More on Vector Spaces and Linear transformations

1.1 Let X be a vector space over the field $\mathbb{F}$ and given a set I, let us write

$$X^I := I \xrightarrow{x} X \text{ is a function.}$$

For $x \in X^I$, we shall write $x(i) \in X$ as $x^i \in X$ at each $i \in I$. For $x, u \in X^I$, $\lambda \in \mathbb{F}$, define $x + u\lambda \in X^I$ by $(x + u\lambda)^i := x^i + u^i\lambda$.

Verify that this turns X^I into a vector space(over $\mathbb{F}$).

1.2 Write $(X^I)_0 := \{x \in X^I | x^i = 0 \text{ for all but finitely many } i \in I\}$

Verify that $(X^I)_0$ is a vector subspace of X^I.

1.3 Take $X = \mathbb{F}$ and show that I is (in bijective correspondence with) a basis of $(\mathbb{F}^I)_0$

Hint:

Define functions $I \xrightarrow{e_i} \mathbb{F}$ by $e_i(j) = \delta_{ij} = \begin{cases} 1, if & i = j; \\ 0, if & i \neq j. \end{cases}$

Then each $e_i \in (\mathbb{F}^{\mathbb{I}})_0$ and if $i \neq j$, we have $e_i(i) = 1$ while $e_j(i) = 0$ so that $e_i \neq e_j$. Thus $I \xrightarrow{e} (\mathbb{F}^I)_0$ defined by $e(i) := e_i$ is injective which means that $I \xrightarrow{e} e(I) \subseteq (\mathbb{F}^I)_0$ is bijective. We thus identify $i \in I$ with $e_i \in (\mathbb{F}^I)_0$. If $x \in (\mathbb{F}^I)_0$, we have $x^i = 0$ for all but finitely many $i \in I$, let us say $x^{k_1}, \ldots, x^{k_n}$ are these nonzero elements of $\mathbb{F}$. Then $(e_{k_i} x^{k_i})(i) = (e_{k_i}(i))x^{k_i} = \begin{cases} x^{k_i}, if & k_i = i; \\ 0, if & k_i \neq i. \end{cases}$

And thus $x^i = x(i) = \sum_{l=1}^{n} (e_{k_l} x^{k_l})(i)$ at each $i \in I$ which means $x = \sum_{l=1}^{n} e_{k_l} x^{k_l}$ i.e. each $x \in (\mathbb{F}^I)_0$ can be expressed as a finite linear combination of distinct vectors $e_{k_l} \in e(I) \subseteq (\mathbb{F}^I)_0$ in exactly one way.

Thus the set $e(I)$ is a basis of $(\mathbb{F}^I)_0$, and since we agreed to identify I with $e(I)$, via $I \xrightarrow{e} e(I)$, we can say I is a basis of $(\mathbb{F}^I)_0$; this of course means that dim $(\mathbb{F}^I)_0 = |I|$ where $|I|$ is the (finite or infinite) number of elements of I. the vector space $(\mathbb{F}^I)_0$ is called the vector space freely generated by I.

To sum up:

Given any finite or infinite number d, take a set I with $|I| = d$. The vector space $(\mathbb{F}^I)_0$ has dimension

d and $e(I) \subseteq (\mathbb{F}^I)_0$ supplies one explicit basis for $(\mathbb{F}^I)_0$; this will also be expressed by saying that $(\mathbb{F}^I)_0$ is freely generated by I as well as by saying that I is a basis for $(\mathbb{F}^I)_0$.

1.4 As instances of (1.3) above,

(i) The vector space of dimension 0 is $\{0\}$, freely generated by ϕ and ϕ is a basis.

(ii) The vector space of dimension 1 is $\mathbb{F}$, freely generated by a singleton $\{1\}$ with $e_1 = 1 \in \mathbb{F}$ as a basis.

(iii) The vector space of dimension $n \in \mathbb{N}$ is $\mathbb{F}^n$, freely generated by some n-element set $\{1, 2, \ldots, n\}$, with $\left\{ e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} \\ \\ \longrightarrow i-th\ row \\ \\ \\ \end{bmatrix} \right\}$ as a basis. We have, for $x \in \mathbb{F}^n$, $x = \sum_{i=1}^{n} e_i x^i = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}$, $x^i \in \mathbb{F}$ obtained for a unique $n-$tuple $(x^1, \ldots, x^n)$ of scalars.

Note that $\{1, 2, \ldots, n\} \xrightarrow{e_i} \mathbb{F}$, $e_i(j) = \delta_{ij}$ for $1 \leq i \leq n$ and $1 \leq j \leq n$ completely described by the column $n \times 1$ display as above; further, that for $n = 0$ we have (i), for $n = 1$, we have (ii).

1.5 A different display style is in vogue for the case $I = \mathbb{N} = \{0, 1, 2, \ldots\}$. $\mathbb{N} \xrightarrow{x} X$, is of course just a sequence $(x(0), x(1), \ldots)$ of vectors $x(i) \in X$; they will be written x_i now and the sequence will be displayed as $x = \sum_{n=0}^{\infty} x_n \theta^n$. The vector $X^{\mathbb{N}}$ will be denoted by $X[\theta]$ with $(X^{\mathbb{N}})_0$ being denoted by $X[\theta]$. Thus for $x,\ u\ \in X[\theta],\ \lambda \in \mathbb{F}$, we write $x + u\lambda = \sum_{n=0}^{\infty} (x + u\lambda)_n \theta^n = \sum_{n=0}^{\infty} (x_n + u_n \lambda)\theta^n = \sum_{n=0}^{\infty} x_n \theta^n + \sum_{n=0}^{\infty} u_n \lambda \theta^n$

For $X = \mathbb{F}$ therefore, an element $\alpha \in \mathbb{F}[\theta]$ is written as $\alpha = \sum_{n=0}^{\infty} \alpha_n \theta^n$; this will be called a power series in the *(single)* indeterminate θ *(one also says 'variable' for 'indeterminate')* with coefficient $\alpha_n \in \mathbb{F}$.

An element $a = \sum_{n=0}^{m} a_n\theta^n \in (\mathbb{F}^{\mathbb{N}})_0$ will be called a polynomial of degree m if $a_m \neq 0$; the zero polynomial for which $a_n = 0$ at each n will be said to be of degree $-\infty$ *(some authors would say the degree is 'undefined').* We note that each nonzero $\lambda \in \mathbb{F}$ can be regarded as a polynomial of degree 1; when we wish to think of $\lambda \in \mathbb{F}$ as $\lambda\mathbb{F}[\theta] = (\mathbb{F}^{\mathbb{N}})_0$ in this way, we say it is a constant. Thus $\mathbb{F}$ is a subspace of $\mathbb{F}[\theta]$ which is in turn a subspace of $\mathbb{F}[\theta]$.

An explicit basis of $\mathbb{F}[\theta]$ is obtained by the set $\{1, \theta, \theta^2, \ldots\}$. Thus dimension of $\mathbb{F}[\theta]$ is not finite.

Verify that $P_n[\theta]$, the collection of polynomials of degree at most n, (thus $a = a(\theta) \in P_n[\theta]$ iff $a = a_0 + a_1\theta + \ldots + a_m\theta^m$ with $0 \leq m \leq n$) is a subspace of $\mathbb{F}[\theta]$ and that $\{1, \theta, \theta^2, \ldots, \theta^n\}$ is a basis of $P_n[\theta]$. Thus dim $P_n[\theta] = n + 1$ and as such, $P_n[\theta] \cong \mathbb{F}n + 1$.

1.6 (i) Given $\alpha = \sum_{n=0}^{\infty} \alpha_n\theta^n \in \mathbb{F}[\theta]$, and $x = \sum_{n=0}^{\infty} x_n\theta^n \in X[\theta]$, define $x\alpha := \sum_{n=0}^{\infty} (x\alpha)_n\theta^n$ where $(x\alpha)_n := \sum_{p+q=n} x_p\alpha_q$. Verify that $x\alpha \in X[\theta]$.

Hint:

For any $n \in \mathbb{N}$, only finitely many $p, q \in \mathbb{N}$ exist with $p + q = n$ so that $\sum_{p+q=n} x_p\alpha_q$ is a finite sum of vectors in X and thus is in X; let us remind ourselves that each $x_p\alpha_q \in X$ since $x_p \in X,\ \alpha_q \in \mathbb{F}$.

(ii) In particular, with $\alpha = \sum_{n=0}^{\infty} \alpha_n\theta^n \in \mathbb{F}[\theta]$ and $\beta = \sum_{n=0}^{\infty} \beta_n\theta^n \in \mathbb{F}[\theta]$, we obtain $\gamma = \sum_{n=0}^{\infty} \gamma_n\theta^n \in \mathbb{F}[\theta]$, where $\gamma_n = (\alpha\beta)_n := \sum_{p+q=n} \alpha_p\beta_q \in \mathbb{F}$ and thus $\gamma \in \mathbb{F}[\theta]$; we say that γ is the convolution of α and β. Verify that with convolution as multiplication, $\mathbb{F}[\theta]$ becomes a ring with identity *(what is the identity?)* of which $\mathbb{F}[\theta]$ is a subring *(well known to you as the 'polynomial ring').* Further, verify that these are commutative rings.

Hint:

Repeat what you did for the polynomial ring $\mathbb{F}[\theta]$ earlier; the same proof works for $\mathbb{F}[\theta]$ as well.

1.7 For vector spaces X and Y over $\mathbb{F}$, we shall also say $X \xrightarrow{A} Y$ is $\mathbb{F}$ linear instead of simply saying A is 'linear' when we wish to emphasize the role of $\mathbb{F}$.

(i) Verify that $X[\theta] \xrightarrow{\theta} Y[\theta]$ defined by $\theta(x) := x\theta$ (that is, $x = \sum_{n=0}^{\infty} x_n\theta^n \mapsto \sum_{n=0}^{\infty} x_n\theta^{n+1} = x\theta$) is $\mathbb{F}-$linear. *(Since it shifts the sequence $(x_0, x_1, \ldots)$ to $(0, x_0, x_1, \ldots)$, it is called the 'right shift operator').*

Thus if $X[\theta] \xrightarrow{A} Y[\theta]$ is $\mathbb{F}-$linear, it raises two $\mathbb{F}-$ linear transformations:

$X[\theta] \xrightarrow{A\theta} Y[\theta] = X[\theta] \xrightarrow{\theta} X[\theta] \xrightarrow{A} Y[\theta]$,

and $X[\theta] \xrightarrow{\theta A} Y[\theta] = X[\theta] \xrightarrow{A} Y[\theta] \xrightarrow{\theta} Y[\theta]$,

(ii) Show that $A\theta = \theta A$ iff $A_\theta k = \theta^k A$ for each $k \in \mathbb{N}$.

Hint:

Clearly, for $k = 0$, the result holds with θ^0 being identity. If $A\theta^k = \theta^k A$, we have $(A\theta^k)\theta = \theta(A\theta^k) = \theta(\theta^k A)$ i.e. $A\theta^{k+1} = \theta^{k+1}A$.

(iii) Prove that for $X[\theta] \xrightarrow{A} Y[\theta]$, the following are equivalent:

(a) A is $\mathbb{F}[\theta]-$linear (i.e. $A(x + u\alpha) = A(x) + (A(x))\alpha$ for $x, u \in X[\theta]$ and $\alpha \in \mathbb{F}[\theta]$; this does not mean of course that $X[\theta] and Y[\theta]$ are vector spaces over $\mathbb{F}[\theta]$ since $\mathbb{F}[\theta]$ is not a field but does mean that they are 'modules' over the ring $\mathbb{F}[\theta]$, a concept which we do not cover in this course).

(b) A is $\mathbb{F}[\theta]-$linear (i.e. $A(x + ua) = A(x) + A(u)a$ for $x, u \in X[\theta]$ and $a \in \mathbb{F}[\theta]$, again, $\mathbb{F}[\theta]$ is not a field).

(c) A is $\mathbb{F}-$linear with $A\theta = \theta A$.

Hint:

$\mathbb{F}[\theta]-$ linearity of A certainly entail the $\mathbb{F}[\theta]-$linearity of A which in turn entails the $\mathbb{F}-$linearity of A since $\mathbb{F}$ is a subring of $\mathbb{F}[\theta]$ and $\mathbb{F}[\theta]$ is a subring in turn of $\mathbb{F}[\theta]$. Further, if A is $\mathbb{F}[\theta]-$linear, we do have $A\theta = \theta A$ since $\theta \in \mathbb{F}[\theta]$ and this $(a) \Rightarrow (b) \Rightarrow (c)$ is established. If (c) is true, we have $A\theta^k = \theta^k A$ for each $k \in \mathbb{N}$ and if $a = a_0 + a_1\theta + \ldots + a_n\theta^n \in \mathbb{F}[\theta]$, we find

$$\begin{aligned}
(Aa)(x) &= A(a_0 + a_1\theta + \ldots + a_n\theta^n)(x) \\
&= (Aa_0)(x) + (Aa_1\theta)(x) + \ldots + A(a_n\theta^n)(x) \\
&= a_0A(x) + a_1(A\theta)(x) + \cdots + a_n(A\theta^n)(x) \\
&= a_0A(x) + a_1\theta(A(x)) + \cdots + a_n\theta^n A(x) \\
&= aA(x) \text{ at each } x \in A
\end{aligned}$$

Which means that A is in fact $\mathbb{F}[\theta]-$linear. Thus $(c) \Rightarrow (b)$ is also true.

Now assume (b). Given $\alpha = \sum_{n=0}^{\infty} \alpha_n\theta^n \in \mathbb{F}[\theta]$, write it as $\alpha = (\sum_{n=0}^{p-1} \alpha_n\theta^n) + (\sum_{n=0}^{\infty} \alpha_{n+p}\theta^n)\theta^p = a + \beta\theta^p$ with $a = \sum_{n=0}^{p-1} \alpha_n\theta^n \in \mathbb{F}[\theta]$, $\beta = \sum_{n=0}^{\infty} \alpha_{n+p}\theta^n \in \mathbb{F}[\theta]$ and $\theta^p \in \mathbb{F}[\theta]$. Then for $x \in X[\theta]$,

we have

$$\begin{aligned} A(x\alpha) &= A(xa + x\beta\theta^p) && [\because\ x \in X[\theta],\ a \in \mathbb{F}[\theta] \subseteq \mathbb{F}[\theta] \\ & && \beta \in \mathbb{F}[\theta],\ \theta^p \in \mathbb{F}[\theta] \subseteq \mathbb{F}[\theta],\ \text{so that} \\ & && \beta\theta^p \in \mathbb{F}[\theta],\ \text{and for } \gamma \in \mathbb{F}[\theta], \delta \in \mathbb{F}[\theta], \\ & && \text{we have } x(\gamma+\delta) = x\gamma + x\delta] \\ &= A(xa + u\theta^p) && [\text{with } u = x\beta \in X[\theta] \text{ since } x \in X[\theta] \text{ and} \\ & && \beta \in \mathbb{F}[\theta]\ ] \\ &= A(x)a + A(u)\theta^p && [\because\ X[\theta] \xrightarrow{A} Y[\theta] \text{ is } \mathbb{F}[\theta] - \text{linear by} \\ & && \text{assumed (b)}] \end{aligned}$$

Therefore $(Ax)\alpha$

$$\begin{aligned} \text{while } (Ax)\alpha &= (Ax)(a + \beta\theta^p) \\ &= (Ax)a + (Ax)(\beta\theta^p) && [\because\ a \in \mathbb{F}[\theta], \beta\theta^p \in \mathbb{F}[\theta],\ A(x) \in Y[\theta], \\ & && \text{and for } y \in Y[\theta], \gamma, \delta \in \mathbb{F}[\theta], \\ & && y(\gamma+\delta) = y\gamma + y\delta.] \end{aligned}$$

and $A(x\alpha)$ have the same coefficients for powers k of θ upto $k < p$. But p was arbitrary, therefore $(Ax)\alpha$ and $A(x\alpha)$ have the same coefficient for all powers of θ and thus $A(x\alpha) = (Ax)\alpha$ for any $x \in X[\theta], \alpha \in \mathbb{F}[\theta]$ i.e. A is $\mathbb{F}[\theta]-$linear and (a) holds.

This proves the advertized result

$$(a) \Rightarrow (b) \Rightarrow (c).$$

2. Finding the dual basis to a given basis $\{e_1, \ldots, e_n\}$ of X will be done by applying the defining formula $\langle e^i | e_j \rangle = \delta_{ij}$. Thus suppose we want to find the basis dual to $L = \{L_0, \ldots, L_n\}$ of P_{n+1} (= the space of polynomials with degree at most n) where $\{L_i | 0 \le i \le n\}$ are the lagrange polynomials (refer to Handout-I, page (7)). If the required basis (for X') is given by $\{L^i | 0 \le i \le n\}$ then we have $P_{n+1} \xrightarrow{L^i} \mathbb{F}$ given by $\langle L^i | a \rangle := a(\lambda_i)$ $(\because\ a(\theta) = \sum_{i=0}^{n} a(\lambda_i) L_i(\theta)$ as proved on Handout-I, page (7)).

(i) Find the basis dual to $\{e_i = \theta^i | 0 \le i \le n\}$ of P_{n+1}

Hint:

Since by the Taylor formula $a(\theta) = \sum_{i=0}^{n} \frac{(D^i(a))(0)}{i!} \theta^i$, we shall have $\langle e^i | a \rangle = \frac{(D^i(a))(0)}{i!}$ for any $a = a(\theta \in P_{n+1})$.

Since we did not say $\mathbb{F} = \mathbb{R}$, you may worry about the validity of this Hint. However, the

following should make you happy.

Let us say a vector space $\mathcal{A}$ is an (associative) algebra (with an identity) iff the vectors actually form a (associative) ring (with identity) and further

$$\lambda(xu) = (\lambda x)u = x(\lambda u)$$

holds for all $x, u \in \mathcal{A}, \lambda \in \mathbb{F}$.

Example 0.1. $\mathcal{A} = L(X, X)$ is an algebra when BA for $X \xrightarrow{A,B} X$ is regarded as multiplication of matrices; because of the correspondence $A \leftrightarrow a$ between $L(X, X)$ and $\mathrm{Mat}_n(\mathbb{F})$ for $\dim X = n < \infty$, these two are actually the same.

Example 0.2. The polynomials $\mathbb{F}[\theta]$ form an algebra since they form a ring under convolution as multiplication.

Now given an algebra $\mathcal{A}$, say that a linear transformation $\mathcal{A} \xrightarrow{D} \mathcal{A}$ is a derivation if it satisfies

$$D(xu) = (Dx)u + x(Du) \text{ for all } x, u \in \mathcal{A}$$

and prove that Taylor formula copying it from your calculus text, assuming $\mathrm{char}\ \mathbb{F} = 0$.

This may still leave you unsatisfied since P_{n+1} is not an algebra: $x, u \in P_{n+1}$ does not mean $xu \in P_{n+1}$. However, convince yourself that the dual basis can still be found by the trick suggested.

If you don't believe that 'Derivation is a concept of Linear Algebra', you may assume for this problem that $\mathbb{F} = \mathbb{R}$

(i) Take $\{e_0 = 1, e_1 = 1 + \theta, e_2 = \theta + \theta^2\}$. Prove it is a basis for $P_2[\theta]$ and find the dual basis $\{e'_0, e'_1, e'_2\}$.

Hint:

Any $a(\theta) \in P_2[\theta]$ is uniquely given as $a_0 + a_1\theta + a_2\theta^2$ with $a_0 = x^0 + x^1,\ a_1 = x^1 + x^2,\ a_2 = x^2$ when we write

$$\begin{aligned} a(\theta) &= x^0 e_0 + x^1 e_1 + x^2 e_2 \\ &= (x^0 + x^1) + (x^1 + x^2)\theta + x^2\theta^2 \end{aligned}$$

Thus $X^0 = a_0 - a_1 + a_2,\ x^1 = a_1 - a_2,\ x^2 = a_2$

Since $\langle e^i | a \rangle = x^i$, we see that the required $P_2[\theta] \xrightarrow{e^i} \mathbb{F}$ are given by $\langle e^0 | a \rangle = \langle e^0 | a_0 + a_1\theta +$

$a_2\theta^2\rangle = a_0 - a_1 + a_2 = x^0$

$\langle e^1|a\rangle = x^1 = a_1 - a_2$

$\langle e^2|a\rangle = x^2 = a_2$

3. We already know that a vector space X is accompanied by its twin, X'. However, complex vector spaces, i.e. vector spaces over $\mathbb{F} = \mathbb{C}$, actually come as quadruplets. To be explicit, suppose X is a complex vector space. Then there is another scalar multiplication given by $\lambda * x := x\overline{\lambda}$ where the scalar multiplication on the RHS is the scalar multiplication in X. The abelian group X of the vectors then turns into another vector space, call it $\overline{X}$, under this new scalar multiplication. We shall say $\overline{X}$ is the 'conjugate-space' of X. Then there is naturally $(\overline{X})'$; let us write $X^* = (\overline{X})'$. So we have $X, \overline{X}, X', X^*$; these are the four vector spaces signaled by saying 'the complex vector space X'.

3.1 When we say $X \xrightarrow{A} Y$ is a linear map, we mean that it is a homomorphism of the abelian groups:

$A(x+u) = Ax + Au$ (the addition in X is preserved by A into the addition in Y), and,

$A(x\lambda) = (Ax)\lambda$ (the scalar multiplication in X is preserved by A into the scalar multiplication in Y)

Thus a linear map $\overline{X} \xrightarrow{A} Y$ would be such that

$$A(x+u) = Ax + Au \text{ and } A(\lambda * x) = (Ax)\lambda \text{ i.e. } A(x\overline{\lambda}) = (Ax)\lambda.$$

Such a map is called a semi-linear map $X \to Y$ (rather than a linear map $\overline{X} \xrightarrow{A} Y$); other names for 'semi-linear' are 'conjugate-linear' and 'anti-linear'. Thus what is called 'sesqui[$=1\frac{1}{2}$]-linear form' on X in Handout-II, is simply a bilinear form $\overline{X} \times X \xrightarrow{s} \mathbb{C}$.

Note:

If $X \xrightarrow{A} Y \xrightarrow{B} Z$ is a scheme of conjugate-linear maps,$X \xrightarrow{BA} Y$ is linear ($BA(x+u\lambda) = B(\overline{\lambda}A(u) + A(x)) = BA(x) + BA(u)\lambda$)

To sum up:

In a linear map, the scalar emerges unscathed, in a conjugate-linear map, the scalar is flipped into its conjugate.

3.2 We equip $\overline{X}$ with the inner product with $\langle x|u\rangle := \langle u|x\rangle$ (the RHS being the inner-product in X)

Since we have $\langle u|x\rangle = \overline{\langle x|u\rangle}$, we obtain

$\langle \lambda x|u\mu\rangle = \langle \mu u|\lambda x\rangle$ $(=\overline{\langle \lambda x|u\mu\rangle})$

$= \overline{\mu}\langle u|x\rangle\lambda$

i.e. the inner product $\langle -|-\rangle_*$ is linear in the first variable and conjugate-linear in the second variable while the inner product on X was conjugate-linear in the first and linear in the second variable) if considered as an inner product on X; note however, that the definition is quite consistent since we do have

$$\langle \lambda x|u\mu\rangle_* = \overline{\mu}\langle u|x\rangle\lambda = \overline{\lambda} * (\langle x|u\rangle_*) * \mu$$

So that with the scalar multiplication as defined on $\overline{X}$, we indeed have conjugate-linearity in the first variable and linearity in the second variable.

Note:

It is important to be clear about these things. Thus some text-books assert that the import of the Riesz Representation theorem is to say that if X is a Hilbert space, it is isomorphic to its dual (= transpose) X'; this is false since the bijection displayed is conjugate-linear and not linear.

4. Prove that $\mathbb{R}$ is a vector space of infinite dimension over $\mathbb{F} = \mathbb{Q}$.

Hint:

Since $\pi \in \mathbb{R}$ is transcendental, there is no polynomial $a = a_0 + a_1\theta + \ldots + a_n\theta^n$ with $a_i \in \mathbb{Q}$ for which $a(\pi) = 0$, i.e. $0 = a_0 + a_1\pi + \ldots + a_n\pi^n$ with $a_i \in \mathbb{Q}$ forces $a_i = 0$. This means $\{1, \pi, \pi^2, \ldots, \pi^n\}$ is linearly independent over $\mathbb{Q}$ for any n and thus we cannot have a space over $\mathbb{Q}$ and must have therefore a basis which is seen now to have an infinite number of elements. Thus linear independence is tied to the field $\mathbb{F}$ under consideration.

5. Given $a = a(\theta) \in \mathbb{F}[\theta]$, $a = a_0 + a_1\theta + \ldots + a_m\theta^m$, $a_m \neq 0$ (i.e. deg $(a) = m$) and $X \xrightarrow{A} X$, dim $(X) = n < \infty$, we get $X \xrightarrow{a(A)} X$ defined by

$$a(A) := a_0 Id + a_1 A + \ldots + a_m A^m \ (A^0 = \text{Id on } X)$$

5.1 Verify

(i) $a(\theta) = b(\theta) \Rightarrow a(A)b(A)$,

(ii) $a(\theta) + b(\theta) = c(\theta) \Rightarrow a(A) + b(A) = c(A)$,

(iii) $a(\theta)b(\theta) = c(\theta) \Rightarrow a(A).b(A) = c(A)$,

(iv) $a(\theta) = d(\theta)q(\theta) + r(\theta)$ ($q(\theta)$ is the quotient and $r(\theta)$ is the remainder on dividing $a(\theta)$ by $d(\theta)) \Rightarrow a(A) = d(A)q(A) + r(A)$ and illustrate this by

$a(\theta) = \theta^3 + 3\theta + 2 + \lambda$ ($\lambda \in \mathbb{F}$), $a(\theta) = d(\theta)q(\theta) + r(\theta)$ for

$d(\theta) = \theta + 1$, $A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$

Hint:

$q(\theta) = \theta + 2, r(\theta) = \lambda, d(A) = A + Id_2 = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$, $a(A) = \begin{bmatrix} 8+\lambda & 8 \\ 4 & 4+\lambda \end{bmatrix}$,

$q(A) = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}$, $r(A) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$.

Here $\mathbb{F}^2 \xrightarrow{A} \mathbb{F}^2$ is given by $Ax = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \in \mathbb{F}^2$ for $x = \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \in \mathbb{F}^2$.

(v) Prove that if $\left(X, \underline{\underline{e}}\right) \xrightarrow[a]{A} \left(X, \underline{\underline{e}}\right)$ (i.e. $a = \begin{bmatrix} a_i{}^j \end{bmatrix} = a = \begin{bmatrix} e^j(e_i) \end{bmatrix} = [\langle e^j | e_i \rangle]_{n \times n}$ is the matrix of A with respect to the basis $\underline{\underline{e}} = (e_1, \ldots, e_n)$) and $p(\theta) = p_0 + p_1(\theta) + \ldots + p_m\theta^m \in \mathbb{F}[\theta]$ then $\left(X, \underline{\underline{e}}\right) \xrightarrow[P]{p(A)} \left(X, \underline{\underline{e}}\right)$ with $p = \begin{bmatrix} p_i{}^j \end{bmatrix}$ where $p_i{}^j$ is the $j \times i$ entry of the matrix $p(a) = p_0 + p_1 a + \ldots + p_m a^m$ in $\mathrm{Mat}_n(\mathbb{F})$. Illustrate this for $X = \mathbb{F}^2$, $p(\theta) = \theta^2 + 3$, $Ax = \begin{bmatrix} x^1 + x^2 \\ x^1 - 2x^2 \end{bmatrix}$.

Hint:

For $\underline{\underline{e}} = \{e_1, e_2\}$, $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, we have $Ae_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $Ae_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ and thus the matrix of A is $a = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$ while $(p(A))(x) = (A^2 + 3Id)(x) = \begin{bmatrix} 5x^1 - x^2 \\ -x^1 + 8x^2 \end{bmatrix}$ supplying $(p(A))(e_1) = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$, $(p(A))(e_2) = \begin{bmatrix} -1 \\ 8 \end{bmatrix}$ and thus $\begin{bmatrix} 5 & -1 \\ -1 & 8 \end{bmatrix}$ as the matrix of $p(A)$ (with respect to $\underline{\underline{e}}$) which is indeed $a^2 + 3Id_2$. Note further that with $d(\theta) = \theta - 1$, we have $q(\theta) = \theta + 1$, $r(\theta) = 4$ and $(p(A)) \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} = \begin{bmatrix} 5x^1 - x^2 \\ -x^1 + 8x^2 \end{bmatrix} = \begin{bmatrix} x^2 \\ x^1 - 3x^2 \end{bmatrix} + \begin{bmatrix} 2x^1 + x^2 \\ x^1 - x^2 \end{bmatrix} + \begin{bmatrix} 4x^1 \\ 4x^2 \end{bmatrix} = (d(A).q(A) + r(A)) \begin{bmatrix} x^1 \\ x^2 \end{bmatrix}$ exactly as expected, in analogy with (iv) and thus $p(A) = d(A).q(A) + r(A)$ because $p(\theta) = d(\theta)q(\theta) + r(\theta)$.

(vi) Show that $p(a)$ of (v) above has determinant

$$\det\, p(a) = p(\lambda_1).p(\lambda_2)\dots p(\lambda_n)$$

for $\mathbb{F} = \mathbb{C}$ where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $a = \left[\ a_i{}^j\ \right]_{n\times n}$

Hint:

$p(\theta)$ can be resolved into linear factors:

$p(\theta) = \alpha(\mu_1 - \theta)\dots(\mu_m - \theta)$ for $\alpha, \mu_1, \dots, \mu \in \mathbb{C}$ and then the matrix polynomial $p(a) = \alpha_0(\mu_1 I_n - a)\dots(\mu_m I_n - a)$. So if χ_a is the characteristic polynomial of a, we get $\det(a) = \lambda_1 \dots \lambda_n,\ p_0 = (-1)^n \lambda_1 \dots \lambda_n,\ p_{n-1}$=-trace$(a) = -(\lambda_1 + \dots + \lambda_n)$ and thus

$$\begin{aligned}
\det\, p(a) &= \alpha^n(\det(\mu_1 I_n - a))\dots(\mu_m I_n - a) \\
&= \alpha^n \chi_a(\mu_1)\dots\chi_a(\mu_m) \\
&= \prod_{i=1}^{n} \alpha(\mu_1 - \lambda_1)(\mu_2 - \lambda_2)\dots(\mu_m - \lambda_m) \\
\text{since } \chi_a(\theta) &= \theta I_n - a = (\theta - \lambda_1)\dots(\theta - \lambda_n), \text{ and thus} \\
\det\, p(a) &= p(\lambda_1)p(\lambda_2)\dots p(\lambda_n)
\end{aligned}$$

(vii) Prove that in (vi), the eigenvalues of $p(a)$ are $p(\lambda_1), \dots, p(\lambda_n)$ and if the eigenvector w_i corresponds to the eigenvalues λ_i for a, the eigenvector w_i corresponds to the eigenvalues $p(\lambda_i)$ for $p(a)$ also, $1 \le i \le n$.

Hint:

We get

$$\begin{aligned}
\det\, \chi_{p(a)} &= \det[\theta I_n - p(a)] = \det[\theta I_n - \alpha(\mu_1 - \theta)\dots\alpha(\mu_m - \theta)] \\
&= [\theta - p(\lambda_1)]\dots[\theta - p(\lambda_n)]
\end{aligned}$$

supplying the eigenvalues $p(\lambda_i),\ 1 \le i \le n$, for $p(a)$, given that λ_i are the eigenvalues of a. Further,

$$\begin{aligned}
(p(a))(w_i) &= (p_0 I_n + p_1 a + \dots + p_m a^m)(w_i) \\
&= (p_0 + p_1\lambda_i + \dots + p_m(\lambda_i)^m)(w_i) \\
&= p(\lambda_i)w_i
\end{aligned}$$

so that $p(\lambda_i)$ corresponds to the eigenvector w_i for $p(a)$.

(viii) In (vii) above, consider arbitrarily given polynomials $g(\theta),\ f(\theta)$ with $g(\lambda_i) \in 0$ for $1 \le i \le n$. Prove that the eigenvalues of $h(a) = f(a)g^{-1}(a) = f(a)(g(a))^{-1} = (g(a))^{-1}f(a)$ are given by $h(\lambda_i) := \frac{f(\lambda_i)}{g(\lambda_i)}$ with corresponding eigenvectors w_i for $1 \le i \le n$.

Hint:

We have $\det(g(a)) = g(\lambda_1)\ldots g(\lambda_n) \neq 0$ so $(g(a))^{-1}$ exists and $\det(h(a)) = (\det(f(a)))[\det g(a)]^{-1} = \prod_{i=0}^{n} \frac{f(\lambda_i)}{g(\lambda_i)} = \prod_{i=0}^{n} h(\lambda_i)$ which means $\det[\theta I_n - h(a)] = \prod_{i=0}^{n}[\theta - h(\lambda_i)]$ and provides $h(\lambda_i)$ as the eigenvalues. Further, $(g(a))(w_i) = g(\lambda_i)w_i$ and $(f(a))w_i = f(\lambda_i)w_i$ which means $h(a)g(a)(w_i) = f(a)(w_i)$ forces $(h(a))(w_i)\frac{f(\lambda_i)}{g(\lambda_i)}$.

(ix) In (viii) above, do we have $g^{-1}(a) = (g(a))^{-1}$ as a polynomial in the matrix a?

Hint:

Writing $B = g(a)$, we know that B is invertible. Suppose it has the minimum polynomial $\mu_B(\theta) = \mu_0 + \mu_1\theta + \ldots + \theta^r$. Then $\mu_0 \neq 0$ since otherwise $\lambda = 0$ is an eigenvalue of $B = g(a)$ and B is not invertible. Now this means $\mu_0 I_n + \mu_1 B + \ldots + B^r = 0$ which supplies $B^{-1} = -\frac{1}{\mu_0}[\mu_{I_n} + \ldots + B^{r-1}]$ which means that $(g(a))^(-1) = -\frac{1}{\mu_0}[\mu_1 I_n + \mu_2(g(a)) + \ldots + (g(a))^{r-1}]$ which is a polynomial in the matrix a since $g(a)$ is a polynomial in the matrix a.

6. Suppose $d_{n-1}(\theta)$ is the gcd of all minors of order $n-1$ in the matrix $a = \left[\ a_i{}^j\ \right]_{n\times n}$ being thus the matrix polynomial $\det(\theta I_n - a) = \chi_a(\theta)$. Prove that the minimal polynomial of the matrix a is given by

$$\mu_A(\theta) = \frac{\chi_a(\theta)}{d_{n-1}(\theta)}.$$

Hint:

Since $\chi_a(\theta) = \det(\theta I_n - a)$ can be expressed as the sum of elements in the first row of the matrix $\theta I_n - a$ multiplied by their corresponding cofactors, we have $d_{n-1}(\theta)$ as a divisor of $\chi_a(\theta)$ which means $\mu_a(\theta)$ is a polynomial which must be monic since both $\chi_a(\theta)$ and $d_{n-1}(\theta)$ are monic. Since $d_{n-1}(\theta)$ is a divisor of all the elements in $\mathrm{adj}(\theta I_n - a)$ (the adjutant of $\theta I_n - a$, not the adjoint), we can write $d_{n-1}(\theta)r(\theta) = adj(\theta I_n - a)$ where the gcd of all the elements in the matrix $r(\theta)$ (entries from $\mathbb{F}[\theta]$) is 1.

Now $(\theta I_n - a)\ adj\ (\theta I_n - a) = \chi_a(\theta)I_n$ so that, on division by $d_{n-1}(\theta)$, we get $(\theta I_n - a)r(\theta) = \mu(\theta)I_n$ and thus $\mu(a) = 0$ (remainder theorem). Therefore, we do have $\mu(\theta)$ as an annihilating polynomial of a. If $p(\theta)$ is some annihilating polynomial of a with deg $p(\theta) <$ deg $\mu(\theta)$, the remainder theorem supplies $(\theta I_n - a)q(\theta) = p(\theta)I_n$; with $p(\theta)\widetilde{q}(\theta) = \mu(\theta)$ which must exist since $p(a) = 0 = \mu(a)$, we have now

$$(\theta I_n - a)\widetilde{q}(\theta)q(\theta) = \mu(\theta)I_n$$

which supplies $r(\theta) = \widetilde{q}(\theta)q(\theta)$ and contradicts the fact that the gcd of all the elements in the matrix $r(\theta)$ is 1. Thus $\mu(\theta)$ must be the annihilating polynomial of a having least degree i.e. $\mu(\theta)$ is the minimal polynomial.

Example 0.3. For $a = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ we have $\theta I_3 - a = \begin{bmatrix} \theta-1 & 0 & -1 \\ 0 & \theta-1 & 0 \\ 0 & 0 & \theta+1 \end{bmatrix}$, $\text{adj}(\theta I_n - a) = \begin{bmatrix} \theta^2-1 & 0 & \theta-1 \\ 0 & \theta^2-1 & 0 \\ 0 & 0 & (\theta-1)^2 \end{bmatrix}$ with $d_2(\theta) = \theta - 1$, $\chi_a(\theta) = (\theta-1)(\theta^2-1)$, and $\mu_a(\theta) = \theta^2 - 1$; verify that $\mu_a(a) = a^2 - I_2 = 0$.

7. If W is a subspace of a vector space over $\mathbb{F}$, the collection $X/W := \{x + W | x \in X\}$ is turned into a vector space (over $\mathbb{F}$, the same field) under the operation, given $x, u \in X,\ \lambda \in \mathbb{F}$,

$$(x + W) + (u + W)\lambda := (x + u\lambda) + W$$

Hint:

X/W is certainly an abelian group, seeing that X is an abelian group and W is hence automatically a normal subgroup (before becoming a subspace). We note that $W = 0 + W$ is the additive identity for this group X/W. Writing $[x]$ for $x + W$ therefore, we see that elements of X/W are $[x]$ for $x \in X$ and the operation supplied is

$$[x] + [u]\lambda := [x + u\lambda].$$

Since $[u] = [u']$ means $u - u' \in W$ and this ensures $(u-u')\lambda \in W$ i.e. $[(u-u')\lambda] = [u\lambda - u'\lambda] = [0]$, we get $[u\lambda] - [u'\lambda] = 0$ which means the scalar multiplication inbuilt in this is well defined; the addition fragment is well defined anyway (from the theory of abelian groups). It is now almost trivially verified that X/W becomes a vector space. X/W is called the quotient space of X by W; its dimension is called the codimension of W.

7.1 $\dim X = \dim W + \dim(X/W)$

Hint:

For a proof, see the Appendix Handout, page(7).

7.2 The first isomorphism theorem for vector spaces says:

If $X \xrightarrow{A} Y$ is a linear transformation,

$$A(X) \cong X/kerA$$

Prove it.

Hint:

Just repeat the proof of the first isomorphism theorem for groups.

8. Consider the vector space $\{\mathbb{Z} \xrightarrow{f} \mathbb{F}\}$ of all functions for which $f(j) = f(j+n)$ i.e. all functions 'with period n' for a fixed integer n. Writing any such $f(j)$ as $z^j \in \mathbb{F}$, prove that this can be written as $z := \begin{bmatrix} z^{j+1} \\ \vdots \\ z^{j+n} \end{bmatrix}$ as long as any n consecutive integers $j+i, \ldots, j+n$ are fixed.

In particular, this can be written as $z := \begin{bmatrix} z^0 \\ z^1 \\ \vdots \\ z^{n-1} \end{bmatrix}$

8.1 Show that this space can be regarded as the space of all functions $\mathbb{Z}_n \to \mathbb{F}$ and that it is isomorphic to $\mathbb{F}^n$.

Hint:

Take $e_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, $e_1 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$, $\ldots$, $e_{n-1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$ as basis.

8.2 Fix a positive integer N. Prove that $E_0(j) := \frac{1}{\sqrt{N}}, \ldots, E_{N-1}(j) := \frac{1}{\sqrt{n}} e^{2\pi i(N-1)j}/N$, $(i = \sqrt{-1})$ for $j = 0, \ldots, N-1$ supplies an orthonormal basis for $\mathbb{C}^N$ (in the sense of (8.1)) with respect to the inner product $(u|v) := \sum_{n=0}^{N-1} \overline{u^n} v^n$.

Hint:

$$
\begin{aligned}
(E_k|E_m) &= \sum_{j=0}^{N-1} E_m(j)\overline{E_k(j)} \\
&= \sum_{j=0}^{N-1} (\frac{1}{\sqrt{N}} e^{2\pi imj/N})\overline{(\frac{1}{\sqrt{N}} e^{2\pi ikj/N})} \quad [\because \ E_l \text{ has } j-\text{th component } \frac{1}{\sqrt{N}} e^{2\pi ilj/n} \text{where } i \text{ is the complex number} \sqrt{-1}] \quad \text{Thus } \|E_m\|^2 = \\
&= \sum_{j=0}^{n-1} (\frac{1}{N} e^{\frac{2\pi i(m-k)j}{N}^j})
\end{aligned}
$$

1 ($\because \ m-k=0$) and for $m \neq k$, we have a geometric progression with sum $\frac{1-e^{(\frac{2\pi i(m-k)}{N})^N}}{1-e^{\frac{2\pi i(m-k)}{N}}} = 0$ since $m-k$ is an integer.

This means $\{E_0, \ldots, E_{N-1}\}$ is a set of N distinct orthonormal vectors in $\mathbb{C}^N$ and provides therefore an orthonormal basis for $\mathbb{C}^N$.

8.3 In the Hilbert space $\mathbb{C}^N$, prove the Parseval's relation

$$(v|w) = \sum_{k=1}^{N} \overline{(v|b_k)}(w|b_k)$$

where $\{b_1, \ldots, b_N\}$ is some orthonormal basis (just calculate directly) and derive the Pancherel's formula $\|v\|^2 = \sum_{k=1}^{N} |(v|b_k)|^2$ (just put $v = w$ in Parseval's relation).

8.4 Apply the formulas of (8.3) to the basis found in (8.2) and calculate $\|z\|^2 = \sum_{m=0}^{N-1} |(z|E_m)|^2$ where $\overline{(z|E_m)} = \sum_{n=0}^{N-1} z(n)\frac{1}{\sqrt{N}} e^{-2\pi imn/N}$ (Recall: $z(n) = z^n$ is the $n-$th component of z i.e. the value at $n \in \mathbb{Z}_N$ of $\mathbb{Z}_N \xrightarrow{z} \mathbb{C}$; we know now that $z = \begin{bmatrix} z^0 \\ \\ z^{N-1} \end{bmatrix} \in \mathbb{C}^N$).

8.5 Write $\widehat{z}(m) := \sum_{n=0}^{N-1} z(n)e^{-2\pi imn/N}$ so that $\mathbb{Z}_N \xrightarrow{\widehat{z}} \mathbb{C}$ given by $(\widehat{z}(0), \ldots, \widehat{z}(N-1)) \in \mathbb{C}^N$ and prove that $z \mapsto \widehat{z}$ defines a linear transformation $\mathbb{C}^N \xrightarrow{DFT} \mathbb{C}^N$. (Just verify).

Note:

In this context, one usually writes $l^2(\mathbb{Z}_N)$ for $\mathbb{C}^N$ to be consistent with the notation $l^2(\mathbb{N})$ for the space of square-summable complex sequences.

$$l^2(\mathbb{N}) := \{z(j) \in \mathbb{C} | \sum_{j=1}^{\infty} |z(j)|^2 < \infty\}$$

with the inner product $(z|w) := \sum_{j=1}^{\infty} \overline{z(j)}w(j)$ (which is an infinite-dimensional Hilbert space).

—— The linear transformation $l^2(\mathbb{Z}_N) \xrightarrow{DFT} l^2(\mathbb{Z}_N)$ is known as the discrete fourier transform.

9. Recall that (Handout-I, page (18), (1.8.3)) that if $X \xrightarrow{A} Y$ is a linear transformation, dim $X = n \ < \infty$, dim $Y = m \ < \infty$, then there exists a basis $\underline{e}$ for X and a basis $\underline{d}$ for Y such that the matrix representation for A with respect to these bases is of the form $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ where I_r is the identity matrix of order r, r being the rank of A. Let us make a definition: say that two linear transformations $X \xrightarrow{A,B} Y$ are equivalent iff there exist invertible operators $P \in L(X,X),\ Q \in L(X,X)$ with $PAQ = B$. This is then an equivalence relation on $L(X,X)$ ($A = Id_Y AID_X$ shows reflexivity, $PAQ = B \Leftrightarrow A = P^{-1}BQ^{-1}$ shows symmetry, and $PAQ = B,\ RBS = C \Rightarrow C + (RP)A(QS)$ shows transitivity) which should remind us of an equivalence relation on $L(X,X)$ introduced earlier (change-of-Basis, page(10)) two linear transformations $X \xrightarrow{A,B} X$ are similar iff there exists one invertible $X \xrightarrow{P} X$ with $PAP^{-1} = B$. There exists in fact a result which enables us to slide similarity into equivalence and vice versa:

$$X \xrightarrow{A,B} X \ \text{ are similar iff}$$
$$X[\theta] \xrightarrow{\chi_A(\theta),\chi_B(\theta)} X[\theta] \ \text{ are equivalent}$$

(stated in terms of matrices, two $n \times n$ matrices a, b are similar iff their characteristic polynomials $\chi_a(\theta)$ and $\chi_b(\theta)$ are equivalent; this is the more popular version. For $X[\theta]$, refer to page III-(3) of this handout)

Which is in fact a special case of the following:

For a ring $\mathbb{K}$ and $\alpha, \beta \in \mathbb{K}$, the following statements are equivalent:

(i) There is an invertible $y \in \mathbb{K}$ such that $\beta = \gamma\alpha\gamma^{-1}$

(ii) There are invertible $p(\theta), q(\theta) \in \mathbb{K}[\theta]$ with $\theta - \beta = p(\theta)(\theta - \alpha)q(\theta)$. When $\mathbb{K} = L(X,X)$; note that we are talking about a noncommutative ring here.

9.1 From this perspective every linear transformation $X \xrightarrow{A} Y$ has a very simple representative out of its equivalence class under the relation termed 'equivalence' introduced above, namely, the linear transformation raised by the matrix $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ (of course, we are talking about dim $X < \infty,\ dim\ Y < \infty$). This is frequently expressed by saying that "every $m \times n$ matrix has a normal or canonical form supplied by this matrix".

However, the similarity relation is not simply a special case of equivalence relation. When talking about equivalence, we are talking about different coordinatization $\underline{e}$ and $\underline{d}$, the invert-

ibles P and Q introduce new coordinatization $\underline{\underline{e'}}$, $\underline{\underline{d'}}$ and 'A becomes B'. But $PAP^{-1} = B$ is different: whatever change of basis is done by P is undone by P^{-1}; we are back to the same coordinatization of the vector space X.

The problem here is to find, for each $X \xrightarrow{A} X$, just one basis for X relative to which A is cast into some 'as simple as possible' form.

9.2 We now state the problem in a more precise form.

(i) To begin with the (left) action of a group G an a set X is defined to be a function $G \times X \to X$, written $(g, x) \mapsto gx$ satisfying

$$1_G x = x \text{and } g(g'x) = (gg')x \text{ for all } x \in X, g, g' \in G$$

we say X is a $G-$set.

Writing $x \sim y$ (for $x, y \in X$) iff there is $g \in G$ such that $gx = y$ supplies an equivalence relation on X ($1_G x = x$ ensures reflexivity $gx = y \Leftrightarrow x = g^{-1}y$ ensures symmetry, $gx = y,\ g'y = z \Rightarrow (g'g)x = z$ ensures transitivity); the equivalence class of x [x]:={y $\in X | gx = y$ for some $g \in G$}is called the orbit of x.

(ii) A vector space X is made into a $GL(x)-$ set (where $GL(X)$ is the general linear group of X consisting under composition, of all invertibles $A \in L(X, X)$) via the action $x \mapsto Ax$, $\{0\}$ and $X \setminus \{0\}$ being the two orbits. Call this the natural action.

(iii) Now consider the vector space $L(X, X)$. Each invertible P in $L(X, X)$ raises a linear transformation

$$L(X, X) \xrightarrow{Ad\ P} L(X, X)$$

given by $(Ad\ P)A := PAP^{-1}$ and thus $L(X, X)$ becomes a $G-$ set

$$GL(X) \times L(X, X) \xrightarrow{(P,A) \mapsto (AdP)A} L(X, X)$$

($Ad(P_1P_2) = (AdP_1)(ADP_2)$, $(AdId) = Id$ are immediate); the orbit of $A \in L(X, X)$ is the similarity class of A,

$$\text{orbit}(A) = \{B \in L(X, X) | B = (Ad\ P)A = PAP^{-1} \text{ for some } P \in GL(X)\}$$

A canonical form for $X \xrightarrow{A} X$ is a representative from its orbit by a unique choice i.e. given by a function which selects exactly one member from each $GL(X)-$ orbit.

To give an illustration, suppose X is a complex vector space, dim $X = n < \infty$. Then there are two results (we prove neither here; this is just for illustration)

(i) Each $A \in L(X, X)$ has a 'triangular form' i.e. there is some basis for X such that with respect to that, A can be represented by an upper triangular matrix.

This does not provide a canonical form, since two distinct triangular forms $\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$, $\begin{bmatrix} \lambda & 2 \\ 0 & \lambda \end{bmatrix}$ correspond to the same $GL(V)-$ orbit.

(The result is : each$A \in L(X, X)$ is given by $\begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$ where $\lambda_1, \ldots, \lambda_n$ are the n eigenvalues of A with possible repetitions and $*$ means 'anything')

(ii) Each $A \in L(X, X)$ has a 'Jordan Canonical Form' i.e there is some basis for X such that with respect to that,A can be represented by

$\begin{bmatrix} J_\lambda & & 0 \\ & \ddots & \\ 0 & & J_{\lambda_k} \end{bmatrix}$ where $J_{\lambda_k} = \begin{bmatrix} \lambda_k & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_k \end{bmatrix}$ each λ_k on the diagonal n_k times ,$n_1 + n_2 + ... + n_k = n$, $\lambda_1, \lambda_2, ..., \lambda_k$ distinct eigen values of A. This provides a 'canonical form' since *(but for the ordering of the eigenvalues)* this is uniquely given from each $GL(V)-$orbit. Thus, while both these statements are "equivalent" in the sense that both the triangular and the jordan forms are always available for any $X \xrightarrow{A} X$ as long as its characteristic polynomials splits into linear factors

$$X_A(\theta) \;=\; (\theta - \lambda_1)^{n_1} \ldots (\theta - \lambda_k)^{n_k},\; n_1 + n_2 + \ldots + n_k = n$$

for $dim_{\mathbb{F}} X = n$ which happens whenever $\mathbb{F}$ is algebraically closed *(and in particular for $\mathbb{F} = \mathbb{C}$)*, one of them provides a canonical form, the other does not.

(iii) For X equipped with an inner product *(indeed equipped with some bilinear form; remember from III-8 that the inner product is a special kind of bilinear form $\overline{X} \times X \longrightarrow \mathbb{C}$)* we ask for more; since we are not interested here in 'canonical form theory', we halt this discussion at this point and proceed to the two situations which are somewhat special. But note that here the interest is in the $G-$space X where G is the subgroup of $GL(X)$ preserving the inner product and thus one $GL(X)-$orbit may split into several $G-$orbits.

9.4 suppose we have n distinct eigenvectors, i.e.,in the factorization

$$X_A(\theta) = (\theta - \lambda_1)^{n_1} \dots (\theta - \lambda_k)^{n_k},\ n_1 + \dots + n_k \ = \ n,$$

we have $1 = \ n_1 = \dots = \ n_k$, $k = n$. Then the eigenvectors corresponding to these are automatically linearly independent. But we do know that if there are n linearly independent eigenvectors $x_1 , \dots , \ x_n$ then regardless of the n eigenvalues $\lambda_1 , \dots , \ \lambda_n$ being repeated or not,we can form the model matrix $P \ = \ \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}_{n \times n}$ with $a = PDP^{-1}$ where (i) $D \ = \ diag(\lambda_1, \dots, \lambda_n)$ *(each λ_i repeated as many times as it occurs in $X_A(\theta)$ on the diagonal)*, (ii) a is same matrix of A_k to start with; further, we know that this amounts to $X \ = \ \oplus_{j=1}^{k} E_j$ where E_j=ker$(\lambda_j id - A)$=ker$(A - \lambda_j id)$ is the eigenspace corresponding to the eigenvalue λ_j, and (iii) such matrices are called 'diagonalizable' *(eigenspectrum page 7)*. We examine these diagonalizable matrices. Noting that whenever we can find r linearly independent eigenvectors $\{x_1, \dots, x_r\}$, by extending this to a basis $\{x_1, \dots, x_r, x_{r+1}, \dots, x_n\}$ we have a representation $\left[\begin{array}{ccc|c} \lambda_1 & & 0 & * \\ & \ddots & & \\ 0 & & \lambda_r & \\ \hline 0 & & & * \end{array}\right]$ *(* means 'anything')* the diagonalizable matrix is a particularly favourable instance of this: $X \ \xrightarrow{T} \ X$ is diagonalizable iff $X \ = \ \bigoplus_{\lambda \in \Lambda} E_\lambda(T)$ iff $T \ = \ P^{-1}DP, D \ = \ diag(\lambda_1, ..., \lambda_n)$ P is modal matrix iff there is some basis such that T has the representation $\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$ iff there are n linearly independent eigenvectors *(dim X=n)*

9.5 illustrations

9.5.1 if $A = \begin{bmatrix} 1 & \dots & 1 \\ & & \\ 1 & \dots & 1 \end{bmatrix}$, with $v_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ for and $v_i = e_i - 1 - e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$ for $2 \leq i \leq n$; then $\{v_1, \dots, v_n\}$ are linearly independent *(prove it)* and with respect to this basis, A has the matrix $diag(n, 0, \dots)$.Thus A is diagonalizable.

9.5.2 Suppose $X \xrightarrow{T} X$ $X = T(X) \bigoplus kerT$ and $T(X) = \{x \in X | Tx = x\} = \{a + b | a \in T(X), b \in kerT, a = a + b\} = \{a + b | a \in T(X), b = 0\}$ then choose a basis $\{x_1, x_2, \ldots, x_r\}$ of $T(X)$ extending it to a basis $\{x_1, \ldots, x_r, x_r + 1, \ldots, x_n\}$ where $\{x_r + 1, \ldots, x_n\}$ is some basis of ker$T$, then since $T(X) = \{x \in X | Tx = x\}$, we have that with respect to this basis of X, T has the representation $\left[\begin{array}{c|c} id_r & 0 \\ \hline 0 & r \end{array}\right]$ One says that T is a projection of X onto $T(X)$ along ker$T$. Prove that this is equivalent to $T^2 = T$ and compare with 1.8.3 on page 18 of handout-I.

9.5.3 Prove that if $X = ker(T - id_x) \bigoplus ker(T + id)$ then there is a basis of X such that with respect to this, T has the representation $\left[\begin{array}{c|c} id & 0 \\ \hline 0 & -id \end{array}\right]$ *(choose a basic of $ker(T - id_X)$ and it to a basic of X by choosing some basic of $ker(T + id_X)$* one says that T is a reflection at $ker(T - id_X)$ along B. Prove that this is equivalent to $T^2 = id_X$. *(Note that we are requiring that each $x \in X$ can be written as $\frac{1}{2}[(id_X - T)x + (id_X + T)x]$ and thus we need characteristic of to be $\neq 2$).*

9.5.4 Let us recaptitulate : if $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of $X \xrightarrow{A} X$ and $E_i = ker(A - \lambda_i)$, for the polynomials $L_i(\theta) = \prod_{j \neq i} \frac{\theta - \lambda_i}{\lambda_i - \lambda_j}$, then given $0 = v_1 + ... + v_k = \sum_{j=1}^{k} v_j$, $v_j \in E_i$, we have $0 = L_i(A) \sum_j v_j = \sum_j L_i(\lambda_j) v_j = \sum_j \delta_j^i v_j = v_i$ for $1 \leq i \leq k$ which answers that $W := E_1 \bigoplus ... \bigoplus E_k$ is a distinct sum *(and also that the eigenvectors corresponding to $\lambda_1, ..., \lambda_k$ are linearly independent;a bitslicker than the arguments supplied for 1. $\rightharpoonup$ .1,1. $\rightharpoonup$ .2 on page 6 of the eigenspectrum handout)*; One says that A is diagonalizable iff $X = E_i \bigoplus ... \bigoplus E_k$. Now if $x \in X$,$x \neq 0$,we have $x \in E_j$ for exactly one j; then $Ax = \lambda_j x$ and thus $Ax \in E_j$ again. All told, we have $A = \sum \lambda_i P_i$ where $X \xrightarrow{P_i} X$ is given by $P_i(x) = x_i$ for $x = x_1 + \ldots + x_k \in X$, $1 \leqslant i \leqslant k$. Clearly, we have $(P_1 + \ldots + P_k)(x) = x_1 + \ldots + x_k = x$ so that $P_1 + \ldots + P_k = id$, and also, $P_i P_j = 0$ for $i \neq j$ which means $P_i = P_i id = P_i(P_1 + \ldots + P_k) = P_i^2$ so that P_i is actually a projection *(sec 9.5.2, III − 22)*.

9.5.5 Now suppose $X \xrightarrow{P_i} X$, $1 \leqslant i \leqslant k$ are given with the properties $id = \sum_{j=1}^{k} P_i$, $P_i P_j = 0$ for $i \neq j$ *(so that in particular $P_i^2 = P_i$ P_i is a projection)* we have $X = E_1 \bigoplus ... \bigoplus E_k$ with $E_i \neq 0$, P_i projection of X onto E_i along $\sum_{j \neq i} E_i$, $E_i = P(X)$. If now we construct the operator $A := \sum_{i=1}^{k} \lambda_i P_i$, $\lambda_i \in F$ all distinct; we shall have, for $x_j \in E_j$, $Ax_j = (\sum \lambda_i P_i) x_j = \lambda_j x_j$ which means $x_j = 0 + \ldots + x_j + 0 + \ldots \in X$ is an eigen vector of A with λ_j as eigen value; further, the entire spectrum is $\{\lambda_1, \ldots, \lambda_k\}$ since if an additional distinct eigenvalue λ_{k+1} exits,

$E_{k+1} = ker(A - \lambda_{k+1}id)$ is another new summand and $X = E_1 \bigoplus \ldots \bigoplus E_k$ is not adequate. But this means that A is diagonalizable.

9.5.6 suppose $X \xrightarrow{A} X$ has the minimal polynomial $\mu_A(\theta) = \prod_{i=1}^{k} (\theta - \lambda_i)$ so that $o = \mu_A(A) = (\lambda_1 id - A)...(\lambda_k id - A) \in L(X, X)$ then if we write $P_i = L_i(A)$ for the lagrange polynomials L_i, we get $P_i = \prod_{j \neq i} \frac{(\lambda_i id - A)}{\lambda_i - \lambda_j}$ and clearly,

(a) $P_1 + \ldots + P_k = id$

($P_1 + P_2 = -\frac{\lambda_1 id - A}{\lambda_2 - \lambda_1} + \frac{\lambda_2 id - A}{\lambda_1 - \lambda_2} = \frac{\lambda_1^2 id - \lambda_1\lambda_2 id - \lambda_1 A + \lambda_2 A + \lambda_2^2 id - \lambda_2\lambda_1 id - \lambda_2 A + \lambda_1 A}{(-\lambda_2 - \lambda_1)^2} = \frac{(\lambda_1 - \lambda_2)^2 id}{(\lambda_1 - \lambda_2)^2} = id$ and so on)

(b) $P_i P_j = 0$, and

(c) $A = \sum \lambda_i P_i$ as simple algebraic consequences so that we are in the (9.5.5) situations.

9.5.7 The form $A = \sum_{i=1}^{k} \lambda_i P_i$ obtained in (9.5.4) as well as (9.5.5) above,is called the spectral form of the operator $X \xrightarrow{A} X$. Noting that this means $A^2 = (\sum_i \lambda_i P_i)(\sum_j \lambda_j P_j) = \sum_i \sum_j \lambda_i \lambda_j P_i P_j = \sum_i \lambda_i^2 P_i^2 = \sum_i \lambda_i^2 P_i$ *($\because$ P_i is a projection and then $P_i^2 = P_i$)* and that in general $A^n = \sum_i \lambda_i^n P_i$ similarly we, observe that for any polynomial $a(\theta) := a_0 + a_1\theta + \ldots + a_n\theta^n$, we have $a(A) = \sum_i a_i A_i^n = \sum_i a(\lambda_i) P_i$ and summarize this discussion as follows:

Theorem 0.1. If $X \xrightarrow{A} x$ is diagonalizable with eigenspaces $E_i, \ldots, E_k$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ with projections $P_1, \ldots, P_k$ corresponding to the decomposition $X = E_1 \bigoplus \ldots \bigoplus E_k$, we get

(i) $P_i^2 = P_i \neq 0$,

(ii) $P_i P_j = 0$ for $i \neq j$,

(iii) $id_X = \sum_{i=1}^{k} P_i$,

(iv) $ImP_i = E_i = ker(\lambda_i id_X - A)$,

(v) $A = \sum_{i=1}^{k} \lambda_i P_i$,

(vi) $a(\theta) = \sum_{i=i}^{k} a(\lambda_i) P_i$ for $a(\theta) \in F[\theta]$,

(vii) $P_i = \prod_{j \neq i} \frac{\lambda_i id - A}{\lambda_j - \lambda_i}$, and

(viii) the minimal polynomial $\mu_A(\theta)$ of A is $\mu(\theta) = \prod_{i=1}^{k} (\theta - \lambda_i)$ Conversely: if $X \xrightarrow{A} X$ satisfies (v) i.e. $A = \sum_{i=1}^{k} \lambda_i P_i$ where $\lambda_1, \ldots, \lambda_k$ are distinct scalars and $P_i : X \to X$ are non zero operators satisfying (ii) and (iii) then A is diagonalizable with spectrum $\{\lambda_1, \ldots, \lambda_k\}$ and properties $(i) \to (viii)$ all hold. Further, given only that (viii) holds, i.e. $\mu_A(\theta) = \prod_{i=1}^{k} (\theta - \lambda_i)$, $\theta - \lambda_i$all distinct linear factors, A is diagonalizable with spectrum $\{\lambda_1, \ldots, \lambda_k\}$ and properties $(i) \to (vii)$ hold with P_i defined by (vii) i.e. $P_i := \prod_{j \neq i} \frac{\lambda_i id - A}{\lambda_j - \lambda_i}$. *(please rework and fill in any details you think are lacking).*

Illustration Show that the operator $X \xrightarrow{P_i} X$ given by $P_i := |e_i >< e^i|$ for a basis $\{e_1, \ldots, e_n\}$ of X, dual basis $\{e^1, \ldots, e^n\}$, are all projections.

9.6 For X equipped with an inner product, $F = C$, we seek that (i) the basis for diagonalization be orthonormal, and (ii) the modal matrix P used in the diagonalization preserve the inner product *(for modal matrix refer to III − 22,(9.40), or the eigen spectrum page 7, as well as page 9 para 2.2).*

9.6.1 To generate practice in the quite common usage of the 'mathematicans convention', we take the inner product to be conjugate-linear in the second variable and linear in the first variable and for the same reason, use A^* for A^+; thus our definition is generated from $(Av|w) = (v|A^*w)$ for all $v \in V,\ w \in W$ for the Hilbert Adjoint $W \xrightarrow{A^*} V$ of $V \xrightarrow{A} W$ and A^* has the matrix $[\bar{a}^i_j]_{n \times m}$ for A having matrix $[\bar{a}^j_i]_{m \times n}$, $1 \leq i \leq n$, $1 \leq j \leq m$, with respect to chosen orthonormal bases e of V, d of W; this is of course simply the tranjugate *(= conjugate - transpose =transpose - conjugate)* of the matrix $a = [\bar{a}^j_i]_{m \times n}$ introduced previously of the eigenspectrum, then $(AB)^* = B^*A^*$, $det(A^*) = det(A)$, $(Id)^*_n = Id_n$.

9.6.2 Thus given $V \xrightarrow{A} V$, we have $V \xrightarrow{A^*} V$. We say A is

(a) normal iff $AA^* = A^*A$ *(this is not in conflict with the definition given on page 9 of the eigenspectrum as we shall soon show)*,

(b) self-adjoint iff $A = A^*$

(c) skew-adjoint iff $A = -A^*$, and

(d) unitary iff $A^* = A^{-1}$.

9.6.3 since $(Ax|Ay) = (x|A^*Ay) = (x|y)$ holds iff $A^* = A^{-1}$, we see that A is unitary iff A preserves the inner product.

9.6.4 Since$|A^* - \lambda id_V| = |(A - \bar{\lambda} id_V)^*| = |A - \bar{\lambda} id_V|$, λ is an eigenvalue of A iff $\bar{\lambda}$ is an eigenvalue of A^*, *(with the same multiplicity).*

9.6.5 if $Ax = \lambda x$, $A^*y = \mu y$, we get $\lambda(x|y) = (\lambda x|y) = (Ax|y) = (x|A^*y) = (x|\mu y) = \mu(\bar{x}|y)$ *(remember the 'mathematicians convention')* so that $(\lambda - \bar{\mu})(x|y) = 0$ forcing $x \perp y$ if $\lambda \neq \bar{\mu}$.

9.6.6 Suppose $X \xrightarrow{A} X$ is normal, $dimX = n < \infty$. If $n = 1$, an orthonormal basis of eigenvectors of A trivially exists *(here $X = \mathbb{C}$, the vector 1 is a basis, and since A is simply $\lambda \in \mathbb{C}$ for some λ as per the remark on page 2 of the first handout, we do have $Ax = \lambda x$ with $x = 1$).* Assume now that $n \geq 2$. Then $\mathbb{C}$ is algebrically closed, some eigenvalue λ and a corresponding eigenvector x with $\| x \| = 1$ can be found. The subspace $x^\perp := \{u \in X | (u|x) = 0\}$ is then A - invariant. *($\because (Au|x) = (u|A^*x) = (u|\bar{\lambda}x) = \lambda(u|x) = \lambda.0 = 0$ so $Au \in x^\perp$)* But $x^\perp \xrightarrow{A} x^\perp$ is again normal *(prove this)* and $dimx^\perp < n$ so by induction hypothesis, an orthonormal basis of $x^\perp$ can be surely found, consisting of eigenvectors of $x^\perp \xrightarrow{A} x^\perp$ which are of course eigenvectors of $X \xrightarrow{A} X$ *(verify this).* We simply add x to this basis, obtaining a basis of X and conclude: If $X \xrightarrow{A} X$is normal, an orthonormal basis of eigenvectors of A exists for X.

If conversely, X has an orthonormal basis $\{e_1, ..., e_n\}$, $Ae_i = \lambda_i e_i, 1 \leq i \leq n$ then each $x \in X$ can be written as $\sum_{i=1}^{n} e_i x^i$, $x^i \in \mathbb{C}$ and we have $AA^*(x) = A[\sum(A^*e_i)x^i] = A[\sum \bar{\lambda} e_i x^i] = \sum \bar{\lambda_i}(Ae_i)x^i = \sum \lambda_i(\bar{\lambda_i}e_i)$
$A^*(\sum \lambda_i e_i x^i) = A^* \sum(Ae_i)x^i = A^*A[\sum e_i x^i] = A^*A(x)$. This being true at each $x \in X$, we conclude that $AA^* = A^*A$ i.e. A is normal.

To sum up:

$X \xrightarrow{A} X$ is normal iff X possesses an orthonormal basis consisting of eigenvectors of A.

9.6.7 Suppose $X \xrightarrow{A} X$ is normal, $dim_{\mathbb{C}}X = n < \infty$. Then there exist an orthonormal basis $\{e_1, \ldots, e_n\}$ consisting of eigenvectors of A, *say*$Ae_i = \lambda_i e_i$ for $1 \leq i \leq n$. Then $A^* = \pm A$ iff $A^*e_i = \pm Ae_i$ iff $\bar{\lambda_i}e_i = \pm\lambda_i e_i$ iff $(\bar{\lambda_i} \mp \lambda_i)e_i = 0$ iff $\bar{\lambda_i} = \pm\lambda_i$ for each i, $1 \leq i \leq n$; that is,

A is $\frac{self-adjoint}{skew-adjoint}$ iff all eigenvalues of A are $\frac{real}{purelyimaginary}$

and $A^* = A^{-1}$ iff $A^*e_i = A^{-1}e_i$ iff $\bar{\lambda_i}e_i = \lambda_i^{-1}e_i$ iff $\bar{\lambda_i} = \lambda_i^{-1}$ iff $|\lambda_i|^2 = 1$ for each i, $1 \leq i \leq n$, that is, A is unitary iff all eigenvalues of A have absolute value 1.

(i) This is a result about normal operators; if some linear operator $X \xrightarrow{A} X$ has real eigenvalues only, it does not follow that T is self-adjoint.

(ii) Thus if A is normal, we can find a unitary matrix P such that $P^*DP = A$, D is diagonal and note that unitary matrices are exactly those which preserve the inner product.

10 We shall now provide some illustrations and close this discussion

10.1 The matrix $A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$ is normal with $\chi_A(\theta) = \theta^3 - 5\theta^2 + 9\theta - 9$ supplying eigenvalues $\lambda_1 = 3, \lambda_2 = 1+\sqrt{2}, \lambda_3 = 1-\sqrt{2}$; for which we have eigenvectors $x_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $x_2 = \begin{bmatrix} 1 \\ -\frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{bmatrix}$, $x_3 = \begin{bmatrix} 1 \\ \frac{i}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} \end{bmatrix}$ which are orthogonal with norm $\| x_i \| = \sqrt{2}$ ensuring that $\frac{1}{\sqrt{2}}x_i$ are orthogonal.
So we take $U := \frac{1}{\sqrt{2}}[x_1\ x_2\ x_3]_{3\times 3}$ which is unitary and $U^*AU = diag(1, 1+\sqrt{2}i, 1-\sqrt{2}i)$.

verify all these statements and carry out a similar discussion for

(a) $\begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ -1 & -1 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 1 & 1 \\ -1 & 2 & -2 \\ -1 & 2 & 2 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$

Ans (a) $U = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ -1 & \alpha & \bar{\alpha} \\ 1 & -\bar{\alpha} & -\alpha \end{bmatrix}$ $\alpha = \frac{1}{2}(1+\sqrt{3}i)$

(b) $U = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 & -1 & -1 \\ 1 & \alpha & \bar{\alpha} \\ -1 & -\bar{\alpha} & -\alpha \end{bmatrix}$ $\alpha = \frac{1}{2}(-2+\sqrt{6}i)$

$(c) U = \begin{bmatrix} 0 & \gamma & \delta \\ \varepsilon & \alpha & \beta \\ -\varepsilon & \alpha & \beta \end{bmatrix}$ $\alpha = -\frac{\gamma}{2}(1+\sqrt{3}$, $\beta = \frac{\delta}{2}(-1+\sqrt{3})$, $\gamma = (3+\sqrt{3})^{\frac{-1}{2}}$ $\delta = (3-\sqrt{-3})^{\frac{-1}{2}}$,
$\varepsilon = \frac{1}{\sqrt{2}}$.

10.2 Take $X = C^2$ with $e_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $e_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ as basis. Show that 'the mirror' $U_M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, 'the beam splitter' $U_B = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, and 'the phase shift' $U_P = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{bmatrix}$ are all unitary. If $X \xrightarrow{\rho_{in}} X$ is given by $|e_0 >< e^0| = \rho_{in}$, calculate $\rho_{out} := U_B\ U_M\ U_P\ U_B\ \rho_{in}\ U_B^+\ U_P^+\ U_M^+\ U_B^+$ *(A^+ stands for the Hilbert Adjoint of A).*

Hint: *ρ_{in} is just the operator* $\begin{bmatrix} 1 \\ 0 \end{bmatrix} [1\ 0] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, and $U_B\ U_M\ U_P\ U_B = \frac{1}{2} \begin{bmatrix} e^{i\theta}+1 & e^{i\theta}-1 \\ -e^{i\theta}+1 & -e^{-i\theta}-1 \end{bmatrix}$ by direct computation; A^+ is just tranjugate of A. Now find $\rho_{out} = \frac{1}{2} \begin{bmatrix} 1+\cos(\theta) & i\sin(\theta) \\ -i\sin(\theta) & 1-\cos(\theta) \end{bmatrix}$

10.3 Consider the 'Pauli matrices' $\sigma_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\sigma_2 := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$, $\sigma_3 := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ Show that they are all Hermitian *(=self-adjoint)* and traceless. For a given unit vector $\underline{n} \in \mathbb{R}^3$, define $\underline{n}.\underline{\sigma} := n_1\sigma_1 + n_2\sigma_2 + n_3\sigma_3 =: \sum$ and calculate $\sum^2 = I_2$ concluding that $\sum$ is unitary. Find the eigenvalues of $\sum$, calculate $< \sum |e_0 >= \sum(e_0)$ and $|(\psi|\sum(e_0)|^2$. Writing $\prod(\underline{n}) := \frac{1}{2}[\sum^2 + \sum]$, Show that $\prod(\underline{n})$ is selfadjoint, a projection, has trace 1 and find $\prod(\underline{n}) \begin{bmatrix} e^{i\phi} & \cos(\theta) \\ \sin(\theta) & \end{bmatrix}$.

Hint:

Direct calculations.'traceless' means has trace 0. $\sum$ is selfadjoint and $\sum^2 = I_2$ so $\sum$ is $\sum^{-1}$ and thus unitary; its eigenvalues can thus be only real with modulus 1 but since the trace is 0, they must be +1 and -1. $e_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *as in 10.2; then* $\sum(e_0) = n_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + n_2 \begin{bmatrix} 0 \\ i \end{bmatrix} + n_3 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *and hence* $|(\psi|\sum(e_0))|^2 = n_3^2$. *You will need* $n_1^2 + n_2^2 + n_3^2 = 1$ *somewhere in all this. It is helpful to note that if* $[A\ ,\ B]_+ := AB + BA$, *then* $0 = [\sigma_1\ ,\ \sigma_2]_+ = [\sigma_2\ ,\ \sigma_3] = [\sigma_3\ ,\ \sigma_1]_+$ *in calculating* $(\prod(\underline{n}))^2 = \prod(\underline{n})$ *proving thus that is projection.*

$\prod(\underline{n}) \begin{bmatrix} e^{i\phi} & \cos(\theta) \\ \sin(\theta) & \end{bmatrix} = \frac{1}{2} \begin{bmatrix} (1+n_3)e^{i\phi}\cos\theta + (n_1 - in_2)\sin\theta \\ (n_1 + in_2)e^{i\phi}\cos\theta + (1-n_3)\sin\theta \end{bmatrix}$.

10.4 If the minimum polynomial $\mu_A(\theta)$ of $A \in L(X, X)$ has degree m, prove that for any $f(\theta) \in \mathbb{F}[\theta]$ with $deg(f(\theta)) \geq m$, we have $f(A) = r(A)$ for some $r(\theta) \in \mathbb{F}(\theta)$ with $deg(r(\theta)) < m$. Calculate

$f(A)$ for

$f(\theta) = 2\theta^5 + 5\theta^3 + 7\theta$, $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

Hint:

$f(\theta) = q(\theta)\mu_A(\theta) + r(\theta)$ with $deg(r(\theta)) < m$ (division algorithm, $deg(f(\theta)) \geq deg(\mu_A(\theta)) = m$ given). Then since $\mu_A(\theta) = 0$, we have $f(A) = r(A)$; for the given minimal, $\mu_A(\theta) = \theta^2 - 1, q(\theta) = 2\theta^3 + 7\theta$; $r(\theta) = 14\theta$, $f(A) = 14A$).

10.5 Show that (i) self adjoint $\Rightarrow$ Normal $\Leftarrow$ Unitary, but (ii) $A = \begin{bmatrix} 1 & 1 \\ i & 3+2i \end{bmatrix}$ is normal while neither selfadjoint nor unitary so that the implications are not reversible

Hint:

$AA^* = \begin{bmatrix} 1 & 1 \\ i & 3+2i \end{bmatrix} \begin{bmatrix} 1 & -i \\ 1 & 3-2i \end{bmatrix} = \begin{bmatrix} 2 & 3-2i \\ 3+3i & 14 \end{bmatrix} = A^*A$

and, (iii) $B = \begin{bmatrix} 1 & i \\ 0 & 1 \end{bmatrix}$ is not normal.

10.6 Show that T is normal $\Rightarrow T - \lambda id$ is normal.

Hint: *direct calculation.*

10.7 Show that $A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & -5 \\ 0 & 1 & -2 \end{bmatrix}$ is not digonalizable over $\mathbb{R}$ but is digonalizable over $\mathbb{C}$.

Hint:

Over $\mathbb{R}$, there is a single eigenvalue $\lambda = 3$ with exactly one independent eigenvector; over $\mathbb{C}$, 3, i, -i are distinct eigenvalues of this 3×3 matrix. Find the modal matrix P to calculate $P^{-1}AP = \begin{bmatrix} 3 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{bmatrix}$

10.8 For $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $g(\theta) = \theta^2 + 1$, Show that $g(A)$ has $\{e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\}$ as eigenvectors and e_2 is not an eigenvector of A though e_1 is ; revisit III $-$ 11(vii) to see why this needs to be noted.

10.9 (a) Show that for $A, B \in Mat_n(F)$, AB and BA have the same eigenvalues.

Hint:

Since the product of two invertible matrices is again invertible, 0 is an eigenvalue of AB iff AB is non invertible iff A or B is noninvertible iff BA is noninvertible iff 0 is an eigenvalue of BA.

A non zero λ is an eigenvalue of $AB \Rightarrow ABx = \lambda x$ with $x \neq 0$; then $ABx \neq 0$ but since $BAx = B \Rightarrow x = \lambda Bx$, we have Bx is an eigenvector of BA with λ as eigenvalue. Now interchange A and B.

(b) Show that λ $X \xrightarrow{T} X$ iff $\lambda I\alpha - T$ is not invertible.

(c) Prove that if $id - AB$ is invertible then $id - BA$ is invertible and $(id - BA)^{-1} = id + b(id - AB)^{-1}A$.

(d) Examine whatever interconnection among (a), (b), and (c) you think might exist.

10.10 Find a matrix B whose minimal polynomial is $\theta^4 - 5\theta^3 - 2\theta^2 + 7\theta + 4$

Hint: *How about the companion matrix ?*

10.11 Find a matrix B whose characteristic polynomial is $\theta^4 - 5\theta^3 - 2\theta^2 + 7\theta + 4$

Hint: *How about the companion matrix ?.*

10.12 The companion matrix of $a(\theta) = \theta^n + a_{n-1}\theta^{n-1} + \ldots + a_1\theta + a_0 \in F[\theta]$ was defined as $C_a = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ -a_0 & -a_1 & -a_2 & \ldots & -a_{n-1} \end{bmatrix}$ on page 8 of the eigenspectrum (2.1) where it was proved that $\chi(C_a(\theta)) = a(\theta) = \mu(C_a(\theta))$. Some author prefer to define the companion matrix as $\begin{bmatrix} -a_{n-1} & 1 & 0 & \ldots & 0 \\ -a_{n-2} & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ -a_1 & 0 & 0 & \ldots & 1 \\ -a_0 & 0 & 0 & \ldots & 0 \end{bmatrix}$, some others as $\begin{bmatrix} -a_{n-1} & -a_{n-2} & \ldots & -a_1 & -a_0 \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$ still others as $\begin{bmatrix} 0 & \ldots & 0 & -a_0 \\ 1 & \ldots & 0 & -a_1 \\ \vdots & \ldots & \vdots & \vdots \\ 0 & \ldots & 1 & -a_{n-1} \end{bmatrix}$. What happens to the result?

Hint:

$\chi(\theta) = \mu(\theta)$ is still true; just verify.

10.13 Find the similarity classes of 2×2 matrix A over $\mathbb{C}$.

Hint:

There must be two eigenvalues, say α, β. *If there are two independent eigenvectors, the matrix is similar to the diagonal matrix,* $A = P^{-1}\begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} P$ *with* $\alpha = \beta$ *possible. If there is only one independent eigenvector, we have* $A = Q^{-1}\begin{bmatrix} \alpha & 0 \\ 1 & \alpha \end{bmatrix} Q$ *since* $A - \alpha I$ *is similar to* $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ *If there is only one eigenvector in which case* $A - \alpha I = Q^{-1}\begin{bmatrix} \alpha & 0 \\ 1 & \alpha \end{bmatrix} Q \Rightarrow A = Q^{-1}\left\{\begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right\} Q = Q^{-1}\begin{bmatrix} \alpha & 0 \\ 1 & \alpha \end{bmatrix} Q$ *; note that eigenvector means nonzero vector so at least one nonzero eigenvector must be there. Note further that if* $\alpha = \beta$, *we have* $(A - \alpha I)^2 = N^2 = 0$ *i.e.* N *has the single eigenvalue 0 so that the eigenspace of* N *is ker* N *with dim(kerN)= 2; dim(kerT) = 2 corresponds to* $N = 0$ ($\because$ *ker* $N = \mathbb{C}^2$ *then) and we must have* A *diagonalizable with* $A = \begin{bmatrix} \alpha & 0 \\ 1 & \alpha \end{bmatrix}$ *while* $dim(kerN) = 1$ *means that there is just one-element basis for ker* N *which is our situation* $A - \alpha I = Q^{-1}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} Q$ *above.*

10.14 When is $A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \end{bmatrix}$ diagonalizable ?

Hint:

0 is the only eigenvalue so dim(kerA) = 4 i.e. nullity A = 4 so that rank A = 0 i.e. A is similar to O i.e. $a = b = c = 0$.

10.15 Find the eigenvalues of $A = \begin{bmatrix} k & 1 & 0 & \dots & 0 \\ 1 & k & 1 & \dots & 0 \\ 0 & 1 & k & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & k \end{bmatrix}_{n \times n}$ if it is given that X_i = column vector with $j - th$ entry $\{\sin \frac{ij\pi}{n+1}\}$ are eigenvectors.

Hint:

$$AX_i = \begin{bmatrix} k\sin\frac{i\pi}{n+1} + \sin\frac{2i\pi}{n+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ k\sin\frac{n-i\pi}{n+1} + \sin\frac{(n-1)i\pi}{n+1} \end{bmatrix} = [k+2\cos\tfrac{i\pi}{n+1}]X_i \text{ so that the eigenvalues are } k+2\cos\tfrac{i\pi}{n+1}, 1 \leq i \leq n.$$

Suggestion: The syllabus is inordinately long for about 14 lectures allotted. Even so, I have tried to include every term mentioned, with comprehensive arguments and illustrations *(= problems with adequate hints)*in the handouts. For the examination which is rather close, since the treatment is necessarily abstract with new ideas and terminology piling up at every stage of the linear development of the subject, here are a few tips :

1) De-emphasize pages 17-21(*ldota*,canonical forms) of this handout, the next 3 sheets, para 11, on Banach spaces, and also bilinear forms *(pages (19-20))* in the first handout.

2) Concentrate on matrices *(I am saying it reluctantly but never mind)*; it is easier to see what the 'tranjugate' is rather than what the 'Hilbert adjoint' is.

3) Some of the proofs are quite lengthy for some theorems. though they form the core of the subject, you should see how they are used in action; it is not necessary to cram the proof; for that matter, it is bad practice to cram anything.

4) Do not ignore the Dirac notation; it is the easiest way to understand the component forms. Nevertheless, remarks like " a change-of-basis matrix is simply an invertible matrix " are key statements rather than just wily summaries.

5) Work out the illustrations using the hints on paper with a pen/pencil; don't read the handouts like a newspaper.

6) Prepare in advance; don't postpone till the day before the exam.

7) There are additional consultation hours, every Thursday and Monday, 1700-1800, in my office. Use them but don't start studying the handout in search of a question after arriving there, they start 22 April come prepared with a list of queries.

8) Be happy; maths is good fun if treated with respect.

11 While it not the intention of this course to go into Banach spaces beyond the definition, it is important to know that

(i) All finite - dimensional normed spaces are Banach spaces, but

(ii) All finite - dimensional normed spaces are not Hilbert however,

(iii) IPS $\Leftrightarrow$ Hilbert is true for finite dimensional spaces,

(iv) any two norms on a finite dimensional space are equivalent as matrices,

(a) topologically, so anything framed in terms of open sets *(continuity, compactness, convergence etc)*, and

(b) uniformly, so that anything framed in terms of 'uniformities' *(uniform continuity, total boundedness, cauchy sequences etc)* remain the same,

(v) Any linear transformation $X \rightarrow Y$ between two normed finite dimensional spaces *(hence Banach also)* is uniformly continuous hence continuous

(vi) nearly all these statements fail for infinitedimensional spaces.

11.1 On any normed space X *(regardless of the dimension)* the norm function $X \xrightarrow{\|.\|} \mathbb{R}$ is continuous and thus attains its maximum as well as minimum on a compact subset. One can prove that the unit disk $\{x \in X | \parallel x \parallel \leq 1\}$ as well as the unit sphere $\{x \in X | \parallel x \parallel = 1\}$ are compact if $dim\ X < \infty$; indeed, the unit disk is compact iff $dim\ X < \infty$. These things enable one to show that if X and Y are normed spaces and a linear transformation $X \xrightarrow{A} Y$ is continuous, one gets

$$\parallel A \parallel := \quad inf_{\alpha>0}\{\parallel A(x) \parallel \leq \alpha \parallel x \parallel, \quad \text{for all } x \in X\}$$
$$= sup_{\|x\|\leq 1} \parallel Ax \parallel = sup_{\|x\|=1} \parallel Ax \parallel < \infty$$

and that with this definition, $L(X,Y)$ turns into a normed space which is Banach iff Y is Banach. In particular, if $\widehat{X}$ stands for the space of all continuous linear forms, $\widehat{X}$ is Banach whether or not X is Banach. Further,

$$\parallel Ax \parallel \leq \parallel A \parallel . \parallel x \parallel \quad \text{clearly holds.}$$

11.2 For $dim\ X < \infty, dim\ Y < infty,\ anyX \xrightarrow{A} Y$ is continuous *(as stated in 9.5(v) above)* and is represented by some $m \times n$ matrix

$a = [a_i^j]$: $(X\ ,\ \underline{e}) \xrightarrow[a=[a_i^j]]{A} (Y\ ,\ \underline{d})$, dim $X = n < \infty$, $dimY = m < \infty$ as seen earlier *(handout I).*

Thus it makes sense to speak of matrix norm , or norm of a matrix. Since $dimL(X,Y) = mn = dimMat_{m\times n}(F) < \infty$, such a thing is part of discussion about finite dimensional spaces.

(i) But of course, the formula $\parallel A \parallel := sup_{\parallel x\parallel \leq 1} parallel Ax \parallel = sup_{\parallel x\parallel =1} \parallel Ax \parallel < \infty = sup_{\parallel x\parallel \neq 0} \frac{\parallel Ax\parallel}{\parallel x\parallel}$ supplied in 9.6 is not expected to be the only norm turning $L(X,Y)$, and in particular $Mat_{m\times n}(\mathbb{F})$, into a normed space *(necessarily Banach also if $dim(L(X,Y)) < \infty$).* This formula supplies what is known as the operator norm and naturally depends on which two norms on X and Y are being used. We illustrate for $m = n = 2,\ X = Y = \mathbb{C}^2,\ with\ \mathbb{C}^2 \xrightarrow{A} \mathbb{C}^2$ being given by $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$.

For $\parallel z \parallel_1 := \sum_{i=1}^{2} |z^i|$ on both X and Y, the operator norm is $\parallel A \parallel_1 := sup_{\parallel x_1\parallel =1} \parallel Ax \parallel_1 = sup_{\parallel x\parallel_1} [|x_1 - x_2| + x_1 + x_2|] = 1 + 1 = 2$ *(attained at $x^1 = 0, x^2 = 1$ for instance)*

For $\parallel z \parallel_\infty := max_i|z^i|$ on both X and Y, the operator norm is $\parallel A \parallel_\infty := sup_{\parallel x\parallel_\infty =1} \parallel Ax \parallel_\infty = max_{\parallel x\parallel_\infty =1} \begin{pmatrix} |x^1 - x^2|, \\ |x^1 + x^2| \end{pmatrix} = 2$

For $m = n,\ X = Y = \mathbb{C}^n$, with $\parallel z \parallel_2 := +\sqrt{(z|z)}$ on both X and Y, the operator norm $\parallel A \parallel := sup_{\parallel x\parallel_2 =1} \parallel Ax \parallel_2$ is also given by $\parallel A \parallel_2 := \sqrt{\lambda_{max}}$ where λ_{max} is the maximum eigenvalue of A i.e.$\lambda_{max} := \max\limits_{1\leq i\leq n} \{\lambda_i | \lambda_i \text{ is an eigen value of } A\}$. For $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ of (i), that is $\parallel A \parallel_2 = \sqrt{2}$.

We know that $L(X,X)$ is an algebra *(sec III-7)* and in particular, so is therefore $Mat_n(F)$ where $dimX = n < \infty$. Any normed vector space $\mathcal{A}$ which is also an algebra when $\parallel ab \parallel \leq \parallel a \parallel\parallel b \parallel$ is satisfied, and not otherwise, called a normed algebra. Thus while $L(X,X)$ is a normed vector space in various possible ways, it is a normed algebra iff the norm chosen on $L(X,X)$ satisfies $\parallel BA \parallel \leq \parallel B \parallel\parallel A \parallel$ for $A, B \in L(X,X)$. The term matrix norm is applied in general to only those norms on $Mat_n(\mathbb{F})$ which satisfy this inequality

$$\parallel ba \parallel \leq \parallel b \parallel\parallel a \parallel \quad for a, b \in \ \mathrm{Mat}_n(\mathbb{F})$$

.

Addendum: While we promised to say no more on'direct sum', you may still wonder how $\mathbb{F}^2 = \mathbb{F} \bigoplus \mathbb{F}$ is claimed where $\mathbb{F}$ and $\mathbb{F}$ are not disjoint. Actually there is an 'external direct sum' and an 'internal

direct sum', both sliding into each other in the finite - dimensional situation but rather than worrying about all this at present, note that in $\mathbb{F}^2 = \mathbb{F} \bigoplus \mathbb{F}$, we are saying $\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \beta \end{bmatrix}$ so that $\begin{bmatrix} \alpha \\ \alpha \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \alpha \end{bmatrix}$, thus $\begin{bmatrix} \alpha \\ 0 \end{bmatrix} \in \mathbb{F}$, the first summand, $\begin{bmatrix} 0 \\ \alpha \end{bmatrix} \in \mathbb{F}$, the second summand and the two summands are disjoint. Similarly, for $\mathbb{F}^3 = \mathbb{F} \bigoplus \mathbb{F} \bigoplus \mathbb{F}$, we have $\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \beta \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \gamma \end{bmatrix}$ so that the three copies of $\mathbb{F}$ are disjoint. $\mathbb{F}[\theta]$ has infinite sequences $\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ . \\ . \end{bmatrix} = a(\theta)$ and this is why we say $\mathbb{F}[\theta] = \mathbb{F} \bigoplus \mathbb{F} \bigoplus ...$