Linear Algebra

Alec Zabel-Mena

February 26, 2022

Contents

Chapter 1

Systems of Linear Equations.

1.1 Linear Equations.

Definition. Let F be a field with $x_1, \ldots, x_n \in F$. let $y_1, \ldots, y_m, A_{ij} \in F$ for $1 \leq i \leq m$ and $1 \leq j \leq n$. We call a set of equations a **system of m linear equations in n variables** if it has the form:

$$\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= y_1 \\ &\vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &= y_m \end{aligned} \tag{1.1}$$

We call an n-tuple $(a_1, \ldots, a_n)$ a **solution** to the system if it satisfies the set of equations in (1.1). If $y_1 = \cdots =_n = 0$, then we call the system **homogeneous**.

Example 1.1. Consider the following homogeneous system of 2 linear equations in 3 variables.

$$\begin{aligned} 2x_1 - x_2 + x_3 &= 0 \\ x_1 + 3x_2 + 4x_3 &= 0 \end{aligned}$$

adding -2 times the first equation to the second we get $-7x_2 - 7x_2 = 0$, so $x_2 = -x_3$. Adding 3 times the first equation to the second, we get $7x_1 + 7x_3 = 0$, so $x_1 = -x_2$. SO $(-x_3, -x_3, x_3)$ is a solution to the system. In fact, the system has as solutions all triples of the form $(-a, -, a, a)$.

Definition. Given a system of linear equations of the form (1.1), we call the equation:

$$(c_1A_{11} + \cdots + c_1A_{m1}) + \cdots + (c_1A_{1n} + \cdots + c_1A_{mn}) = c_1y_1 + \cdots + c_my_m \tag{1.2}$$

a **linear combination** of equations of the system.

Definition. If

$$\begin{aligned} B_{11}x_1 + \cdots + B_{1n}x_n &= z_1 \\ &\vdots \\ B_{k1}x_1 + \cdots + B_{kn}x_n &= z_k \end{aligned} \tag{1.3}$$

is a system of k linear equations in n variables, we says the systems described by (1.1) and (1.3) are **equivalent** if each equation in each system is a linear combination of equations in the other system.

Theorem 1.1.1. *Equivalent systems of linear equations have exactly the same solutions.*

Proof. Let

$$\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= y_1 \\ &\vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &= y_m \end{aligned} \tag{1}$$

and

$$\begin{aligned} B_{11}x_1 + \cdots + B_{1n}x_n &= z_1 \\ &\vdots \\ B_{k1}x_1 + \cdots + B_{kn}x_n &= z_k \end{aligned} \tag{2}$$

be equivalent systems of linear equations. let $(a_1, \ldots, a_n)$ be a solution to (1). Then $A_{i1}a_1 + \cdots + A_{in}a_n = y_i = (c_1B_{11} + \cdots + c_kB_{k1})a_1 + \cdots + (c_1B_{1n} + \cdots + c_kB_{kn})a_n = c_1z_1 + \cdots + c_kz_k$. Which means that $(a_1, \ldots, a_n)$ is also a solution to (2). ■

1.2 Matrices and Elementary Row Operations.

Working with systems of linear equations can be difficult or tedious. One thing to note is when taking linear combinations of equations in a given system, what we're really working with are with the coefficients of the system. We can then abbreviate the following system:

$$\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= y_1 \\ &\vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &= y_m \end{aligned} \tag{1}$$

as an equation $AX = Y$, where:

$$A = \begin{pmatrix} A_{11} & \ldots & A_{1n} \\ & \vdots & \\ A_{m1} & \ldots & A_{mn} \end{pmatrix} \tag{2}$$

where $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $Y = \begin{pmatrix} y_1 \\ \vdots \\ x_m \end{pmatrix}$. We call A, X, and Y are called "matrices". We can define a matrix more formally as follows.

Definition. Given a system of m linear equations in n variables of the form (1), we define an $m \times n$ **matrix** over a field F as a map $A : (i, j) \to A_{ij}$, where $A_{ij} \in F$ for $1 \leq i \leq m$ and $1 \leq j \leq n$. We call the image $A(i, i) = A_{ij}$ an **entry** of the matrix and we call i a **row** of the matrix, and j a **column** of the matrix. Here, m is the number of rows of a matrix, and n is the number of columns of a matrix. We denote the space of all $m \times n$ matrices over a given field F as $F^{m\times n}$.

This brings us to a method of formulating analogs to taking linear combinations on systems of linear equations; this will allow us to study systems of linear equations using matrices.

Definition. Let $AX = Y$ be a system of linear equations where A is an $m \times n$ matrix of coefficients, X is an $n \times 1$ matrix of variables, and Y is an $m \times 1$ matrix. We defin an **elementary row operation** on the system to be a map $e : F^{m\times n} \to F^{m\times n}$ by taking $e : A \to e(A)$. We define 3 elementary row operations as:

(1) $e(A)_{ij} = A_{ij}$ if $i \neq r$, and $e(A)_{ij} = cA_{ij}$ for some $c \in F$.

(2) $e(A)_{ij}) = A_{ij}$ if $i \neq r$ and $e(A)_{ij} = A_{rj} + cA_{sj}$, where $r \neq s$.

1. $e(A)_{ij}$ if $i \neq r$, and $e(A)_{ij} = A_{sj}$, where $r \neq s$.

These operations are multiplication by a "scalar" in F, the addition of the $c \in F$ times row s to row r, and a swap of rows r and s.

Theorem 1.2.1. *To each type of elementary row operation of type* (1), (2), *and* (3), *e, there is an elementary row operation e', of the same type as e such that $e'(e(A)) = e(e'(A)) = A$ for any $m \times n$ matrix $A \in F^{m\times n}$.*

Proof. Let e be an elementary row operation of type (1). If $i \neq r$, we are done, define $e' = e$. Otherwise, we have $e(A)_{ij} = cA_{ij}$ for some $c \in F$. Define then, $e'(A)_{ij} = c^{-1}A_{ij}$. Then $e'(e(A_{ij})) = A_{ij}$ and $e(e'(A_{ij})) = A_{ij}$.

Now let e be of type (3). If $i \neq r$, we are done. Otherwise, $e(A)_{rj} = A_{sj}$. Define then $e'(A)_{sj} = A_{rj}$. Then $e'(e(A)_{rj}) = e(e'(A)_{sj}) = A_{rj}$.

Now let e be of type (2); again if $i \neq r$, we are done. We have $e(A)_{ij} = A_{ij} + cA_{rj}$ for some $c \in F$. Define $e'(A)_{ij} = A_ij + (-c)A_{rj}$. Then by similar reasoning above, $e(e'(A)) = e'(e(A)) = A$. ■

Remark. What this theorem says is that for any given elementary row operation e of any given type, its inverse e^{-1} is also an elementary row operation of the same type. This also implies that elementary row operations of types $(1) - (3)$ are $1 - 1$ from $F^{m\times n}$ onto itself; which can be verfied easily.

Definition. If A and B are $m \times n$ matrices over a field F, we call B **row equivalent** to A if there exists a finite sequence of elementary row operations $\{e_i\}_{i=1}^k$ such that $A \to e_1(A) = A_1 \to \cdots \to e_k(A_{k-1}) = A_k = B$.

Lemma 1.2.2. *Row equivalence defines an equivalence relation.*

Proof. Clearly A is row equivalent to itself just take $e_1(A)_{ij} = 1 \cdot A_{ij} = A_{ij}$. Now if a matrix B is row equivalent to A via the sequence $\{e_i\}$ of row operations, then taking the sequence of row operations $\{e_i^{-1}\}$ from B back to A shows that A is row equivalent to A. Now for matrices C, B, and A: if C is row equivalent to B via $\{e_i\}$ and B is row equivalent to A via the sequence $\{e'_j\}_{j=1}^l$, then C can be taken onto A via the sequence of row operations $\{e_h\}_{h=1}^j = \{e_i\} \cup \{e'_j\}$, where $e_h = e_i$ for $h <= i$ and $e_h = e'_j$ for $i < h <= j$. So C is row equivalent to A. ■

Theorem 1.2.3. *If A and B are row equivalent $m \times n$ matrices, the homogeneous systems $AX = 0$ and $BX = 0$ have exactly the same solutions.*

Proof. Take $A = A_0 \to \cdots \to A_k = B$ by a sequence $\{e_i\}_{i=1}^k$ of row operations were $e_i(A) = A_i$. Consider the systems $A_iX = 0$ and $A_{i+1}X = 0$. By theorem 1.1.1, each equation in $A_iX = 0$ is a linear combination of equations in $A_{i+1}X = 0$, hence they have equivalent solutions. Hence so do $AX = 0$ and $BX = 0$. ■

Remark. This means that if two matrices are in the same equivalence class, with respect to row equivalence, then any two systems defined by those matrices, respectively have the same solution set. This makes to computation of solutions for systems of equations much easier.

Example 1.2. (1) Let $F = \mathbb{Q}$. and $A = \begin{pmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & 1 & 5 \end{pmatrix}$. take the following sequence of row operations:

$$\begin{pmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & 1 & 5 \end{pmatrix} \to_2 \begin{pmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & 1 & 5 \end{pmatrix} \to_2 \begin{pmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & 6 & 1 & 7 \end{pmatrix} \to_1$$
$$\begin{pmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & 6 & \frac{1}{2} & -\frac{7}{2} \end{pmatrix} \to_2 \begin{pmatrix} 0 & -9 & 3 & 4 \\ 1 & 0 & -2 & 13 \\ 0 & 6 & \frac{1}{2} & -\frac{7}{2} \end{pmatrix} \to_2 \begin{pmatrix} 0 & 0 & \frac{15}{2} & \frac{55}{2} \\ 1 & 0 & -2 & 13 \\ 0 & 6 & \frac{1}{2} & -\frac{7}{2} \end{pmatrix} \to_1$$
$$\begin{pmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & -2 & 13 \\ 0 & 6 & \frac{1}{2} & -\frac{7}{2} \end{pmatrix} \to_2 \begin{pmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 6 & \frac{1}{2} & -\frac{7}{2} \end{pmatrix} \to_2 \begin{pmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & \frac{5}{3} \end{pmatrix}$$

So the system $AX = 0$ has the same solution as the system $BX = 0$, where $B = \begin{pmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & \frac{5}{3} \end{pmatrix}$, so $AX = 0$ has as solutions all 4-tuples of the form $(-\frac{17}{3}c, \frac{5}{3}, \frac{11}{3}c, c)$.

(2) Let $F = \mathbb{C}$ and consider $A = \begin{pmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{pmatrix}$. Take:

$$A \to_2 \begin{pmatrix} 0 & 2+i \\ 0 & 3+i2 \\ 1 & 2 \end{pmatrix} \to_1 \begin{pmatrix} 0 & 1 \\ 0 & 3+i2 \\ 1 & 2 \end{pmatrix} \to \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}$$

Thus $AX = 0$ is equivalent to the system of equations:

$$x_1 = 0$$
$$x_2 = 0$$

So $AX = 0$ has all triples of the form $(0, c, 0)$ as its solutions.

Definition. We call an mn matrix **row reduced** if the following hold:

(1) The first nonzero entry in each row of R is 1.

(2) Each column of R containing a nonzero leading entry of each row has all its other entries 0.

Example 1.3. Define the matrix $I_{n\times n}$, over an arbitrary field F by

$$T_{ij} = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & ,i \neq j \end{cases} \tag{1.4}$$

I is a row reduced matrix. We call I the $n \times n$ **Identity matrix** over F, and we call δ_{ij} the **Kronecker delta** function.

Theorem 1.2.4. *Every $m \times n$ matrix over a field F is row equivalent to a row reduced matrix.*

Proof. Let A be an mn matrix over F. If every entry in the row $A_{1i} = 0$, for all $1 \leq i \leq n$, we satisfy (1), else if $k \in \mathbb{Z}^+$ is the smallest pssible integer for which $A_{1j} \neq 0$, multiply row 1 by A_{1k}^{-1} and we get (1) for row 1. Now for row $i \geq 2$, add $-A_{ik}$ times row 1 to row i and we get 2 for column k. Continuing along this method, we get a row reduced matrix. ■

Remark. We also have by row equivalence and theorem 1.2.3, we get the result. The proof above, is more illustrative, however.

1.3 Row Reduces Echelon Matrices.

We cann go even farther with row reduces matrices.

Definition. An $m \times n$ matrix R a **row reduced echelon** matrix, or said to be in **row reduced echelon form** if:

(1) R is row reduced.

(2) Every row of R with all entries 0 occur below every row with nonzero entries.

(3) If rows $1, \ldots, r$ are the nonzero rows, and if every leading nonzero entry of row i occurs in column k_i, then $k_1 < \cdots < k_n$.

Lemma 1.3.1. *If R is an $m \times n$ row reduced echelon matrix, then every entry in R is 0, or there is an $r \in \mathbb{Z}^+$ with $1 \leq r \leq m$, and r positive integers $k_1, \ldots, k_r$ with $1 \leq k_i \leq n$, such that:*

(1) $R_{ij} = 0$ for $i > r$ and $R_{ij} = 0$ if $j < k_i$.

(2) $R_{ik_j} = \delta_{ij}$ for $1 \leq i, j \leq r$.

(3) $k_1 < \cdots < k_r$.

Example 1.4. (1) The matrix $0^{m\times n} = (0)_{ij}$ defined by having all its entries 0 is in row reduced echelon form. We call $0^{m\times n}$ the **zero** matrix and denote it 0 when context is clear.

(2) The matrix $\begin{pmatrix} 0 & 1 & -3 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ is a row reduced echelon matrix.

Theorem 1.3.2. *Every $m \times n$ matrix over a field is row equivalent to a row reduced echelon matrix.*

Theorem 1.3.3. *If A is an $m \times n$ matrix, and $m < n$, then the homogeneous system of linear equations $AX = 0$ has a nontrivial solution.*

Proof. By the previous theorem, take $A \to R$ where R is a row reduced echelon matrix. The systems $AX = 0$ and $RX = 0$ are equivalent, by theorem 1.2.3. Now if r is the number of nonzero rows of R, then $r \leq m$, hence $r < n$, so let us choose some x_i not in the r variables $x_{k_1}, \ldots, x_{k_r}$, where $1 \leq k_i \leq n$. We can then construct a solution in which $x_j = 1$, hence $RX = 0$ has a non trivial solution, therefore, so does $AX = 0$. ■

Definition. We define an $m \times n$ matrix over a field F to be **square** if and only if $m = n$.

Theorem 1.3.4. *If A is an $n \times n$ square matrix, then A is row equivalent to the indentity matrix I if and only if the system $AX = 0$ has only the trivial solution.*

Proof. If A is row equivalent to I, then $AX = 0$ and $IX = 0$ have the same solutions, which is $X = 0$. Now suppose that both $AX = 0$ has only the trivial solution. Take $A \to R$ where R is an $n \times n$ row reduced echelon matrix. Let r be the number of nonzero rows of R, certainly $RX = 0$ has no nontrivial solution, so $r \geq n$, and since $r \leq n$, we get $r = n$. By lemma 1.3.1, we have $R = I$. ■

Definition. Let $AX = Y$ be a system of linear equations where A is an $m \times n$ matrix and Y is an $m \times 1$ matrix. We form the **augmented** matrix A' by appending Y to the matrix A after column n. That is A' is the $m \times n+1$ matrix where $A'_{ij} = A_{ij}$ if $j \leq n$, and $A'_{ij} = y_i$ for $j = n + 1$. We denote $A' = (A|Y)$.

Lemma 1.3.5. *Let $AX = Y$ be a system of linear equations. Take $A \to R$ a row reduced echelon matrix, and form the system $RX = Z$. Form the augmented matrices $A' = (A|Y)$ and $R' = (R|Z)$ respectively. Then the systems $A'X = 0$ and $R'X = 0$ are equivalent if and only if $AX = 0$ and $RX = 0$ are equivalent.*

Proof. Taking $A \to R$, we take $A' \to R'$ by the same sequence of elementary row operations. It should then be noted that in the systems $AX = Y$ and $RX = Z$, that $Y \to Z$. ■

Example 1.5. let $F = \mathbb{Q}$ and $A = \begin{pmatrix} 1 & -2 & 1 \\ 2 & 1 & 1 \\ 0 & 5 & -1 \end{pmatrix}$ and consider the system $AX = Y$, form the augmented matrix $A' = (A|Y)$, then through a sequence of elementary row operations, we take:

$$A \to \begin{pmatrix} 1 & 0 & \frac{3}{5} & \frac{1}{5}(y_1 + 2y_2) \\ 0 & 1 & \frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & y_3 - y_2 - 2y_1 \end{pmatrix}$$

Then the system has solutions if, and only if $y_3 - y_2 - 2y_1 = 0$, and such solutions have the form (x_1, x_2, c) where $x_1 = -\frac{3}{5}c + \frac{1}{5}(y_1 + 2y_2)$ and $x_2 = \frac{1}{5}c + \frac{1}{5}(y_2 - 2y_1)$, where $c \in \mathbb{Q}$.

Lemma 1.3.6. *Let F be a field and $F \subseteq K$ a field extension of F. If the system of linear equations $AX = 0$ has solutions over K, then it has solutions over F.*

1.4 Matrix Multiplication.

Definition. Let A and B be $m \times n$ and $n \times p$ matrices over a field F. We define the **product** of A and B to be the binary operations $* : F^{m\times n} \times F^{n\times p} \to F^{m\times p}$ where $* : (A, B) \to AB$ and AB is the $m \times p$ matrix whose $ij^{th} entry is defined to be (AB)_{ij} = \sum_{r=1}^{n} A_{ir}B_{rj} (1.5)$

Example 1.6. 0 Consider the following matrices over $\mathbb{Q}$.

(1) $\begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} 5 & -1 & 2 \\ 15 & 4 & 8 \end{pmatrix} = \begin{pmatrix} 5 & -1 & 2 \\ 0 & 7 & 2 \end{pmatrix}$

(2) $\begin{pmatrix} 1 & 0 \\ -2 & 3 \\ 5 & 4 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 6 & 1 \\ 3 & 8 & -2 \end{pmatrix} = \begin{pmatrix} 0 & 6 & 1 \\ 9 & 12 & 4 \\ 12 & 62 & -3 \\ 3 & 8 & -2 \end{pmatrix}$

(3) $\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 8 \\ 29 \end{pmatrix}$

(4) $\begin{pmatrix} -1 \\ 3 \end{pmatrix} \begin{pmatrix} 2 & 4 \end{pmatrix} = \begin{pmatrix} -2 & -4 \\ 6 & 12 \end{pmatrix}$

(5) $\begin{pmatrix} 2 & 4 \end{pmatrix} \begin{pmatrix} -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 12 \end{pmatrix}$

It should be noted that matrix multiplication is not commutative. It can be shown that the field F^{mp} (taking certain liberties) under the operation $*$ forms a group. it also tunrs out that matrix multiplication gives consistency to the abbreviation of the system of linear equation $AX = Y$.

Example 1.7. Let $A = \begin{pmatrix} A_{11} & \dots & A_{1n} \\ & \vdots & \\ A_{m_1} & \dots & A_{mn} \end{pmatrix}$, $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $Y = \begin{pmatrix} y_1 \\ \vdots \\ Y_m \end{pmatrix}$ We get then:

$$\begin{pmatrix} A_{11} & \dots & A_{1n} \\ & \vdots & \\ A_{m1} & \dots & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} A_{11}x_1 + \cdots + A_{1n}x_n \\ \vdots \\ A_{m_1}x_1 + \cdots + A_{mn}x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

which gives the system

$$A_{11}x_1 + \cdots + A_{1n}x_n = y_1$$
$$\vdots$$
$$A_{m_1}x_1 + \cdots + A_{mn}x_n = y_m$$

Theorem 1.4.1. *Matrix multiplication is associative.*

Proof. Let A be an $m \times n$ matrix, B an $n \times p$ matix, and C a $p \times q$ matrix. Then:

$$\begin{aligned}(A(BC))_{ij} &= \sum_r A_{ir}(B_{rj}C_{rj}) \\ &= \sum_r A_{rj} \sum_s B_{rj}C_{rj} \\ &= \sum_r \sum_s A_{rj}B_{rj}C_{rj} \\ &= \sum_s \sum_r A_{rj}B_{rj}C_{rj} \\ &= \sum A_{ij}B_{rj} \sum C_{rj} \\ &= ((AB)C)_{ij}\end{aligned}$$

■

Definition. Let A be an $n \times n$ square matrix. We define the **square** of A to be the $n \times n$ matrix $A^2 = AA$. Similarly we define the **cube** of A as the $n \times n$ matrix $A^3 = AA^2 = A^2A = AAA$. We can define the n^{th} **power** of A recursively as:

(1) $A^0 = A$.

(2) $A^{n+1} = AA^n = A^nA$.

Definition. We call an $m \times m$ matrix an **elementary** matrix if it can be obtained from the $m \times m$ identity matrix via a single elementary row operation.

Example 1.8. There are only 5 2×2 elementary matrices. They are: $\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$, $\begin{pmatrix}0 & c\\ 1 & 0\end{pmatrix}$, $\begin{pmatrix}0 & 1\\ c & 0\end{pmatrix}$, $\begin{pmatrix}c & 0\\ 0 & 1\end{pmatrix}$, $\begin{pmatrix}1 & 0\\ 0 & c\end{pmatrix}$ where $c \neq 0 \in F$ in the latter two.

Theorem 1.4.2. *Let e be an elementary row operation and let E be the $m \times m$ elementary matrix corresponding to e, that is $E = e(I)$. Then for every $m \times n$ matrix A, $e(A) = EA$.*

Proof. Suppose that e is of type (2), and suppose $r \neq s$. Then

$$E = \begin{cases}\delta_{ik}, \text{ if } i \neq r \\ \delta_{rk} + c\delta_{sk}, \text{ if } i = r\end{cases}$$

where $c \in F$. Then $(EA)_{ij} = \sum E_{ik}A_{kj} = \begin{cases}A_{ik}, \text{ if } i \neq r \\ A_{rk} + cA_{sj} \text{ if } i = r\end{cases}$.

If e is of type (1), then it is necessarily a special case in where $\delta_{rk} = 0$ whenever $i = r$ above. Now if e is of type (3), then we have

$$E = \begin{cases}\delta_{ik}, \text{ if } i \neq r \\ \delta_{sj}, \text{ if } i = r\end{cases}$$

then we get $(EA)_{ij} = \sum E_{ik}A_{kj} = \begin{cases} A_{ik}, \text{ if } i \neq r \\ A_{rk} + cA_{sj} \text{ if } i = r \end{cases}$. ■

Corollary. *Let A and B be $m \times n$ matrices, respectively, over a field F. Then B is row equivalent to A if, and only if $B = PA$, where $P = E_s \dots E_2E_1$ is an $m \times m$ matrix and, where E_i is an elementary matrix for $1 \leq i \leq s$.*

Proof. Let E_i be an elementary matrix of any type for $1 \leq i \leq s$. Suppose that $B = PA$. Then $B = (E_s s \dots E_2E_1)A$. By definition, E_1A implies that $A \to_{e_1} E_1A$, so we get $A \to_{e_1} \cdots \to_{e_s \dots e_1e_2} PA = B$ making B row equivalent to A.

Conversely, if B is row equivalent to A, then $A \to_{e_1} \cdots \to_{e_s \dots e_1e_2} PA = B$ via a sequence $\{e_i\}_{i=1}^{s}$ of elementary row operations, and where P is some $m \times m$ matrix. Take $E_i = e_i(i)$, and we see that $P = E_s \dots E_2E_1$, hence $B = (E_s \dots E_2E_1)A$. ■

1.5 Invertible Matrices.

Definition. Let A be an $n \times n$ square matrix over a field F. We call an $n \times n$ matrix B a **left inverse** of A if $BA = I$, we call B a **right inverse** if $AB = I$. We call B the **inverse** of A if it is both a left and right inverse, and we say A is invertible. We denote the inverse of A as A^{-1}.

We can now state a fundamental property of all $n \times n$ matrices.

Lemma 1.5.1. *If A has a left inverse B, and a right inverse C, then $B = C$ and A is invertible.*

Proof. We have $BA = I$ and $AC = I$. Then $B = BI = B(AC) = (BA)C = IC = C$. ■

Theorem 1.5.2. *Let $*$ be the matrix product. Then $F^{n \times n}$ forms a group over $*$.*

Corollary. *Let A and B be $n \times n$ matrices. Then:*

(1) If A is invertible, then so is A^{-1}, and $(A^{-1})^{-1} = A$.

(2) If A and B are invertible, then so is AB, and $(AB)^{-1} = B^{-1}A^{-1}$.

Proof. This follows from group theory. ■

Corollary. *Products of invertible matrices are invertible.*

Theorem 1.5.3. *Elementary matrices are invertible.*

Proof. Let e be an elementary row operation. Let $E = e(I)$ and $E' = e^{-1}(I)$. Then $EE' = e(e^{-1}(I)) = I$ and $E'E = e^{-1}(e(I)) = I$; so E is invertible and $E' = E^{-1}$. ■

Example 1.9. (1) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

(2) $\begin{pmatrix} 0 & c \\ 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & -c \\ 1 & 0 \end{pmatrix}$

(3) $\begin{pmatrix} 0 & 1 \\ c & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & -c \\ -c & 0 \end{pmatrix}$

(4) For $c \neq 0$, $\begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} c^{-1} & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & c^{-1} \end{pmatrix}$

Theorem 1.5.4. *If A is an $n \times n$ matrix, the following are equivalent:*

(1) A is invertible.

(2) A is row equivalent to I.

(3) A is a product of elementary row matrices.

Proof. Suppose that A is invertible. Let R be a row reduced echelon matrix, row equivalent to A. Then $R = E_k \dots E_2 E_1 A$. By the corolloray to theorem 1.5.3, we have E_i is invertible for each $1 \leq i \leq k$. Then $A = E_1^{-1} E_2^{-1} \dots E_k^{-1} R$. So A is invertible if and only if R is invertible. Since R is square, it is invertible if and only if $R = I$. Thus by hypothesis, $A = E_1^{-1} E_2^{-1} \dots E_k^{-1}$. This also implies that A is row equivalent to I.

Now if A is row equivalent to I, then $I = PA$ where $P = E_k \dots E_1$. We have then that $A = AI = A(PA) = (AP)A$, which makes $AP = I$, hence A is invertible. ■

Corollary. *If A is invertible, the sequence of elementary matrices $\{E_i\}_{i=1}^k$ taking $A \to I$ has product A^{-1}; that is $A^{-1} = E_k \dots E_1$.*

Corollary. *If A and B are $m \times n$ matrices, B is row equivalent to A if, and only if $B = PA$, where P is an $m \times m$ invertible matrix.*

Theorem 1.5.5. *If A is an $n \times n$ matrix, the following are equivalent:*

(1) A is invertible.

(2) The system $AX = 0$ has only the trivial solution.

(3) The system $AX = Y$ has a solution X for each $m \times 1$ matrix Y.

Proof. That (1) implies (2) implies (3) is clear and follows from theorems 1.3.4 and 1.5.4. Now suppose that $AX = Y$ has a solution for each Y:

$$E = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

If $RX = E$ can be solved for X, the least row is not 0. Now $R = PA$ with P invertible, thus $RX = E$ if and only if $AX = P^{-1}E$, which has a solution. ■

Corollary. *A square matrix with either left or right (or both) inverses is invertible.*

Proof. If A has both left and right inverses, we are done. Suppose then that A is square with left inverse B, then $BA = I$. Now $AX = 0$ has only the trivial soltion, since $X = IX = (BA)X = B(AX)$, making A invertible. The same holds for right inverses by similar reasoning. ■

Corollary. *Let $A = A_1A_2 \dots A_k$, where A_i is an nn matrix for each $1 \leq i \leq k$. Then A is invertible if and only if each A_i is invertible.*

Proof. If A_i is invertible for each $1 \leq i \leq k$, then so are the products, so A is also invertible.

Now suppose that A is invertible. suppose that X is an $n \times 1$ matrix and that $A_kX = 0$. Then $AX = (A_1A_2 \dots A_{k-1})A_kX = 0$, implying that $A_kX = 0$ has only the trivial solution. So A_k is invertible, hence $A_1 \dots A_{k-1} = AA_k^{-1}$ is invertible. Extending this reasoning to A_i for $1 \leq i \leq k-1$, we get A_i invertible. ■

Example 1.10. Let $F = \mathbb{Q}$ and consider $A = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix}$. Consider $AX = Y$, then:

$$\begin{pmatrix} 2 & -1 & y_1 \\ 1 & 3 & y_2 \end{pmatrix} \to_3 \begin{pmatrix} 1 & 3 & y_2 \\ 2 & -1 & y_1 \end{pmatrix} \to_2 \begin{pmatrix} 1 & 3 & y_2 \\ 0 & -7 & y_1 - 2y_2 \end{pmatrix} \to_1$$
$$\begin{pmatrix} 1 & 3 & y_2 \\ 0 & 1 & \frac{1}{7}(y_1 - 2y_2) \end{pmatrix} \to_2 \begin{pmatrix} 1 & 0 & \frac{1}{7}(y_2 + 3y_1) \\ 0 & 1 & \frac{1}{7}(y_1 - 2y_2) \end{pmatrix}$$

We get that A is invertible and that $A = \begin{pmatrix} \frac{3}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{2}{7} \end{pmatrix}$.

This example showds us that we can find the inverse of any invertible matrix A by forming the system $AX = Y$, and taking $(A|Y) \to (I|Y')$, where $Y' = A^{-1}Y$. Even better, we can compute A^{-1} by taking $(A|I) \to (I|A^{-1})$.

Chapter 2

Vector Spaces.

2.1 Definitions and Examples

Definition. We call a nonemty set V a **vector space** over a field F, if given a binary operation $+ : V \times V \to V$ called **vector addition** and an operation $\cdot : F \times V \to V$ called **scalar multiplication**, we have that $(V, +)$ forms an abelian group, and for all $v, w \in V$ and $\alpha, \beta \in F$:

(1) $\alpha(v + w) = \alpha v + \alpha w$.

(2) $(\alpha + \beta)v = \alpha v + \beta v$.

(3) $\alpha(\beta v) = (\alpha\beta)v$.

(4) $1v = v$, where 1 as the identity element of F under its multiplication.

Lemma 2.1.1. *Let V be a vector space over a field F. Then the operation $\cdot : F \times V \to V$ of scalar mutliplication as a group homomorphism of V into V.*

Proof. Taking $\cdot : F \times V \to V$ by $(\alpha, v) \to \alpha v$, restrict $\cdot$ to V, i.e. consider $\cdot|_V : V \to V$ by $v \to \alpha v$ for $\alpha \in F$. By (1) of the scalar multiplication rules, we get that $\cdot|_V$ as a homomorphism; which makes $\cdot$ a homomorphism. ■

Example 2.1. (1) Let F be a field and $F \subseteq K$ a field extension of F. Then K as a vector space over F with $+$ the usual addition of K and $\cdot$ the multiplication of K restricted to F by the first part, i.e. the product $\cdot : v \to \alpha v$ with $\alpha \in F$.

(2) Let F be a field and consider F^n the set of ordered n-tuples of elements of F, for some $n \in \mathbb{Z}^+$. Take $+ : (v, w) \to v+w$ by $(v_1, \ldots, v_n)+(w_1, \ldots, w_n) = (v_1+w_1, \ldots, v_n+w_n)$, where $v = (v_1, \ldots, v_n), w = (w_1, \ldots, w_n) \in F^n$, and $\cdot : (\alpha, v) \to \alpha v$ by $\alpha(v_1, \ldots, n_n) = (\alpha v_1, \ldots, \alpha v_n)$. Then F^n as a vector space over F.

(3) Let F be any field and let $F[x]$ be the polynomial field over F. Take $+$ to be polynomial addition, and $\cdot$ the multiplication of a constant in F by a polynomial in $F[x]$. Then $F[x]$ as a vector space over F.

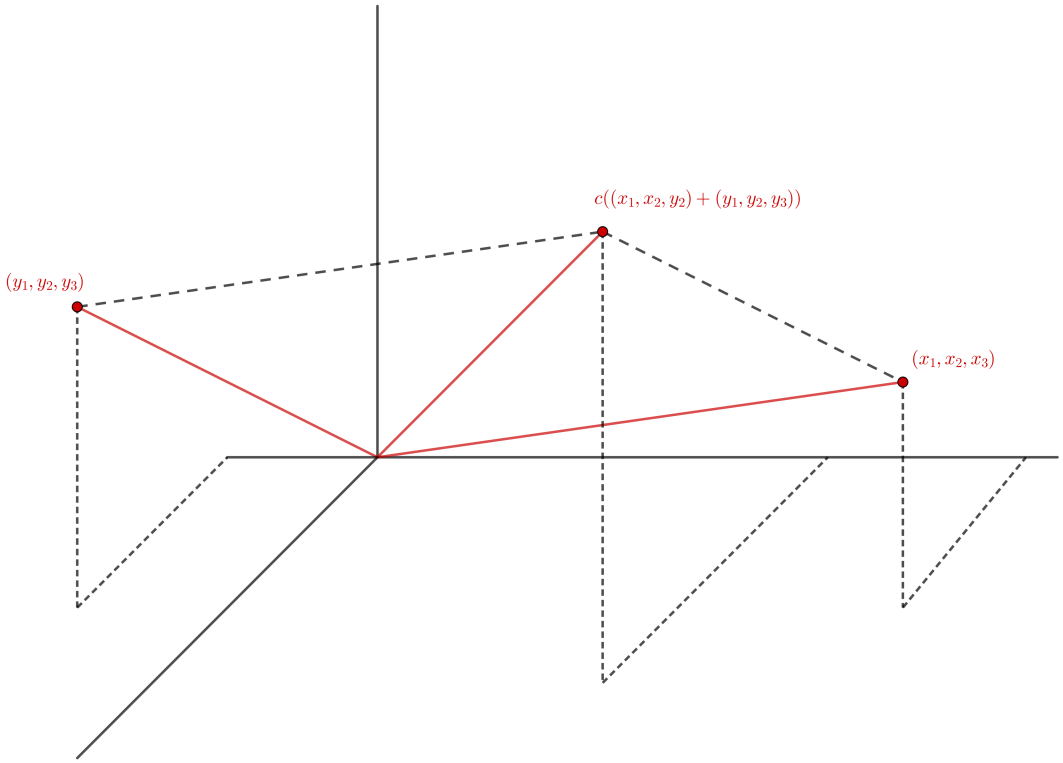

Figure 2.1: The vector sum and scalar product of vectors in $\mathbb{R}^3$.

(4) The space of $m \times n$ matrices over a field F, $F^{m\times n}$ forms a vector space under matrix addition and scalar multiplications. Naturally, the vectors in this space are $m \times n$ matrices and the scalars are elements of F.

(5) The space of all functions over a field F is a vector space. Observe that for functions $f(x), g(x)$, with $x \in F$, and $c \in F$, that $f + g(x)$ is a function and so is $cf(x)$.

(6) Let $F[x]$ be the polynomial field over a field F and consider the set $P_n = \{f \in Fx : \deg f < n\}$. Then P_n as a subset of $F[x]$ forms a vector space over F under the same operations $+$ and $\cdot$ (this last example motivates the following definition).

Definition. Let V be a vector space over a field F. We say a subset $W \subseteq V$ as a **subspace** of V if W as also a vector space over F.

Example 2.2. (1) For any vector space V, the subset (0) is a proper subspace of V.

(2) The set of n-tuples of the form $(0, x_2, \ldots, x_n)$ is a subspace of F^n, but the set of n-tuples of the form $(1 + x_2, x_2, \ldots, x_n)$ for $n \geq 2$ is not a subspace.

(3) $F[x]$ is a subspace of the space of space of all functions as a vector space over F. Moreover, the polynomial space P_n is a subspace of $F[x]$.

(4) We say an $n \times n$ matrix A is **symmetric** over the field F if $A_{ij} = A_{ji}$ for each i, j. The space of symmetric matrices forms a subspace of $F^{n\times n}$.

(5) We call an $n \times n$ matrix A **Hermitian** (or **self adjoint**) over $\mathbb{C}$ if $A_{ij} = \overline{A_{ji}}$ for each i, j. As an example, a 2×2 matrix is Hermitian if, and only if it has the form

$$\begin{pmatrix} z & x+iy \\ x-iy & w \end{pmatrix}$$

The set of all Hermitian matrices is not a subspace of $\mathbb{C}^{n\times n}$, for if A is Hermitian, its diagonal entries $A_{11}, A_{22}, \dots$ are all real numbers, but the diagonal of the matrix iA is not necessarily real. However, the set of all $n \times n$ complex Hermitian matrices is a vector space over $\mathbb{R}$.

(6) We define the **solution space** of a system of homogeneous linear equations to be the set of all $n \times 1$ matrices X such that $AX = 0$ for some $m \times n$ matrix A over F. This space is a subspace of of $F^{m\times 1}$, which can be considered the space of all values of the system AX.

Lemma 2.1.2. *Let V be a vector space over a field F, and let $W \subseteq V$ be a subspace of V. Then for all $w_1, w_2 \in W$ and $\alpha, \beta \in F$, $\alpha w_1 + \beta w_2 \in W$.*

Proof. Since W as a vector space we have that $\alpha w_1, \beta w_2 \in W$; then by closure of vector addition, $\alpha w_1 + \beta w_2 \in W$. ■

Definition. Let U and V be vector spaces over a filed F. We call a mapping $T : U \to V$ a **homomorphism** of U into V if:

(1) $T(u_1 + u_2) = T(u_1) + T(u_2)$.

(2) $T(\alpha u_1) = \alpha T(u_1)$.

for all $u_1, u_2 \in U$ and $\alpha \in F$. If T as $1-1$ from U onto V, then we call T an **isomorphism** and we say U as **ismorphic** to V and write $U \simeq V$. We define the **kernel** of T to be $\ker T = \{u \in U : T(u) = 0\}$. We call the set of all homomorphism of U into V $\hom(U, V)$. In linear algebra, we also call a homomorphism a **linear transformatio**; in which case we can refer to $\hom(U, V)$ as the set of all linear transformations from U to V.

Example 2.3. Let F be a field and consider the vector spaces F^n and P_n defined in examples (2) and (4). Then $P_n \simeq F^n$. Take the map $a_0 + a_1x + \cdots + a_nx^{n-1} \to (a_0, \dots, a_{n-1})$, which defines an isomorphism.

Lemma 2.1.3. *2.1.3 If V as a vector space over a field F, then for all $\alpha \in F$ and $v \in V$:*

(1) $\alpha 0 = 0$.

(2) $0v = 0$.

(3) $(-\alpha)v = -(\alpha v)$.

(4) $\alpha v = 0$ *and* $v \neq 0$ *implies* $\alpha = 0$.

Proof. (1) $\alpha 0 = \alpha(0+0) = \alpha 0 + \alpha 0$, hence $\alpha 0 = 0$.

(2) $0v = (0+0)v = 0v + 0v$, hence $0v = 0$.

(3) He have $0 = 0v$, that as $0 = (\alpha + (-\alpha))v = \alpha v + (-\alpha)v$. Adding both sided by $-(\alpha v)$ we get the desired result.

(4) If $\alpha \neq 0$ and $v \neq 0$, then $0 = \alpha^{-1}0 = \alpha^{-1}(\alpha v) = 1v = v$ which makes $v = 0$, which cannot happen. So $\alpha = 0$.

■

Lemma 2.1.4. *2.1.4 Let V be a vector space over a field F and let $W \subseteq V$ be a subspace of V. Then V/W as a vector space over F where for $v_1 + W, v_2 + W \in V/W$ and $\alpha \in F$ we have:*

[label=(0)]

1. $(v_1 + W) + (v_2 + W) = (v_1 + v_2 + W)$.
2. $(v_1 + W) = \alpha v_1 + W$.

Proof. Since V as an abelian group, and W a subgroup of V under $+$, we get that V/W as the quotient group of V over W; which as abelian since W as abelian.

Suppose now that for $v, v' \in V$ that $v+W = v'+W$, then for $\alpha \in F$ we have $\alpha(v+W) = \alpha(v'+W)$, and by hypotheses, we have $v-v' \in W$. Now since W as a subspace, $\alpha(v-v') \in W$ as well, so $\alpha v + W = \alpha v' + W$, so the product as well defined.

Now consider $v, v' \in W$ and $\alpha, \beta \in F$. By our product we have that $\alpha(v + w + W) = \alpha(v+w)+W = (\alpha v + \alpha w)+W = (\alpha v + W)+(\alpha v' + W)$, $(\alpha+\beta)(v+W) = (\alpha+\beta)v+W = (\alpha v + \beta v) + W = \alpha(v+W) + \beta(v+W)$, $\alpha(\beta v + W) = \alpha\beta v + W = (\alpha\beta)v + W$, and finally, $1(v+w) = 1v + W = v + W$. Therefore V/W as a vector space over F. ■

Definition. Let V be a vector space over F and let $W \subseteq V$ be a subspace of V. We call the vector space formed by taking the quotient group of V over W, V/W the **quotient space** of V over W.

Theorem 2.1.5 (The First Homomorphism Theorem for Vector Spaces)**.** *2.1.5 If $T : U \to V$ as a homomorphism of U onto V, and $W = \ker T$, then $V \simeq U/W$. If U as a vector space and $W \subseteq U$ as a subspace of U, then there as a homomorphism of U onto U/W.*

Proof. By the fundamental theorem of homomorphisms, we have that, as groups, $V \simeq U/W$. That there as a homomorphism from U onto U/W follows immediately. ■

Definition. Let V bhe a vector space over a field F and let $\{U_i\}_{i=1}^n$ be a collection of subspaces of V. We call V the **internal direct sum** of $\{U_i\}$ if every element of V can be written uniquely as a vector sum of elements of each U_i for $1 \leq i \leq n$; That as for $v \in V$, $v = u_1 + \cdots + u_n$ as unique where $u_i \in U_i$.

Lemma 2.1.6. *2.1.6 Let $\{V_i\}_{i=1}^n$ be a collection of vector spaces over a field F and let $V = \prod_{i=1}^n V_i$ and define $+ : V \times V \to V$ by $(v_1, \ldots, v_n) + (v_1', \ldots, v_n') = (v_1 + v_1', \ldots, v_n + v_n')$ and define $\cdot : F \times V \to V$ by $\alpha(v_1, \ldots, v_n) = (\alpha v_1, \ldots, \alpha v_n)$. Then V as a vector space over F.*

Proof. Since V_i as a vector space for all $1 \leq i \leq n$, they are all abelian groups, hence V as closed under $+$, and inherits associativity, as well a s commutativity. Now letting $0 = (0_1, \ldots, 0_n)$, where 0_i as the identity of V_i, we get for any $v \in V$ that $v + 0 = o + v = v$, so 0 as the identity. Likewise for any $v \in V$, $-v = (-v_1, \ldots, -v_n)$ serves as the inverse for v. So $(V, +)$ forms an abelian group.

Now by the axioms of scalar multiplication on each of the V_i, let $v = (v_1, \ldots, v_n), w = (w_1, \ldots, w_n) \in V$ and $\alpha, \beta \in F$. We get $\alpha(v + w) = \alpha(v_1 + w_1, \ldots v_n + w_n) = (\alpha(v_1 + w_1), \ldots, \alpha(v_n + w_n)) = (\alpha v_1 + \alpha w_1, \ldots, \alpha v_n + \alpha w_n) = (\alpha v_1, \ldots, \alpha v_n) + (\alpha w_1, \ldots, \alpha w_n) = \alpha v + \alpha w$. We also get $(\alpha + \beta)v = ((\alpha + \beta)v_1, \ldots, (\alpha + \beta)v_n) = (\alpha v_1 + \beta v_1, \ldots, \alpha v_n + \beta v_n) = (\alpha v_1, \ldots, \alpha v_n) + (\beta v_1, \ldots, \beta v_n) = \alpha v + \beta v$. Through similar calculation, we get that $\alpha(\beta v) = (\alpha\beta)v$ and $1v = v$; which makes V into a vector space. ■

Definition. Let $\{V_i\}_{i=1}^n$ be a collection of vector spaces over a field F and let $V = \prod_{i=1}^n V_i$ and define $+ : V \times V \to V$ by $(v_1, \ldots, v_n) + (v_1', \ldots, v_n') = (v_1 + v_1', \ldots, v_n + v_n')$ and define $\cdot : F \times V \to V$ by $\alpha(v_1, \ldots, v_n) = (\alpha v_1, \ldots, \alpha v_n)$. We call V, as a vector space over F the **external direct sum** of $\{V_i\}$ and write $V = V_1 \oplus \cdots \oplus V_n$, or $V = \bigoplus_{i=1}^n V_i$.

Theorem 2.1.7. *2.1.7 Let V be a vector space and let $\{U_i\}_{i=1}^n$ be a collection of subspaces of V. If V is the internal direct sum of $\{U_i\}$ then V is isomorphic to the external direct sum of $\{U_i\}$; that is: $V \simeq \bigoplus_{i=1}^n U_i$.*

Proof. Let $v \in V$. By hypothesis $v = u_1 + \cdots + u_n$ with $u_i \in U_i$ for $1 \leq i \leq n$, and it is a unique representation of v. Define then, the map $T : V \to \bigoplus_{i=1}^n U_i$ by the map $v = v = u_1 + \cdots + u_n \to (u_1, \ldots, u_n)$. Since v has a unique representation by definition, T is well defined; moreover it is $1 - 1$, as $(u_1, \ldots, u_n) = (w_1, \ldots w_n)$ implies $u_i = w_i$ for all $1 \leq i \leq n$, hence $u_1 + \cdots + u_u = w_1 + \cdots + w_n$, and since this sum is unique, they both represent a vector $v \in V$. That T is onto follows directly from definition.

Finally, let $v, w \in V$, then $v = u_1 + \cdots + u_n$ and $w = w_1 + \cdots + w_n$. Hence $T(v + w) = T(u_1 + w_1 + \cdots + u_n + w_n) = (u_1 + w_1, \ldots, u_n + w_n) = (u_1, \ldots, u_n) + (w_1, \ldots, w_n) = T(v) + T(w)$. Similarly, $T(\alpha v) = (v)$. ■

Remark. That V is the internal direct sum of $\{U_i\}$ and that $V \simeq U_1 \oplus \cdots \oplus U_n$ by the above theorem permits us to write $V = U_1 \oplus \cdots \oplus U_n$, or $V = \bigoplus_{i=1}^n U_i$.

2.2 Linear Independence and Bases.

Definition. If V is a vector space over a field F and ive $v_1, \ldots, v_n \in V$, then we call any element $v \in V$ of the form $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ for $\alpha_1, \ldots, \alpha_n \in F$ a **linear comination** of $v_1, \ldots, v_n$ over F.

Definition. Let V be a vector space. We call the set of all linear combinations of finite sets of elements of a nonempty subset $S \subseteq V$ the **linear span** of S; and we write span S.

Lemma 2.2.1. *2.2.1 If V is a vector space, and $S \subseteq V$ is nonempty, then* $\operatorname{span} S$ *is a subspace of V.*

Proof. Since $\operatorname{span} S$ is the set of all linear combinations of finite sets of elements of S, it is clear that $\operatorname{span} S \subseteq V$. Now let $v, w \in \operatorname{span} S$, then $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$ and $w = \mu_1 w_1 + \cdots + \mu_m w_m$; where $\lambda_i, \mu_j \in F$ and $v_i, w_j \in S$ for $1 \leq i \leq n$ and $1 \leq j \leq m$. Now consider $\alpha, \beta \in F$, then $\alpha v + \beta w = \alpha(\lambda_1 v_1 + \cdots + \lambda_n v_n) + \beta(\mu_1 w_1 + \cdots + \mu_m w_m) = (\alpha\lambda_1)v_1 + \cdots + (\alpha\lambda_n)v_n + (\beta\mu_1)w_1 + \cdots + (\beta\mu_m)w_m$ which is a linear combination of the finite set $\{v_1, \ldots, v_n, w_1, \ldots, w_m\}$ of elements of S. Therefore $\alpha v + \beta w \in \operatorname{span} S$. ■

Lemma 2.2.2. *2.2.2 If $S, T \subseteq V$, then:*

(1) $S \subseteq T$ implies $\operatorname{span} S \subseteq \operatorname{span} T$.

(2) $\operatorname{span}(S \cup T) = \operatorname{span} S + \operatorname{span} T$.

(3) $\operatorname{span}(\operatorname{span} S) = \operatorname{span} S$.

Proof. (1) Let $v \in \operatorname{span} S$, then $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$, with $v_1, \ldots, v_n \in S$. Since $S \subseteq T$, $v_1, \ldots, v_n \in T$, hence $v \in \operatorname{span} T$.

(2) Let $v \in \operatorname{span}(S \cup T)$, then $v = \lambda_1 v_1 + \cdots + \lambda_n v_n + \mu_1 w_1 + \cdots + \mu_m w_m = (\lambda_1 v_1 + \cdots + \lambda_n v_n) + (\mu_1 w_1 + \cdots + \mu_m w_m)$, where $v_i \in S$ and $w_j \in T$. Then $v \in \operatorname{span} S + \operatorname{span} T$.

Now for $v \in \operatorname{span} S + \operatorname{span} T$, $v = u + w$ with $u \in \operatorname{span} S$ and $w \in \operatorname{span} T$, hence v is a linear combination of the finite set $\{u_1, \ldots, u_n, w_1, \ldots, w_n\}$ of elements of $S \cup T$, hence $v \in \operatorname{span}(S \cup T)$.

(3) Clearly $\operatorname{span} S \in \operatorname{span}(\operatorname{span} S)$. Suppose then that $v \in \operatorname{span}(\operatorname{span} S)$. Then $v =_1 v_1 + \cdots + \alpha_n v_n$ where $v_i = \beta_{i1} v_{i1} + \cdots + \beta_{im} v_{im}$ where $v_{ij} \in S$. Hence $v = ((\alpha_1\beta_{11})v_{11} + \cdots + (\alpha_1\beta_{1m})v_{1m}) + \cdots + (\alpha_n\beta_{n1})v_{n1} + \cdots + (\alpha_n\beta_{nm})v_{nm}$. Therefore $\operatorname{span}(\operatorname{span} S) \subseteq \operatorname{span} S$. ■

Definition. We call a vector space V over a field F **finite dimensional** over F of there is a finite subset $S \subseteq V$ whose linear span is V; that is $\operatorname{span} S = V$.

Example 2.4. F^n is finite dimensional. Let $S = \{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)\}$. Then $\operatorname{span} S = F^n$.

Definition. Let V be a vector space over a field F. We say that a set of $\{v_1, \ldots, v_n\}$ of elements of V **linearly dependent** over F if there exist $\lambda_1, \ldots, \lambda_n \in F$, not all 0 such that $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$. We call $\{v_1, \ldots v_n\}$ **linearly independent** over F if it is not linearly dependent over F; that is $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ implies $\lambda_1 = \cdots = \lambda_{n=0}$.

Example 2.5. (1) In F^3, the vectors $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ are linearly independent, where as $(1, 1, 0)$, $3, 1, 3$, $(5, 3, 3)$ are linearly dependent.

(2) Consider the set $\mathbb{C}$ of complex numbets as a vector space over $\mathbb{R}$. The vectors $1, i$ are linearly independent over $\mathbb{R}$ since $i \notin \mathbb{R}$. However, $1, i$ is not linearly independent over $\mathbb{C}$, as $i^2 + 1 = 0$ by definition; where $\lambda_1 = i$ and $\lambda_2 = 1$.

Lemma 2.2.3. *If $v_1, \dots, v_n \in V$ are linearly independent, then every element in* span $\{v_1, \dots, v_n\}$ *can be represented unqiuely as a linear combination of $v_1, \dots, v_n$.*

Proof. Let $v \in \text{span}\{v_1, \dots, v_n\}$ such that $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$ and $v = \mu_1 v_1 + \cdots + \mu_n v_n$. Then $\lambda_1 v_1 + \cdots + \lambda_n v_n = \mu_1 v_1 + \cdots + \mu_n v_n$, then $(\lambda_1 - \mu_1)v_1 + \cdots + (\lambda_n - \mu_n)v_n$. By linear independence, this implies that $\lambda_i - \mu_i = 0$, for all $1 \leq i \leq n$. Therefore v is uniquely represented. ■

Theorem 2.2.4. *2.2.4 If $v_1, \dots, v_n \in V$, then they are linearly independent, or v_k is a linear combination of $v_1, \dots, v_{k-1}$ for $1 \leq k \leq n$.*

Proof. If $v_1, \dots, v_n$ are linearly independent, then we are done. Now suppose that they are linearly dependent, then $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ for $\lambda_1, \dots, \lambda_n$ not all 0. Let k be the largest such integer for which $\lambda_k \neq 0$, and $\lambda_i = 0$ for all $k < i$. Then $\lambda_1 v_1 + \cdots + \lambda_n v_n = \lambda_1 v_1 + \cdots + \lambda_k v_k$ where $\lambda_1, \dots, \lambda_k$ are not all 0 for $1 \leq i \leq k$. Then we have that $v_k = (\lambda_k^{-1}\lambda_1)v_1 + \cdots + (\lambda_k^{-1}\lambda_{k-1})v_{k-1}$ which is a linear combination of $v_1, \dots, v_{k-1}$. ■

Corollary. *If $v_1, \dots, v_n \in V$ have W as a linear span, and if $v_1, \dots, v_k$ are linearly independent, then there is a linearly independent subset of $\{v_1, \dots, v_n\}$ of the form $\{v_1, \dots v_k, v_{i_1}, \dots, v_{i_r}\}$ which span W.*

Proof. If $v_1, \dots, v_n$ are linearly independent, then we are done. If not, let j be the smallest such integer for which v_j is a linear combination of its predecessors. Since $v_1, \dots, v_k$ are linearly independent, we get $k < j$. then consider the set $S = \{v_1, \dots, v_n\} \backslash v_j = \{v_1, \dots, v_k, \dots, v_{j-1}, v_{j+1}, \dots, v_n\}$ which has $n-1$ elements. Clearly, span $S \subseteq W$.

Now let $w \in W$, then $w = \lambda_1 v_1 + \cdots + \lambda_n v_n$. Since v_j is a linear combination of $v_1, \dots, v_{j-1}$, we get that $w = \lambda_1' v_1 + \cdots + \lambda_k' v_k + \cdots + \lambda_{j-1}' v_{j-1} + \lambda_{j+1} v_{j+1} + \cdots + \lambda_n v_n$ which makes $W \subseteq \text{span}\, S$.

Now if we proceed by removing all voctors which are linear combinations of their predecessors, we get a set $\{v_1, \dots, v_k, v_{i_1}, \dots, v_{i_r}\}$ with span S; by the preceding argument, we get again that $W \subseteq \text{span}\, S$. ■

Corollary. *If V is a finite dimensional vector space, then there is a finite set of linearly independent vectors $\{v_1, \dots, v_n\}$ such that* span $\{v_1, \dots, v_n\} = V$.

Proof. By definition, since V is finite dimensional, there is a finite set of vectors $\{u_1, \dots, u_m\}$ with linear span V. Then by the previous corollary, there is a subset $\{v_1, \dots, v_n\}$ of linearly independent vectors whose span is also V. ■

Definition. We call a subset S of a vector space V a **basis** if S consists of linearly independent vectors, and span $S = V$. We call the basis $\{e_1, \dots, e_n\}$ where $e_i = (0_1, \dots, 1_i, \dots, 0_n)$ the **standard basis**.

What the above corollary states, is that if V is a finite dimensional vector space, and $u_1, \dots, u_m$ (not necessarily independent), span V, then $u_1, \dots, u_m$ contain a basis of V.

Example 2.6. (1) A basis need not be finite. Consider the polynomial field $F[x]$ the set $\{1, x, x^2, \dots, x_n, \dots\}$ forms a basis of $F[x]$. However, the set $\{1, x, x^2, \dots, x^n\}$ span the subspace P_n of $F[x]$.

(2) Let P be an invertible $n \times n$ matrix with entries in F. then the columns $P_1, \dots P_n$ of P form a basis for the space of column matrices $F^{n\times 1}$. The proof is left as an exercise.

Lemma 2.2.5. *2.2.5 If V is a finite dimensional vector space, then $V \simeq F^n$ for some $n \in \mathbb{Z}^+$.*

Proof. By lemma 1.2.3 and the above corollary, any $v \in V$ is the unique combination of basis elements $v_1, \dots, v_n$; that is $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$. Now take the map $v \to (\lambda_1, \dots, \lambda_n)$ is well defined, $1-1$ by linear independence and onto. Hence $V \simeq F^n$. ■

Remark. In fact if $\{v_1, \dots, v_n\}$ is a basis for V, then $|\{v_1, \dots, v_n\}| = n$.

Lemma 2.2.6. *2.2.6 If $v_1, \dots, v_n \in V$ forms a basis, and $w_1, \dots, w_m \in V$ are linearly independent, then $m \leq n$. Moreover, the set $\{v_1, \dots, v_n\}$ is maximally linearly independent.*

Proof. For any arbitrary vector $v \in V$, v is a linear combination of $v_1, \dots, v_n$ by lemma 1.2.3, hence $\{v_1, \dots, v, v\}$ is linearly dependent. This makes $\{v_1, \dots, v_n\}$ maximally independent.

Now $w_m \in V$ is a linear combination of $v_1, \dots, v_n$; moreover they span V by theorem 1.2.4, therefore, by the previous corollary there is a subset $\{w_m, v_{i_1}, \dots, v_{i_k}\}$ with $k \leq n-1$ which is a basis of V.

Repeating by taking $w_{m-1}, w_m, \dots, v_{i_k}$; we get, eventually, a basis $\{w_{m-1}, w_m, \dots, v_{j_1}, dots, v_{j_s}\}$, with $s \leq n-1$. Repeating then of the vectors $w_2, \dots, w_{m-2}$, we get a basis $\{w_2, \dots, w_{m-1}, \dots, v_\alpha\}$. Since $w_1, \dots, w_m$ are linearly independent, w_1 is not a linear combiantion of the others, hence the basis contains some v. Now the basis above has $m-1$ w_i's, at the cost of one $v \in V$, hence $m-1 \leq n-1$; thus $m \leq n$. ■

Corollary. *Any two bases have the same number of elements.*

Proof. Let $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_m\}$ be bases with n and m elements respectively. Since they are both linearly independent, by above we get $m \leq n$ and $n \leq m$. Therefore $m = n$. ■

Corollary. *$F^n \simeq F^m$ if and only if $n = m$.*

Proof. F^n has the basis $\{(1, 0, \dots, 0)_n, \dots, (0, 0, \dots, 1)_n\}$ and F^m has basis $\{(1, 0, \dots, 0)_m,$ $\dots, (0, 0, \dots, 1)_m\}$, and any isomorphism must map a basis to a basis. ■

Corollary. *If V is finite dimensional over F, with $V \simeq F^n$ for some unque n, then any basis in V has exactly n elements.*

Definition. If V is a finite dimensional vector space over a field F, with a basis $\{v_1, \dots, v_n\}$ of n elements, we call the n **dimension** of V over F and write $\dim_F V = n$ or $\dim V = n$.

Example 2.7. (1) $\dim F^n = n$.

(2) $\dim_F P_n = n$, and $\dim F[x] = \infty$ (since $F[x]$) is infinite dimensional.

(3) $\dim_{\mathbb{R}} \mathbb{C} = 2$.

Corollary. *If V and U are finite dimensional vector spaces over a field F, with $\dim_F V = \dim_F U$, then $V \simeq U$.*

Proof. $V \simeq F^n$ and $F^n \simeq U$. By transitivity, we get $V \simeq U$. ■

Lemma 2.2.7. *2.2.7 If V is a finite dimensional vector space over F and of $u_1, \ldots, u_m \in V$ are linearly independent, then there exis $u_{m+1}, \ldots, u_{m+r} \in V$ such that $\{u_1, \ldots, u_m, u_{m+1}, u_{m+r}\}$ is a basis of V.*

Proof. By finite dimensionality, there is a basis $v_1, \ldots, v_n$ of V, which span V. Hene span $\{u_1, \ldots, u_m, v_1, \ldots, v_n\} = V$, therefore by theorem 1.2.4 there is a subset $\{u_1, \ldots, u_m, v_{i_1}, \ldots, v_{i_r}\}$ which is a basis of V. Now just map $v_{i_j} \to u_{m+j}$ for each $1 \leq j \leq r$. ■

Remark. This gives us a method for constructing bases of vector spaces.

Lemma 2.2.8. *2.2.8 If V is finite dimensional, and if W is a subspace of V, then W is also finite dimensional. Moreoverm* $\dim W \leq \dim V$ *and* $\dim V/W = \dim V - \dim W$.

Proof. If $\dim V = n$, then any set of $n+1$ vectors in V is linearly dependent, by maximality, hence so is any set of $n+1$ vectors in W. Then there exists a maximal set of linearly independent elements in W, $w_1, \ldots, w_m$, with $m \leq n$. If $w \in W$, then $w_1, \ldots, w_m, w$ are linearly dependent with $\lambda_1 w_1 + \cdots + \lambda_m w_m + \lambda w = 0$. Now $\lambda \neq 0$, for that would imply $w_1, \ldots, w_n, w$ linearly independent. Hence $w = \mu_1 w_1 + \cdots + \mu_m w_m$ where $\mu_i = \lambda^{-1}\lambda_i$. Thus we get $w \in \text{span}\{w_1, \ldots, w_m\}$, i.e. $W = \text{span}\{w_1, \ldots, w_m\}$, thus $w_1, \ldots w_m$ form a basis of W. Therefore $m = \dim W \leq \dim V = n$.

Now take $V \to V/W$ by $v_1, \ldots, v_r \to v_1', \ldots, v_r'$. By lemma **??**, if $\{w_1, \ldots, w_m\}$ form a basis of W, then there exist $v_{m+1}, \ldots, v_{m+r}$ such that $\{w_1, \ldots, wm, v_{m+1}, v_{m+r}\}$ form a basis for V. That is, for any $v \in V$, $v = \lambda_1 w_1 + \cdots + \lambda_m w_m + \mu_1 v_1 + \cdots + \mu_r v_r$. Then we get that $v' = \mu_1 v_1' + \cdots + \mu_r v_r'$, hence $\text{span}\{v_1', \ldots, v_r'\} = V/W$. Now if $\gamma_1 v_1' + \cdots + \gamma_r v_r' = 0$, then $\gamma_1 v_1' + \cdots + \gamma_r v_r \in W$, making $\gamma_1 v_1' + \cdots + \gamma_r v_r = \lambda_1 w_1 + \cdots + \lambda_m v_m$. By linear independence, $\gamma_i, \lambda_j = 0$ for all $1 \leq i \leq r$ and $1 \leq j \leq m$. This V/M has a basis of $r = \dim V - \dim W$ elements. Therefore $\dim V/W = \dim v - \dim W$. ■

Corollary. *If U and W are finite dimensional subspaces of a vector space V, then $U + W$ is finite dimensional, and* $\dim(A + B) = \dim A + \dim B - \dim_A \cap B$.

Proof. We have $U + W/W \simeq U/U \cap W$. Hence we get that $\dim U + W/W = \dim(U + W) - \dim W = \dim U/U \cap W = \dim U - \dim U \cap W$. Then $\dim(U + W) = \dim W = \dim U - m_9 U \cap W$. ■

2.3 Dual Spaces.

Lemma 2.3.1. *2.3.1 Let V and W be vector spaces over a field F. Then* $\hom(V, W)$ *is a vector space over F.*

Proof. First, let $T, L \in \hom(V, W)$, and $\alpha, \beta \in F$. Then $T + L(\alpha v + \beta u) = \alpha T(v) + \beta T(u) + \alpha L(v) + \beta L(u) = \alpha(T + L) + \beta(T + L)$, so $T + L \in \hom(V, W)$. Since $+$ is just function addition, it is associative. Likewise, the zero map $0 : V \to W$ by $v \to 0$ and the map $-T : V \to W$ by $v \to -v$ define the identity of $\hom(V, W)$ and the inverse of T respectively. This makes $(\hom(V, W), +)$ into a group. Now by the properties of homomorphisms, we also see that $\alpha(T + L) = \alpha T + \alpha L$, $(\alpha + \beta)T = \alpha T + \beta T$, $\alpha(\beta T) = (\alpha\beta)T$ and $T(1v) = 1T(v)$. This makes $\hom(V, W)$ a vector space. ■

Lemma 2.3.2. *2.3.2 If $S, T \in \hom(V, W)$ such that $S(v_i) = T(v_i)$ for all v_i in a basis $\{v_1, \ldots, v_n\}$ of V, then $S = T$.*

Proof. Since $\{v_1, \ldots, v_n\}$ is a basis of V, we have for every $v \in V$, $v = \lambda_1 v_1 + \cdots + \lambda_n + v_n$ for unique $\lambda_1, \ldots, \lambda_n \in F$. Then we get $S(v) = \lambda_1 S(v_1) + \cdots + \lambda_n + S(v_n) = \lambda_1 T(v_1) + \cdots + \lambda_n + T(v_n) = T(v)$. Thus $S(v) = T(v)$ for all $v \in V$. ■

Theorem 2.3.3. *2.3.3 If V and W are vector spaces with $\dim V = m$ and $\dim W = n$, then $\dim \hom(V, W) = mn$.*

Proof. Let $\{v_1, \ldots, v_m\}$ and $\{w_1, \ldots, w_n\}$ be bases for V and W, respectively. Then for any $v \in V$, $v = \lambda_1 v_1 + \cdots + \lambda_m v_m$ for unique $\lambda_1, \ldots, \lambda_n \in F$. Now let $T_{ij} \in \hom(V, W)$ be defined such that $T_{ij}(v_i) = 0$ for $i \neq j$ and $T_{ij}(v) = \lambda_i w_j$; for $1 \leq i \leq m$ and $1 \leq j \leq n$. We see there are mn possible such T_{ij}. Now let $S \in \hom(V, W)$, then $S(v_i) \in W$, hence $S = \mu_{11} w_1 + \cdots + \mu_n w_{1n}$ for unique $\mu_{11}, \ldots, \mu_{1n} \in F$. Then $S(v_i) = \mu_{i1} w_1 + \cdots + \mu_{in} w_n$ for unique $\mu_{i1}, \ldots, \mu_{in} \in F$. Now let $S_0 = \mu_{11} T_{11} + \cdots + \mu_{1n} T_{1n} + \cdots + \mu_{m1} T_{m1} + \cdots + \mu_{mn} T_{mn}$. Then $S_0(v_k) = (\mu_{11} T_{11} + \cdots + \mu_{1n} T_{1n} + \cdots + \mu_{m1} T_{m1} + \cdots + \mu_{mn} T_{mn})(v_k) = \mu_{11} T_{11}(v_k) + \cdots + \mu_{1n} T_{1n}(v_k) + \cdots + \mu_{m1} T_{m1}(v_k) + \cdots + \mu_{mn} T_{mn}(v_k)$. Since $T_{ij}(v_k) = 0$ for $i \neq k$ we get $S_0(v_k) = \alpha_{k1} w_1 + \cdots + \alpha_{kn} wn$. So $S_0(vk) = S(v_k)$ for the basis $\{v_1, \ldots, v_m\}$ of V; this makes $S_0 = S$.

Now since $S = S_0$ is arbitrary, and subsequentially a linear combination of the T_{ij}, we get that span $\{T_{11}, \ldots, T_{1n}, \ldots, T_{m1}, \ldots, T_{mn}\} = \hom(V, W)$. Now suppose for $\beta_{11}, \ldots, \beta_{1n}$, $\ldots, \beta_{m1}, \ldots, \beta_{mn} \in F$ that $\beta_{11} T_{11} + \cdots + \beta_{1n} T_{1n} + \cdots + \beta_{m1} T_{m1} + \cdots + \beta_{mn} T_{mn} = 0$. Then we get that $(\beta_{11} T_{11} + \cdots + \beta_{1n} T_{1n} + \cdots + \beta_{m1} T_{m1} + \cdots + \beta_{mn} T_{mn})(v_k) = \beta_{k_1} w_1 + \cdots + \beta_{kn} w_n = 0$. Since $\{w_1, \ldots, w_n\}$ is a basis of W, this makes $\beta_{kj} = 0$ for all $1 \leq n$. Thus $\{T_{11}, \ldots, T_{1n}, \ldots, T_{m1}, \ldots, T_{mn}$ linearly independent, and hence a basis of $\hom(V, W)$. Therefore, $\dim \hom(V, W) = mn$. ■

Corollary. $\dim \hom(V, V) = m^2$.

Corollary. $\dim \hom(V, F) = m$.

Definition. Let V be a vector space over a field F. We call the vector space $\hom(V, F)$ the **dual space** of V and denote it $\operatorname{dual} V$. We call elements of $\operatorname{dual} V$ **linear functionals** on V into F.

If V is an infinite dimensional vector space, the $\operatorname{dual} V$ is very big and of no interest. In these cases, we use properties of other possible structures of $\operatorname{dual} V$ to find a restricted subspace. If V is finite dimensional, then $\operatorname{dual} V$ is finite and always defined.

Lemma 2.3.4. *2.3.4 If V is a finite dimensional vector space, and $v \neq 0 \in V$, then there is a linear functional $\hat{v} \in \operatorname{dual} V$ such that $\hat{v}(v) \neq 0$.*

Proof. Let $\{v_1, \ldots, v_n\}$ be a bases of V and let $\hat{v}_i \in \operatorname{dual} V$ be defined by $\hat{v}_i(v_j) = 0$ whenever $i \neq j$ and $\hat{v}_i(v_j) = 1$ otherwise. Then if $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$, $\hat{v}_i(v) = \lambda_i$. Then $\{\hat{v}_i, \ldots, \hat{v}_n\}$ forms a basis of $\operatorname{dual} V$. Now if $v \neq 0 \in V$. by lemma **??**, we get a basis $v_1 = v, v_2, \ldots, v_n$. Thence there is a linear functional $\hat{v}_1(v_1) = \hat{v}_1(v) = 1$. ■

Definition. Let V be a finite dimensional vector space with basis $\{v_1, \ldots, v_n\}$. We define the **dual basis** of $\{v_1, \ldots, v_n\}$ to be a basis of linear functionals $\{\hat{v}_1, \ldots, \hat{v}_n\}$ of $\operatorname{dual} V$ such that $\hat{v}_i(v_j) = 0$ wheberver $i \neq j$ and $\hat{v}_i(v_i) = 1$ otherwise.

Lemma 2.3.5. *2.3.5 If V is a finite dimensional vector space, and $T \in \operatorname{dual} V$ such that $T(v)$ is fixed, then the map $\psi : v \to T_v$, where $T_v(T) = T(v)$ defines an isomorphism of V onto* $\operatorname{dual}(\operatorname{dual} V)$.

Proof. Let $v_0 \in V$. Let $T \in \operatorname{dual} V$ be a linear functional such that $T(v_0)$ is fixed. Then $T(v_0)$ defines a linear functional of $\operatorname{dual} V$ into F. Let $T_{v_0} : \operatorname{dual} V \to F$ be defined by $T_{v_0}(T) = T(v_0)$, for any $T \in \operatorname{dual} V$. Notice that for $T, L \in \operatorname{dual} V$ and $\alpha, \beta \in F$, we have $T_{v0}(\alpha T + \beta L) = \alpha T(v_0) + \beta L(v_0) = \alpha T_{v0}(T) + \beta T_{v0}(L)$, which makes $T_{v0} \in \operatorname{dual}(\operatorname{dual} V)$.

Now given any $v \in V$, we can associate it with a $T_v \in \operatorname{dual}(\operatorname{dual} V)$. Now define $\psi : V \to \operatorname{dual}(\operatorname{dual} V)$ by $\psi : v \to T_v$. Then for $v, w \in V$ and $\alpha, \beta \in F$ we have $T_{\alpha v + \beta w}(T) = \alpha T(v) + \beta T(w) = \alpha T_v(T) + \beta T_w(T)$, so ψ is a homomorphism of V onto $\operatorname{dual}(\operatorname{dual} V)$; ψ is onto by definition.

Now let $v \in \ker \psi$. So $\psi(v) = 0$; that means $t_v(T) = T(v) = 0$ for all $T \in \operatorname{dual} V$. However, by lemma 1.3.3, there must be a $T \in \operatorname{dual} V$ for which $T(v) \neq 0$ when $v \neq 0$. Therefore, if $v \in \ker T$, it must be that $v = 0$, that is $\ker T = (0)$. Thus ψ is $1 - 1$, which makes it an isomorphism. ■

Definition. Let W be a subspace of a vector space V. We denote the **annihilator** of W to be $A(W) = \{T \in \operatorname{dual} V : T(v) = 0\}$.

Example 2.8. (1) Let $W_1, W_2 \subseteq V$ be subspaces of a finite dimensional vector space. Let $T \in A(W_1 + W_2)$. Then $T(w) = 0$ for $w \in W_1 + W_2$, hence $w = w_1 + w_2$ where $w_i \in W_i$ for $1 \leq i \leq 2$. So we get $T(w_1) + T(w_2) = 0$ which makes either both $T(w_1), T(w_2)$ 0 or inverses of each other. IN either case, $T(w_1) + T(w_2) \in A(W_1) + A(W_2)$ or $T(w_1) + T(w_2) \in A(W_1) \cap A(W_2) \subseteq A(W_1) + A(W_2)$. So $A(W_1 + W_2) \subseteq A(W_1) + A(W_2)$. On the other hand we have $A(W_1), A(W_2) \subseteq A(W_1 + W_2)$, hence $A(W_1) + A(W_2) \subseteq A(W_1 + W_2)$. Hence we have $A(W_1 + W_2) = A(W_1) + A(W_2)$.

(2) Similarly, if $T \in A(W_1 \cap W_2)$, then $T(w) = 0$ for $w \in W_1 \cap W_2$, making $T(w) \in A(W_1) \cap A(W_2)$. By similar reasoning to before, we also get that $A(W_1) \cap A(W_2) \subseteq A(W_1 \cap W_2)$. So we get $A(W_1 \cap W_2) = A(W_1) \cap A(W_2)$

Let $\tilde{T} \in \operatorname{dual} W$ such that $\tilde{T}(w) = T(w)$ for any $w \in W$; where $T \in \operatorname{dual} V$. Now define the map $\psi : \operatorname{dual} V \to \operatorname{dual} W$ by $\psi : T \to \tilde{T}$. Then we see that $A(W) = \ker \psi$, which makes it a subspace.

Lemma 2.3.6. *2.3.6 If $S \subseteq V$ is a subset of a finite dimensional vector space, then $A(S) \subseteq A(\operatorname{span} S)$.*

Proof. Since $S \subseteq \operatorname{span} S$, it is clear that $A(S) \subseteq A(\operatorname{span} S)$. Now let $v \in \operatorname{span} S$. Then $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$ where $v_1, \ldots, v_n \in S$. Let $\psi : \operatorname{dual} V \to \operatorname{dual} S$ by $T \to \tilde{T}$ where $\tilde{T}(s) = T(s)$ for all $s \in S$. Then $\psi(v) = \lambda_1 \psi(v_1) + \cdots + \lambda_n \psi(v_n) = \psi(v) = \lambda_1 T(v_1) + \cdots + \lambda_n T(v_n)$. Since $\ker \psi = A(S)$, and $T(v_i) = 0$ for all $v_i \in S$ for $1 \leq i \leq n$, we get $\psi(v) = 0$ hence $v \in A(S)$; which puts $A(\operatorname{span} S) \subseteq A(S)$. ■

Theorem 2.3.7 (The Second Homomorphism Theorem for Vector Spaces)**.** *2.3.7 If V is a finite dimensional vector space, and $W \subseteq V$ is a subspace of V, then* $\operatorname{dual} W \simeq \operatorname{dual} V/A(W)$, *and* $\dim A(W) = \dim V - \dim W$.

Proof. Consider again the map $\psi : \operatorname{dual} V \to \operatorname{dual} W$ by $T \to \tilde{T}$, where $\tilde{T}(w) = T(w)$ for all $w \in W$; and recalling above that $A(W) = \ker T$.

Let $h \in \operatorname{dual} W$. By lemma **??**, if $\{w_1, \ldots, w_m\}$ is a basis of W, then there is a basis $\{w_1, \ldots, w_m, v_1, \ldots, v_r\}$; hence $\dim V = r + m$. Let W_1 be a subspace of V such that $Span\{v_1, \ldots, v_r\} = W_1$. Then $V = W \oplus W_1$. Now if $h \in \operatorname{dual} W$, let $f \in \operatorname{dual} V$ be defined by $f(v) = w$ where $v = w + w_1 \in W \oplus W_1$. By definition, we have that $f \in \operatorname{dual} V$ and $f = h$. So $\psi(f) = h$ making ψ onto. Since $A(W) = \ker \psi$, by the first homomorphism theorem for vector spaces, we get $\operatorname{dual} W \simeq \operatorname{dual} V/A(W)$.

Moreover, we get $\dim \operatorname{dual} W = \dim \operatorname{dual} V/A(W) = \dim \operatorname{dual} V - \dim A(W)$, and since $\dim \operatorname{dual} V = \dim V$ and $\dim \operatorname{dual} W = \dim W$; we get $\dim A(W) = \dim V - \dim W$. ■

Corollary. $A(A(W)) = W$.

Proof. Notice that $A(A(W)) \subseteq \operatorname{dual}(\operatorname{dual} V)$. Clearly, $W \subseteq A(A(W))$, for if $\psi(w) = T_w$ by $T_w(f) = f(w)$ and $T_w = 0$ for all $f \in A(W)$. Now by above we get $\dim A(A(W)) = \dim \operatorname{dual} V - \dim A(W) = \dim V - (\dim V - \dim W) = \dim W$. This makes $W \simeq A(A(W))$; and since $W \subseteq A(A(W))$, we get $W = A(A(W))$. ■

Theorem 2.3.8. *2.3.8 The system of homogeneous linear equations:*

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned} \tag{2.1}$$

where $a_{ij} \in F$ is of rank r, then there are $n - r$ linearly independent solutions in F^n.

Proof. Consider the system described by equation 2.1, with $a_{ij} \in F$ for $1 \leq i \leq m$ and $1 \leq j \leq n$. Let U be a subspace of m vectors generated by $\{(a_{11}, \ldots, a_{1n}), \ldots, (a_{m1}, \ldots, a_{mn})\}$. Consider the basis $\{(1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)\}$ of F^n and let $\{\hat{v}_1, \ldots, \hat{v}_n\}$ be its dual basis. Then $T \in \operatorname{dual} F^n$ has the form $T = x_1\hat{v}_1 + \cdots + x_n\hat{v}_n$, with $x_i \in F$. for $1 \leq i \leq n$.

Now for $(a_{11}, \ldots, a_{1n}) \in U$, $T(a_{11}, \ldots, a_{1n}) = (x_1\hat{v}_1 + \cdots + x_n\hat{v}_n)(a_{11}, \ldots, a_{1n}) = a_{11}x_1 + \cdots + a_{1n}x_n$, since $\hat{v}_i(v_j) = 0$ for $i \neq j$. Conversely, every solution $(x_1, \ldots, x_n)$ gives an element of the form $x_1\hat{v}_1 + \cdots + x_n\hat{v}_n$ in $A(U)$. Therefore, the number of linearly independent solutions of equation 2.1 is $\dim A(U) = \dim V - \dim U = n - r$. ■

Corollary. *If $n > m$, then there is a solution $(x_1, \ldots, x_n)$ where not all x_i is 0.*

2.4 Inner Product Spaces.

Definition. We define a vector space V over $\mathbb{C}$ to be an **inner product space** if there exists a binary operation $\langle , \rangle : V \times V \to \mathbb{C}$ such that for all $v, u, w \in V$ and $\alpha, beta \in \mathbb{C}$:

(1) $\langle u, v\rangle = \overline{\langle v, u\rangle}$.

(2) $\langle u, u\rangle \geq 0$ and $\langle u, u\rangle = 0$ if and only if $u = 0$.

(3) $\langle \alpha u + \beta v, w\rangle = \alpha\langle u, w\rangle + \beta\langle v, w\rangle$.

Example 2.9. (1) In $\mathbb{C}^n$, let $u = (\alpha_1, \ldots, \alpha_n)$ and $v = (\beta_1, \ldots, \beta_n)$ and define $\langle u, v\rangle = \sum_{i=1}^n \alpha_i\overline{\beta_i}$. Notice that $\sum \alpha_i\overline{\beta_i} = \sum_{i=1}^n \overline{\beta_i}\alpha_i = \overline{\sum \overline{\alpha_i}\beta_i}$; so $\langle u, v\rangle = \overline{\langle v, u\rangle}$. We also have that $\langle u, u\rangle \geq 0$ and is 0 only when $u = 0$. Moreover, if $w = (\gamma_1, \ldots, \gamma_i)$ and $\alpha, \beta \in \mathbb{C}$, then $\langle \alpha u + \beta v, w\rangle = \sum(\alpha\alpha_i + \beta\beta_i)\overline{\gamma_i} = \alpha\sum \alpha_i\overline{\gamma_i} + \beta\sum \beta_i\overline{\gamma_i} = \alpha\langle u, w\rangle + \beta\langle v, w\rangle$. So $\langle , \rangle$ defines an inner product over $\mathbb{C}^n$.

(2) Let $\mathbb{C}^{[0,1]}$ be the set of all complex valued functions continous on the domain $[0, 1]$. If $f, g \in \mathbb{C}^{[0,1]}$, define $\langle f, g\rangle = \int_0^1 f(t)\overline{g(t)}t$. Then $\langle , \rangle$ defines an inner product over $\mathbb{C}^{[0,1]}$. Let $f, g, h \in \mathbb{C}^{[0,1]}$ and $\alpha, \beta \in \mathbb{C}$. We have then that $\langle f, g\rangle = \int f\overline{g} = \overline{\int \overline{f}g} = \overline{\int \overline{f}g} = \overline{\langle g, f\rangle}$. Moreover, $\int_0^1 f\overline{f}t \geq 0$; now $\langle f, f\rangle = 0$ if $f = 0$. Now if $\int f\overline{f}t = 0$, letting $f(t) = x(t) + iy(t)$, by the product of conjugates, and the sum rule, $x(t) = y(t) = 0$, i.e. $f = 0$. Again, by the rules of complex integras, $\langle \alpha f + \beta g, h\rangle = \int(\alpha f + \beta g)\overline{h} = \alpha\int f\overline{h} + \beta\int gh$.

Definition. Let V be an inner product space over $\mathbb{C}$. The **norm** of $v \in V$ is the map $\|\cdot\| : V \to \mathbb{R}$ by $\|v\| = \sqrt{\langle v, v\rangle}$.

Lemma 2.4.1. *2.4.1 If V is an inner product space, with $u, v \in V$ and $\alpha, \beta \in \mathbb{C}$, then $\langle \alpha u + \beta v, \alpha u + \beta v\rangle = \alpha\overline{\alpha}\langle u, u\rangle + \alpha\overline{\beta}\langle u, v\rangle + \overline{\alpha}\beta\langle v, u\rangle + \beta\overline{\beta}\langle v, v\rangle$.*

Proof. Take (3) on the inner product $\langle \alpha u + \beta v, \alpha u + \beta v\rangle$ to get: $\langle \alpha u + \beta v, \alpha u + \beta v\rangle = \alpha\langle u, \alpha u + \beta v\rangle + \beta\langle v, \alpha u + \beta v\rangle = \alpha\overline{\langle \alpha u + \beta v, u\rangle} + \beta\overline{\langle \alpha u + \beta v, v\rangle} = \alpha\overline{\alpha}\langle u, u\rangle + \alpha\overline{\beta}\langle u, v\rangle + \overline{\alpha}\beta\langle v, u\rangle + \beta\overline{\beta}\langle v, v\rangle$. ■

Corollary. $\|\alpha u\| = |\alpha|\|u\|$.

Proof. We have $||\alpha u||^2 = \langle \alpha u, \alpha u\rangle = \alpha\overline{\alpha}\langle u, u\rangle$. Since $\alpha\overline{\alpha} = |\alpha|^2$ we have $||\alpha u|| = |\alpha|^2||u||^2$ which gives us the result. ■

Lemma 2.4.2. *2.4.2 If $a, b \in \mathbb{R}$ such that $a > 0$ and $a\lambda^2 + 2b\lambda + c \geq 0$ for all $\lambda \in \mathbb{R}$, then $b^2 \leq ac$.*

Proof. We complete the squares. $a\lambda^2 + 2b\lambda + c = \frac{1}{a}(a\lambda + b)^2 + (c - \frac{b^2}{a}) \geq 0$. Choosing $\lambda = -\frac{b}{a}$, we get $c - \frac{b^2}{a} \geq 0$. ■

Theorem 2.4.3 (The Cauchy-Schwarz Inequality)**.** *2.4.3 If V is an inner product space over $\mathbb{C}$ with $u, v \in V$, then $|\langle u, v\rangle| \leq \|u\|\|v\|$.*

Proof. If $\langle u, v\rangle \in V = \mathbb{R}$, and $u \neq 0$, then for any $\lambda \in \mathbb{R}$, $\langle u\lambda + v, u\lambda + v\rangle = \lambda^2\langle u, u, \rangle + 2\lambda\langle u, v\rangle + \langle v, v\rangle \geq 0$. Letting $a = \langle u, u\rangle$, $b = \langle u, v\rangle$ and $c = \langle v, v\rangle$ we get $a\lambda^2 + 2b\lambda + c \geq 0$. By the above lemma, then $b^2 \leq ac$; i.e. $|\langle u, v\rangle|^2 \leq ||u||^2||v||^2$.

Now take $\alpha = \langle u, u\rangle \in V \neq \mathbb{R}$. Then $\alpha \neq 0$. Now we observe that $\langle \frac{u}{\alpha}, v\rangle = \frac{1}{\alpha}\langle u, v\rangle = \frac{1}{\langle u,v\rangle}\langle u, v\rangle = 1$; so $\langle \frac{u}{\alpha}, v\rangle \in \mathbb{R}$. Then by above, we have $1 = |\langle \frac{u}{\alpha}, v\rangle| \leq ||\frac{u}{\alpha}||||v|| = \frac{1}{|\alpha|}||u||||v||$, that is $1 \leq \frac{\|u\|\|v\|}{|\alpha|}$; giving the desired result. ■

Example 2.10. (1) Let $V = \mathbb{C}^n$ and $\langle u, v\rangle = \sum_{i=1}^n \alpha_i, \overline{\beta_i}$ with $u = (\alpha_1, \ldots, \alpha_n)$ and $v = (\beta_1, \ldots, \beta_n)$. Then we have $|\sum \alpha_i \overline{\beta_i}| \leq \sum |\alpha_i|^2 \sum |\beta_i|^2$.

(2) If $V = \mathbb{C}^{[0,1]}$ with $\langle f, g\rangle = \int_0^1 f(t)\overline{g(t)}t$, then we have $|\int_0^1 f\overline{g}| \leq \int_0^1 |f|^2 \int_0^1 |g|^2$.

Definition. If V is an inner product space, we say that $u, v \in V$ are **orthogonal** (or that u is **orthogonal** to v) if $\langle u, v\rangle = 0$.

Example 2.11. If u is orthogonal to v, thejn $\langle u, v\rangle = \overline{\langle v, u\rangle} = \overline{0} = 0$, making v orthogonal to u.

Definition. If V is an inner product space, and $W \subseteq V$ is a subsapce of V we call the **orthogonal complement** of W the space $W^\perp = \{x \in V : \langle x, w\rangle = 0, \text{ for all } w \in W\}$.

Lemma 2.4.4. *2.4.4 $W^\perp$ is a subspace of V.*

Proof. Clearly $W^\perp \subseteq V$. Moreover, let $a, b \in W^\perp$ and $\alpha, \beta \in \mathbb{C}$, then $\langle \alpha a + b, w\rangle = \alpha\langle a, w\rangle + \beta\langle b, w\rangle = 0$, so $\alpha a + \beta b \in W^\perp$. ■

Example 2.12. Note that $W \cap W^\perp = \{x \in V : \langle x, w\rangle = 0\}$. If $w \in W^\perp$, then $\langle w, w\rangle = 0$ making $w = 0$, hence $W \cap W^\perp = 0$.

Definition. We call a set of vectors $\{v_i\}_{i\in\mathbb{Z}^+}$ of an inner product space V **orthonormal** if:

(1) $\langle v_i, v_i\rangle = 1$.

(2) $\langle v_i, v_j\rangle = 0$ whenever $i \neq j$.

Lemma 2.4.5. *2.4.5 If $\{v_i\}$ are a set of orthonormal vectors of V, then $\{v_i\}$ is also linearly independent. Moreover, if $\{v_i\}$ is finite and $w = \alpha_1 v_1 + \cdots + \alpha_n v_n$, then $\alpha_i = \langle w, v_i\rangle$ for each $1 \leq i \leq n$.*

Proof. Supposse that $\alpha_1 v_1 + \cdots + \alpha_n v_n + \cdots = 0$, then $\langle \alpha_1 v_1 + \cdots + \alpha_n v_n + \ldots, v_i\rangle = \alpha_1\langle v_1, v_i\rangle + \cdots + \alpha_n\langle v_n, v_i\rangle + \cdots = 0$. Since $\{v_i\}$ is orthonormall, we get that $\alpha_i = 0$ for each i, implying linear independence. Now if $\{v_i\}_{i=1}^n$ is finite, letting $w = \alpha_1 v_1 + \cdots + \alpha_n v_n$; by above we get that $\langle w, v_i\rangle = \alpha_i$ by orthonormality. ■

Lemma 2.4.6. *2.4.6 If $\{v_1, \ldots, v_n\}$ are orthonormal in V, and $w \in V$, then $u = w - \langle w, v_1\rangle v_1 - \cdots - \langle w, v_n\rangle v_n$ is orthogonal to each v_i for $1 \leq i \leq n$.*

Proof. $\langle u, v_i\rangle = \langle w - \langle w, v_1\rangle v_1 - \ldots \langle w, v_n\rangle v_n, v_i\rangle = \langle w, v_i\rangle - \langle w, v_i\rangle\langle v_1, v_i\rangle + \cdots + \langle w, v_n\rangle\langle v_n, v_i\rangle = 0$, making u orthogonal to v_i. ■

Theorem 2.4.7 (The Gram-Schmidt Orthogonalization Theorem)**.** *2.4.7 Let V be a finite dimensional inner product space. Then V has an orthonormal set as a basis.*

Proof. Let $\dim V = n$ and let $\{v_1, \ldots, v_n\}$ be a basis of V. Take $w_1, \ldots, w_n$ as follows: $v_1 | w_1$, $w_2 \in \text{span}\{w_1, v_2\}$ and $w_3 \in \text{span}\{w_1, w_2, v_3\}$; in general take $w_i \in \text{span}\{w_1, \ldots, w_{i-1}, v_i\}$. Let $v_1 = \|v_1\| w_1$, then $\langle w_1, w_1\rangle = \langle \frac{v_1}{\|v_1\|}, \frac{v_1}{\|v_1\|}\rangle = \frac{1}{\|v_1\|^2}\langle v_1, v_1\rangle = 1$; hence $\|w_1\| = 1$. Now consider $\langle \alpha w_1 + v_2, w_1\rangle = 0$. Then $\alpha\langle w_1, w_1\rangle + \langle v_2, w_1\rangle = 0$; since $\|w_1\| = 1$, then $\alpha =$

$-\langle v_2, w_1\rangle$. Now let $u_2 = -\langle v_2, w_1\rangle w_1 + v_2$. u_2 is orthogonal to w_1 by lemma **??** and since v_1 and v_2 are linearly independent, so must w_1 and v_2. So $u_2 \neq 0$. Now let $\|u_2\|w_2 = u_2$. We have then by above that, $\{w_1, w_2\}$ is orthonormal. Continuing along, suppose then that $\{w_1, \ldots, w_i\}$ are orthonormal, where $\|u_i\|w_i = u_i$, and where $u_i = -\langle v_i, w_1\rangle - \cdots - \langle v_i, w_{i-1}\rangle w_i + v_i$. Take $u_{i+1} = -\langle v_{i+1}, w_1\rangle - \cdots - \langle v_{i+1}, w_i\rangle + v_{i+1}$. By the above and lemma **??**, $w_1, \ldots, w_i, v_{i+1}$ are linearly independent, so $u_{i+1} \neq 0$. Putting $\|u_{i+1}\|w_i = u_i$, clearly $\langle w_{i+1}, w_{i+1}\rangle = 1$. We also have, by the construction, that $\langle u_{i+1}, w_1\rangle = \cdots = \langle u_{i+1}, w_i\rangle = 0$. So $w_1, \ldots w_n$ are orthonormal.

Constructin $\{w_1, \ldots, w_n\}$ from the basis $\{v_1, \ldots, v_n\}$ this way gives an orthonormal set of n linearly independent vectors; i.e. a basis. ■

Corollary (Bessel's Inequality)**.** *For all $v \in V$:*

$$\sum_{i=1}^{m} |\langle w_i, v\rangle|^2 \leq \|v\|^2. \tag{2.2}$$

Example 2.13. Let $V = \mathbb{R}_3[x]$ be the real field of all polynomials of deg < 3. Define for $p(x), q(x) \in \mathbb{R}_3[x]$

$$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)x.$$

Now consider the basis $\{1, x, x^2\}$ of $\mathbb{R}_3[x]$. Take $w_1 = \frac{1}{\|1\|} = \frac{1}{\sqrt{\int_1^1 x}} = \frac{1}{\sqrt{2}}$. Take $u_2 = -\langle x, w_1\rangle w_1 + x = -\frac{\langle x, w_1\rangle}{\sqrt{2}} + x = x \neq 0$. Now take $w_2 = \frac{u_2}{\|u_2\|} = \frac{x}{\sqrt{\int_{-1}^1 x^2 x}} = \frac{\sqrt{3}}{2}x$. Taking $u_3 = -\langle x^2, w_1\rangle w_1 - \langle x^2, w_2\rangle w_2 + x^2 = -\frac{1}{3} + x^2 \neq 0$; so taking $\frac{u_2}{\|u_3\|} = \frac{-\frac{1}{3}+x^2}{\sqrt{\int_{-1}^1 (-\frac{1}{3}+x^2)x}} = \frac{\sqrt{10}}{4}(-1+3x^2)$, we get the orthonormal basis $\{x, -\frac{1}{3} + x^2, \frac{\sqrt{10}}{4}(-1 + 3x^2)\}$.

Theorem 2.4.8. *2.4.8 If V is a finite dimensional inner product space, and if $W \subseteq V$ is a subspace of V, then $V = W \oplus W^\perp$.*

Proof. Since $W \subseteq V$ is a subsapace of V, W inherits the inner product of V (restrict $\langle, \rangle$ to $W \times W$); similarly, $W^\perp$ also inherits the inner product. By the Gram-Schmidt orthogonaliztion theorem, there is an orthonormal set of vectors $\{w_1, \ldots, w_r\}$ which is a basis of W. Now if $v \in V$, by lemma **??** take $v_0 = v - \langle v, w_1\rangle - \cdots - \langle v, w_r\rangle w_r$ and $\langle v_0, w_i\rangle = 0$ for each $1 \leq i \leq r$. Then $v = v_0 + \langle v, w_1\rangle + \cdots + \langle v, w_r\rangle w_r \in W + W^\perp$.*Since*W $\cap W^\perp = 0$, we get $V = W \oplus W^\perp$. ■

Corollary. $(W^\perp)^\perp = W$.

Proof. If $w \in W$, then for any $u \in W$, $\langle u, w\rangle = 0$, hence $W \subseteq (W^\perp)^\perp$. Now $V = W^\perp \oplus (W^\perp)^\perp$ and we have $\dim W = \dim (W^\perp)^\perp$, which gives us $W = (W^\perp)^\perp$. ■

2.5 Modules.

Definition. Let R be a ring. We say a nonempty set M is a **left module** over R (or a **left R-module**) if there are operations $+ : M \times M \to M$ and $\cdot : R \times M \to M$ such that $(M, +)$ is an abelian group, and for any $r, s \in R$ and $a, b \in M$:

(1) $r(a+b) = ra + rb$.

(2) $r(sa) = (rs)a$.

(3) $(r+s)a = ra + sa$.

Similarly, we call M a **right module** (or **right R-module**) over R if $(a+b)r = ar + br$, $(as)r = a(sr)$, and $a(r+s) = ar+as$. We call M **unital** if R ha a unit element, and $1m = m$ for all $m \in M$.

We focus on left modules.

Example 2.14. (1) All vector spaces are unital left modules over any field F.

(2) Let G be a group together with an arbitrary operation $+$ and define an action $\cdot : \mathbb{Z} \times G \to G$ by $(n, a) \to na \in G$. Then the properties of exponents in groups gives $r(a+b) = ra + rb$, $r(sa) = (rs)a$, and $(r+s)a = ra + sa$. This makes every group a left $\mathbb{Z}$-module.

(3) Let R be a ring, and let M be a left ideal of R. Take $r, m \to rm$. Since M is an ideal, $rm \in M$, and by the multiplicative associative, and distributive laws, M is a left R-module.

(4) Any ring R is a left (and right) module over itself.

(5) Let R be a ring, and (λ) a left ideal of R. Consider the quotient ring $R/(\lambda)$. define $+$ by $(a+\lambda)+(b+\lambda) = (a+b)+\lambda$ and $r(a+\lambda) = ra+\lambda$. Clearly these operations are well defined, and $(R/(\lambda), +)$ forms a group; moreover, $(a+\lambda)+(b+\lambda) = (a+b)+\lambda = (b+\lambda)+(a+\lambda)$, so $R/(\lambda)$ is abelian under $+$. Now notice that $r(a+b+\lambda) = r(a+b)+\lambda = ra+rb+\lambda = (ra+)+(rb+) = r(a+\lambda)+r(b+\lambda)$, $r(sa+\lambda) = rsa+\lambda = rs(a+\lambda)$, and $(r+s)(a+\lambda) = (r+s)a+\lambda = ra+rs+\lambda = r(a+\lambda)+s(a+\lambda)$. This makes $R/(\lambda)$ a left R-module. We call this module the **left quotient module** of R by (λ).

Definition. Let M be an R-module (left or right) and $A \subseteq M$, we call A a **submodule** of M is $A \leq M$ and whenever $r \in R$ and $a \in A$, $ra \in A$, or $ar \in A$.

Definition. If M is an R-module with a collection of submodules $\{M_i\}_{i=1}^s$. We call M the **direct sum** of $\{M_i\}$ if for every $m \in M$, there are uniquely determined $m_i \in M_i$ for $1 \leq i \leq s$, such that $m = m_1 + \cdots + m_s$. We write $M = M_1 \oplus \cdots \oplus M_s$, or $M = \bigoplus_{i=1}^s M_i$.

Definition. An R-module is **cyclic** if there exists $m_0 \in M$ such that $m = rm_0$ (or $m = m_0 r$) for all $m \in M$ and some $r \in R$.

Definition. We say an R-module is **finitely generated** if there exists $a_1, \ldots, a_n \in M$ such that for every $m \in M$, $m = r_1 a_1 +_n a_n$ (or $m = a_1 r_1 + \cdots + a_n r_n$) for $r_1, \ldots, r_n \in R$. We call $\{a_i\}_{i=1}^n$ the **generating set**; and we call it a **minimal generating set** if $\{a_i\} \backslash a_j$ does not generate M, for $1 \leq i, j \leq n$. We call the size of a minimal generatng set the **rank** of M and denote it $\operatorname{rank} M$.

Most of the definitions are stated for both left and right R-modules. However, we consider the following theorems only for left R-modules.

Theorem 2.5.1 (The Fundamental Theorem on Finite Modules)**.** *2.5.1 Let R be a Euclidean domain; then any finitely generated module M is the direct sum of a finite number of cyclic submodules.*

Proof. By definition, if M is finitely generated, then there are $a_1, \ldots, a_n \in M$ for which every element of M is of the form $r_1a_1 + \cdots + r_na_n$, for $r_1, \ldots, r_n \in R$. If M is indeed a direct sum of a finite collection of cyclic submodules, then each r_ia_i is uniquely determined.

By induction in the rank of M; if $\operatorname{rank} M = 1$, then M is generated by a single element m_0. That is, for some $r \in R$, every element of M has the form rm_0; this makes M cyclic by definition, and hence the direct sum of itself.

Now suppose for $\operatorname{rank} M = q$, that $M = \bigoplus_{i=1}^{q} M_q$, where M_i is a cyclic submodule. Suppose now that $\operatorname{rank} M = q+1$ and let $\{a_i\}_{i=1}^{q+1}$ be a minimal generating set for M. Then there are $r_1, \ldots, r_{q+1} \in R$ for which $r_1a_1 + \cdots + r_{q+1}a_{q+1} = 0$ (the identity of $(M, +)$). If $r_1a_1 = \cdots = r_{q+1a_{q+1}} = 0$, then $M = \bigoplus_{i=1}^{q+1} M_i$ and we are done.

Now suppose that not all the r_ia_i are 0. Since R is a Euclidean domain, with a degree function deg, there is an element s_1 of minimum degree occuring as a coefficient in a relation of $\{a_i\}_{i=1}^{q+1}$. Thenn $s_1a_1 + \cdots + s_{q+1}a_{q+1} = 0$, where $\deg s_1 \leq \deg s_i$ for all $1 < i \leq q+1$. Now if $r_1a_1 + \cdots + r_{q+1}a_{q+1} = 0$, then $s_1 | r_1$, for if $r_1 = ms_1 + t$ with $t = 0$ or $\deg t < \deg s_1$, then taking $(ms_1)a_1 + \ldots (ms_{q+1})a_{q+1} = 0$ and subtracting $r_1a_1 + \cdots + r_{q+1}a_{q+1}$, we get $ta_1 + (r_2 - ms_2)a_2 + \cdots + (r_{q+1} - ms_{q+1})a_{q+1} = 0$, since $\deg t < \deg s_1$, and s_1 has minimum such degree, this makes $t = 0$.

We also have $s_1 | s_i$ for all $1 \leq i \leq q+1$ (obviously $s_1 | s_1$). For, suppose that $s_1 \not| s_i$ for all $1 < i \leq q+1$, then $s_2 = m_2s_1 + t$ with $\deg t < \deg s_1$. Now $a_1' = a_1 + m_2a_2 + \cdots + m_{q+1}a_{q+1}, m_2a_2, \ldots, m_{q+1}a_{q+1}$ also generate M; however, $s_1a_1' + ta_2 + s_3a_3 + \cdots + s_{q+1}a_{q+1} = 0$, so t is a coefficient occuring in some relation of $\{a_i\}$. But $\deg t < \deg s_1$, which contradicts that s_1 has minumum such degree, so $t = 0$ and hence $s_1 | s_2$. Similarly we get $s_1 | s_i$.

Now consider $a_1^* = a_1 + m_2a_2 + \cdots + m_{q+1}a_{q+1}, a_2, \ldots, a_{q+1}$. They generate M; moreover $s_1a_1^* = s_1a_1 + (s_1m_2)a_2 + \cdots + (s_1m_{q+1})a_{q+1} = s_1a_1 + \cdots + s_{q+1}a_q + 1 = 0$. If $ra_1^* = ra_1 + (rm_2)a_2 + \cdots + (rm_{q+1})a_{q+1} = 0$, then there is some relation on $\{a_i\}$ for which a_1 has coefficent r, i.e. $s_1 | r$, so $r_1a_1^* = 0$. Letting M_1 the cyclic submodule generated by a_1^*, and M_2 the submodule finitely generated by $\{a_i\}_{i=2}^{q+1}$, we have $M_1 \cap M_2 = 0$ and $M = M_1 + M_2$; hence $M = M_1 \oplus M_2$. Now by hyothesis, we get $M_2 = M_2' \oplus M_3 \oplus \cdots \oplus M_{q+1}$, each of which is a cyclic submodule of M; which completes the proof. ■

Corollary. *Any finite abelian group is the direct product of cyclic groups.*

Proof. Consider the finite abelian group G as a $\mathbb{Z}$-module. ■

Theorem 2.5.2. *2.5.2 The number of non-isomorphic finite abelian groups of order p^n is $p(n)$; where $p(n)$ is the number of partitions of n.*

Proof. Let G be a finite abelian group of order $\operatorname{ord} G = p^n$, for $n, p \in \mathbb{Z}^+$ and p prime. By the corollary to the fundamental theorem, $G = G_1 \times \cdots \times G_k$, where G_i is a cyclic group of

order $\operatorname{ord} G_i = p^{n_i}$, where $n_k \leq \cdots \leq n_1 \leq n_1$. Then

$$G_1 \times G_2 = \frac{\operatorname{ord} G_1 \operatorname{ord} G_2}{\operatorname{ord}(G_1 \cap G_2)}.$$

Since $G_1 \times G_2$ is a direct product, $\operatorname{ord}(G_1 \cap G_1) = (e)$, so $\operatorname{ord} G_1 \times G_2 = \operatorname{ord} G_1 \operatorname{ord} G_2 = p^{n_1} p^{n_2} = p^{n_1+n_2}$. Continuing this way we get $p^n = \operatorname{ord} G = \operatorname{ord}(G_1 \times \cdots \times G_k) = p^{n_1+\cdots+n_k}$, hence $n = n_1 + \cdots + n_k$ making $\{n_i\}_{i=1}^k$ a partition of n.

On the other hand, if $\{n_i\}_{i=1}^k$ is a partition of n, then we construct G of ord $= p^n$ as follows: for $1 \leq i \leq k$, let G_i be a cyclic group of order $\operatorname{ord} G_i = p^{n_i}$ and let G be the external direct product of $\{G_i\}_{i=1}^k$. G is an abelian group of order p^n. Hence for each partition of n, there isa abelian group of order p^n, if we take p^{n_i} for $1 \leq i \leq k$, characterizing G up to isomorphism, we get a $1-1$ correspondence ofo non-isomorphic finite abelian groups of order p^n and partitions if n. ■

Corollary. *The number of non-isomorphic finite abelian groups of order $p_1^{n_1} \ldots p_k^{n_k}$ for p_i distinct primes is $p(n_1) \ldots p(n_k)$.*

We now observe R-modules in the context of homomorphisms.

Definition. Let R be a ring, and let M and N be left R-modules. We define a map $T : M \to N$to be a **left R-homomorphism** if

(1) $(m_1 + m_2)T = m_1T + m_2T$.

(2) $(rm_1)T = r(m_1)T$.

We define the **kernel** of T to be $\ker T = \{x \in M : xT = 0\}$. We define the **image** of T to be $\operatorname{Im} T = \{xT : x \in M\}$.

Here we mean xT to be $T(x)$ to reduce notational encumberance. In the case of composition of R-homomorphisms, we mean TS $S \circ T$.

Example 2.15. Let $T : M \to N$ and $S : N \to Q$ be left R-homomorphisms. Define $TS : M \to Q$ by $xTS = (xT)S$. Then for $r, s \in R$ and $m_1, m_2 \in M$, we have that $(rm_1 + sm_2)TS = r((m_1T)S) + s((m_2T)S)$. Which makes TS into an R-homomorphism. It is easy to see then that $\ker TS = \{xT : xTS = 0\}$.

Lemma 2.5.3. *2.5.3 Let M and N be left R-modules, and let $T : M \to N$ be a left R-homomorphism. Then $\ker T$ and $\operatorname{Im} T$ are submodules of M and N respectively.*

Proof. Since T is a R-homomorphism, it is a group homomorphism; hence $\ker T \leq M/$ Now letting $r \in R$ and $x \in \ker T$, $(rx)T = r(xT) = r0 = 0$, putting $rx \in \ker T$. Similarly, by the bilinearity of T, $\operatorname{Im} T \leq N$ and $xT \in \operatorname{Im} T$. since $rx \in M$, and $r(xT) = (rx)T$, we get that $r(xT) \in \operatorname{Im} T$. ■

Lemma 2.5.4. *2.5.4 Let T be an R-homomorphism. Then T is $1-1$ if and only if $\ker T = 0$.*

Proof. Suppose that T is $1-1$. Then $xT = yT$ implies $x = y$, this makes $\operatorname{ord}(\ker T) = 1$, hence $\ker T = 0$. Now suppose that $kerT = 0$, and let $xT = yT$. Then $xT - yT = (x-y)T = 0$, so $x - y \in \ker T$. This makes $x - y = 0$, hence $x = y$ which makes T $1-1$. ■

Definition. Let M and N be R-modules. We say that an R-homomorphism $T : M \to N$ is an **R-isomorphism** if T is $1-1$ from M onto N. In this case, we say that M and N are **R-isomorphic**, and write $M \simeq_R N$.

We would also like to define what a "left quotient module" much in the same manner we described the left quotient module" of a ring R by a left ideal (λ). Our motivation is the fact that if M is a left R-module, and $A \subseteq M$ is a submodule, then since $(r, a) \in R \times A$ implies $ra \in A$, this makes A into a left ideal of R. So already we have that R/A is a left quotient module of R by A.

We would like to take this same quotient, restricting R to M. Define the operations $+ : M/A \times M/A \to M/A$ by $(a + A) + (b + A) = (a + b) + A$ and $\cdot : R \times M/A \to M/A$ by $r(a + A) = ra + A$. Like in the case of quotient modules by ideals, these operations are well defined, and make $(M/A, +)$ into a group; moreover they satisfy the rest of the axioms for modules. Thus we then have the following definition.

Definition. Let M be a left R-module and $A \subseteq M$ a submodule. Define the operations $+$ and $\cdot$ by $(a + A) + (b + A) = (a + b) + A$ and $r(a + A) = ra + A$, respectively. We call the module M/A the **left quotient module** of M by A.

Lemma 2.5.5. *2.5.5. Let M be an a left R-module, and let $A \subseteq M$ be a submodule. Then there exists a left R-homomorphism from M onto M/A.*

Proof. Take the map $m \to m+A$ which defines a left R- homomorphism for $(rm+sn)+A = r(m + A) + s(n + A)$; this map is also onto by definition. ■

Theorem 2.5.6. *2.5.6 Let M and N be R-modules. If $T : M \to N$ is an R-homomorphism from M onto T, then $N \simeq_R M$.*

Proof. By the fundamental theorem for group homomorphisms, we have that as groups, $N \simeq M/\ker T$. By the aximos of modules, this makes $N \simeq_R M/\ker T$. ■

Definition. We call an R-module M **irreducible** if its only submodules are 0 and M.

2.6 Coordinates.

We return to our discussion on vector spaces, specially concerning bases and row equivalence (covered in next section). Here, we introduce the concept of a "coordinate space".

Definition. If V is a finite dimensional vector space, we define an **ordered basis** of V to be a finite sequence of vectors, which form a basis for V.

Let $\mathcal{B} = \{\alpha_1, \ldots \alpha_n\}$ be an ordered basis of a vector space V of $\dim V = n$. Then there is a unique n-tuple of scalars such that $\alpha = \sum_{i_1}^{n} \mathrm{x}_i\alpha_i.Wcall\mathrm{x}_i$ the i-th **coordinate** of α **relative** to $\mathcal{B}$.

Now let $\beta = \sum y_i\alpha_i$ be another vector of V and define $+ : V \times V \to V$ and $\cdot"F \times V \to V$ by $\alpha + \beta = \sum (x_i + y_i)\alpha_i$ and $c\alpha = \sum (cx_i)\alpha_i$. We see then that $\mathcal{B}$ determines a $1-1$ mapping $\alpha \to (x_1, \ldots, x_n)$ of V onto F^n.

Definition. Let V be a vector space of $V = n$, and $\mathcal{B} - \{\alpha_1, \ldots, \alpha_n\}$ a coordinate basis. Let $\alpha = (x_1, \ldots, x_n)$. We call the $n \times 1$ matrix

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \tag{2.3}$$

the **coordinate matrix** of α **relatve** to $\mathcal{B}$. We also write X as $X = (x_1, \ldots, x_n)^T$.

Theorem 2.6.1. *Let V be an n-dimensional vector space over a field F. let $\mathcal{B}$ and $\mathcal{B}'$ be ordered bases of V. Then there is a unique $n \times n$ invertible matrix P over F such that:*

(1) $(\alpha)_{\mathcal{B}} = P(\alpha)_{\mathcal{B}'}$.

(2) $(\alpha)_{\mathcal{B}'} = P^{-1}(\alpha)_{\mathcal{B}}$.

for all $\alpha \in V$. The columns of P are $P_j = (\alpha'_j)_{\mathcal{B}}$.

Proof. Let $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ and $\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$. There are unique scalars $P_{ij} \in F$ such that $\alpha'_j = \sum_{i=1}^n P_{ij}\alpha_i$ for $1 \leq i \leq n$. Now let $(x'_1, \ldots, x'_n)$ be the coordinates of α with respect to $\mathcal{B}'$. Then $\alpha = \sum_{j=1}^n x'_j\alpha'_j = \sum x_j \sum P_{ij}\alpha_i = \sum\sum (P_{ij}x'_j)\alpha_i$. Since $(x_1, \ldots, x_n)$ are uniquely determined in $\mathcal{B}$, we get $x_i = \sum_{j=1}^n P_{ij}x'_j$ for $1 \leq i \leq n$.

Now define the matrix $P = (P_{ij})_{n\times n}$ and let $X = (x_1, \ldots, x_n^T)$ and $X' = (x'_1, \ldots, x'_n)^T$. Then we have $X = PX'$. Since $\mathcal{B}$ and $\mathcal{B}'$ are bases, and hence linearly independent, we get $X = 0$ if and only if $X' = 0$, so $X' = P^{-1}X$ by row equivalence. ■

Theorem 2.6.2. *Suppose P is an $n\times n$ invertrible matrix over F. let V ne an n-dimensional vector space over F and let $\mathcal{B}$ be an ordered basis of V. Then there is a unique ordered basis $\mathcal{B}'$ of V such that:*

(1) $(\alpha)_{\mathcal{B}} = P(\alpha)_{\mathcal{B}'}$.

(2) $(\alpha)_{\mathcal{B}'} = P^{-1}(\alpha)_{\mathcal{B}}$.

for all $\alpha \in V$.

Proof. Let $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$. if $\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$, then from the previous theorem, we get $\alpha'_j = \sum P_{ij}\alpha_i$. Notice then that $\sum_j P_{ij} = \sum_j P^{-1}_{jk} \sum_i P_{ij}\alpha_i = \sum_i \sum_j P_{ij}P^{-1}_{jk}\alpha_i = \alpha_k$; which spans $\mathcal{B}$, and hence V. So $\mathcal{B}'$ is a basis. That $\mathcal{B}'$ is unique comes from linear independence. ■

Example 2.16. (1) Let F be an arbitrary field, and let $\alpha = (x_1, \ldots, x_n) \in F^n$. Let $\mathcal{B} = \{e_1, \ldots, e_n\}$ be the standard basis of F^n. $\mathcal{B}$ is an ordered basis, and we have $\alpha = (x_1, \ldots, x_n)^T$.

(2) let $F = \mathbb{R}$ and $\theta \in \mathbb{R}$. Consider the matrix $P = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. P is invertible with $P^{-1} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, so for all $\theta \in \mathbb{R}$, the subspace $\mathcal{B}' = \{(\cos\theta, \sin\theta), (-\sin\theta, \cos\theta)\}$

is an ordered basis for $\mathbb{R}^2$. We visualize this basis by rotating the standard basis anti-clockwise by an angle of θ. Let $\alpha = (x_1, x_2)$, then:

$$(\alpha)_{\mathcal{B}'} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

which gives

$$\begin{aligned} x_1' &= x_1\cos\theta + x_2\sin\theta \\ x_2' &= -x_1\sin\theta +_2 \cos\theta \end{aligned}$$

(3) Let $F = \mathbb{C}$. Consider the matrix $P = \begin{pmatrix} -1 & 4 & 5 \\ 0 & 2 & -3 \\ 0 & 0 & 8 \end{pmatrix}$ which has inverse $P = \begin{pmatrix} -1 & 4 & 5 \\ 0 & \frac{1}{2} & \frac{3}{16} \\ 0 & 0 & \frac{1}{8} \end{pmatrix}$. So the vectors form a basis $\mathcal{B}' = \{(-1,4,5),(0,2,-3),(0,0,8)\}$ and the coordinates (x_1', x_2', x_3') of $\mathcal{B}'$ are given by:

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} -1 & 4 & 5 \\ 0 & \frac{1}{2} & \frac{3}{16} \\ 0 & 0 & \frac{1}{8} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

2.7 Row Equivalence.

We can use the properties of vector spaces and bases to summarise the row equivalence of matrices.

Definition. Let A be an $m \times n$ matrix over a field F. We define the **i-th row vector** of A to be the vector $\alpha_i = (A_{i1}, \dots A_{in})$ in the vector space F^n. We call the subspace of F^n spanned by the row vectors of A the **row space** of A and denote it $\operatorname{row} A$. We call the dimension of $\operatorname{row} A$ the **rank** of A and denote it $\operatorname{rank} A$.

Theorem 2.7.1. *Row equivalent matrices have the same row space.*

Proof. If A is an $m \times n$ matrix and P is a $k \times m$ matrix whose row vectors are defined by $\beta_i = P_{i1}\alpha_1 + \cdots + P_{im}\alpha_m$, where α_i are the row vectors of A for $1 \le i \le m$. That is $\operatorname{row} B$ is a subspace of $\operatorname{row} A$. Now if P is $m \times m$ and invertible, then B is row equivalent to A, so $A = P^{-1}B$ makes $\operatorname{row} A$ a subspace of $\operatorname{row} B$. That is $\operatorname{row} A = \operatorname{row} B$. ■

Theorem 2.7.2. *Let R be a nonzero row reduced echelon matrix. Then the nonzero rows of R form a basis for* $\operatorname{row} R$.

Proof. Let $\rho_1, \dots, \rho_r \in \operatorname{row} R$ be the nonzero row vectors of R. Clearly $\operatorname{row} R = \{\rho_1, \dots, \rho_r\}$. All that is left is linear independence. Since R is a row reduced echelon matrix, we have by definition that there are $k_1, \dots k_r \in \mathbb{Z}^+$ such that:

(1) $R_{ij} = 0$ if $j < k$.

(2) $R_{ik_j} = \delta_{ij}$ (the Kronecker delta).

(3) $k_1 < \cdots < k_r$.

Now suppose that $\beta = (b_1, \ldots, b_r) \in \operatorname{row} R$. Then we have $\beta = c_1\rho_1 + \ldots_r \rho_r$ for $c_i \in F$ for $1 \leq i \leq r$. Expanding ρ_i and by above, we see that $c_j = b_{k_j}$. In particular, $\beta = 0$ implies that $\beta = c_1\rho_1 + \ldots_r \rho_r = 1$, which implies that c_j is the k_j-th coordinate of the 0 vector. That is $c_j = 0$ for all j. ■

Theorem 2.7.3. *Let $m, n \in \mathbb{Z}^+$ and let F be a field. Suppose that W is a subspace of F^n and that $\dim W \leq m$. Then there is precisely one $m \times n$ row reduced echelon matrix over F with W as its row space.*

Proof. There is at least one such matrix as described above. Now select m vectors $\alpha_1, \ldots, \alpha_m \in W$ for which $\operatorname{span} W = \{\alpha_i\}_{i=1}^m$. Let A be the $m \times n$ matrix with $\operatorname{row} A = \{\alpha_i\}$, and let R be a row reduced echelon matrix row equivalent to A. then $\operatorname{row} R = \operatorname{row} A = W$ by theorem 2.7.1.

Let R be any row reduced echelon matrix with $\operatorname{row} R = W$, and let $\rho_1, \ldots, \rho_r$ be the nonzero vectors of R; and suppose theat the leading nonzero entry of ρ_i occurs in column k_i with $1 \leq i \leq r$.. By theorem **??**, these vectors form a basis for W, and by the proof of the same theorem, we had $\beta = \sum_{i=1}^r b_{k_i}\rho_i$. Thus any vector β is determined by the coordinates b_{k_i}. Now suppose that $\beta \neq 0 \in W$. By above, we get that $R_{ij} = 0$ for $i > s$ adn $j \leq k_s$, thus $\beta = (0, \ldots, 0, b_{k_s}, b_n)$ where $b_{k_s} \neq 0$. Also note that there exists a vector in W with nonzero k_s-th coordinate; namely ρ_s.

Now that R is uniquely determined by W, let us describe it in terms of W. Consider $\beta \in W$. If $\beta \neq 0$, then $\beta = (0, \ldots, 0, b_t, \ldots, b_n)$ where $b_t \neq 0$. Let $k_{1,r} \in \mathbb{Z}^+$ be those t for which there is some $\beta \neq 0$. Take $k_1 < \cdots < k_r$. For each k_s, there is one and only one ρ_s for which the k_s-th coordinate is 1 and every ther coordinate is 0. Then R is the $m \times n$ matrix whose row space is $\operatorname{row} R = \{\rho_1, \ldots, \rho_r, 0, \ldots, 0\}$. ■

Corollary. *Each $m \times n$ matrix A over F is equivalent to precisely one row reduced echelon matrix.*

Corollary. *Let A and B be $m \times n$ matrices over F. Then B is row equivalent to A if and only if* $\operatorname{row} A = \operatorname{row} B$.

Theorem 2.7.4. *Let A and B be $m \times n$ matrices over a field F. The following are equivalent:*

(1) B is row equivalent to A.

(2) $\operatorname{row} B = \operatorname{row} A$.

(3) $B = PA$ where P is an $m \times m$ invertible matrix.

(4) The homogeneous systems $AX = 0$ and $BX = 0$

Chapter 3

Linear Transformations.

3.1 Linear Transformations.

When we studied vector spaces, we introduced the definition of a "vector space homomorphism", or (better know as) a linear transformation. It would be interesting to study the space of such linear transformations. Let V be a vector space over a field F and consider the space $\mathrm{Hom}\,(V,V)$ of all linear transformations from V into itself. It was shown that $\mathrm{Hom}\,(V,V)$ forms a vector space over F. This was done with the property of linearity. Now for $T_1, T_2 \in \mathrm{Hom}\,((V,V))$, consider $T_2 \circ T_1(v)$ for $v \in V$. Now let $\alpha, \beta \in F$ and $u, v \in V$. Then:

$$\begin{aligned} T_2 \circ T_1(\alpha v + \beta u) &= T_2(T_1(\alpha v + \beta u)) \\ &= T_2(\alpha T_1(v) + \beta T_1(u)) \\ &= \alpha T_2(T_1(v)) + \beta T_2(T_1(u)) \\ &= \alpha T_2 \circ T_1(v) + \beta T_2 \circ T_1(u) \end{aligned}$$

This makes $T_2 \circ T_1$ a linear transformation, so $T_2 \circ T_1 \in \mathrm{Hom}\,(V,V)$. We can speculate on some properties of $T_2 \circ T_1$.

Lemma 3.1.1. *Let V be a vector space and $T_1, T_2, T_3 \in \mathrm{Hom}\,(V,V)$, and consider the composition in $\mathrm{Hom}\,(V,V)$. The following hold:*

(1) $T_3 \circ (T_1 + T_2) = T_3 \circ T_1 + T_3 \circ T_2$.

(2) $(T_1 + T_2) \circ T_3 = T_1 \circ T_3 + T_2 \circ T_3$.

(3) $(T_3 \circ T_2) \circ T_1 = T_3 \circ (T_2 \circ T_1)$.

(4) $\alpha(T_2 \circ T_1) = T_2 \circ (\alpha T_1) = (\alpha T_2) \circ T_1$.

The follwong are some well known examples of linear transformations.

Example 3.1. (1) For any vector space V, the **zero** transform $0 : V \to V$ defined by $0 : \alpha \to 0$ is a linear transformation.

(2) Let F be a field and let $V = F[x]$. Take the map $D : F[x] \to F[x]$ by taking $f \to f'$ where f' is the derivative of the polynomial f. That is if $f(x) = c_0 + c_1x + \cdots + c_nx^n$, then $Df(x) = c_1 + 2c_2x + \cdots + nc_nx^{n-1}$. The map D is a linear transformation called the **differentiation** transform.

(3) Let A be an $m \times n$ matrix over a field F and let T be defined by $T(X) = AX$. Then T is a linear transformation from $F^{n\times 1}$ into $F^{m\times 1}$. The map $U : \alpha \to \alpha A$ is also a linear transformation from $F^n \to F^m$.

(4) Let P and Q be $m \times m$ and $n \times n$ matrices over F Define the map $T : F^{m\times n} \to F^{mn}$ by $T : A \to PAQ$. Notice that for $x, y \in F$ that $T(xA + yB) = P(xA + yB)Q = xPAQ + yPBQ = xT(A) + yT(B)$ so T is a linear transformation.

(6) Let $V = C(\mathbb{R})$ the space of all continuous funtions from $\mathbb{R}$ to $\mathbb{R}$. Define the map $T : \mathbb{R} \to \mathbb{R}$ by $Tf(x) = \int_0^x f\ dt$. By the properties of integration from real analysis, it is easy to see that T is a linear transformation. This linearity if a fundamental property of the integral. Moreover, we can see that Tf is continuous and has a continuous first derivative.

We should notice lemma 3.1.1 makes $\mathrm{Hom}\,(V, V)$ into an associative ring. We can also see there is an identity $I \in \mathrm{Hom}\,(V, V)$, so $\mathrm{Hom}\,(V, V)$ is an associative ring with unit. We also notice that for any T, $\alpha T = T \circ (\alpha I) = (\alpha I) \circ T$, so αI commutes with every linear transformation in the space. This motivates the following definition.

Definition. We call an associative ring A an **algebra** over a field F if A is a vector space over F such that for al $a, b \in A$ and $\alpha \in F$, $\alpha(ab) = (\alpha a)b = a(\alpha b)$.

Example 3.2. $\hom\,(V, V)$ is an algebra over F.

Let us repeat the definition for a linear transformation in a more restricted sense.

Definition. Let V be a vector space over a field F. A **linear transformation** on V is an element of $\hom\,(V, V)$.

Lemma 3.1.2. *If A is an algebra over a field F, with unit element, then A is isomprphic to a subalgebra of* $\hom\,(A, A)$.

Proof. If A is an algebra over F, it is a vector space over F by definition.

Now let $a \in A$ and define $T_a : A \to A$ by $v \to va$, for every $v \in A$. We have that $v + u \to (v + u)a = va + ua$ and $\alpha v \to (\alpha v)a = \alpha va = \alpha(va)$, which makes T_a a linear transformation on A. Notice that T_a is onto.

Now define $\psi : A \to \hom\,(A, A)$ by $a \to T_a$ for all $A \in A$. We notice that since T_a is onto, then so is ψ. Now notice that $a + b \to T_{a+b} = T_a + T_b$ and $\alpha a \to T_{\alpha a} = \alpha T_a$, which makes ψ into a homomorphism, also notice that $ab \to T_{ab} = T_aT_b$, which makes ψ a ring homomorphism.

Now suppose that $a \in \ker\psi$, then $\psi(a) = T_a = 0$, that is for every $v \in A$, $v \to va = 0$. Then $1a = a = 0$, which makes $\ker\psi = \langle 0\rangle$. So ψ is $1-1$ and onto, making it an isomorphism. This makes $A \simeq \hom\,(A, A)$. ■

Lemma 3.1.3. *Let A be an algebra with unit element, over a field F. Suppose that $\dim_F A = m$, then every element of A satisfies some nontrivial polynomial in $F[x]$ of $\deg \leq m$.*

Proof. For $a \in A$, consider the $m+1$ elements $1, a, \ldots, a^m \in A$. Then since $\dim A = m$, the set $\{1, a, \ldots, a^m\}$ must be linearly dependent; that is there are $\alpha_i \in F$, not all 0, for $0 \leq i \leq m$ for which $\alpha_0 1 + \alpha_1 a + \cdots + \alpha_m a^m = \alpha_0 + \alpha_1 a + \cdots + \alpha_m a^m = 0$. This implies that a is the root of the polynomial $p(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_m x^m \in Fx$. ■

Theorem 3.1.4. *If V is a vector space of dimension n over a field F, then for any linear transformation $T \in \hom(V, V)$, there exists a nontrivial polynomial $q \in F[x]$ of $\deg \leq n^2$ such that $q(T) = 0$.*

Proof. Since $\dim V = n$, by theorem **??** we have $\dim \hom(V, V) = n^2$. The rest follows from the above lemma. ■

Definition. Let V be a finite dimensional vector space, and let $T \in \hom(V, V)$. We call a polynomial $p \in F[x]$ a **minimal polynomial** for T if $p(T) = 0$, and for any other polynomial $q \in F[x]$ with $q(T) = 0$, then $p|q$.

Definition. We call a linear transformation T on a vector space V **right invertible** if there is a linear transformation S on V for which $ST = 1$. We call T **left invertible** if there exists a linear transformation U on V with $TU = 1$. If T is both left and right invertible, then we call T **invertible**, or **regular**; otherwise we call T **singular**.

Example 3.3. Let $F = \mathbb{R}$ and let $V = \mathbb{R}[x]$. Define linear transformations S and T by $S(q(x)) = q'(x)$ (the first derivative of q) and $T(q(x)) = \int_1^x q(x)\, dx$. Then $TS \neq 1$ but $ST = 1$. This means that real integrals on polynomials are right invertible, but not left invertible.

Theorem 3.1.5. *If V is a finite dimensional vector space over a field F, then $T \in \hom(V, V)$ is invertible if, and only if the constant in the minimal polynomial of T is nonzero.*

Proof. Let $p(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_k x^k$ be the minimal polynomial of $\deg = k$ for T. If $\alpha_0 \neq 0$, then $p(T) = \alpha_k T^k + \ldots + \alpha_1 T + \alpha_0$. We then get:

$$1 = T(-\frac{1}{\alpha_0}\alpha_k T^{k-1} + \cdots + \alpha_1).$$

Likewise, if T is invertible, but $\alpha_0 = 0$, then we have $p(T) = \alpha_1 T + \cdots + \alpha_k T^k = (\alpha_1 + \alpha_2 T + \cdots + \alpha_k T^{k-1})T = 0$. Multyplyuing by T^{-1}, we get $\alpha_1 + \alpha_2 T + \cdots + \alpha_k T^{k-1} = 0$. So T satisfies the polynomial $q(x) = \alpha_1 + \alpha_2 x + \cdots + \alpha_k x^{k-1}$ which has $\deg = k - 1$. This contradicts that p is the minimal polynomial. ■

Corollary. *If T is invertible, then T^{-1} is a polynomial in $F[T]$.*

Corollary. *If T is singlular, then there is an $S \in \hom(V, V)$ with $TS = St = 0$.*

Corollary. *If T is right invertible, then T is invertible.*

Proof. Suppose that T is right invertible, but singular. Then there is an S with $TS = 0$. Then for some linear transformation U, $U(TS) = (UT)S = 1S = S \neq 0$ which is a contradiction. ■

Theorem 3.1.6. *If V is a finite dimensional vector space over a field F, then $T \in \hom(V, V)$ is singlular if and only if there exists a $v \neq 0 \in V$ with $T(v) = 0$.*

Proof. If T is singular, then there is a linear transformation $S \neq 0$ with $TS = 0$. Then there is some $w \in V$ with $S(w) \neq 0$. Now let $v = S(w)$, then $T(v) = T(S(w)) = TS(w) = 0(w) = 0$. ■

Lemma 3.1.7. *Let V be a vector space over a field F, and let $T \in \hom(V, V)$, then $T(V) \subseteq V$ and $T(V) = V$ if, and only if T is onto.*

Theorem 3.1.8. *If V is a finite dimensional vector space over a field F, then $T \in \hom(V, V)$ is invertible if, and only if $T(V) = V$.*

Proof. If T is invertible, then for $v \in V$, $T(T^{-1}(v)) = v$ which makes $T(V) = V$, and T onto.

On the other hand, let $T(V) = V$ and suppose that T is singular. Then there is some $v_1 \neq 0 \in V$ such that $T(v_1) = 0$. By lemma **??**, we can find $v_1, \ldots, v_n \in V$. We can see then that $T(V)$ consists of the linear combinations of $w_1, \cdots w_n$, where $w_i = T(v_i)$ for $1 \leq i \leq n$. Sine $w_1 = 0$ we get $\operatorname{span}\{w_2, \ldots, w_n\} = T(V)$. We get then $\dim T(V) \leq n - 1 < \dim V$; this contradicts that $T(V) = V$. ■

Corollary. *T is invertible if and only if $\dim T(V) = \dim V$.*

The above corollary motivates the definition of the "rank" of a linear transformation, which will, in turn, motivate another definition of the rank of a matrix.

Definition. Let V be a finite dimensional vector space over a field F. We define the **rank** of a linear transformation $T : V \to V$ to be:

$$\operatorname{rank} T = \dim_F V(T) \tag{3.1}$$

We add the F subscript as a reminder that the rank of T is dependent on the field F.

Lemma 3.1.9. *If V is a finite dimensional vector space over a field F, and $S, T \in \hom(V, V)$ are linear transformations, then:*

(1) $\operatorname{rank} TS \leq \operatorname{rank} T$.

(2) $\operatorname{rank} ST \leq \operatorname{rank} T$.

(3) $\operatorname{rank} TS = \operatorname{rank} ST = \operatorname{rank} T$ *for S invertible.*

Proof. We have first that $S(V) \subseteq V$, so $TS(V) = T(S(V)) \subseteq T(V)$. Now let $\operatorname{rank} T = m$, then $T(V)$ has a basis of m vectors $\{w_1, \ldots, w_m\}$, and so $\operatorname{span}\{S(w_1), \ldots, S(w_m)\} = ST(V) = S(T(V))$ which makes $\operatorname{rank} ST \leq \operatorname{rank} T$.

Finally, if S is invertible, we have that $S(V) = V$, so $TS(V) = T(S(V)) = T(V)$ and $ST(V) = S(T(V)) = T(V)$. The corresponding equalities follow from above. ■

Corollary. *If S is invertible, then $T = \operatorname{rank} STS^{-1}$.*

3.2 Characteristic Roots.

Definition. Let V be a finite dimensional vector space over a field F, and let T be a linear transformation on V. We an element $\lambda \in F$ an **Eignevalue**, or **characteristic root** of T if $\lambda 1 - T = -T$ is singular.

Theorem 3.2.1. *Let V be a finite dimensional vector space over a field F, and let T be a linear transformation on V. Then $\lambda \in F$ is an Eigenvalue of T if, and only if $T(v) = \lambda v$, for some $v \neq 0 \in V$.*

Proof. If λ is is an Eigenvalue, then $-T$ is singular, and hence there is some $v \in V$ for which $\lambda - T(v) = \lambda v - T(v) = 0$. On the otherhad, if $T(v) = \lambda v$ for some $v \neq 0 \in V$, then $\lambda v - T(v) = 0$ which makes $\lambda - T$ singular. ■

Lemma 3.2.2. *If $\lambda \in F$ is an Eigenvalue for a linear transformation on a vector space over a field F, then for any $q \in F[x]$, $q(\lambda)$ is an Eigenvalue of $q(T)$.*

Proof. Let $\lambda \in F$ be an Eigenvalue, then there is a $v \neq 0 \in V$ for which $T(v) = \lambda v$, by the above theorem. Then $T^2(v) = T(T(v)) = \lambda(\lambda v) = \lambda^2 v$. By induction, we can show that $T^n(v) = \lambda^n v$ for $n \in \mathbb{Z}^+$. Now let $q(x) = \alpha_0 x^m + \cdots + \alpha_{m-1}x + \alpha_m$ with $\alpha_i \in F$. Then we get :

$$\begin{aligned} q(T(v)) &= \alpha_0 T^m(v) + \cdots + \alpha_{m-1}T(v) + \alpha_m \\ &= \alpha_0 \lambda^m v + \cdots + \alpha_{m-1}\lambda v + \alpha_m \\ &= (\alpha_0 \lambda^m + \cdots + \alpha_{m-1}\lambda + \alpha_m)v \end{aligned}$$

So we have $q(T(v)) = q(T)(v) = q(\lambda)v$ ■

Theorem 3.2.3. *Let V be a vector space over a field F and let T be a linear transformation on V If $\lambda \in F$ is an Eigenvalue of T, then λ is a root of the minimal polynomial for T. In particular, there are finitely many Eigenvalues of T in F.*

Proof. Let $p \in F[x]$ be the minimal polynomial for T. By the above lemma, we have that if $\lambda \in F$ is an Eigenvalue, then $p(\lambda)$ is an Eigenvalue for $p(T)$; that is $p(T(v)) = p(\lambda)v$ for some $v \neq 0 \in V$. We also have that $p(T) = 0$, therefore we have that $p(T(v)) = p(\lambda)v = 0$, and since $v \neq 0$, it must be that $p(\lambda) = 0$. Therefore λ is a root of p. Moreover, since there are atmost $\deg p$ roots of p, there must be at most $\deg p$ Eigenvalues for T. ■

Remark. Since $\dim V = n$ for some n, notice that $\deg p \leq n^2$, thus there are atmost n^2 Eigenvalues for T.

Remark. Also notice that since λ is a root of a given polynomial, then it makes sense to also call it the characteristic **root** of the linear transformation T. We may use these terms interchangibly.

Lemma 3.2.4. *If S and T are linear transformations on a vector space V over a field F, and S is invertible, then T and STS^{-1} share the same minimal polynomial.*

Proof. If S is invertible, then $(STS^{-1})^i = ST^iS^{-1}$. Now for $q \in F[x]$, we have $q(STS^{-1}) = Sq(T)S^{-1}$ and if $q(T) = 0$, the equality is 0. Now if p is the minimal polynomial for T, then $p(T) = 0$, and we get $p(STS^{-1}) = Sp(T)S^{-1} = 0$, making p the minimal polynomial for STS^{-1}. ■

Corollary. *T and STS^{-1} share the same Eigenvalues.*

Definition. Let V be a vector space over a field F and let T be a linear transformation on V. We call a $v \neq 0 \in V$ an **Eigenvector**, or **characteristic vector** of T, belonging to the Eigevalue $\lambda \in F$ if $T(v) = \lambda v$.

It is worth noting that the above theorems and lemmas for Eigenvalues can be reformulated for Eiegenvectors.

Theorem 3.2.5. *Let V be a vector space over a field F. If $\lambda_1, \ldots, \lambda_n$ are distinct Eigenvalues of a linear transformation T on V, and if $v_1, \ldots, v_n$ are Eigenvectors for T belonging to $\lambda_1, \ldots, \lambda_n$, respectively, then $\{v_1, \ldots, v_n\}$ are linearly independent over F.*

Proof. If $n = 1$, we are done, so suppose that $n > 1$. Let $v_1, \ldots, n_n$ be linearly dependent Eigenvectors, then there are $\alpha_1, \ldots, \alpha_n \in F$ not all 0 for which $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$. We can reformulate this equation such that it is the shortest possible relation:

$$\beta_1 v_1 + \cdots + \beta_j v_j = 0 \tag{3.2}$$

for which $\beta_i \neq 0$, for $1 \leq i \leq j$. Since $T(v_i) = \lambda_i v_i$, we get:

$$\lambda_1 \beta_1 v_1 + \cdots + \lambda_j \beta_j v_j = 0 \tag{3.3}$$

Multiplying by equation (3.2) and then subtracting from equation (3.3) we get:

$$({}_2 - \lambda_1)\beta_2 v_2 + \cdots + (\lambda_j - \lambda_1)\beta_j v_j = 0$$

Then $\lambda_i - \lambda_1 \neq 0$ for $i > 0$ and ${}_i \neq 0$ so we have $(\lambda_i - \lambda_1)_i \neq 0$, which would give a shorter relation than that of equation (3.2), a contradiction! ■

Corollary. *If T is a linear transformation and $\dim V = n$, then T has atmost n distinct Eigenvalues.*

Corollary. *If T is a linear transformation $\dim V = n$ and T has exactly n distinct Eigenvlues, then there is a basis of V over F consisting solely of Eigenvalues.*

3.3 Matrices.

Let V be a vector space of $\dim = n$ over a field F. Let $\{v_i\}_{i=1}^n$ be a basis of V. If T is a linear transformation on V, then since $T : V \to V$, $T(v_i) \in V$ for all i. We can then represent this as a linear combination of the basis elements:

$$\begin{aligned} T(v_1) &= \alpha_{11} v_1 + \cdots + \alpha_{1n} v_n \\ T(v_2) &= \alpha_{21} v_1 + \cdots + \alpha_{2n} v_n \\ &\vdots \\ T(v_n) &= \alpha_{n1} v_1 + \cdots + \alpha_{nn} v_n \end{aligned}$$

So we can determine T on any vector by knowing its action on a given basis. This then gives a nice way to represent a linear transformation as the entries of a matrix.

Definition. Let V be a vector spac e over a field F with $\dim_F V = 1$, and let $\{v_i\}_{i=1}^n$ be a basis of V. If T is a linear transformation on V, then the **matrix** of T in $\{v_i\}$ is the $n \times n$ matrix

$$m(T) = (\alpha_i j)) \tag{3.4}$$

where $T(v_i) = \sum_j \alpha_i j v_j$ and $\alpha_{ij} \in F$ for $1 \leq i, j \leq n$.

Example 3.4. (1) Let F be a field and let $V = P_n[x]$, and let D be defined by $D : f \to f'$ where f' is the first derivative of f. Now D is the differentiation transform, so it is a linear transformation. Now given the basis $\{x^i\}_{i=1}^{n-1}$ of $F_n[x]$, we get:

$$\begin{aligned}
D(1) &= 0 \cdot 1 + \cdots + 0x^{n-1} \\
D(x) &= 1 \cdot 1 + \cdots + 0x^{n-1} \\
&\vdots \\
D(x^i) &= 0 \cdot 1 + \ldots ix^{i-1} + \cdots + 0x^{n-1} \\
&\vdots \\
D(x^{n-1}) &= 0 \cdot 1 + \cdots + (n-1)x^{n-2} + \cdots + 0x^{n-1}
\end{aligned}$$

then by definition, the matrix for D is:

$$\begin{pmatrix}
0 & 0 & \ldots & 0 & 0 \\
1 & 0 & \ldots & 0 & 0 \\
0 & 2 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & n-1 & 0
\end{pmatrix}$$

Now if we are given the basis $\{1 + x^i\}_{i=1}^n$ we get by similar reasoning:

$$\begin{pmatrix}
0 & 0 & 0 & \ldots & 0 & 0 \\
1 & 0 & 0 & \ldots & 0 & 0 \\
-2 & 2 & 0 & \ldots & 0 & 0 \\
-3 & 0 & 3 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
-(n-1) & 0 & 0 & \ldots & n-1 & 0
\end{pmatrix}$$

So the matrix of the linear transformation depends on the choice of basis. One should keep in mind that these two matrices represent the same transformation D.

(2) If F is a field and V is a vector space of $\dim V = n$, and if T is a linenar transformation with n distinct eigenvalues $\lambda_1, \cdots \lambda_n$, then by the corollaryto theorem 3.2.5 T has the

matrix:

$$\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \tag{3.5}$$

Lemma 3.3.1. *Let V be a vector space of* $\dim = n$ *over a field F and let $m(\hom(V,V))$ be the set of all matrices of linear trasnformations on V over F. Then $m(\hom(V,V))$ is a subspace of $F^{n\times n}$.*

Proof. For any linear transformation T on V, $m(T)$ has entries in F. ■

Theorem 3.3.2. *Let V be a vector space over a field F with* $\dim V = n$. *Then the space of all matrices of linear transformations on V over F is an algebra over F.*

Proof. Let $m(\hom(,V,V))$ be the space of all linear transformations on V. By the above lemma, it is a vector space. Now let $\{v_1,\dots,v_n\}$ be a basis for V, and let $m(T) = (t_{ij})$, $m(S) = s_{ij}$ be matrices of the linear transformations T and S. Then $T(v_i) = \sum_j t_{ij}v_j$, and $S(v_i) = \sum_j s_{ij}v_j$. Notice that for $v_i \in V$, we have $TS(v_i) = \sum_k t_{kj}\sum_j s_{ij}v_j = \sum_j(\sum_k t_{ik}s_{kj})v_j$ So we get $m(TS) = (a_{ij})$ where $a_{ij} = \sum_k t_{ik}s_{kj}$. That is $m(TS) = m(T)m(S)$. Then it is a straightforward computation to check, for $\alpha \in F$, that $\alpha m(TS) = \alpha m(T)m(S) = (\alpha m(T))m(S) = m(T)(\alpha m(S))$. This makes $m(\hom(V,V))$ into an algbera over F. ■

Corollary. *$m(\hom(V,V))$ is an associative Algebra.*

Lemma 3.3.3. *Let V be a vector space over a field F with* $\dim V = n$. *Then the algebra of all matrices linear transformations on V is equal to $F^{n\times n}$ the space of all $n\times n$ matrices.*

Proof. Let $m(\hom(V,V)$ be the algebra of all matrices of linear transformations on V. We have shown in lemma 3.3.1 that $m(\hom(V,V)) \subseteq F^{n\times n}$. Now, let $\{v_i\}_{i=1}^n$ be a basis of V over F, and let $(a_{ij}) \in F^{n\times n}$. Construct the a map $T : V \to V$ by $T : v_i \to \sum_j a_{ij}v_j$. Then T is a linear transformation on V, and by definition we get $m(T) = (aij)$. Thus $F^{n\times n} \subseteq m(\hom(V,V))$ ■

Corollary. *$F^{n\times n}$ is an associative algebra.*

Remark. We will now just denote the space of matrices of linear transformations by $F^{n\times n}$.

Theorem 3.3.4. *Let V be a vector space over a field F with* $\dim_F V = n$, *then $F^{n\times n} \simeq \hom(V,V)$.*

Proof. Given the map $\phi : T \to m(T)$ where $m(T)$ is the matrix of the linear transformation T on V. By definition, since we can find a corresponding matrix associated with a given linear transformation, we get that ϕ is onto. Moreover, we have that $\phi(TS) = \phi(T)\phi(S)$ and $\phi(T+S) = \phi(T) + \phi(S)$, which makes ϕ into a homomorphism. Now, notice that if $T \in \ker\phi$, then $m(T) = 0$ the $n\times n$ 0 matrix. Then necesarrily, we must have that $T = 0$, for if $T \neq 0$, then for some basis $\{v_i\}_{i=1}^n$ of V, there is some v_i for which $T(v_i) = \sum t_{ij}v_j \neq 0$, making $t_{ij} \neq 0$ for atleast one t_{ij}, for $1 \leq i,j \leq n$. So we get that $\ker\phi = 0$, and so we have established an isomorphism. ■

Remark. What this theorem tells us is that we can represent linear transformations on a given vector space as matrices over the ground field. This will allow us to prove results about linear transformations using matrices and vice versa.

Theorem 3.3.5. *Let V be a vector space over a field F with $\dim_F V = n$, and let T be a linear transformation on V. Let $\{v_i\}_{i=1}^n$ and $\{w_i\}_{i=1}^n$ be bases on V, and $m_{\{v_i\}}(T)$ the matrix of T under the first basis and $m_{\{w_i\}(T)}$ the matrix of T under the second basis. Then there is an $n \times n$ matrix $C \in F^{n\times n}$ with $m_{\{w_i\}}(T) = C^{-1}m_{\{v_i\}}(T)C$. In particular, $C = m_{\{v_i\}}(S)$ for some linear transformation S on V.*

Proof. Let $m_{\{v_i\}}(T) = (a_{ij})$ and $m_{\{w_i\}}(T) = (b_{ij})$ Then $T(v_i) = \sum a_{ij}v_j$ and $T(w_i) = \sum b_{ij}v_j$. Let S be a linear transformation on V defined by $S : v_i \to w_i$. Since $\{v_i\}$ and $\{w_i\}$ are bases, S is onto, hence by theorem 3.1.8 S is invertible.

Notice then, that $TS(v_i) = \sum b_{ij}S(v_j) = S(\sum b_{ij}v_j)$, thus $S^{-1}TS = \sum b_{ij}v_j$. Then we get that $m_{\{v_i\}}(S^{-1}TS) = (b_{ij}) = m_{\{w_i\}}(T)$. Moreover, notice that $m_{\{v_i\}}(S^{-1}ST) = m_{\{v_i\}}(S)^{-1}m_{\{v_i\}}(T)m_{\{v_i\}}(S)$; which completes the proof. ■

Example 3.5. Let F be a field and $V = P_4[x]$ and consider the bases $\{x_i\}_{i=0}^3$ and $\{1+x^i\}_{i=0}^3$. Then for the differential transformation $D : f \to f'$ we get:

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ -3 & 0 & 3 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

$$m_{\{1+x^i\}}(D) = C^{-1}m_{\{x^i\}}(D)C$$
$$m_{\{1+x^i\}}(D) = m_{\{x^i\}}(S)^{-1}m_{\{x^i\}}(D)m_{\{x^i\}}(S)$$

3.4 Canonical Forms.

For this section, let V be a vector space over a field F with $\dim V = n$. We begin with the following definition.

Definition. We call two linear transformations, $T, S \in \hom(V, V)$ **similar** if there is an invertible linear transformation $C \in \hom(V, V)$ for which $T = C^{-1}SC$. We denote similarity by writing $T \simeq S$.

Lemma 3.4.1. *The relation of similarity of linear transformations on V is an equivalence relation.*

Proof. Let S and T be linear transformations on V. We note that $T \simeq T$, for take $I \in \hom(V, V)$, the identity transformation, then $T = I^{-1}TI$.

Now, notice that if $T \simeq S$, then there is a linear transformation C on V for which $T = C^{-1}SC$. Then $S = CTC^{-1} = (C^{-1})^{-1}TC^{-1}$.

Finally, let U be a linear transformation on V, and suppose that $T \simeq S$ and $S \simeq U$. Then $T = C^{-1}SC$ and $S = B^{-1}UB$ where C and B are linear transformations on V. By above, $S = CTC^{-1}$, so we get $CTC^{-1} = B^{-1}UB$, then $T = C^{-1}B^{-1}UBC = (BC)^{-1}U(BC)$. This makes $T \simeq U$. ■

Definition. We say two matrices $A, B \in F^{n\times n}$ are **similar**, if there exists a matrix $S \in F^{n\times n}$ such that $A = S^{-1}BS$. We write $A \simeq B$.

Lemma 3.4.2. *The similarity of matrices is an equivalence relation.*

Proof. We have that the similarity of linear transformations is an equivalence relation, and since $\hom(V, V) \simeq F^{n\times n}$ (isomorphism), we can conclude that the similarities of matrices is isomorphic to the similarity of linear transformations. ■

Corollary. $\hom(V, V)/\simeq_T \simeq F^{n\times n}/\simeq_M$. *Where* $\simeq_T$ *and* $\simeq_M$ *are the similarity of linear transformations and the similarity of matrices respectively.*

Definition. Consider $\hom(V, V)$, and let $\simeq$ be the equivalence of similar linear transformations. We call the equivalence classes of $\hom(V, V) \simeq$ **similarity classes** of $\hom(V, V)$. Likewise, if $\simeq$ is the similarity of matrices, then the equivalence classes of $F^{n\times n}$ are called **similarity classes** of $F^{n\times n}$.

The introduction of similarity classes leads to the problem of finding whether two linear transformations (or two matrices) lie in the same class, this then leads to the so called "cononical forms", which are matrices which have a particularly nice structure. We begin their treatment now. From now on, let us also make the following convention for linear transformations:

$$T(W) \equiv TW.$$

Where W maybe another linear transformation, a subspace, or a vector.

Definition. We call a subspace $W \subseteq V$ **invariant** under a linear transformation T on V if $TW \subseteq W$.

Lemma 3.4.3. *If* $W \subseteq V$ *is invariant under a linear transformation* T *on* V*, then* T *induces a linear transformation* $\overline{T}$ *on* V/W *defined by* $T : v+W \to Tv+W$*, for* $v \in V$*. If* T *satisfies the polynomial* $q \in F[x]$*, then so does* $\overline{T}$*. If* $p_1, p \in F[x]$ *are the minimal polynomials for* $\overline{T}$ *and* T*, respectively, then* $p_1|p$.

Proof. Consider V/W whose elements are the cosets $v + W$, of W in V. Define a map $\overline{T}$ on V/W by $\overline{T} : v + W \to Tv + W$; notice that for $\alpha, \beta \in F$, and $u, v \in V$, $\overline{T}(\alpha u + v) = \alpha Tu + \beta Tv + W = \alpha\overline{T}(u + W) + \beta\overline{T}(v + w)$, so $\overline{T}$ is a linear transformation on V/W.

Now for $v_1, v_2 \in V$, suppoe that $v_1 + W = v_2 + W$, then $v_1 - v_2 \in W$ and since $TW \subseteq W$, $T(v_1 - v_2) \in W$. Thus $Tv_1 + W = Tv_2 + W$.

Noq if $q \in F[x]$ is satisfied by T, i.e. $q(T) = 0$, then notice that $q(\overline{T}(v + W)) = q(Tv)+W = W$, so $\overline{T}$ also satisfies q; in particular, if $p_0, p \in F[x]$ are the minimal polynomials of $\overline{T}$ and T, respectively, then we have $p_0|p$. ■

Definition. We call a matrix $A \in F^{n\times n}$ **triangular** if $A = (a_{ij})$ where $a_{ij} = 0$ for all $j > i$. That is all the entries above the diagonal of A are 0.

Definition. We call a linear transformation T on V **triangular** if the matrix representing T is triangular. That is T is triangular if given the basis $\{v_1, \ldots, n_n\}$ of V:

$$\begin{aligned} Tv_1 &= t_{11}v_1 \\ Tv_2 t_{21}v_1 &+ t_{22}v_2 \\ &\vdots \\ Tv_i &= t_{i_1}v_1 + t_{i_2}v_2 + \ldots t_{ii}v_i \\ &\vdots \\ Tv_n &= t_{n1}v_1 + t_{n2}v_2 + \cdots + t_{nn}v_n \end{aligned} \tag{3.6}$$

where $T = (t_{ij})$.

Theorem 3.4.4. *If T is a lineaer transformation on V with its Eigenvalues in F, then there is a basis of V in which the matrix of T is trangular.*

Proof. By induction on $\dim V$, if $\dim V = 1$, then every linear transformation on V is a scalar, and we are done.

Now suppose the theorem is true for vector spaces of $\dim = n$, and suppose that $\dim V = n + 1$. Let $\lambda_1 \in F$ be an Eigenvalue of T, then ther is a $v_1 \neq 0 \in V$ for which $Tv_1 = \lambda_1 v_1$. Take, then $W = \{\alpha v_1 : \alpha \in F\}$. Then W is a subspace of V and $\dim W = 1$, thus $\dim V/W = (n+1) - 1 = n$. By lemma 3.4.3, T "induces" a linear transformation on V/W, whose minimal polynomial p_0 divides the minimal polynomial p of T. That is the roots of p_0 are alos roots of p, and so the roots of p lie in F, so by hypothesis, there is a basis $\{\overline{v}_2, \ldots, \overline{v}_{n+1}\}$ of V/W over F such that:

$$\begin{aligned} T\overline{v}_2 &= \alpha_{22}\overline{v}_2 \\ &\vdots \\ T\overline{v}_i &= \alpha_{i2}\overline{v_2} + \cdots + \alpha_{ii}\overline{v}_i \\ &\vdots \\ T\overline{v}_{n+1} &= \alpha_{n+12}\overline{v}_2 + \cdots + \alpha_{n+1n+1}\overline{v}_{n+1} \end{aligned}$$

Now take $v_1, \ldots, v_{n+1} \in V$ and map them into $\overline{v}_2, \ldots, \overline{v}_{n+1}$, respectively. Then $\{v_1, \ldots, v_{n+1}\}$ forms a basis for V; moreover, $Tv_2 - \alpha_{22}v_2 \in W$, so $Tv_2 - \alpha_{22}v_2$ is a multiple of v_1, similarly, $Tv_i - \alpha_{1=i2}v_2 - \cdots - \alpha_{ii}v_i \in W$, thus the basis $\{v_1, \ldots, v_{n+1}\}$ provides us with a basis where all Tv_i is a linear combination of the v_i, and its predecessors (i.e. all v_j where $j < i$). Therefore the matrix of T in this basis is trangular. ■

Theorem 3.4.5. *If the matrix $A \in F^{n\times n}$ has all its Eigenvalues in F, then there is a matrix $C \in F^{n\times n}$ similar to A.*

Definition. If T is a linear transformations satisfying the condition in theorem 3.4.4, then we say that T can be brought to **triangular form** over F. Similarly if $A \in F^{n\times n}$ satisfies theorem 3.4.5, then we say that A can be brought to **triangular form** over F.

Theorem 3.4.6. *If* $\dim_F V = n$ *and if* T *is a linear transformation on* V *which has all its Eigenvalues in* F*, then* T *satisfies a polynomial of* $\deg = n$ *over* F*.*

Proof. By the aboce theorem, there is a basis $\{v_1, \ldots, v_n\}$ of V over F with $Tv_i = \alpha_{i1}v_1 + \cdots + \alpha_{nn}v_n$ for all $1 \leq i \leq n$. Equivalently, $(T - \lambda_i)v_i = \alpha_{i1}v_1 + \cdots + \alpha_{ii-1}v_{i-1}$. We can obtain then, via computation that $(T - \lambda_1)\ldots(T - \lambda_i) = 0$ for all $1 \leq i \leq n$; in particular, the matrix $S = (T - \lambda_1)\ldots(T - \lambda_n)$, satisfies $Sv_1 = \cdots = Sv_n = 0$, so S annaihilates the basis $\{v_1, \ldots, v_n\}$. Therefore S annihilates all of V, thus $S = 0$. Consequently, T satisfies the following polynomial $p(x) = (x - \lambda_1)\ldots(x - \lambda_n)$ where $\deg p = n$. ■

Example 3.6. Let $F = \mathbb{R}$ and let

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

Then A has the minimal polynomial $x^2 + 1$, which means that A has no Eigenvalues in $\mathbb{R}$, hence it cannot be brought to triangular form. However if we take $F = \mathbb{C}$, then A can be brought to triangular form as its Eigenvalues are $i, -i \in \mathbb{C}$.

The example above shows that not every matrix, or linear transformation has its Eigenvalues in the ground field F, and hence, not every matrix can be brought to triangular form. However notice that in the above example, A has its Eigen values in $\mathbb{C} = \operatorname{cl}\mathbb{R}$, the algebraic closure of $\mathbb{R}$. So it is possible that if a matrix A does not have all of its Eigenvalues in F, it may indeed have them in $\operatorname{cl} F$.

Lemma 3.4.7. *Let* V *be a vector space of* $\dim V = n$ *over a field* F*. Let* $F \subseteq K$ *be an extension of* F*, then there is a vector space* V_K *of* $\dim_K V_K = n$ *such that* $V \subseteq V_K$*.*

Proof. If $F \subseteq K$ is any extension field, then notice that $F^{n\times n} \subseteq K^{n\times n}$, so any $n \times n$ matrix over F can be considered as an $n \times n$ matrix over K. Let $T \in F^{n\times n}$ have the minimal polynomial $p \in F[x]$ over F. If we take $T \in K^{n\times n}$, then there is a polynomial $p_0 \in K[x]$ for which p_0 is the minimal polynomial of T over K ; then $p_0|p$. Therefore there is a finite extension K of F in where the minimal polynomial p of T over F has all its roots in K. Therefore, T has all its characteristic roots in K. ■

Corollary. *Let* F *be a field and let* $T \in F^{n\times n}$*. If* $p \in F[x]$ *is the minimal polynomial of* T *over* F*, then* T *has all its characteristic roots in* $\operatorname{cl} F$*.*

Remark. It should be noted that although we largely consider roots in the algebraic closure, $\operatorname{cl} F$ of F, that lemma 3.4.7 was proved for any arbitrary extension field.

One class of linear transformations with all their characteristic roots in F is the class of **nilpotent** transformations, which have all their characteristic roots 0. So these nilpotent transformations can aloways be brought to triangular form over F.

Definition. Let V be a finite dimensional vector space over a field F. We call a linear transformation T on V **nilpotent** if all its characteristic roots are 0.

Lemma 3.4.8. *Let V be a vector space for dimension $\dim V = n$ and let $V_i \subseteq V$ be a subspace of dimension $\dim V_i = n_i$ for $1 \leq i \leq k$. If $V = V_1 \oplus \cdots \oplus V_k$ and is invariant under a liniear transformation T on V, then a basis of V can be found so that the matrix of T in this basis has the form:*

$$\begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & A_k \end{pmatrix} \tag{3.7}$$

Proof. Choose the basis of V such that $\{v_1^{(i)}, \dots, v_{n_i}^{(i)}\}$ is a basis of V_i for each $1 \leq i \leq k$. Since V_i is invariant under T, we have $TV_i \subseteq V_i$, so $Tv_j \subseteq V_i$ making it a linear combinattion of the basis elements $\{v_j^{(i)}\}_{j=1}^{n_i}$, and of only these. Then T under this basis is of the form. That each A_j is the matrix of Tv_j, this makes the matrix with all A_j in the diagonal the matrix of a linear transformation. ■

Lemma 3.4.9. *If T is a nilpotent transformation on an n-dimensional vector space V over a field F, then $\alpha_0 + \alpha_1 T + \cdots + \alpha_m T^m$, with $\alpha_i \in F$ for all $1 \leq i \leq n$ is invertible if $\alpha_0 \neq 0$.*

Proof. If S is nilpotent and $\alpha_0 \neq 0$, then:

$$(\alpha_0 + S)(\frac{1}{\alpha_0} - \frac{S}{\alpha_0^2} + \cdots + (-1)^r \frac{S^{r-1}}{\alpha_0^r}) = 1$$

If $S^r = 0$. Now if $T^r = 0$, then $S = \alpha_1 T + \cdots + \alpha_m T^m$, and so $S^r = (\alpha_1 T + \cdots + \alpha_m T^m)^r = 0$, by the binomial theorem, which makes $\alpha + S$ invertible. ■

Define $M_t = (m_{ij})$ the $t \times t$ matrix where:

$$m_{ij} = \begin{cases} 1, & \text{if } i = j + 1 \\ 0 & \text{everywhere else.} \end{cases}$$

i.e. M_t has all its entries 1 on the superdiagonal (the diagonal above the diagonal of the matrix.).

Definition. Let V be a finite dimensional vector space over a field F and let T be a nilpotent transformation on V. We call an integer $k \in Z^+$ the **index of nilpotence** (or the **nilpotent index**) if $T^k = 0$ but $T^{k-1} \neq 0$; that is it is the minimum such positive integer for which $T^k = 0$.

Lemma 3.4.10. *Let V be a finite dimensional vector space with subspace $W \subseteq V$, and let T be a nilpotent transformation on V. If for some $n \in \mathbb{Z}^+$ there is a $w \in W$ such that $T^{n-k} w = 0$, where k^+, then $w = T^k w_0$ for some $w_0 \in W$.*

Proof. Since $v_1 \in V_1$, take $u = \alpha_1 v_1 + \alpha_2 T v_2 + \cdots + \alpha_{k+1} T v_{k+1} + \alpha_{n_1} T v_{n_1}$. Then, $T^{n_1 - k} i = 0$, however $T^{n_1 - k} v, \dots, T^{n_1 - 1} v$ are linearly indepenent over F, thus $\alpha_1 = \cdots = \alpha_{n_1} = 0$; so $u = T^k u_0 \in V_1$. ■

Lemma 3.4.11. *Let V be a finite dimensional vector space and let T be a nilpotent transformation on V. There exists a subspace $W \subseteq V$ invariant undet T such that $V = V_1 \oplus W$, for some other subspace $V_1 \subseteq V$.*

Proof. Let W be a subspace of V of largest pissible dimension such that $V_1 \cap W = \langle 0 \rangle$ and W is invariant under T. Suppose then that $V \neq V_1 + W$, then there is an element $z \in V$ for which $z \notin V_1 + W$. Since $T^{n_1} = 0$, ther is a $0 \leq k \leq n_1$ such that $T^k z \in V_1 + W$ byt $T^i z \notin V_1 + W$, for all $i < k$. Thus $T^k z = v + w$ where $v \in V_1$ and $w \in W$, but then $T^{n_1} z = T^{n_1-k} T^k z = T^{n_1-k} v + T^{n_1-k} w = 0$. Since both V_1 and W are invariant under T we have both $T^{n_1-k} u, T^{n_1-k} w \in V_1, W$, repspectively. Thus since $V_1 \cap W = \langle 0 \rangle$, we get that $T^{n_1-k} u = -T^{n_1-k} w \in V_1 \cap W = 0$, thus $T^{n_1-k} u = 0$. By lemma 3.4.10, $u = T^k u_0$ for some $u_0 \in V_1$; therefore $T^k z = u + w = T^k u_0 + w$. Let $z_1 = z - u_0$, then $T^k z_1 = T^k z - T^k u_0 = w \in W$. Thus by the invariance of W, we have $T^m z_1 \in W$ for all $k \leq m$. On the other hand, $i < k$, and $T^i z_1 \notin V_1 + W$, for otherwise $T^i z \in V_1 + W$, contradicting our choice of k.

Low let W_1 be the subspace of V spanned by W, and $z_1, Tz_1, \ldots T^{k-1} z_1$. Since $z_1 \notin W$ and $W \subseteq W_1$, $\dim W_1 \geq \dim W$. Moreover, since $T^k z \in W$ and by the invariance of W under T, $TW_1 \subseteq W_1$, making W_1 invariant under T. By the maximality of W, there is an element of teh form $w_0 + \alpha_1 z_1 + \cdots + \alpha_k T^{k-1} z_1 \neq 0$ in $V_1 \cap W$, with $w_0 \in W$. Now not all α_i can be 0, otherwise $w_0 \neq 0 \in \langle 0 \rangle$, which cannot happen. So let α_s be the first nonzero α, then $w_0 + T^{s-1} z_1 + (\alpha_s + \alpha_{s+1} T + \cdots + \alpha_k T^{k-s}) \in V_1$. By the above lemma, we have that $\alpha_s + \alpha_{s+1} T + \cdots + \alpha_k T^{k-s}$ is invertible, and its inverse, R is a polynomial in T. Thus W and V_1 are invariant under R; however, we also have $Rw_0 + T^{s-1} z_1 R V_1 \subseteq V_1$, this forces $T^{s-1} z_1 \in V_1 + RW$, and since $s - 1 < k$, this is impossible. Thereofre $V = V_1 + W$, because $V_1 \cap W = \langle 0 \rangle$, this makes $V = V_1 \oplus W$. ∎

Theorem 3.4.12. *Let V be a finite dimensional vector space over a field F and let T be a nilpotent transformation on V with nilpotent index n_1. Then a basis of V can be found such that the matrix of T in this basis has the form:*

$$\begin{pmatrix} M_{n_1} & 0 & \ldots & 0 \\ 0 & M_{n_2} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & M_{n_r} \end{pmatrix} \tag{3.8}$$

Where $n_r \leq \ldots n_2 \leq n_1$ and $\dim V = \sum n_i$.

Proof. By definition of the nilpotent index, there is a $v \in V$ for which $T^{n_1-1} v \neq 0$. Now consider the vectors $Tv, \ldots, T^{n_1-1} v$, and let $\alpha_1 v + \alpha_2 Tv + \cdots + \alpha_{n_1} T^{n_1-1} v = 0$; and let α_s be the first nonzeo α, then,

$$T^{s-1} v = (\alpha_s + \alpha_{s+1} T + \cdots + \alpha_{n_1} T^{n_1-s}) = 0$$

and since $\alpha_s \neq 0$, then by lemma 3.4.9, $\alpha_s + \alpha_{s+1} T + \cdots + \alpha_{n_1} T^{n_1-s}$ is invertible, so $T^{s-1} v = 0$; however, since $s < n_1$ this contradicts that $T^{n_1-s} v \neq 0$, thus no such α_s exists making $\{Tv, \ldots T^{n_1-1} v\}$ linearly indepenent over F.

Now let V_1 be the subspace spanned by $\{v, Tv, \ldots, T^{n_1-1}v\}$; then V_1 is invariant under T, and the linear transformation on V_1 induced by T has the matrix M_{n_1}.

Now, by lemma 3.4.11, there is a subspace $W \subseteq V$ invariant under T, for which $V = V_1 \oplus W$. Using the basis $\{v, Tv, \ldots, T^{n_1-1}v\}$ of V_1, and appending any basis of W to get a basis of V, we have that the matrix of T has the form:

$$\begin{pmatrix} M_{n_1} & 0 \\ 0 & A_2 \end{pmatrix}$$

by lemma 3.4.8. Here, A_2 is the matrx of T_2, the linear transformation on W induced by T. Then $T^{n_1} = 0$, $T_2^{n_2} = 0$, for some $n_2 \leq n_1$. Repeating the argument and taking $W = V_2 \oplus W_2$, and so on, we gaet a basis of V for which the matrix of T in V is of the form found in equation (3.8). ■

Definition. Let V be a fintite dimensional vector space over a field F and let T be a nilpotent transformation. Let $V_i \subseteq V$ be a subspace of V for $1 \leq i \leq r$ and let $\{T_i\}_{i=1}^r$ be a collection of nilpotent transformations induced on V_i, respectively by T, with indices of nilpotnence $\{n_i\}_{i=1}^r$, respectivly. Then we call each n_i an invariant of the nilpotent transformation T.

Definition. Let V be a finite dimensional vector space and let T be a nilpotent transformation on V, we call a subspace $M \subseteq V$ of dimension $\dim M = m$, invariant under T, **cyclic** with respect to T if:

(1) $T^m M = \langle 0 \rangle$, and $T^{m-1} M \neq \langle 0 \rangle$.

(2) There is a $z \in M$ such that $z, Tz, \ldots, T^{m-1}z$ form a basis of M.

Lemma 3.4.13. *Let V be a finite dimensional vector space with nilpotent transformation T. If $M \subseteq V$ is cyclic with respect to T, then $\dim T^k M = m - k$ for all $k \leq m$.*

Proof. We can get a basis of $T^k M$ by taking the image of any basis of M under T^k. Let $\{z, Tz, \ldots, T^{m-1}z\}$ be this basis. Then we can get a basis $\{T^k z, T^{k+1}z, \ldots, T^{m-1}z\}$ of M. This basis has $m - k$ elements, and so we are done. ■

Theorem 3.4.14. *Two nilpotent transformations are similar if, and only if they have the same invariants.*

Proof. Let T be a nilpotent transformation on V, then we can find $n_i \in Z^+$ such that $n_1 \geq n_2 \geq, \ldots, n_r$ and subspaces of V, $V_1, \ldots, V_r$, cyclic with respect to T and of $\dim V_i = n_i$ for all $1 \leq i \leq n$, such that $V = V_1 \oplus \cdots \oplus V_r$.

Suppose we can also find $m_j \in Z^+$ and $U_j \subseteq V$ for which $n_{j+1} \geq n_j$ and $\dim U_j = n_j$, and where U_i is cyclic with respect to T such that $V = U_1 \oplus \cdots \oplus U_s$. This is, of course for all $1 \leq j \leq s$. Suppose as well that $s \neq r$ and all m_j are distinct from n_i. Then assume that $m_j < n_i$.

Cinsider then $T^{m_j}V$, since $V = \bigoplus V_i$, we have $T^{m_j}V = \bigoplus T^{m_j}V_i$. Since $\dim T^{m_j}V_i = n_i - m_j$ for all i, we get $\dim T^{m_j}V \geq (n_1 - m_j) + \cdots + (n_i - m_j)$. On the other hand, since $V = \bigoplus U_j$ and since $T^{m_j}U_k = \langle 0 \rangle$ for all $k \leq j$, we have $T^{m_j}V = \bigoplus T^{m_j}U_k$. Thus:

$$\dim T^{m_j}V = (m_1 - m_j) + \cdots + (m_k - m_j)$$

so by our choice of i, j and k, $n_i = m_j$ and $r = s$. Thus

$$\dim T^{m_j} V = (n_1 - m_j) + \cdots + (n_k - m_j)$$

but this contradicts our previous value. This makes the n_i we chose unique. Therefore, the invariants of T are also unique.

Now, by the above discussion, if two nilpotent transformations have different invaraints, they cannot be similar, hence their respective matrices (both of the form in equation (3.8)) cannot be similar either.

Conversely, if S and T are two nilpotent transformations on V with the same invatants. Therefore there are invariants $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_n\}$ of V such that the matrix of S in the first basis and the matrix of T in the second basis is equal to:

$$\begin{pmatrix} M_{n_1} & 0 & \ldots & 0 \\ 0 & M_{n_2} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & M_{n_r} \end{pmatrix}$$

Now if A is the linear transformation on V defined by $A : v_i \to w_i$, then $S = A^{-1}TA$, making $T \simeq S$. ■

Remark. If we let $p(n)$ be the number of partitions of a positive integer n, then we notice that the invariants of T determine a partition of $\dim V = n$. On the otherhand, any partion of n determines the invariants of nilpotent transformation. Thus the number of distinct similarity classes of $n \times n$ nilpotent matrices is precisely $p(n)$.

Example 3.7. Let

$$T = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in F^3$$

act on F^3 with the basis $\{(1,0,0), (0,1,0), (0,0,1)\}$. Let $v_1 = (1,0,0), v_2 = T(1,0,0), v_3 = (0,0,1)$. Then in the basis $\{v_1, v_2, v_3\}$, we have that the matrix of T is:

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

So the invariants of T are 1 and 2. If we take:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

then:

$$A^{-1}TA = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Definition. Let V be a finite dimensional subspace over a field F and let $W \subseteq V$ be a subspace of V invariant under a linear transformation T on V. Let T_W be a linear transformation on W be defined by $T_W : w \to Tw$ for all $w \in W$. Then we say that T **induces** the linear transformation T_W on W, or that T_W is the linear transformation **induced** on W by T.

Lemma 3.4.15. *Let V be a finite dimensional vector space over a field F and let V_1 be a subspace of V invariant under a linear transformaion T on V. Let T_1 be the linear transformation on V_1 induced by T. If $q \in F[x]$ is a polynomial, then the linear transformation on V_1 induced by $q(T)$ is precisely $q(T_1)$; in particular, if $q(T) = 0$, then $1(T_1) = 0$.*

Remark. The last statement of this lemma means that T_1 satisfies any polynomial satisfied by T.

Lemma 3.4.16. *Suppose that $V = V_1 \oplus V_2$ be a finite dimensional subspace over a field F, where $V_1, V_2 \subseteq V$ are subspaces invariant under a linear transformation T in V. Let T_1 and T_2 be linear transformations induced by T on v_1 and V_2, respectively. Then if the minimal polynomials of T_1 and T_2 over F are p_1 and p_2, respectively, then the minimal polynomial of T over F is the least common multiple $[p_1, p_2]$.*

Proof. If p is the minimal polynomial of T over F, then $p(T_1) = p(T_2) = 0$, hence $p_1|p$ and $p_2|p$. Thus $[p_1, p_2]|p$.

Now, let $q(x) = [p_1, p_2](x)$, and consider $q(T)$. For $v_1 \in V_2$, since $p_1|q$ and $p_2|q$, we have $q(T)v_1 = q(T_1)v_1 = 0$, similarly for $v_2 \in V_2$, $q(T)v_2 = 0$. Now given any $v \in V$, $v = v_1 + v_2$ for $v_1, v_2 \in V_1, V_2$, respectively. Consequently then, we have $q(T)v = q(T)v_1 + q(T)v_2 = 0$, thus T satisfies q. ∎

Corollary. *If $V = \bigoplus_{i=1}^{k} V_i$ where $V_i \subseteq V$ is a subspace invariant under T, and if T_i is the linear transformation induced by T on V_i, and p_i is the minimal polynomial of T_i; then the minimal polynomial of T over F is the least common multiple $[p_1, \ldots, p_k]$.*

Theorem 3.4.17. *Let $V = \bigoplus_{i=1}^{k} V_i$ be a finite dimensional vector space over a field F where V_i is a subspace invariant under a linear transformation T on V, for each $1 \leq i \leq k$. Let T_i be the linear transformation induced by T on V_i, then the minimal polynomial of T_i is of the form $q_i^{l_i}(x)$, where $l_i \in \mathbb{Z}^+$.*

Proof. If $k = 1$, then $V = V_1$ and $T = T_i$ and $p = q_i^1$. We are done.

Now suppose that $k > 1$, and consider the polynomials $h_i(x) = \prod_{j \neq i} q_j^{l_j}(x)$, where $1 \leq i \leq k$. Since $k > 1$, we have $h_i \neq p$, thus $h_i(T) \neq 0$. That is, there is a $v \in V$ such that $h_i(T)v = w \neq 0$; however, $q_i^{l_i}(T)w = h(T)q_i^{l_1}(T)v = p(T)v = 0$, thus $w \neq 0 \in V_i$ hence $V_i \neq \langle 0 \rangle$. Notice then, that for $v_j \in V_j$, $j \neq i$, $h_i(T)v_j = 0$ since $p_j|h_i$.

Moreover, the polynomials h_i are all coprime, for $1 \leq i \leq k$, i.e. $(h_1, \ldots, h_k) = 1$. Therefore there are polynomials $a_1, \ldots, a_k \in F[x]$ for which $a_1h_1 + \cdots + a_kh_k = 1$. Thus $a_1(T)h_1(T) + \cdots + a_k(T)h_k(T) = 1$, hence given $v \in V$, $v = v_1 = v(a_1(T)h_1(T) + \cdots + a_k(T)h_k(T)) = a_1(T)h_1(T)v + \cdots + a_k(T)h_k(T)v$. Now, we have that each $a_i(T)h_i(T)v \in h_i(T)V$, so $h_i(T)V \subseteq V$; i.e., $h_i(T)$ is invariant under T. Therefore, $v = v_1 + \cdots + v_k$ where each $v_i = a_i(T)h_i(T)v$. Therefore $V = V_1 + \cdots + V_k$ where each $V_i = h_i(T)V$. ∎

Theorem 3.4.18. *Let V be a finite dimensional vector space over a field F, and let T be a linear transformation on V. Suppose that $p \in F[x]$ is the minimal polynomial of T over F. Then we can factor p in the following way:*

$$p(x) = q_1^{l_1}(x) \dots q_k^{l_k}(x) \tag{3.9}$$

where each of the $q_i \in F[x]$ are distinct irreducible polynomials and $l_i \in \mathbb{Z}^+$, for all $1 \leq i \leq k$.

Proof. If $k = 1$, we are done. Now suppose that $k > 1$. Let q_i be the minimal polynomial for T_i. Let $V_i = \{v \in V : q^{l_i}(V_i)v = 0\}$, for all $1 \leq i \leq k$. We can see that $V_i \subseteq V$ is a subspace of V. Now if $v \in V_i$, since $q_i(T)$ and T commute, we have $q^{l_i}(T)Tv = Tq^{l_i}(T)v = 0T = 0 \in V_i$; i.e. $TV_i \subseteq V_i$. This makes V_i invariant under T. Notice then that $V = \bigoplus V_i$ (why?), then by theorem 3.4.17 we have that $p(x) = q_1(x) = q_1^{l_1}(x) \dots q_k^{l_k}(x)$. ■

Corollary. *If the distinct Eigenvalues $\lambda_1, \dots \lambda_k$ of T lie in F, then $p(x) = (x - \lambda_1)^{l_1} \dots (x - \lambda_k)^{l_k}$.*

Remark. We make the comvention that for any polynomial $q \in F[x]$, and $l \in Z^+$, that $q^l(x) = q(x)^l = q(x) \dots q(x)$ l times. This should not be confused with the l-th derivative $q^{(l)}$ of q.

Definition. The matrix:

$$\begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 0 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & & & \dots & \lambda \end{pmatrix} \tag{3.10}$$

is called the **basic Jordan block** belonging to λ.

Theorem 3.4.19. *Let V be a finite dimensional vector space over a field F and let T be a linear transformation on V. If T has all its distinct Eigenvalues $\lambda_1, \dots, \lambda_k \in F$, then a basis of V can be found in which T has a matrix of the form:*

$$\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix} \tag{3.11}$$

where each:

$$J_i = \begin{pmatrix} B_{i1} & & & \\ & B_{i2} & & \\ & & \ddots & \\ & & & B_{ir_i} \end{pmatrix} \tag{3.12}$$

and where each B_{ij} are basic Jordan blocks belonging to λ_i, for each $1 \leq j \leq r_i$.

Proof. Notice that the $m \times m$ basic Jordan block belonging to λ is the matrix $\lambda + M_m$. ■

Definition. Let V be a finite dimensional vector space over a field F, and let T be a linear transformation on V with a matrix J of the form in equation (3.11), where $\dim B_{i_1} \geq \dim B_i 2 \geq \cdots \geq \dim B_{ir_i}$. Then we call J the **Jordan form** of T.

Lemma 3.4.20. *Let V be a finite dimensional vector space over a field F. If T_1 and T_2 are linear transformations on V, with all their distinct characteristic roots in F, then $T_1 \simeq T_2$ if, and only if they can be brought to the same Jordan form.*

Proof. ■

Theorem 3.4.21. *Let F be a field and let $A \in F^{n\times n}$, and suppose that K is the splitting field of the minimal polynomial of A over F. Then an invertible matrix $C \in^{n\times n}$ can be found so that $C^{-1}AC$ is in Jordan form.*

Lemma 3.4.22. *Let V be a finite dimensional vector space over a field F and let T be a linear transformation on V. Suppose T has the minimal polynomial: $p(x) = \gamma_0 + \gamma_1 x + \cdots + \gamma_{r-1}x^{r-1} + x^r$, and that V, as a module, is cyclic relatie to T. Then there is a basis of V over F such that the matrix of T in this basis has the form:*

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -\gamma_0 & -\gamma_1 & -\gamma_2 & \cdots & -\gamma_{r-1} \end{pmatrix} \tag{3.13}$$

Proof. ■

Definition. Let F be a field. If $f \in F[x]$ has the form:

$$f(x) = \gamma_0 + \gamma_1 x + \cdots + \gamma_{r-1}x^{r-1} + x^r \tag{3.14}$$

Then the $r \times r$ matrix $Cf \in F^{r\times r}$, of the form:

$$Cf = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -\gamma_0 & -\gamma_1 & -\gamma_2 & \cdots & -\gamma_{r-1} \end{pmatrix} \tag{3.15}$$

is called the **companion matrix** of f.

Remark. The companion matrix can also be written as $C(f(x))$, however we write it Cf to denote C as a linear operator acting on f; i.e. it is a linear transformation on $F[x]$.

Remark. We also make note that in lemma 3.4.22, if V is cyclic relative to T, and if the minimal polynomial in $F[x]$ of T is of the form found in equation (3.14), then there is a basis of V for which the matrix of T is the companion matrix for p; i.e. $m(T) = Cp$.

Also notice that for any $f \in F[x]$ monic, Cf satisfies f and has f as its minimal polynomial. That is $f(Cf) = 0$.

Theorem 3.4.23. *Let V be a finite dimensional vector space over a field F, and let T be a linear transformation on V. If T has the minimal polynomial $p(x) = q^e(x)$, where $q \in F[x]$ is monic irreducible; then a basis of V over F can be found such that the matrix of T in this basis has the form:*

$$\begin{pmatrix} Cq^{e_1} & & & \\ & Cq^{e_2} & & \\ & & \ddots & \\ & & & Cq^{e_r} \end{pmatrix} \tag{3.16}$$

where $e = e_1 \geq e_2 \geq \cdots \geq e_r$, and Cq^{e_i} is the companion matrix for q^{e_i} for all $1 \leq i \leq r$.

Proof. ■

Corollary. *If T has the minimal polynomial $p(x) = q_1^{n_1}(x) \dots q_k^{n_k}(x)$ over F, where each q_i are distinct irreducible polynomials over F, then there is a basis of V for which the matrix of T has the form:*

$$\begin{pmatrix} R_1 & & & \\ & R_2 & & \\ & & \ddots & \\ & & & R_k \end{pmatrix} \tag{3.17}$$

where

$$\begin{pmatrix} Cq_i^{e_{i1}} & & & \\ & Cq_i^{e_{i2}} & & \\ & & \ddots & \\ & & & Cq_i^{e_{ir_i}} \end{pmatrix} \tag{3.18}$$

where $e_i = e_{i1} \geq e_{i2} \geq \cdots \geq e_{ir_i}$.

Definition. Let F be a field. If a matrix $T \in F^{r \times r}$ can be brought to the form of equation (3.17), then we say that it can be brough to **rational form** over F. We call the matrix in (3.17) the **rational canonical form** of T.

Definition. Let F be a field and let $T \in F^{r \times r}$. We call the polynomials $q_1^{e11}, \dots, q_k^{e_{k1}} \in F[x]$ that make up the minimal polynomial of T the **elementary divisors** of T.

Definition. Let V be a finite dimensional vector sapce over a field F. If $\dim V = n$, then the **characteristic polynomial** of a linear transformation T on V is the product of its elementary divisors.

Remark. We refer to the elementrayr divisors of a linear transformation by making note that matrices are just representations of linear transformations.

Lemma 3.4.24. *Let V be a finite dimesnional vector space over a field F and let T be a linear transformation on V. If $p_T \in F[x]$ is the characteristic polynomial of T, then the Eigenvalues of T are roots of p_T and T satisfies p_T; i.e. $p_T(T) = 0$. Conversly, every root of p_T is an Eigenvalue of T; in particular the multiplicity of any root of p_T equals the multiplicity of that root as an Eigenvalue of T.*

Proof. ■

Theorem 3.4.25. *Let V and W be finite dimensional vector spaces over a field F, and suppose that $\psi : V \to W$ is an isomorphism of V onto W. Let S and T be linear transforamtions on V and W, respectively, such that $\psi(Sv) = T\psi(v)$, for any $v \in V$. Then S and T have the same elementary divisors.*

Proof. ■

Theorem 3.4.26. *Let V be a finite dimensional vector space. If T and S are linear transformations on V, then $T \simeq S$ if, and only if they have the same elementary divisors.*

Proof. ■

Corollary. *Let F be a field. If $A, B \in F^{n\times n}$ such that $A \simeq B$ in $L^{n\times n}$, where K is an extension of F, then $A \simeq B$ in F.*

Proof. ■

3.5 The Trace and Tranpose.

Definition. Let F be a field and let $A \in F^{n\times n}$ be an $n \times n$ matrix. We define the **trace** of A to be the sum of the elements on the diagonal. That is:

$$\operatorname{trace} A = \sum_{i=1}^{n} a_{ii} \tag{3.19}$$

Where $A = (a_{ij})$.

Bibliography

[1] I. N. Herstein, *Topics in algebra*. New York: Wiley, 1975.

[2] K. Hoffman and R. Kunze, *Linear algebra*. Englewood Cliffs, NJ: Prentice-Hall, 1971.