathbb{R}$. Then,

$$\begin{aligned} A(a\mathbf{u} + b\mathbf{v}) &= A\left(a(u_1, \ldots, u_k)^T + b(v_1, \ldots, v_k)^T \right) \\ &= A\left((a \cdot u_1, \ldots, a \cdot u_k)^T + (b \cdot v_1, \ldots, b \cdot v_k)^T \right) \\ &= A(a \cdot u_1 + b \cdot v_1, \ldots, a \cdot u_k + b \cdot v_k)^T \end{aligned}$$

%TODO: change this vector notation and finish the proof..

[3]i.e., They can *always* perform the operations that we want

2.2.19. **Proposition.** Let $\alpha = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for V and $\beta = \{\mathbf{w}_1, \ldots, \mathbf{w}_l\}$ be a basis for W, and let $\mathbf{v} = x_1\mathbf{v}_1 + \ldots + x_k\mathbf{v}_k \in V$.

(a) If A is an $l \times k$ matrix, then the function $T(\mathbf{v}) = \mathbf{w}$, where $[\mathbf{w}]_\beta A [\mathbf{v}]_\alpha$ is a linear transformation.

(b) If $A = [S]_\alpha^\beta$ is the matrix of a transformation $S : V \to W$, then the transformation T constructed from $[S]_\alpha^\beta$ is equal to S.

(c) If T is the transformation of (a) constructed from A, then $[T]_\alpha^\beta = A$.

2.2.20. **Proposition.** Let V and W be finite-dimensional vector spaces. Let α be a basis for V and β a basis for W. Then the assignment of a matrix to a linear transformation from V to W given by T goes to $[T]_\alpha^\beta$ is bijective[4].

2.3 Kernel and image

2.3.1. **Definition of Kernel.** The *kernel* of T, denoted $\mathrm{Ker}(T)$, is the subset of V consisting of all vectors $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{0}$. Writing in familiar set builder notation:

$$\mathrm{Ker}(T) := \{\mathbf{v} \in V | T(\mathbf{v}) = \mathbf{0}\}$$

One should notice the difference between the familiar Null Space of a transformation and the Kernel here. Kernel is defined for all vector spaces, however, null-spaces are for $\mathbb{R}^n$ only.

2.3.2. **Proposition: Kernel is a subspace.** Let $T : V \to W$ be a linear transformation. $\mathrm{Ker}(T)$ is a subspace of V.
Proof:
Since $\mathrm{Ker}(T) \subset V$, it suffices to show that $\mathrm{Ker}(T)$ is closed under addition and scalar multiplication. Since T is linear, $\forall \mathbf{u}, \mathbf{v} \in \mathrm{Ker}(T)$ and $a \in \mathbb{R}$, we have

$$T(\mathbf{u} + a\mathbf{v}) = T(\mathbf{u}) + aT(\mathbf{v}) = \mathbf{0} + a\mathbf{0} \implies \mathbf{u} + a\mathbf{v} \in \mathrm{Ker}(T)$$

$\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

[4] Injective and surjective

2.3.7. **Proposition.** Let $T : V \to W$ be a linear transformation of finite-dimensional vector spaces, and let α, β be bases for V, W respectively. Then $\mathbf{x} \in \text{Ker}(T)$ if and only if the coordinate vector of $\mathbf{x}$, $[\mathbf{x}]_\alpha$, satisfies the system of equations

$$\begin{cases} a_{11}x_1 + \ldots + a_{1k}x_k & = 0 \\ & \vdots \\ a_{l1}x_1 + \ldots + a_{lk}x_k & = 0 \end{cases}$$

where the coefficient a_{ij} are the entries of the matrix $[T]_\alpha^\beta$

2.3.8. **Independence is basis-independent.** If $\alpha = \{v_1, \ldots, v_k\}$ is a basis for V, then $\mathbf{x}_1, \ldots, \mathbf{x}_m \in V$ are independent if and only if $[\mathbf{x}_1]_\alpha, \ldots, [\mathbf{x}_m]_\alpha$ are independent.

2.3.10. **Definition of Image.** The subset of W constsing of vectors $\mathbf{w} \in W$ for which there exists a $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$ is called the *image* of T and is denoted by $\text{Im}(T)$. In set builder notation we have

$$\text{Im}(T) := \{\mathbf{w} \in W | T(\mathbf{v}) = \mathbf{w} \text{ for some } \mathbf{v} \in V\} \text{ where } T : V \to W$$

2.3.11. **Proposition: Image is subspace.** Let $T : V \to W$ be a linear transformation. The image of T is a subspace of W, the co-domain.
Proof:
Let, $\mathbf{w}_1, \mathbf{w}_2 \in \text{Im}(T)$, and let $a \in \mathbb{R}$. Since $\mathbf{w}_1$ and $\mathbf{w}_2 \in V$ with $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. Then we have

$$\begin{aligned} a\mathbf{w}_1 + \mathbf{w}_2 &= aT(\mathbf{v}_1) + T(\mathbf{v}_2) \\ &= T(a\mathbf{v}_1 + \mathbf{v}_2) \implies a\mathbf{w}_1 + \mathbf{w}_2 \in \text{Im}(T) \text{ by linearity} \end{aligned}$$

Hence, by the "quick check rule", we know that image is a subspace of the co-domain. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

2.3.12. **Proposition.** If $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is ant set that spans V (in particular, it could be a basis of V), then $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_m)\}$ spans[5] $\text{Im}(T)$.
Proof($\supseteq$):

[5]By "spans" we mean equal to each other

Let $\mathbf{w} \in \text{Im}(T)$, then $\exists \mathbf{v} \in V$ with $T(\mathbf{v}) = \mathbf{w}$. Since $Span\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} = V$, then $\exists a_1, \ldots, a_m$ s.t. $a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m = \mathbf{v}$. Then,

$$\begin{aligned}\mathbf{w} &= T(\mathbf{v}) \\ &= T\left(\sum_{i=1}^{m} a_i\mathbf{v}_i\right) \\ &= \sum_{i=1}^{m} a_i T(\mathbf{v}_i) \qquad \text{// by linearity}\end{aligned}$$

Therefore, $\text{Im}(T)$ is contained in $Span\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_m)\}$.
Proof($\subseteq$):
Let $\mathbf{w} \in Span\{T(\mathbf{v}_1, \ldots, \mathbf{v}_m)\}$, then (reversing what we did previously) we have

$$\begin{aligned}\mathbf{w} &= \sum_{i=1}^{m} a_i T(\mathbf{v}_i) \\ &= T\left(\sum_{i=1}^{m} a_i\mathbf{v}_i\right) \qquad \text{// by linearity} \\ &= T(\mathbf{v}) \in \text{Im}(T)\end{aligned}$$

Hence mutual inclusion yields us the wanted result, and this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

2.3.13. **Corollary.** If $\alpha = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis for V, and $\beta = \{\mathbf{w}_1, \ldots, \mathbf{w}_l\}$ is a basis for W, then the vectors in W whose coordinate vectors (in terms of β) are the column of $[T]_\alpha^\beta$ span $\text{Im}(T)$.

2.3.17. **Rank-Nullity Theorem.**[6] If V is a finite-dimensional vector space and $T : V \to W$ is a linear transformation, then

$$\dim(\text{Ker}(T)) + \dim(\text{Im}(T)) = \dim(V)$$

Or equivalently[7],

$$\dim(\text{Ker}(T)) + \text{Rank}(T) = \dim(V)$$

[6]Known as The Dimension Theorem in book
[7]also, $\dim(\text{Im}(T)) = \dim(\text{Rol}(T)) = \dim(\text{Col}(T))$ = #pivot in r.r.e.f

2.4 Applications of Rank-Nullity Theorem

2.4.2. **Proposition.** A linear transformation $T : V \to W$ is injective if and only of $\dim(\text{Ker}(T)) = 0$. Informally speaking, we can think of this as "No information is lost during the linear transformation".
Proof($\Longrightarrow$):
If T is injective, then by definition, there $\exists! \mathbf{v} \in V$ with $T(\mathbf{v}) = \mathbf{0}$. Since we know that $T(\mathbf{0}) = 0, \forall$ linear mappings, the zero vector is the unique vector $\mathbf{v}$ satisfying $T(\mathbf{v}) = \mathbf{0}$. Thus the kernel of T consists of only the zero vector. Therefore, $\dim(\text{Ker}(T)) = 0$.
Proof($\Longleftarrow$):
Conversely, assume that $\dim(\text{Ker}(T)) = 0$. Let $\mathbf{v}_1, \mathbf{v}_2 \in V$ with $T(\mathbf{v}_1) = T(\mathbf{v}_2)$. We want to show that $\mathbf{v}_1 = \mathbf{v}_2$. Since $T(\mathbf{v}_1) = T(\mathbf{v}_2)$, $T(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}$, so $\mathbf{v}_1 - \mathbf{v}_2 \in \text{Ker}(T)$. But if $\dim(\text{Ker}(T)) = 0$, it follows that $\text{Ker}(T) = \{\mathbf{0}\}$. It follows that $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0} \implies \mathbf{v}_1 = \mathbf{v}_2$ as wanted. Thus, T is injective. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$
Remark. This proposition is so important that we summarize it as follows, again. Let $T : V \to W$, with $\dim V = \dim W$, then

$$\begin{aligned} & T \text{ is injective} \\ \iff & T \text{ is surjective} \\ \iff & \ker(T) \text{ is trivial} \\ \iff & \text{Im}(T) = W \end{aligned}$$

2.4.3. **Corollary.** A linear mapping $T : V \to W$ on a finite-dimensional vector space V is injective if and only if $\dim(\text{Im}(T)) = \dim(V)$

2.4.4. **Corollary.** If $\dim(W) < \dim(V)$ and $T : V \to W$ is a linear mapping, then T is not injective.

2.4.5. **Corollary.** If V and W are finite dimensional, then a linear mapping $T : V \to W$ can be injective only if $\dim(W) \geq \dim(V)$

2.4.7. **Proposition.** If W is finite-dimensional, then a linear mapping $T : V \to W$ is surjective if and only if $\dim(\text{Im}(T)) = \dim(W)$.
Remark. Since $\dim(\text{Im}(T)) \leq \dim(V)$ by the theorem, if $\dim(V) < \dim(W)$, then we have $\dim(\text{Im}(T)) < \dim(W)$, hence T is not surjective.

2.4.8. **Corollary.** If V and W are finite-dimensional, with $\dim(V) < \dim(()W)$, then there is no surjective linear mapping $T : V \to W$.

2.4.9. **Corollary.** A linear mapping $T : V \to W$ can be surjective only if $\dim(V) \geq \dim(W)$.

2.4.10. **Proposition.** Let $\dim(V) = \dim(W)$. A linear transformation $T : V \to W$ is injective if and only if it is surjective.
Proof($\Longrightarrow$):
If T is injective, then $\dim(\text{Ker}(T)) = 0$ by proposition (2.4.2). By Theorem (2.3.17), we have $\dim(\text{Im}(T)) = \dim(V)$. Therefore, by proposition (2.4.7), T is injective
Proof($\Longleftarrow$):
If T is surjective, then by Proposition (2.4.7), $\dim(\text{Im}(T)) = \dim(W)$ $= \dim(V)$. Therefore, by Theorem (2.3.17), $\dim(\text{Ker}(T)) = 0$. Hence, by proposition (2.4.2), T is injective. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

2.4.11. **Proposition.** Let $T : V \to W$ be a linear transformation, and let $\mathbf{w} \in \text{Im}(T)$. Let $\mathbf{v}_1$ be any fixed vector with $T(\mathbf{v}_1) = \mathbf{w}$. Then every vector $\mathbf{v}_2 \in T^{-1}(\{\mathbf{w}\})$ can be written uniquely as $\mathbf{v}_2 = \mathbf{v}_1 + \mathbf{u}$, where $\mathbf{u} \in \ker(T)$

2.4.15. **Corollary.** Let $T : V \to W$ be a linear transformation of finite-dimensional vector spaces, and let $\mathbf{w} \in W$. Then $\exists!\mathbf{v} \in V \; s.t. \; T(\mathbf{v}) = \mathbf{w}$ if and only if

(a) $\mathbf{w} \in \text{Im}(T)$, and
(b) $\dim(\text{Ker}(T)) = 0$

2.4.16. **Proposition.**

(a) The set of solutions of the system of linear equations $A\mathbf{x} = \mathbf{b}$ is the subset $T^{-1}(\{\mathbf{b}\})$ of $V = \mathbb{R}^n$
(b) The set of solutions of the system of linear equations $A\mathbf{x} = \mathbf{b}$ is a subspace of V if and only if the system is homogeneous, in which case the set of solutions is $\text{Ker}(T)$.

2.4.17. **Corollary.**

(a) The number of free variables in the homogeneous system $A\mathbf{x} = \mathbf{0}$ (or its echelon form equivalent) is equal to $\dim(\text{Ker}(T))$.

(b) The number of basic variables of the system is equal to $\dim(\mathrm{Im}(T))$.

Now, if the system is *not homogeneous*, then from Proposition (2.4.16) we see that $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b} \in \mathrm{Im}(T)$. Let us assume then that $\mathbf{b} \in \mathrm{Im}(T)$, then this yields us the following terminology:

2.4.18. **Definition of particular solution of a SLE (in-homo).** Given an in-homogeneous system of equations, $A\mathbf{x} = \mathbf{b}$, any single vector $\mathbf{x}$ satisfying the system (necessarily $\mathbf{x} \neq \mathbf{0}$) is called a *particular solution* of the system of equations.

2.4.19. **Proposition.** Let $\mathbf{x}_p$ be a particular solution of the system $A\mathbf{x} = \mathbf{b}$. Then every other solution to $A\mathbf{x} = \mathbf{b}$ is of the form $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, where $\mathbf{x}_h$ is a solution of the corresponding homogeneous system of equations $A\mathbf{x} = \mathbf{0}$. Furthermore, given $\mathbf{x}_p$ and $\mathbf{x}$, there is a unique $\mathbf{x}_h$ such that $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$.

2.4.20. **Corollary.** The syetem $A\mathbf{x} = \mathbf{b}$ has a unique solution if and only if $\mathbf{b} \in \mathrm{Im}(T)$ and the only solution to $A\mathbf{x} = \mathbf{0}$ is the zero vector.

2.5 Composition of Linear Transformations

2.5.1. **Proposition.** If $S : U \to V$ and $T : V \to W$ are linear transformations, then so is TS.
Proof:
Let $\alpha, \beta \in \mathbb{R}$ and let $\mathbf{u}_1, \mathbf{u}_2 \in U$. We must show that TS satisfies Proposition (2.1.2).

$$\begin{aligned} TS(\alpha\mathbf{u}_1 + \beta\mathbf{u}_2) &= T(S(\alpha\mathbf{u}_1 + \beta\mathbf{u}_2)) && \text{// by definition of } TS \\ &= T(\alpha S(\mathbf{u}_1) + \beta S(\mathbf{u}_2)) && \text{// by linearity of } S \\ &= \alpha T(S(\mathbf{u}_1)) + \beta T(S(\mathbf{u}_2)) && \text{// by linearity of } T \\ &= \alpha TS(\mathbf{u}_1) + \beta TS(\mathbf{u}_2) \end{aligned}$$

$\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

2.5.4. **Propositions.**

(a) **Associativity.** Let $R : U \to V$, and $T : W \to X$ be linear transformation of the vectors spaces W, V, W, and X as indicated, then

$$T(SR) = (TS)R$$

(b) **Distributivity I.** Let $R : U \to V$, $S : U \to V$ and $T : V \to W$ be linear transformations of the vectors spaces U, V and W as indicated, then

$$T(R + S) = TR + TS$$

(c) **Distributivity II.** Let $R : U \to V$, $S : V \to W$ and $T : V \to W$ be linear transformations of the vectors spaces U, W and W as indicated, then

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

2.5.6. **Proposition.** Let $S : U \to V$ and $T : V \to W$ be linear transformation, then

(a) $\text{Ker}(S) \subset \text{Ker}(TS)$

(b) $\text{Im}(TS) \subset \text{Im}(T)$

Proof:
We will prove (a) here. If $\mathbf{u} \in \text{Ker}(S)$, $S(\mathbf{u}) = \mathbf{0}$. Then

$$TS(\mathbf{u}) = T(\mathbf{0}) = \mathbf{0}$$

where we notice that the first and last term tells us $\mathbf{u} \in \text{Ker}(TS)$ and this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

2.5.7. **Corollary.** Let $S : U \to V$ and $T : V \to W$ be linear transformations of fin-dim vector spaces, then

(a) $\dim(\text{Ker}(S)) \leq \dim(\text{Ker}(TS))$

(b) $\dim(\text{Im}(TS)) \leq \dim(\text{Im}(T))$

Remark. In naive words, the statement (a) is saying that no transformation can bring what has been smashed to zero back, and thus the composition result of two linear transformations must have a larger kernel. The statement (b) is saying that each time we perform a linear transformation, the "Image Range" of the final output would be restricted, and thus a composition would only result in a tighter restriction.

2.5.9. **Proposition: Compatible Matrix Product.** If $[S]_\alpha^\beta$ has entries a_{ij}, where $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$ and $[T]_\beta^\gamma$ has entries b_{kl}, where $k \in \{1, \ldots, p\}$ and $l \in \{1, \ldots, n\}$, then the entries of $[TS]_\alpha^\gamma$ are

$$\sum_{l=1}^{n} b_{kl} a_{lj}$$

2.5.13. **Proposition: Composition of Transformations Induced By Matrices.** Let $S : U \to V$ and $T : V \to W$ be linear transformation between fin-dim vector spaces. Let α, β, γ be bases for U, V and W, respectively. Then

$$[TS]_\alpha^\gamma = [T]_\beta^\gamma [S]_\alpha^\beta$$

One can comprehend this as: the standard matrix of a transformation that is a composition of two transformations induced by matrices is the matrix product of those matrices. Also note that matrix product *does not* commute.

2.5.14. **Propositions.**

(a) **Associativity.** Let $A \in M_{m\times n}(\mathbb{R}), B \in M_{n\times p}(\mathbb{R}), C \in M_{p\times r}(\mathbb{R})$, then

$$(AB)C = A(BC)$$

(b) **Distributivity I.** Let $A \in M_{m\times n}(\mathbb{R})$ and $B, C \in M_{n\times p}(\mathbb{R})$, then

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

(c) **Distributivity II.** Let $A, B \in M_{m\times n}(\mathbb{R}), C \in M_{n\times p}(\mathbb{R})$, then

$$(A + B)C = AC + BC$$

2.6 The Inverse of A Linear Transformation

2.6.1. **Proposition.** Let $T : V \to W$ be bijective, then the inverse function $S : W \to V$ is a linear transformation.

2.6.2. **Proposition.** A linear transformation $T : V \to W$ has an inverse linear transformation S *if and only if* T is injective and surjective.

2.6.3. **Definition of Inverse and Invertible.** If $T : V \to W$ is a linear transformation that has an inverse transformation $S : W \to V$, we say that T is *invertible*, and we denote the inverse of T here $T^{-1} := S$

2.6.4. **Definition of isomorphism.** If $T : V \to W$ is an invertible linear transformation, T is called an *isomorphism*, and we say V and W are *isomorphic* vector spaces.

2.6.5. **Remark on the T^{-1} notation.** Please note that the inverse here that is acting on a vector is denotes the inverse transformation. For example we have $T : V \to W$, then $w \in W, T^{-1}(w) \in V$ denotes the inverse transformation. We *must* differentiate this from the other notation we encountered earlier which acts on a set rather than on a vector. As an example, we have $T : V \to W$ and $\mathbf{0} \in W$, then

$$T^{-1}(\{\mathbf{0}\}) \equiv \{\mathbf{v} \in V, T(\mathbf{v}) = \mathbf{0}\}$$

Notice that this special example is a.k.a the Kernel of the transformation by definition.

2.6.7. **Proposition.** If V and W are finite-dimensional vector spaces, then there is an isomorphism $T : V \to W$ *if and only if* $\dim(V) = \dim(W)$.
Proof($\Longrightarrow$):
If T is an isomorphism, then we know that T is bijective. So, $\dim(\mathrm{Ker}(T)) = 0$ and $\dim(\mathrm{Im}(T)) = \dim(W)$. Then by the Rank-Nullity Theorem, we conclude that $\dim(V) = \dim(W)$.
Proof($\Longleftarrow$):
If $\dim(V) = \dim(W)$, we must produce an isomorphism $T : V \to W$. Let $\alpha = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for B and $\beta = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be a basis for W. Define T to be the linear transformation from V to W with $T(\mathbf{v}_i) = w_i, i = 1, \ldots, n$. By Proposition (2.1.14), T is uniquely determined by tis choice of values on α. To see that T is injective, notice that if

$$T(a_1\mathbf{v}_1 + \ldots + a_n\mathbf{v}_n) = \mathbf{0}$$

then we have

$$a_1\mathbf{w}_1 + \ldots + a_n\mathbf{w}_n = \mathbf{0}$$

Since the $\mathbf{w}$'s are a basis, we know immediately that $a_1 = \ldots = a_n = 0$. Then $\mathrm{Ker}(T) = \{\mathbf{0}\}$ and T is injective. By proposition (2.4.10) T is also surjective, and then by Proposition (2.6.2) it is an isomorphism.

2.6.9. **The Gauss-Jordan Method of Inverse[8].** Suppose we have the following matrix:

$$A = [T]_{\alpha}^{\beta} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

and we want to find the matrix inverse $B = [T^{-1}]_{\beta}^{\alpha}$. To achieve this, we augment this matrix with the compatible identity map $I_{n\times n}$.

$$\left[\begin{array}{ccc|ccc} a_{11} & \dots & a_{1n} & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} & 0 & \dots & 1 \end{array}\right] = [A|I] \tag{2.2}$$

We row reduce (2.2) *w.r.t* the left half matrix with a result of the following form:

$$\left[\begin{array}{ccc|ccc} 1 & \dots & 0 & b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & b_{n_1} & \dots & b_{nn} \end{array}\right] = [I|B] = \left[I|A^{-1}\right]$$

We then record the $n \times n$ coefficients (b_{ij}) on the right hand side which is our end result of the inverse.

2.6.10. **Definition of invertible matrices.** An $n \times n$ matrix is called invertible if there exists an $n \times n$ matrix B so that $AB = BA = I$. Where the I denotes the compatible identity map and B is called an inverse of A and $A^{-1} := B$

2.6.11. **Proposition.** Let $T : V \to W$ be an isomorphism of finite-dimensional vector spaces. Then for any choice of bases α for V and β for W, we have

$$\left[T^{-1}\right]_{\beta}^{\alpha} = \left[[T]_{\alpha}^{\beta}\right]^{-1}$$

Proof(from the book):
To show the above identity, it suffices to show that $[T^{-1}]_{\beta}^{\alpha}$ is the matrix inverse of $[T]_{\alpha}^{\beta}$. On one hand we have

$$\left[T^{-1}\right]_{\beta}^{\alpha}[T]_{\alpha}^{\beta} = \left[T^{-1}T\right]_{\alpha}^{\alpha} = [I_{n\times n}]_{\alpha}^{\alpha}$$

[8]Extended based on this example

On the other hand, we have

$$[T]_\alpha^\beta \left[T^{-1}\right]_\beta^\alpha = \left[TT^{-1}\right]_\beta^\beta = [I_{n\times n}]_\beta^\beta$$

Then, since the $n \times n$ identity map is unique, $[I_{n\times n}]_\beta^\beta = [I_{n\times n}]_\alpha^\alpha$, we can equate the above two equations, which, by definition tells us that $[T^{-1}]_\beta^\alpha$ is the matrix inverse of $[T]_\alpha^\beta$. We conclude that the above identity is true. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

Proof(from class):
Starting from the identity matrix in the α basis, we have

$$[I]_\alpha^\alpha = \left[T^{-1}T\right]_\alpha^\alpha = \left[T^{-1}\right]_\alpha^\beta [T]_\beta^\alpha \tag{2.3}$$

from where we equate the first and last term of the above line, and notice that by definition $[T]_\beta^\alpha$ is a inverse of $[T^{-1}]_\alpha^\beta$, so $\left[[T]_\beta^\alpha\right]^{-1} = [T^{-1}]_\alpha^\beta$ and this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

2.7 Change of Basis

2.7.2. **Remarks on the notation.** Note that the here the $[I]_\alpha^{\alpha'}$ is *not* the identity map since $\alpha \neq \alpha'$. This notation will come up in the following chapter naturally but it is important to differentiate this notation with the identity map.

2.7.*. **Remark on the identity map.** Note that the inverse of a identity map is itself, that is

$$\left[[I]_\alpha^\beta\right]^{-1} \equiv [I]_\alpha^\beta, \forall\ \alpha, \beta \text{ bases } \forall \text{ vector spaces } V$$

2.7.3. **Proposition.** Let V be a finite-dimensional vector space, and let α, α' be bases for V. Let $\mathbf{v} \in V$. Then the coordinate vector $[\mathbf{v}]_{\alpha'}$ of $\mathbf{v}$ in the basis α' is related to the coordinate vector $[\mathbf{v}]_\alpha$ of $\mathbf{v}$ in the basis α by

$$[I]_\alpha^{\alpha'}[\mathbf{v}]_\alpha = [\mathbf{v}]_{\alpha'}$$

2.7.5. **Theorem.** Let $T : V \to W$ be a linear transformation Between FDVS V and W. Let $I_V : V \to V$ and $I_W : W \to W$ be respective identity

transformations of V and W. Let α, α' be two bases for V and let β, β' be two bases for W, then

$$[T]_{\alpha'}^{\beta'} = [I_W]_{\beta}^{\beta'} \cdot [T]_{\alpha}^{\beta} \cdot [I_V]_{\alpha'}^{\alpha}$$

Remark. In naive words, in order to produce a transformation from bases α' to β' we first make the change of basis from α' to α, then we perform the transformation from α basis to β basis and lastly we change the basis to the wanted β' through a "cross basis identity map". We now present the proof, using the change of basis facts:
Proof:
Since $T = I_W T I_V$, we can write

$$\begin{aligned} [T]_{\alpha'}^{\beta'} &= [I_W T I_V]_{\alpha'}^{\beta'} \\ &= [I_W T]_{\alpha}^{\beta'} [I_V]_{\alpha'}^{\alpha} \\ &= [I_W]_{\beta}^{\beta'} [T]_{\alpha}^{\beta} [I_V]_{\alpha'}^{\alpha} \end{aligned}$$

$\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

2.7.6. **Definition of Similar Matrices.** Let A, B be $n \times n$ matrices. A and B are said to be *similar* if there exists an invertible matrix Q such that:

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

Applying this new definition, we can see the follows: Let $T : V \to V$ is a linear transformation and α, α' are two bases for V, then $A = [T]_{\alpha}^{\alpha}$ is similar to $B = [T]_{\alpha'}^{\alpha'}$, and the invertible matrix Q in the definition is the matrix $Q = [I_V]_{\alpha'}^{\alpha}$

Chapter 3

The Determinant Function

3.1 The Determinant as Area

3.1.1. **Propositions: Geometry of a parallelogram in $\mathbb{R}^2$**

(a) The area of the parallelogram with vertices $\mathbf{0}, \mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{a}_1 + \mathbf{a}_2$ is $\pm(\mathbf{a}_{11}\mathbf{a}_{22} - \mathbf{a}_{12}\mathbf{a}_{21})$

(b) The area is not zero if and only if the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are linearly independent.

3.1.2. **Corollary.** Let $V = \mathbb{R}^2$, $T : V \to V$ is an isomorphism if and only if the area of the parallelogram constructed previously is non-zero

3.1.3. **Proposition.** The function $Area(\mathbf{a}_1, \mathbf{a}_2)$ has the following properties for $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_1'$, and $\mathbf{a}_2' \in \mathbb{R}^2$

(a) $Area(b\mathbf{a}_1 + c\mathbf{a}_1', \mathbf{a}_2) = b\ Area(\mathbf{a}_1, \mathbf{a}_2) + c\ Area(\mathbf{a}_1', \mathbf{a}_2)$ for $b, c \in \mathbb{R}$

(b) [1] $Area(\mathbf{a}_1, b\mathbf{a}_2 + c\mathbf{a}_2') = b\ Area(\mathbf{a}_1, \mathbf{a}_2) + c\ Area(\mathbf{a}_1, \mathbf{a}_2')$ for $b, c \in \mathbb{R}$

(c) $Area(\mathbf{a}_1, \mathbf{a}_2) = -Area(\mathbf{a}_2, \mathbf{a}_1)$, and

(d) $Area((1, 0), (0, 1)) = 1$.

Remark. Recall that (d) is the *n-cube* in $\mathbb{R}^2$, where n-cube is defined as follows: (with $\mathbf{e}_i$'s as components of ξ basis for $\mathbb{R}^n$ Euclidean Space)

$$C_n := \left\{ \mathbf{x} \in \mathbb{R}^n | \mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i \text{ for some } \alpha_i \in [0, 1] \right\} \equiv [0, 1]^n$$

[1] This essentially the same as (a)

3.1.4. **Proposition.** If $B(\mathbf{a}_1, \mathbf{a}_2)$ is any real-valued function of $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^2$ that satisfies Properties (a), (c) and (d) of Proposition (3.1.3), then B is equal to the area function.

3.1.5. **Definition of *determinant* of a 2×2 matrix.**

(a) $\det(b\mathbf{a}_1 + c\mathbf{a}_1', \mathbf{a}_2) = b\det(\mathbf{a}_1, \mathbf{a}_2) + c\det(\mathbf{a}_1', \mathbf{a}_2)$ for $b, c \in \mathbb{R}$

(b) $\det(\mathbf{a}_1, \mathbf{a}_2) = -\det(\mathbf{a}_2, \mathbf{a}_1)$, and

(c) $\det(\mathbf{e}_1, \mathbf{e}_2) = 1$

As a consequence of (3.1.4), $\det(A)$ is given by

$$\det(A) \equiv a_{11}a_{22} - a_{12}a_{21}$$

where a_{ij} refers to the i-th row, j-th column of the matrix A.

3.1.6. **Propositions: Determinant in relations.**

(a) A 2×2 matrix A is invertible if and only if $\det(A) \neq 0$

(b) If $T : V \to V$ is linear transformation of a two-dimensional vector space V, then T is an isomorphism if and only if $\det([T]_\alpha^\alpha) \neq 0$

3.2 The Determinant of an $n \times n$ Matrix

3.2.1. **Definition of Multi-linear.** A function f of the rows of a matrix A is called multi-linear of f is linear function of each of its rows when the remaining rows are held fixed. That is, f is multi-linear if for all $b, b' \in \mathbb{R}$,

$$f(\mathbf{a}_1, \ldots, b\mathbf{a}_i + b'\mathbf{a}_i', \ldots, \mathbf{a}_n) \\ = bf(\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_n) + b'f(\mathbf{a}_1, \ldots, \mathbf{a}_i', \ldots, \mathbf{a}_n)$$

3.2.2. **Definition of Alternating.** A function f of rows of a matrix A is said to be alternating if whenever any two rows of A are interchanged f changes sign. That is, for all $i \neq j, 1 \leq i, j \leq n$, we have

$$f(\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n) = -f(\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_n)$$

Do notice the switched position of $\mathbf{a}_i$ and $\mathbf{a}_j$ in the above equality.

3.2.2. **Lemma.** If f is an alternating real-valued function of the rows of an $n \times n$ matrix A are identical, then $f(A) = 0$.
Proof:
Assume that $\mathbf{a}_i = \mathbf{a}_j$. Then,

$$\begin{aligned} f(A) &= f(\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n) \\ &= -f(\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_n) \\ &= -f(A) \end{aligned}$$

Hence $f(A) = 0$. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$
Remark: Sanity check in $\mathbb{R}^2$. Notice that we have

$$\det \begin{bmatrix} a & b \\ a & b \end{bmatrix} = ab - ab = 0$$

3.2.4. **Definition of Minor.** Let the A be an $n \times n$ matrix with entries a_{ij}, where $i, j = 1, \ldots, n$. The ij-th *minor matrix* of A is defined to be the $(n-1) \times (n-1)$ matrix by deleting the i-th row and j-th column of A. We denote such minor matrix as A_{ij}.

3.2.5. **Proposition.** Let A be a 3×3 matrix, and let f be an alternating multi-linear function. Then

$$f(A) = [a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + a_{13} \det(A_{13})]\, f(I)$$

3.2.6. **Corollary.** There exists exactly one multi-linear alternating function f of the rows of a 3×3 matrix such that $f(I) = 1$.

3.2.7. **The *det* of 3×3 matrices.** The determinant function of a 3×3 matrix is the unique alternating multi-linear function f with $f(I) = 1$. We denote this function by $\det(A)$.

3.2.8. **Theorem.** There exits exactly one alternating multi-linear function $f : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ satisfying $f(I) = 1$, which is called the determinant function $f(A) = \det(A)$. Further, any alternating multi-linear function f satisfies $f(A) = \det(A) f(I)$.

3.2.10. **Proposition: non-invertible square matrix have zero *det*.** If an $n \times n$ matrix A is not invertible, then $\det(A) = 0$.

Proof:
If A is not invertible, then the row space of A has dimension at most $(n-1)$. Therefore, there is a linear dependence among the rows of A,

$$x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n = \mathbf{0} \qquad \exists i \ s.t. \ x_i \neq 0$$

Without loss of generality, we assume that $x_1 \neq 0$, then we can divide the entire equation by x_1. Solving for $\mathbf{a}_1$ yields us

$$\mathbf{a}_1 = -\frac{1}{x_1(x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n)} = \sum_{i=2}^{n}\left(\frac{-x_i}{x_1}\mathbf{a}_i\right)$$

Then we consider the determinant function,

$$\begin{aligned}\det(A) &= \det(\mathbf{a}_1, \ldots, \mathbf{a}_n) \\ &= \det\left(\sum_{i=2}^{n}\left(\frac{-x_i}{x_1}\mathbf{a}_i\right), \mathbf{a}_2, \ldots, \mathbf{a}_n\right) \\ &= \sum_{i=2}^{n}\left(\frac{-x_i}{x_1}\right)\det(\mathbf{a}_i, \mathbf{a}_2, \ldots, \mathbf{a}_n) \\ &= 0\end{aligned}$$

since each term in the final sum is the determinant of a matrix with repeated rows, and by (3.2.2), we conclude the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

3.2.11. **Proposition.** The following equality always holds:

$$\det(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \det(\mathbf{a}_1, \ldots, \mathbf{a}_i + b\mathbf{a}_j, \ldots, \mathbf{a}_n)$$

where $\mathbf{a}_i + b\mathbf{a}_j$ appears in the i-th position.
Proof:
This result could be shown using some algebraic manipulation, we proceed as follows:

$$\begin{aligned}&\det(\mathbf{a}_1, \ldots, \mathbf{a}_i + b\mathbf{a}_j, \ldots, \mathbf{a}_n) = \\ &\qquad \det(\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_n) + b\det(\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n) \\ &\qquad\qquad = \det(\mathbf{a}_1, \ldots, \mathbf{a}_n)\end{aligned}$$

Since the term $(\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n)$ contains repeated rows, hence det evaluates to zero. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

3.2.12. **Lemma: Determinant of a diagonal matrix.** If $A \in M_{n\times n}(\mathbb{R})$ is a diagonal matrix, then $\det(A) = a_{11}a_{22}\cdots a_{nn} = \prod_{i=1}^{n} a_{ii}$

Proof:

We can actually show a more general claim than this.

Claim: If $A \in M_{n\times n}(\mathbb{R})$ is a upper/lower triangular matrix, then $\det(A) = \prod_{i=1}^{n} a_{ii}$. Notice that we will only prove the case for the upper triangular matrix, since the other is basically a mirror image.

We proceed by mathematical induction on n, the size of the matrix. Define predicate $P(n)$: If $T_n \in M_{n\times n}(\mathbb{R})$ is upper triangular than $\det(A) = \prod_{i=1}^{n} a_{ii}$. This is our Induction Hypothesis.

WTS: $\forall n \geq 1, P(n)$.

Basis Case: $n = 1$. It is clearly the case that

$$\det(A) = a_{11} = \left[\prod_{i=1}^{n} a_{ii}\right|_{n=1}$$

so $P(1)$ holds.

Inductive Step: Let $n \in \mathbb{N}^{\geq 1}$. Assume $P(n)$, we must show that $P(n+1)$ follows. Let T_{n+1} be any upper triangular matrix of size $(n+1)\times(n+1)$. Using the minor matrices expansion of the left most column, we have

$$\det(T_{n+1}) = \sum_{i=1}^{n+1}(-1)^{i+1}(T_{n+1})_{i1}\det((T_{n+1})_{i1}) \tag{3.1}$$

$$= [(T_{n+1})_{i1}\det((T_{n+1})_{i1})|_{i=1} + 0 + \ldots + 0 \tag{3.2}$$

where the first $(T_{n+1})_{i1}$ means the $i1$-th element in the matrix T_{n+1} and the second one means the $i1$ minor. We notice that our I.H. tells us that

$$[\det((T_{n+1})_{i1})|_{i=1} = \prod_{i=2}^{n+1} a_{ii}$$

Then, we can rewrite (3.2) as follows

$$\det(T_{n+1}) = (T_{n+1})_{11}\prod_{i=2}^{n+1} a_{ii} = \prod_{i=1}^{n+1} a_{ii}$$

where the a_{ij} represents the ij-th element in T_{n+1}. So $P(n)$ follows, and this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

3.2.13. **Proposition.** If A is invertible, then $\det(A) \neq 0$

3.2.14. **Theorem: Necessary and Sufficient Condition for Invertibility.** Let $A \in M_{n\times n}(\mathbb{R})$. Then, A is invertible if and only if $\det(A) \neq 0$

3.3 Further Properties of The Determinant

3.3.*. **Definition of Cofactor** Given that A_{ij} is the minor matrix for row i and column j in a square matrix A, the *cofactor* at position row i and column j is denoted as C_{ij} and is calculated by the formula

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

Remark. Although the formula seems daunting, it is essentially assigning alternating plus and minus to each slot in each row, going left to right and top to bottom.

3.3.1. **Proposition.** $AA' = \det(A)I$, where $A' := [C_{ij}]_{ij} \in M_{n\times n}(\mathbb{R})$ is the cofactor matrix.

3.3.2. **Corollary.** If A is an invertible $n \times n$ matrix, then A^{-1} is the matrix whose ij-th column is

$$(-1)^{i+j}\frac{\det(A_{ji})}{\det(A)}$$

We can re-write this in notation familiar from FM[2]

$$A^{-1} = \frac{1}{\det(A)}\left[C_A\right]^T$$

where C_A denotes the cofactor matrix, with entires calculated using the formula specified in (3.3.*) above.

3.3.4. **Proposition.** For any fixed j such that $1 \leq j \leq n$, we have

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

3.3.7. **Proposition.** If A and B are $n \times n$ matrices, then

[2]CIE Advanced-Level Further Mathematics

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

(b) If A is invertible, then $\det(A^{-1}) = 1/\det(A)$

3.3.8. **Corollary.** If $T : V \to V$ is a linear transformation, $\dim(V) = n$, then

$$\det([T]_\alpha^\alpha) = \det([T]_\beta^\beta), \forall \text{ bases } \alpha, \beta \text{ for } V$$

3.3.9. **Definition: Determinant of a Linear Transformation.** The determinant of an linear transformation $T : V \to V$ of a finite dimensional vector space is the determinant of $[T]_\alpha^\alpha$ for any choice of α. We denote this with $\det(T)$.

3.3.11. **Proposition: Determinant In Relation w/ Isomorphism.** A linear transformation $T : V \to V$ of a fin-dim vector space is an isomorphism if and only if $\det(T) \neq 0$.
Proof:
Notice that we proved $\det(T) \neq 0 \iff [T]_\alpha^\alpha$ is an isomorphism, so this proof is trivial. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

3.3.12. **Proposition.** Let $S : V \to V$ and $T : V \to V$ be linear transformations of a fin-dim vector space, then

(a) $\det(ST) = \det(S)\det(T)$

(b) T is an isomorphism $\implies \det(T^{-1}) = \det(T)^{-1}$

THIS CHAPTER IS CURRENTLY ARCHIVED...

Chapter 4

Eigen-Problems and Spectral Theorem

4.1 Eigenvalues and Eigenvectors

4.1.2. **Definitions.** Let $T : V \to V$ be any linear mapping, then

(a) A vector $\mathbf{v} \in V$ is called an eigenvector of T if

$$\mathbf{v} \neq \mathbf{0} \wedge \exists \lambda \in \mathbb{R}, T(\mathbf{v}) = \lambda \mathbf{v}$$

4.1.5. **Proposition.** A vector $\mathbf{x}$ is an eigenvector of T with eigenvalue λ if and only if $\mathbf{x} \neq \mathbf{0} \wedge \mathbf{x} \in \ker(T - \lambda I)$

4.1.6. **Definition.** Let $T : V \to V$ be a linear mapping, and let $\lambda \in \mathbb{R}$. The $\lambda-$eigenspace of T, denoted E_λ, is the set

$$E_\lambda = \{\mathbf{x} \in V | T(\mathbf{x}) = \lambda \mathbf{x}\}$$

4.1.9. **Proposition.** Let $A \in M_{n\times n}(\mathbb{R})$. Then $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$

4.1.11. **Definition of Characteristic Polynomial.** Let $A \in M_{n\times n}(\mathbb{R})$. The polynomial $\det(A - \lambda I)$ is called the characteristic polynomial of A.

4.1.12. **Proposition: Similar matrices have equal char-polies**. If two matrices are similar to each other, then they have equal characteristic

polynomial.

Proof:

Suppose A, B are two similar matrices, so that $B = Q^{-1}AQ$ for some invertible matrix Q. Then we have

$$\begin{aligned}\det(B - \lambda I) &= \det(Q^{-1}AQ - \lambda I) \\ &= \det(Q^{-1}AQ - \lambda(Q^{-1}IQ)) \\ &= \det(Q^{-1}AQ - Q^{-1}\lambda IQ) \\ &= \det(Q^{-1}(A - \lambda I)Q) \\ &= \det(Q^{-1})\det(A - \lambda I)\det(Q) \\ &= \frac{\det(A - \lambda I)\det(Q)}{\det(Q)} = \det(A - \lambda I)\end{aligned}$$

and this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

4.1.13. **Alternative form of char-poly of a 2×2 matrix.**[1] For a general 2×2 matrix A, we have

$$Char_A(\lambda) = \det(A - \lambda I) = \lambda^2 - Tr(A) \cdot \lambda + \det A$$

Remark. This form of characteristic polynomial also takes a more general form for larger square matrices, but the expanded formula is messy so we omit it.

A special case of the Cayley-Hamilton Theorem. Consider some 2×2 matrix A, with characteristic polynomial $Char_A(\lambda)$. Then, if we treat A as a number and brutally plug A into its own characteristic polynomial, i.e. $Char_A(A)$, we will find that $Char_A(A) = 0_{2\times 2}$, as if A solves its own characteristic polynomial!

4.1.14. **Corollary.** Let $A \in M_{n\times n}(\mathbb{R})$. Then A has no more than n distinct eigenvalues. In addition, if $\lambda 1, ..., \lambda_k$ are the distinct eigenvalues of A and λ_i is an m_i-fold root[2] of the characteristic polynomial, then $m_1 + ... + m_k \leq n$.

4.1.18. **Statement of the Cayley-Hamilton Theorem.** Let A be any $n \times n$ matrix with characteristic polynomial $Char_A(\lambda)$, then

$$Char_A(A) = 0_{n\times n}$$

[1] Based on example (4.1.13).

[2] Root of the form $(\lambda_i \pm k)^{m_i}$, where $k \in \mathbb{R}$

4.2 Diagonalizability

4.2.1. **Definition of Diagonalizability.** Let V be a finite-dimensional vector space, and let $T : V \to V$ be a linear mapping. T is said to be diagonalizable if there exists a basis V, all of whose vectors are eigenvectors of T.

4.2.2. **Proposition.** $T : V \to V$ is diagonalizable if and only if, for any basis α of V, the matrix $[T]^{\alpha}_{\alpha}$ is similar to a diagonal matrix.
Proof($\Longrightarrow$):
% to appear added later...
Proof($\Longleftarrow$):
If $A = [T]^{\alpha}_{\alpha}$ is similar to a diagonal matrix D, then by definition, there is some invertible matrix P such that $P^{-1}AP = D$. Then, by rearranging, we have $AP = DP$. Notice that since D is diagonal matrix, the columns of PD are just scalar multiples of the columns of P. Hence, the columns of P are all eigenvectors for the matrix A, or equivalently, the columns of P are the coordinate vectors of eigenvectors for the mapping T. On the other hand, since P is invertible, we know its columns are linearly independent, hence those columns are coordinate vectors for a basis of V. Therefore, by definition, T is diagonalizable. **Note** that eigenvalues of T are the diagonal entries of D, and this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$
In naive words, in order for a linear mapping to be diagonalizable, it needs to have enough linearly independent eigenvectors to form a basis of V. To be specific, we need n linearly independent eigenvectors for a n-dimensional to n-dimensional mapping to be diagonalizable.

4.2.4. **Proposition.** Let $\mathbf{x}_i (1 \leq i \leq k)$ be eigenvectors of a linear mapping $T : V \to V$ corresponding to distinct eigenvalues λ_i, then $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$ is linearly independent subset of V.

4.2.5. **Corollary.** For each $i (1 \leq i \leq k)$, let $\{\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,n_i}\}$ be a linearly independent set of eigenvectors of T all with eigenvalue λ_i and suppose the λ_i are distinct. Then,

$$S = \bigcup_{i=1}^{k} \{\mathbf{x}_{i,1}, ..., \mathbf{x}_{i,n_i}\}$$

is linearly independent.

4.2.6. **Proposition.** Let V be finite-dimensional, and let $T : V \to V$ be linear. Let λ be an eigenvalue of T, and assume that λ is an m-fold root of the characteristic polynomial of T. Then, $1 \leq \dim(E_\lambda) \leq m$.

4.2.7. **Theorem.** Let $T : V \to V$ be a linear mapping on a finite-dimensional vector space V, and let $\lambda_1, ..., \lambda_k$ be its distinct eigenvalues. Let m_i be the multiplicity of λ_i as a root of the characteristic polynomial of T. Then T is diagonalizable if and only if

(a) $m_1 + ... + m_k = n = \dim(V)$, ***and***
In naive words, this means T has to have n real eigenvalues

(b) $\forall i, \dim(E_{\lambda_i}) = m_i$
In naive words, this means each of the E_{λ_i} attains its maximum dimension specified in (4.2.6), i.e. $\dim(E_{\lambda_i}) = m_i$

4.2.8. **Corollary.** Let $T : V \to V$ be linear mapping on a finite-dimensional space V, and assume that T has $\dim(V) = n$ distinct real eigenvalues. Then T is diagonalizable.
Proof:
Since T has n distinct real roots, then the algebraic multiplicity associated with each root must be 1. Then notice that $1 \leq \dim(E_{\lambda_i}) \leq m_i = 1, \forall i$, and we conclude the result by applying Theorem (4.2.7). This completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

4.2.9. **Corollary.** A linear mapping $T : V \to V$ on a finite dimensional space V is diagonalizable if and only if the sum of the multiplicities of the real eigenvalues is $n = \dim(V)$, and either

(a) We have $\sum_{i=1}^{k} \dim(E_{\lambda_i}) = n$, where λ_i are the distinct eigenvalues of T, or

(b) We have $\sum_{i=1}^{k} (n - \dim(\mathrm{Im}(T - \lambda_i I))) = n$, where again λ_i's are the distinct eigenvalues.

Remark. The second statement is a direct consequence of the Rank-Nullity Theorem.

4.3 Geometry in Euclidean Space

4.3.1. **Definition: Standard Inner Product.** [3]The standard inner product (or dot product) on $\mathbb{R}^n$ is the function $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined be the following rule: If $\mathbf{x} = (x_1, ..., x_n), \mathbf{y} = (y_1, ..., y_n)$ in the ξ coordinate basis, then

$$\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^{n} x_i y_i$$

4.3.2. **Propositions.** The Inner Product has the following properties:

(a) **Bi-linearity:** $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n, \forall c \in \mathbb{R}, \langle c\mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = c\langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$

(b) **Symmetric:** $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$

(c) **Positive-Definite:** $\forall \mathbf{x} \in \mathbb{R}^n, \langle \mathbf{x}, \mathbf{x} \rangle \geq 0 \wedge \langle \mathbf{x}, \mathbf{x} \rangle = 0 \iff \mathbf{x} = \mathbf{0}$

4.3.3. **Definitions.** We have the following related definitions:

(a) The length, norm, of $\mathbf{x} \in \mathbb{R}^n$ is the scaller defined as

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

(b) $\mathbf{x}$ is called a unit vector if $\|\mathbf{x}\| \equiv 1$, we can normalize a vector $\mathbf{x} \in \mathbb{R}^n$ by $\frac{\mathbf{x}}{\|\mathbf{x}\|}$

4.3.4-(1). **The Cauchy-Schwarz Inequality.**

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$$

This is a historically important inequality and we shall present the proof here:
Proof:
Consider the following inner product, where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n, c \in \mathbb{R}$. We can expand the terms using bilinearity and the symmetric property of of inner product

$$\begin{aligned} \langle \mathbf{x} - c\mathbf{y}, \mathbf{x} - c\mathbf{y} \rangle &= \langle \mathbf{x}, \mathbf{x} \rangle - c\langle \mathbf{x}, \mathbf{y} \rangle - c\langle \mathbf{y}, \mathbf{x} \rangle + c^2 \langle \mathbf{y}, \mathbf{y} \rangle \\ &= \|\mathbf{x}\|^2 + c^2 \|\mathbf{y}\|^2 - 2c\langle \mathbf{x}, \mathbf{y} \rangle \geq 0 \end{aligned}$$

[3]The version of Inner product that appeared earlier in these notes will *always* refer to the standard inner product for the n-dimensional Euclidean Space.

notice that the last "$\geq$" conclusion was drawen based on the positive definite property of the inner product of a vector with itself. Then, we consider

$$c = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{y}\|} \in \mathbb{R}$$

Then our former inequality gives us

$$\begin{aligned} 0 &\leq \|\mathbf{x}\|^2 - \frac{2\langle \mathbf{x}, \mathbf{y} \rangle^2}{\|\mathbf{y}\|^2} + \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\|\mathbf{y}\|^2} \\ &= \|\mathbf{x}\|^2 - \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\|\mathbf{y}\|^2} \\ \implies \|\mathbf{x}\|^2 &\geq \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{y}\|^2} \\ \implies \langle \mathbf{x}, \mathbf{y} \rangle^2 &\leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2 \end{aligned}$$

where we can take the square root on both sides, and this yields us desired $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

4.3.4-(2). **The Triangular Inequality.**

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

This geometrically says that any side of a triangle is less than the sum of the other two sides. This is a consequence of the Cauchy-Schwarz Inequality above.

Proof:

Consider the inequality presented in the proof for Cauchy-Schwarz above, and take the special case of $c = -1 \in \mathbb{R}$

$$\begin{aligned} 0 \leq \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle &= \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 \\ \implies \|\mathbf{x} + \mathbf{y}\|^2 &= \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 \\ &\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 \end{aligned}$$

Taking square root on both sides of the above inequality yields the trigular inequality. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

4.3.5. **Definition of angle between vectors.** The angle between two non-zero vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined to be

$$\theta = \cos^{-1}\left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}\right)$$

Notice that we defined this angle in chapter 2, but we focused on $\mathbb{R}^2$ and $\mathbb{R}^3$. This is an extension to arbitrary n-dimensional Euclidean Space.

4.3.7. **Definition of Orthogonality.** Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are defined to be orthogonal if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$

4.3.9. **Definitions: Orthogonal and Orthonormal**

(a) A set of vectors $S \subset \mathbb{R}^n$ is said to be *orthogonal* if $\forall \mathbf{x}, \mathbf{y} \in S, \mathbf{x} \neq \mathbf{y} \implies \langle \mathbf{x}, \mathbf{y} \rangle = 0$

(b) A set of vectors $S \subset \mathbb{R}^n$ is said to be *orthonormal* if S is orthogonal and, in addition, $\forall \mathbf{x} \in S, \|\mathbf{x}\| = 1$

4.3.10. **Proposition: Orthogonal vectors are linearly independent.** If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are orthogonal, with $\mathbf{x} \neq \mathbf{0} \neq \mathbf{y}$, then $\{\mathbf{x}, \mathbf{y}\}$ is linearly independent.

4.3.10*. **Extension to Proposition.**[4] If we wish, we could extend the Proposition (4.3.10) to the following: Let $S = \{\mathbf{x}_1, ..., \mathbf{x}_k\} \subseteq \mathbb{R}^n$ be any orthogonal set of non-zero vectors. Then, S is linearly independent. To prove this, we can use a simple induction argument that adding a vector that is linearly independent to already existing vectors in a set would not destroy the linear independence ecosystem.

4.3.*. **Polarization Identity.** This is question 6 in the Exercises section of 4.3. We have the following statement

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \langle \mathbf{x}, \mathbf{y} \rangle = (1/4)\left(\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2\right)$$

this shows that we can recover the inner product on $\mathbb{R}^n$ just by measuring lengths of vectors.

[4] Question 8, Exercise 4.3

Proof:
Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then

$$\begin{aligned}\frac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2\right) &= \frac{1}{4}\left(\langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle - \langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y}\rangle\right) \\ &= \frac{1}{4}\left(4\langle \mathbf{x}, \mathbf{y}\rangle\right) \\ &= \langle \mathbf{x}, \mathbf{y}\rangle\end{aligned}$$

and this completes the proof. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

4.4 Orthogonal Projections and The Gram-Schmidt Process

4.4.1. **Definition of Orthogonal Complement.** The orthogonal complement of W, denoted as $W^\perp$, is defined to be

$$W^\perp := \{\mathbf{v} \in \mathbb{R}^n | \langle \mathbf{v}, \mathbf{w}\rangle = 0, \forall \mathbf{w} \in W\}$$

Remark. Notice that the $\mathbf{0}$ vector should be in any such $W^\perp$ sets, since the zero vectors is orthogonal to every vector.

4.4.2. **(Interesting) Examples.**

(a) If we consider $W = \{\mathbf{0}\}$, then we see that every vector in $\mathbb{R}^n$ (including $\mathbf{0}$) is in the orthogonal complement of W. So, $\{\mathbf{0}\}^\perp \equiv \mathbb{R}^n$

(b) On the otherhand, if we consider $W = \mathbb{R}^n$. Then only the zero vector is in the orthogonal complement of W. So, $(\mathbb{R}^n)^\perp \equiv \{\mathbf{0}\}$

4.4.3. **Propositions.**

(a) $\forall$ subspace $W \subset \mathbb{R}^n, W^\perp \subset \mathbb{R}^n$ is also a subspace

(b) $\dim(W) + \dim(W^\perp) = \dim(\mathbb{R}^n) = n$

(c) $\forall\, W \subset \mathbb{R}^n$, W is a subspace $\implies W \cap W^\perp = \{\mathbf{0}\}$

(d) Given a subspace $W \subset \mathbb{R}^n$, $\forall \mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}$ can be written uniquely as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1 \in W, \mathbf{x}_2 \in W^\perp$. In summary, we have $\mathbb{R}^n = W \oplus W^\perp$

4.4.5. **Propositions.**

(a) P_W is a linear mapping

(b) $\text{Im}(P_W) = W, \mathbf{w} \in W \implies P_W(\mathbf{w}) = \mathbf{w}$

(c) $\text{Ker}(P_W) = W^\perp$

4.4.6. **Proposition.** Let $\{\mathbf{w}_1, ..., \mathbf{w}_k\}$ be an orthonormal basis for the subspace $W \subseteq \mathbb{R}^n$

(a) $\forall \mathbf{w} \in W, \mathbf{w} = \sum_{i=1}^k \langle \mathbf{w}, \mathbf{w}_i \rangle \mathbf{w}_i$

(b) $\forall \mathbf{x} \in \mathbb{R}^n, P_W(\mathbf{x}) = \sum_{i=1}^k \langle \mathbf{x}, \mathbf{w}_i \rangle \mathbf{w}_i$

4.4.9. **Theorem: Every Subspace has an Orthonormal Basis.** Let W be a subspace of $\mathbb{R}^n$. Then there exists an orthonormal basis of W.
Remark. A way to think of this theorem and make this trivial is that we first pick some orthogonal basis for such subspace, and then scale them so that they are unit vectors.

4.4.9*. **Gram-Schmidt Process.** Consider $W = span\{\mathbf{w}_j\}_{j=1}^k$, were the $\mathbf{w}$'s are basis vectors for the vector space W. Then we calculate, in sequence

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{w}_1 \\ \mathbf{v}_2 &= \mathbf{w}_2 - P_{span\{\mathbf{v}_1\}}\mathbf{w}_2 \\ &\cdots \\ \mathbf{v}_k &= \mathbf{w}_k - P_{span\{\mathbf{v}_1,...,\mathbf{v}_{k-1}\}}\mathbf{w}_k \end{aligned}$$

after which we normalize $\mathbf{v}_i$ by $\hat{\mathbf{v}}_i = \mathbf{v}/\left\|\mathbf{v}_i\right\|$, then $\alpha = \{\hat{\mathbf{v}}_1, ..., \hat{\mathbf{v}}_k\}$ is an orthonormal basis for the real vector space W.

4.5 Symmetric Matrices

4.5.2a. **Proposition.** Let $A \in M_{n\times n}(\mathbb{R})$.

(a) $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A^T\mathbf{y} \rangle$

(b) A is symmetric if and only if $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A\mathbf{y} \rangle, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

4.5.2b. **Corollary.** Let V be any subspace of $\mathbb{R}^n$, let $T : V \to V$ be any linear transformation, and let $\alpha = \{\mathbf{x}_1, ..., \mathbf{x}_k\}$ be any orthogonal basis for V. Then $[T]_\alpha^\alpha$ is a symmetric matrix if and only if

$$\langle T(\mathbf{x}), \mathbf{y}\rangle = \langle \mathbf{x}, T(\mathbf{y})\rangle, \forall \mathbf{x}, \mathbf{y} \in V$$

4.5.3. **Definition.** Let V be a subspace of $\mathbb{R}^n$. A linear mapping $T : V \to V$ is said to be symmetric if $\langle T(\mathbf{x}), \mathbf{y}\rangle = \langle \mathbf{x}, T(\mathbf{y})\rangle, \forall \mathbf{x}, \mathbf{y} \in V$

4.5.6. **Theorem: Symmetric Matrices have Real Eigenvalues.** Let $A \in M_{n\times n}(\mathbb{R})$ be a symmetric matrix. Then all the roots to the characteristic polynomial $char_A(\lambda)$ are real.

4.5.7. **Theorem.** Let $A \in M_{n\times n}(\mathbb{R})$ be a symmetric matrix, let $\mathbf{x}_1, \mathbf{x}_2$ be eigenvectors of A with eigenvalues λ_1, λ_2 respectively. We have $\lambda_1 \neq \lambda_2 \implies \mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal vectors in $\mathbb{R}^n$.

4.6 The Spectral Theorem

4.6.1. **Theorem.** Let $T : V \to V$ be a symmetric linear mapping. Then there is an *orthonormal* basis of $\mathbb{R}^n$ consisting of eigenvectors of T. In particular, T is diagonalizable.

4.6.3. **Theorem.** Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a symmetric linear mapping, and let $\lambda_1, .., \lambda_k$ be the distinct eigenvalues of T. let P_i be the orthogonal projection of $\mathbb{R}^n$ onto the eigenspace E_{λ_i}. Then

(a) $T = \lambda_1 P_1 + ... + \lambda_k P_k$, and

(b) $I = P_1, ..., P_k$

Chapter 5

Complex Numbers and Complex V.S.

5.1 Complex Numbers

5.1.1. **Definition.** The set of complex numbers, denoted as $\mathbb{C}$, is the set of ordered pairs of real numbers $(a, b) = (a + bi)$ with operations of addition and multiplication defined by

(a) $\forall (a, b), (c, d) \in \mathbb{C}$, the sum of (a, b) and (c, d) is the complex number defined by

$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

(b) $\forall (a, b), (c, d) \in \mathbb{C}$, the product of (a, b) and (c, d) is the complex number defined by

$$(a + ib)(c + id) = (ac - bd) + i(ad + cb)$$

5.1.2. **Definitions.** Let $z = a + bi \in \mathbb{C}$. The real part of z is $Re(z) = a$, while the imaginary part of z is $Im(z) = b$.

5.1.4. **Definition of Field.** A field is a set $\mathbb{F}$ with two operations $+$ and $\cdot$. In addition, it has to satisfy the following axioms:

(a) **Commutativity of Addition:**

$$\forall x, y \in \mathbb{F}, x + y = y + x$$

(b) **Associativity of Addition:**

$$\forall x, y, z \in \mathbb{F}, (x + y) + z = x + (y + z)$$

(c) **Existence of Additive Identity:**

$$\exists! \, 0 \in \mathbb{F}, \forall x \in \mathbb{F}, 0 + x = x$$

(d) **Existence of Additive Inverse:**

$$\forall x \in \mathbb{F}, \exists! - x \in \mathbb{F}, x + (-x) = 0$$

(e) **Commutativity of Multiplication:**

$$\forall x, y \in \mathbb{F}, xy = yx$$

(f) **Associativity of Multiplication:**

$$\forall x, y, z \in \mathbb{F}, (xy)z = x(yz)$$

(g) **Distributivity:**

$$\forall x, y, z \in \mathbb{F}, (x + y)z = xz + yz \wedge x(y + z) = xy + xz$$

(h) **Existence of Multiplicative Identity:**

$$\exists! \, 1 \in \mathbb{F}, \forall x \in \mathbb{F}, 1 \cdot x = x$$

(i) **Existence of Multiplicative Inverse:**

$$\forall x \in \mathbb{F}, x \neq 0 \implies \exists! \, 1/x \in \mathbb{F}, x \cdot 1/x = 1$$

Note that these are very different axioms from those that we encountered when defining a vector space. For example, multiplication wasn't even defined for vectors! Also, there are nine axioms here rather than eight that we used to have.

5.1.4*. **Examples and Counter Examples of Fields.** First, some examples that are not fields.

(a) First let's examine our good old friend: the space of polynomials $P_n(\mathbb{R})$. Note that if a take a polynomial and try to construct a multiplicative inverse for it, the constructed object will no longer be within the field, (a rational function actually) so the set of all polynomials fails to be a field.

(b) The set of integers, $\mathbb{Z}$, is another counter example. Notice that this set fails for exactly the same reason as the previous one, there is no multiplicative inverse for each element.

Next, some examples for fields

(a) One should recognize rather quickly that the set $\mathbb{R}$ is a field! In fact, field is an abstraction originated form the real numbers.

(b) Similarly, the set of algebraic numbers is a field.

$$\text{Algebraic Numbers} := \{x \in \mathbb{R} |\, p(x) = 0 \ \text{ for some } p \in P(\mathbb{R})\}$$

We like this set, because it is closed.

5.1.5. **Proposition: Complex Numbers form a Field** The set of complex numbers is a field with the operations of addition and scalar multiplication as defined previously.
Proof:
All of the axioms except for possibly the axiom (i) are trivial to show to be true, so will neglect them here. Now, if $a + ib \neq 0$, then necessarily $a \neq 0 \wedge b \neq 0$. We then can deduce that $a^2 + b^2 \neq 0$. Consider any real number $a + ib$, let

$$1/(a + ib) := \frac{a - ib}{a^2 + b^2} \in \mathbb{C}$$

It is easy to show that $(a + ib) \cdot \frac{a-ib}{a^2+b^2} = \frac{a^2+b^2}{a^2+b^2} = 1$, so indeed we have a multiplicative inverse for any complex number. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

5.1.7. **Proposition on Uniqueness.** Note that, as indicated in Definition (5.1.4), all of the additive identity in a field, the additive inverse of an element of a field, the multiplicative identity of a field and the multiplicative inverse of a non-zero element of a field are unique.

5.1.8. **Definition of Absolute Value and Arugument.** The absolute value of the complex number $z = a + ib \in \mathbb{C}$ is the non-negative real number defined as

$$r := |z| = \sqrt{a^2 + b^2}$$

and the argument of a the complex number is the angle between the two components, namely a and b on the complex plain. Then we have the following identity

$$z = |z|(\cos\theta + i\sin\theta)$$

Perodicity of the above identity. Notice that if we replace θ with $\theta \pm 2\pi k, k \in mathbbZ$, we defined exactly the same cpmplex number z.

5.1.9. **Statement of De Moivre's Theorem.**[1]

$$\forall x \in \mathbb{R}, n \in \mathbb{Z}, (\cos(x) + i\sin(x))^n = \cos(nx) + i\sin(nx)$$

We can also reformulate this into the familiar notation that we used above, denoting the absolute value, or length, of the complex number, we have

$$z^n = r^n\left(\cos(n\theta) + i\sin(n\theta)\right)$$

5.1.9*. **Relation w/ Euler's Formula.** First, we recall the Euler Formula as below

$$e^{i\theta} = \cos\theta + i\sin\theta$$

Notice that in a special case of $\theta = \pi$, the above identity is a.k.a Euler's Identity[2]. Considering any $z \in \mathbb{C}$, to derive the above identity, we have the following

$$\begin{aligned} z^n &= |z|^n\left(e^{i\pi}\right)^n \\ &= |z|^n e^{i\pi n} \\ &= r^n\left(\cos n\theta + i\sin n\theta\right) \end{aligned}$$

notice that we can now interchange, as we please, $\cos\theta + i\sin\theta$ with $e^{i\theta}$. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

[1] This should be familiar from FM

[2] $e^{i\pi} + 1 = 0$

5.1.11. **Definition of Algebraic Closure.** A field $\mathbb{F}$ is called algebraically closed if

$$p(z) = a_n z^n + ... + a_1 z + a_0 \in P_n(\mathbb{C}) \quad \text{where } a_i \in \mathbb{F} \text{ and } a_n \neq 0$$

has n roots counted with multiplicity in $\mathbb{F}$.

5.1.12. **Theorem: Fundamental Theorem of Algebra.** The set $\mathbb{C}$ is closed and it is the smallest algebraically closed field containing $\mathbb{R}$

5.1.*. **Roots of Unity.** We shall demonstrate this using the following example: suppose that we want to solve $i^{1/5}$. Then, by breaking this down using the Euler's Formula, we have

$$i^{1/5} = \left(e^{i\frac{\pi}{2}+k2\pi}\right)^{1/5} = e^{i\frac{\pi}{10}+\frac{2\pi}{5}k}, k \in \mathbb{Z}^{\geq 0}$$

By stustituting appropriate values of k, we have

$$\begin{aligned}
&\text{when } k = 0, \text{sol} = \cos\frac{\pi}{10} + i\sin\frac{\pi}{10} \\
&\text{when } k = 1, \text{sol} = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = i \\
&\text{when } k = 2, \text{sol} = \cos\frac{9\pi}{10} + i\sin\frac{9\pi}{10} \\
&\qquad \cdots
\end{aligned}$$

by continuing see that the solutions cycles, since we have only 5 unique solutions. $\mathcal{Q}.\mathcal{E}.\mathcal{D}.\dagger$

General result. Now we consider the problem of solving the equality $z^n = 1$, we call this the n-th root of unity.

$$z = \left(e^{2\pi ik}\right)^{\frac{1}{n}} = e^{\frac{2\pi ik}{n}}, k \in \mathbb{Z}^{\geq 0}$$

5.2 (Field) Vector Spaces

5.2.1. **Definition of a Vector Space Over Field**. In Chapter One, we defined what is called a real vector space, we shall extend that idea into a more general one here. A vector space over a field $\mathbb{F}$ is a set V whose elements we call vectors together with, as usual, closure properties under the two operations

(a) **Closure under vector addition:** an operation called vector addition, which for each pair of vectors $\mathbf{x}, \mathbf{y} \in V$ produces another vector in V denoted $\mathbf{x} + \mathbf{y}$, (i.e. $\forall \mathbf{x}, \mathbf{y} \in V, \mathbf{x} + \mathbf{y} \in V$) and

(b) **Closure under scalar multiplication:** an operation called multiplication by a scalar (a field element), which for each vector $\mathbf{x} \in V$, an each scalar $c \in \mathbb{F}$ produces another vector in V denoted $c\mathbf{x}$. (i.e. $\forall \mathbf{x} \in V, \forall c \in \mathbb{F}, c\mathbf{x} \in V$)

and the eight axioms to satisfy

(a) $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in V, (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$

(b) $\forall \mathbf{v}, \mathbf{y} \in V, \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$

(c) $\exists \mathbf{0} \in V s.t. \forall \mathbf{x} \in V, \mathbf{x} + \mathbf{0} = \mathbf{x}$ (Note that this property is a.k.a existence of additive identity)

(d) $\forall \mathbf{x} \in V, \exists (-\mathbf{x}) \in V \ s.t. \ \mathbf{x} + (-\mathbf{x}) = \mathbf{0}$ (Note that this property is a.k.a existence of additive inverse)

(e) $\forall \mathbf{x}, \mathbf{y} \in V, c \in \mathbb{F}, c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$

(f) $\forall \mathbf{x} \in V, c, d \in \mathbb{F}, (c + d)\mathbf{x} = c\mathbf{x} + d\mathbf{x}$

(g) $\forall \mathbf{x} \in V, c, d \in \mathbb{F}, (cd)\mathbf{x} = c(d\mathbf{x})$

(h) $\forall \mathbf{x} \in V, 1\mathbf{x} = \mathbf{x}$

5.3 Geometry in Complex Vector Spaces

5.3.1. **Definition of Hermitian Inner Product.** Let V be a complex vector space. A Hermitian inner product is a comlex valued function on pairs of vectors in V, denoted by $\langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{C}, \forall \mathbf{v}, \mathbf{u} \in V$, which statisfies the following properties

(a) $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V, a, b \in \mathbb{C}, \langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle = a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle$

(b) $\forall \mathbf{u}, \mathbf{v} \in V, \langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{u}, \mathbf{v} \rangle}$

(c) $\forall \mathbf{v} \in V, \langle \mathbf{v}, \mathbf{v} \rangle \geq 0 \wedge \langle \mathbf{v}, \mathbf{v} \rangle = 0 \implies \mathbf{v} = \mathbf{0}$