

MA2101 (Linear Algebra 2)

Jia Cheng

January 2021

1 Definitions and Formula

2 General

Reference Book: Axler's Linear Algebra Done Right

2.1 Well-defined functions

I am inspired to discuss this by the section on quotient spaces.

For a vector space V , subspace U of V , define $V/U = \{v + U : v \in V\}$.

Define vector addition on V/U by $(v + U) + (w + U) = ((v + w) + U)$. Define scalar multiplication on V/U by $\lambda(v + U) = ((\lambda v) + U)$.

One of the steps in proving that our newly defined $+$ is a binary operation on V/U , is to show that it is well-defined. That is, for each distinct element in the domain (V/U), there exists a unique image in the codomain.

For sets like integers, the representation of any element in the domain is unique, so we usually don't have to worry about well-definition of functions. However, in the case of quotient spaces, $v - v' \in U$ implies $(v + U) = (v' + U)$, even though we can have $v \neq v'$. It is because of these multiple representations that we have to worry about "well-defined-ness". We need to make sure when $(v + U) = (v' + U)$, $(w + U) = (w' + U)$, we have $(v + U) + (w + U) = (v' + U) + (w' + U)$.

2.2 Matrices vs Linear Maps

Other than determinants and the section on eigenvectors, MA1101R (Linear Algebra 1) is sufficiently rigorous. The key difference between MA1101R and Axler's Linear Algebra is the approach taken to building up the theory.

An advantage of thinking in terms of linear maps rather than matrices is that linear maps are independent of our choice of bases. The same matrix can do very different things if we decide to change up the bases, but the same linear map remains invariant.

This is especially useful in abstract vector spaces. A generic vector space V does not have a canonical basis (unlike the standard basis in \mathbb{R}^n or polynomials $P(\mathbb{R})$). We can play around with linear maps without having the need to state a basis. (whereas we do if we want to write things in matrix form, or in simultaneous equation form)

An example of the usefulness of this is seen here: <https://math.stackexchange.com/questions/3749/why-do-we-care-about-dual-spaces>. Linear functionals (linear maps that have the field as the codomain) allow the definition of a hyperplane without the need for a basis.

To quote Matt E, the author of the post:

So this gives a reasonable justification for introducing the elements of the dual space to V ; they generalize the notion of linear equation in several variables from the case of \mathbb{R}^n to the case of an arbitrary vector space.

3 Matrices, A revision

3.1 Change of basis

For simplicity, we consider linear operators in \mathbb{R}^2 .

The following notations $M(T, B_1, B_2)$, $M(v, B)$ (respectively matrix representation of a linear transformation T , matrix representation of a vector) are found in Axler's book.

Suppose we have 2 bases, the standard basis (i, j) and another basis (u, v) . Now define the matrix $P = [M(u, (i, j)) \quad M(v, (i, j))]$ Then we will obtain the following equations.

$$\begin{aligned} u &= Pi \\ v &= Pj \end{aligned}$$

There are 2 ways to understand/interpret this equality.

1. First, we "see" P as purely a matrix, and we operator solely in the canonical basis of \mathbb{R}^2 . Then this equality is simply an equality of matrices. i.e. When the standard basis vector i is left multiplied by matrix P , we obtain the vector u , also with respect to the standard basis.
2. Second, we view the equation as a change of basis. To be precise, the first equality can be written as $M(u, (i, j)) = M(I, (u, v), (i, j))M(u, (u, v))$. The second equality is written similarly.
3. Note that regardless of the viewpoint, the above 2 equalities are equalities in matrices! Clearly, i and u are not the same vector.

3.1.1 A possible area of confusion

In this course (or at least chapter 1), Prof Brett usually represents multiplication of matrix P with vector v as Pv . Note that to be fully rigorous, we need to differentiate between an abstract vector and a matrix representation of a vector. For example, if we were to talk about a real polynomial v , which is a vector in $P(\mathbb{R})$, then obviously, matrix multiplication with v is undefined.

So in my opinion, the more rigorous way to write things is $PM(v, B)$, where B is a basis of the vector space v lies in. $M(v, B)$ is that the matrix representation of v w.r.t basis B .

Of course, in chapter 1, we do not deal with abstract vector spaces, so when Prof Brett writes Pv , he means that matrix P is multiplied with the representation of v in the standard basis of \mathbb{R}^n .

3.1.2 MA1101R Recap

Given 2 bases $B_1 = (u_1, u_2)$, $B_2 = (v_1, v_2)$ of \mathbb{R}^2 , then how do we find the change of basis matrix from B_2 to B_1 , $M(I, B_1, B_2)$? (Notice the order "from B_2 to B_1 ". This will be explained in future.)
 $M(I, B_1, B_2) = [M(u_1, B_2), M(u_2, B_2)]$

To find $M(u_i, B_2)$, $i = 1, 2$, we aim to find $a, b \in \mathbb{R}$ s.t. $av_1 + bv_2 = u_i$, i.e. $\begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = (u_i)$

Here, we did not write $M(v_1, \text{basis})$, $M(v_2, \text{basis})$, $M(u_i, \text{basis})$ within the expression, this basis is arbitrary, though ofc it has to be the same for each vector. Usually, this basis would be the canonical

basis of \mathbb{R}^2 .

Next, apply Gaussian Elimination to $\begin{pmatrix} v_1 & v_2 & | & u_1 & u_2 \end{pmatrix}$

3.2 Markov Chains

Given a finite set of states E_1, \dots, E_n , let $A \in \mathbb{R}^{n,n}$ be the matrix where

$$(A)_{i,j} = P(E_{i,m+1} | E_{j,m})$$

where the $E_{j,m}$ denotes the event where state E_j occurs on the m -th trial.

Some properties of the stochastic matrix:

1. Sum of a column: $\sum_{1 \leq i \leq n} (A)_{i,j} = \sum_{1 \leq i \leq n} P(E_i | E_j) = 1$
2. Power of a matrix: $(A^k)_{i,j} = P(E_{i,m+k} | E_{j,m})$

For instance, to use an example from lecture notes:

- $P(\text{rain next day} | \text{rain today}) = 0.6$
- $P(\text{sunny next day} | \text{sunny today}) = 0.7$

The stochastic matrix is then

$$M = \begin{pmatrix} P(R_{i+1} | R_i) & P(R_{i+1} | S_i) \\ P(S_{i+1} | R_i) & P(S_{i+1} | S_i) \end{pmatrix} = \begin{pmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{pmatrix}$$

Then probability theory tells us that

$$\begin{aligned} P(R_{i+2} | R_i) &= P(R_{i+2} | R_{i+1} R_i) P(R_{i+1} | R_i) + P(R_{i+2} | S_{i+1} R_i) P(S_{i+1} | R_i) \\ &= P(R_{i+2} | R_{i+1}) P(R_{i+1} | R_i) + P(R_{i+2} | S_{i+1}) P(S_{i+1} | R_i) = (M^2)_{1,1} \end{aligned}$$

The second equality is due to the assumption that whether it is rainy/sunny on day $i+2$ depends only on whether it rains/is sunny on day $i+1$.

To calculate the actual probabilities: Define $x_n = \begin{pmatrix} P(R_n) \\ P(S_n) \end{pmatrix}$, $\forall n \in \mathbb{N}$

Then

$$\begin{aligned} M^k x_0 &= \begin{pmatrix} P(R_k | R_0) & P(R_k | S_0) \\ P(S_k | R_0) & P(S_k | S_0) \end{pmatrix} \begin{pmatrix} P(R_0) \\ P(S_0) \end{pmatrix} \\ &= P(R_0) \begin{pmatrix} P(R_k | R_0) \\ P(S_k | R_0) \end{pmatrix} + P(S_0) \begin{pmatrix} P(R_k | S_0) \\ P(S_k | S_0) \end{pmatrix} \\ &= \begin{pmatrix} P(R_k \cap R_0) \\ P(S_k \cap R_0) \end{pmatrix} + \begin{pmatrix} P(R_k \cap S_0) \\ P(S_k \cap S_0) \end{pmatrix} \\ &= x_k \end{aligned}$$

3.3 Sums and products of eigenvalues

We claim that trace tr and determinant det are attributes of linear operators, regardless of the choice of basis.

$$\begin{aligned} tr(P^{-1}AP) &= tr(APP^{-1}) = tr(A) \\ det(P^{-1}AP) &= det(P^{-1})det(A)det(P) = det(P^{-1}P)det(A) = det(A) \end{aligned}$$

Hence, if P is the change of basis matrix from eigenvector-basis to standard basis, and A is the matrix of T w.r.t standard basis, we then have

$$\begin{aligned} \text{tr}(D) &= \text{tr}(A) \\ \det(D) &= \det(A) \end{aligned}$$

where D is a diagonal matrix (matrix of T w.r.t eigenvector-basis). This says that $\text{tr}(A)$ is the sum of eigenvalues of linear map T , and $\det(A)$ is the product of eigenvalues of linear map T .

Geometrically (in 2 or 3 dimensions), we can then understand the magnitude of the determinant of a linear map T by the scaling of the area/volume of the parallelogram/parallelepiped bordered by linearly independent eigenvectors after applying linear transformation T on those eigenvectors. Prof Brett calls this the "unit box".

3.4 Useful formula

This result can be proven by expanding the exponential into a sum and applying multivariable calculus. (Simply think of a matrix as a vector)

$$\frac{d}{dt}e^{Bt} = Be^{Bt}$$

Note: Here, $Bt = tB$ (multiplying constant matrix B by scalar t)

Let $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. Then,

$$e^D = \begin{pmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{pmatrix}, e^{Dt} = \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix}$$

where t is a scalar.

4 Vector spaces

In the following chapter, Prof Brett diverges quite significantly from Axler's book. However, what this shows is that Mathematics truly is a "web" as Prof Han Fei says, instead of a series of linear progressions. i.e. There are many places/foundations that we can start from. In the following discussion, I hope to highlight some of the more notable differences from the book, as well as draw some attention to new concepts not discussed in the book.

4.1 Direct Sums

Direct sums are analogous to bases. Whereas a basis allows for the unique representation of vectors in terms of the spans of the basis vectors, a direct sum of vector space allows for the unique representation of a vector by choosing one vector from each vector space.

Let U, V be 2 vector subspaces of some vector space W . The following statements are equivalent.

1. $u_1 + v_1 = u_2 + v_2 \implies u_1 = u_2 \wedge v_1 = v_2$
2. $u + v = 0 \implies u = v = 0$
3. $U \cap V = \{0\}$

Hence any of the 3 statements can be used to define $U \oplus V$. In the case of MA2101, no. 3 is used.

4.1.1 Multiple direct sums

Prof Brett defines

$$\oplus_{i=1}^n U_i = (\dots((U_1 \oplus U_2) \oplus U_3) \oplus \dots) \oplus U_n$$

In particular, $U_1 \oplus U_2 \oplus U_3 = (U_1 \oplus U_2) \oplus U_3$

It turns out that it doesn't matter which order to conduct the direct sum.

The following definition of a multiple direct sum (i.e. direct sum of more than 2 vector spaces) is perhaps more enlightening.

Suppose $\sum_{i=1}^n U_i = \oplus_{i=1}^n U_i$.

Then $\sum_{i=1}^n u_i = 0 \implies \forall 1 \leq i \leq n, u_i = 0$

With this definition in mind, suppose $(U_1 \oplus U_2) \oplus U_3$, and let $u_1 + u_2 + u_3 = 0$. This implies $(u_1 + u_2) + u_3 = 0$, where $u_1 + u_2 \in U_1 + U_2, u_3 \in U_3$. Definition of direct sum for 2 vector spaces says that $u_1 + u_2 = u_3 = 0$.

Now, this reduces to $u_1 + u_2 = 0$. Since $U_1 \oplus U_2$, we then have $u_1 = u_2 = 0$.

Hence, it makes sense to define multiple direct sums in terms of "single" direct sums (i.e. direct sums of 2 vector spaces).

4.2 Vector space isomorphisms

Definition A vector space homomorphism is a linear transformation.

A vector space isomorphism is a bijective linear map.

It is here that the notion of finite-dimensionality is introduced. Note that it is perfectly fine to define finite-dimensional without dimension. We simply need to capture the idea of finiteness.

In Axler's book, a finite-dimensional vector space V is one which has a finite spanning set. i.e. A finite subset of V exists that spans V .

Prof Brett defines finite-dimension via isomorphisms. A finite-dimensional vector space V is one that is isomorphic to \mathbb{F}^n for some $n \in \mathbb{N}$.

Contrary to the book's definition, it is immediately (intuitively) obvious that such an n is unique, if it exists. Suppose not, such that V is isomorphic to both \mathbb{F}^m and \mathbb{F}^n , $m \neq n$. Then since the composition of bijective linear maps gives another bijective linear map, $\mathbb{F}^m \cong \mathbb{F}^n$, which immediately seems absurd. (The rigorous proof for this however, probably still requires the Replacement Theorem. So on a rigorous basis, this probably isn't much easier than Axler's approach.)

4.3 Bases as isomorphisms from \mathbb{F}^n to V

In this course, the convention is to use basis vectors starting from the end of the alphabet. Occasionally, I will also use u, v .

Given a basis $z = (z_1, \dots, z_n)$ of V , we can also define the mapping

$$\begin{aligned} z : \mathbb{F}^n &\rightarrow V \\ z(a) &= \sum_{i=1}^n a_i z_i \end{aligned}$$

where $a = (a_1, \dots, a_n) \in \mathbb{F}^n$. It is easy to verify that z is a bijective linear map, i.e., z is a vector space isomorphism.

Proposition The range/image of a linear transformation $T : U \rightarrow V$ is the span of T applied to the basis vectors of U .

Proposition A vector space isomorphism $\phi : \mathbb{F}^n \rightarrow V$ admits a basis $(\phi(e_1), \dots, \phi(e_n))$.

Proof Since linear map ϕ is injective, the vectors $(\phi(e_1), \dots, \phi(e_n))$ span V . Since ϕ is injective, $\sum_{i=0}^n a_i \phi(e_i) = 0 \implies \phi(\sum_{i=0}^n a_i e_i) = 0 \implies \sum_{i=0}^n a_i e_i = 0 \implies \forall i, a_i = 0$, which proves linear independence. Hence we have a basis.

Hence, "there is a one-one correspondence between isomorphisms of the form $\mathbb{F}^n \rightarrow V$ and bases".

Here is another way to think about a basis $z = (z_1, \dots, z_n)$ of V . In Axler's book, the matrix representation of a vector $v \in V$ with respect to a basis z is denoted as $M(v, z)$. So if we define the linear map $M(\cdot, z) : V \rightarrow \mathbb{F}^{n,1} \cong \mathbb{F}^n$, we then notice that this is the inverse function of $z : \mathbb{F}^n \rightarrow V$.

4.4 Dimension

In Axler's book, dimension of a vector space is defined to be the number of vectors in a basis of V . This is well-defined since by the Replacement Theorem, the number of vectors in any basis of a vector space is the same.

In MA2101, as mentioned previously, defining a vector space V to be finite-dimensional if it is isomorphic to \mathbb{F}^n for some $n \in \mathbb{N}$ suggests a natural definition for dimension – the value of n . (Of course, as mentioned previously, a rigorous treatment of this would still require proving that this n is unique if it exists.) Now, we take this for granted, that n is unique.

Proposition The dimension of a finite-dimensional vector space V is equal to the number of vectors in any basis.

Proof Consider a basis $z = (z_1, \dots, z_n)$ of V . z then corresponds to some vector space isomorphism $\phi_z : \mathbb{F}^n \rightarrow V$ as defined above. (The vectors of basis z are then given by $\phi_z(e_1), \dots, \phi_z(e_n)$.) By definition of dimension, this precisely says that the dimension of V is n .

5 Linear maps

5.1 Change of basis

As mentioned in the previous section, for any basis $z : \mathbb{F}^n \rightarrow V$, we have $z^{-1} = M(\cdot, z) : V \rightarrow \mathbb{F}^n$.

Let y, z be 2 bases of vector space V .

Let $P = z^{-1} \circ y : \mathbb{F}^n \rightarrow \mathbb{F}^n$.

We then have $y = z \circ P$. Here, P is a change of basis from z to y (**note the order!**). In the notation of Axler's book, $P = M(I, y, z)$. (I is the identity linear map.) For any $v \in V$, $M(I, y, z)M(v, z) = M(v, y)$.

Another way to remember the bolded part above is to think of the change of basis of a linear operator $T \in L(V)$.

$$M(T, y, y) = M(I, z, y)M(T, z, z)M(I, y, z) = P^{-1}M(T, z, z)P$$

Hence P changes the basis on which T acts from z to y .

Given a certain basis z of V , every other basis can be expressed in the form of $z \circ P$ for some isomorphism $P : \mathbb{F}^n \rightarrow \mathbb{F}^n$.

5.2 Dual spaces

Given a vector space V over \mathbb{F} . The dual space of V , the set of all linear functionals on V , is denoted $V' = \hat{V} = L(V, \mathbb{F})$.

5.3 Kernel and range

In MA2101, the null space/kernel of a linear map $T : V \rightarrow W$ is denoted $\ker(T)$. $\dim \ker(T) = \text{null}(T)$

The range of T is denoted $\text{rang}(T)$. $\dim \text{rang}(T) = \text{rank}(T)$

In Axler's book, rank is actually defined in terms of the column/row rank of a matrix representation of T . Then, rank is proven to be equal to $\dim \text{rang}(T)$ by establishing an isomorphism between the column space of $M(T, B1, B2)$ and the range of T .

5.3.1 Fundamental Theorem of Linear Maps

Given linear map $T : V \rightarrow W$,

$$\dim V = \text{null}(T) + \text{rank}(T)$$

There are 2 ways to prove the dimension theorem. The first way (Axler) is to construct a basis $B_{\text{null}(T)} = \{y_1, \dots, y_k\}$ of $\ker(T)$, then extend it to a basis $B_V = \{y_1, \dots, y_k, z_1, \dots, z_m\}$ of V . Show that (Tz_1, \dots, Tz_m) is a basis for $\text{range}(T)$. Hence,

$$\dim(V) = k + m = \text{null}(T) + \text{rank}(T)$$

The second way (MA2101) is to find a subspace U of V such that $V = \text{null}(T) \oplus U$. It is possible to find such a subspace via basis extension. Show that $U \cong \text{range}(T)$ by considering the restriction of T to U (also restrict codomain V to $\text{range}(T)$).

$$T|_U : U \rightarrow \text{range}(T)$$

Since isomorphism by definition (in MA2101) implies equal dimension, we are done.

5.4 Dual vectors

Proposition Given a basis $z = (z_1, \dots, z_n)$ of a vector space V , and any choice of vectors $w_1, \dots, w_n \in W$, a linear map $T : V \rightarrow W$ is uniquely identified by $T(z_i) = w_i, i \in \{1, \dots, n\}$.

The proof is trivial.

Corollary Given a basis $z = (z_1, \dots, z_n)$ of a vector space V , and linear maps $T_1, T_2 : V \rightarrow W$. To check if $T_1 = T_2$, it suffices to check that $\forall i \in \{1, \dots, n\}, T_1(z_i) = T_2(z_i)$

Hence we can define dual basis vectors as follows.

Definition $\forall i, j \in \{1, \dots, n\}, \zeta_i(z_j) = \delta_{ij}$.

Prof Brett uses Einstein's notation, however, I decide to stick with the classical subscript notation. However, I will still follow certain conventions for this course, e.g. scalars start from the start of the alphabet; basis vectors start from the end of the alphabet.

To show that ζ is indeed a basis for \hat{V} , we need to check for linear independence and span. Linear independence is easy. As for span, for any $\alpha \in \hat{V}$, define $\beta = \sum_i (\alpha z_i) \zeta_i$. Using the corollary above, it suffices to check for $\forall j \in \{1, \dots, n\}, \alpha z_j = \beta z_j$. Hence $\alpha = \beta$ and ζ spans \hat{V} .

Proposition $\alpha \sum_i p_i \zeta_i \iff p_i = \alpha z_i$

The backward direction is proven above. The forward direction is proven by applying α to z_j for each $j = 1, \dots, n$.

Now, temporarily forget the original set-based definition of the dual basis. We want to see if we can define the dual basis using an isomorphism, i.e. a map-based definition.

So now we don't have a general definition of dual basis. However, we still wish to define a special case for the dual space of $\mathbb{F}^n, \hat{\mathbb{F}}^n$. The dual basis vectors of $\hat{\mathbb{F}}^n$ are the row vectors $(\epsilon_1, \dots, \epsilon_n)$, where $\epsilon_i = (e_i)^T$.

Based on this special case, we now define the following:

Definition Given a basis z of V , the dual basis ζ of \hat{V} is the isomorphism $\zeta : \hat{\mathbb{F}}^n \rightarrow \hat{V}$ such that

$$\forall a \in \mathbb{F}^n, \forall p \in \hat{\mathbb{F}}^n, (\zeta p)(za) = pa$$

Now, I want to recover the original set-based definition of a dual basis. Let $p = \epsilon_i, a = e_j$, then $pa = \delta_{ij}$.

By definition $(\zeta \epsilon_i)(z_j) = \delta_{ij}$. Since ζ is an isomorphism, it corresponds to a set-basis with $(\zeta_1, \dots, \zeta_n)$, where $\zeta_i = \zeta \epsilon_i$. Hence, $\zeta_i z_j = \delta_{ij}$.

And we have recovered the original set definition of a dual basis.

5.4.1 Vector acting on dual vector

I introduce the following notation:

Given $v \in V, \alpha \in \hat{V}$,

$$\alpha(v) = \alpha v = v[\alpha]$$

In MA2101, Prof Brett does not use square brackets, instead just writing $v(\alpha) = v\alpha$, however I find this to be potentially confusing.

The motivation for writing this will come later (e.g. when showing that the matrix representation of the transpose/dual map \hat{T} of T is really the transpose matrix of T).

Definition Given $V, W, T \in L(V, W)$, the **dual map**(Axler)/**transpose**(MA2101) of T is denoted by $\hat{T} \in L(\hat{W}, \hat{V})$.

$$\forall \alpha \in \hat{W}, \hat{T}(\alpha) = \alpha \circ T$$

Hence, $\forall v \in V, v[\hat{T}\alpha] = (\alpha \circ T)v = \alpha T v$. Note that the convention of linear maps is to forego the use of round brackets.

Notice the symmetry here.

5.5 Tensor products

Definition The tensor product \otimes is a bilinear map defined as follows:

$$\begin{aligned} \otimes : V \times \hat{V} &\rightarrow L(V) \\ (v, \alpha) &\mapsto \alpha(\cdot)v \end{aligned}$$

That is, for each $w \in V, (v \otimes \alpha)w = \alpha(w)v$, multiplying v by scalar αw .

The transpose/dual map of $T = (v \otimes \alpha)$ has the following "nice" property.
For all $\beta \in \hat{V}$, $\hat{T}\beta = \beta \circ (v \otimes \alpha) \in \hat{V}$. Applying this to any arbitrary $w \in V$, we get

$$\begin{aligned} (\beta \circ (v \otimes \alpha))w &= \beta(\alpha(w)v) \\ &= \alpha(w)\beta(v) && \text{linearity of } \beta \\ &= \beta(v)\alpha(w) && \text{commutativity of scalar multiplication} \\ &= (\beta(v)\alpha)(w) && \text{scalar multiplication of functions} \end{aligned}$$

Hence, $\hat{T}\beta = (\beta \circ (v \otimes \alpha)) = \beta(v)\alpha$

5.5.1 Basis of $L(V)$

We are given that $z = (z_i)_{1 \leq i \leq n}$ is a basis of V .

Let $\zeta = (\zeta_i)_{1 \leq i \leq n}$ be the dual basis of \hat{V} .

Consider the following set

$$B = \{z_i \otimes \zeta_j : i, j \in \{1, \dots, n\}\}$$

We claim that B is a basis for $L(V)$.

Claim: B spans $L(V)$.

Let $T \in L(V)$.

Define the scalars $T_{i,j}, i, j \in \{1, \dots, n\}$ as follows: For each $1 \leq j \leq n$, $T(z_j) = \sum_{1 \leq i \leq n} T_{i,j} z_i$.

Define

$$S = \sum_{1 \leq i, j \leq n} T_{i,j} z_i \otimes \zeta_j$$

Then for each $1 \leq k \leq n$,

$$\begin{aligned} S(z_k) &= \sum_{1 \leq i, j \leq n} T_{i,j} \zeta_j(z_k) z_i \\ &= \sum_{1 \leq i, j \leq n} T_{i,j} \delta_{jk} z_i \\ &= \sum_{1 \leq i \leq n} T_{i,k} z_i && \text{Substitute } j = k \\ &= Tz_k && \text{definition of } T_{i,j} \end{aligned}$$

Since linear maps are uniquely determined by their actions on basis vectors, $S = T$. This says that B spans $L(V)$.

Claim: B is linearly independent.

Suppose there exists $\lambda_{i,j}, 1 \leq i, j \leq n$ such that

$$\sum_{1 \leq i, j \leq n} \lambda_{i,j} z_i \otimes \zeta_j = 0_{L(V)}$$

Then, for each $1 \leq k \leq n$,

$$\begin{aligned} 0_V = 0_{L(V)}(z_k) &= \sum_{1 \leq i, j \leq n} \lambda_{i,j} \zeta_j(z_k) z_i \\ &= \sum_{1 \leq i, j \leq n} \lambda_{i,j} \delta_{jk} z_i \\ &= \sum_{1 \leq i \leq n} \lambda_{i,k} z_i \end{aligned}$$

Since z is a basis, it is a linearly independent set, hence $\forall 1 \leq i \leq n$, $\lambda_{i,k} = 0$. k runs from 1 to n , so we have $\lambda_{i,k} = 0$ for all $1 \leq i, k \leq n$. This proves linear independence of B .

Definition $T_{i,j}$ are called the components of $T \in L(V)$ with respect to basis z .

Before, the definition of $T_{i,j}$ is implicitly given by $T(z_j) = \sum_{1 \leq i \leq n} T_{i,j} z_i$. Here is a more explicit way to express an individual entry of $T_{i,j}$.

$$\forall 1 \leq i, j \leq n, T_{i,j} = \zeta_i(Tz_j)$$

5.6 Matrices

In MA2101, the (i, j) -th entry of a n times m matrix $M \in \mathbb{F}^{n,m}$ is denoted as M_j^i . I will use the "classical" $M_{i,j}$ instead.

If the entries of a matrix $M_{i,j}$ are known and equal $a_{i,j}$, then M can be denoted as $(a_{i,j})$.

All the matrix notation here onwards will largely follow Axler's book.

Proposition Given $M \in \mathbb{F}^{n,m}$, $N \in \mathbb{F}^{m,p}$,

$$MN = \sum_{1 \leq i \leq m} M_{\cdot, i} N_{i, \cdot}$$

This is covered in MA1101R and is one of the multiple ways with which to view matrix multiplication.

Given a linear operator $T \in L(V)$, we know from the previous section that $T = \sum_{1 \leq i, j \leq n} T_{i,j} (z_i \otimes \zeta_j)$. Its matrix representation with respect to the basis z is $(T_{i,j}) = (\zeta_i(Tz_j))$.

Theorem The matrix representation of \hat{T} with respect to basis ζ is the transpose of the matrix representation of T w.r.t basis z .

Remark Prof Brett actually refers to the matrix representation of \hat{T} relative to basis z , since ζ is derived from z . However, personally, I'd like to emphasise that \hat{T} maps from \hat{V} to \hat{V} , so when referring to the basis I also would refer to the dual basis of \hat{V} .

Proof We have (1)

$$T_{i,j} = \zeta_i(Tz_j) = (\zeta_i \circ T)z_j = (\hat{T}\zeta_i)z_j = z_j[\hat{T}\zeta_i]$$

Notice that as \hat{T} is just another linear operator, we have already defined $\hat{T}_{i,j}$ implicitly in the previous section as

$$\hat{T}\zeta_j = \sum_{1 \leq i \leq n} \hat{T}_{i,j} \zeta_i$$

Hence, (2)

$$\begin{aligned} z_k[\hat{T}\zeta_i] &= \sum_{1 \leq i \leq n} \hat{T}_{i,j}\zeta_i(z_k) \\ &= \hat{T}_{k,j} \end{aligned}$$

Hence, combining (1) and (2), $\forall 1 \leq i, j \leq n$,

$$T_{i,j} = z_j[\hat{T}\zeta_i] = \hat{T}_{j,i}$$

Hence, in matrix form, $(\hat{T}_{i,j}) = (T_{i,j})^t$.

5.7 Linear maps $L(V, W)$, $V \neq W$

As a convention, let $\dim(V) = n, \dim(W) = m$

Define the tensor product:

$$\begin{aligned} \otimes : W \times V' &\rightarrow L(V, W) \\ (w, \alpha) &\mapsto \alpha(\cdot)w \end{aligned}$$

i.e. $\forall v \in V, (w \otimes \alpha)v = \alpha(v)w$

This definition makes a lot more sense when we try to define the basis of $L(V, W)$ in terms of the tensor product.

Let $z = (z_1, \dots, z_n)$ be a basis for V . Let $x = (x_1, \dots, x_m)$ be a basis for W .

Proposition The set

$$\{x_i \otimes \zeta_j : 1 \leq i \leq m, 1 \leq j \leq n\}$$

is a basis for $L(V, W)$.

5.8 Change of Basis

For simplicity, we first consider a single vector space V .

Let $z, y : \mathbb{F}^n \rightarrow V$ be 2 bases of V , and ζ, γ be the corresponding dual bases. A change of basis affects the matrix representation of the following 3 items:

1. Linear maps $T \in L(V)$
2. Vectors $v \in V$
3. Dual vectors $\alpha \in \hat{V}$

Define $P = z^{-1} \circ y$.

Lemma 1 $M(I, y, z) = M(P, e, e)$

For convenience, we write $M(P, e) = M(P, e, e)$, since the two bases are the same.

Proof Consider the (i, j) -th entry of the matrix of I w.r.t bases y, z .

$$\begin{aligned} M(I, y, z)_{i,j} &= \zeta_i I y_j = (\epsilon_i z^{-1}) I (y e_j) \\ &= \epsilon_i (z^{-1} y) e_j = \epsilon_i P e_j = M(P, e)_{i,j} \end{aligned}$$

Since all the entries of the 2 matrices are the same, we have the desired result. **Remark** A similar derivation gives $M(I, z, y) = M(P^{-1}, e)$

5.8.1 Linear maps

I will present 3 different derivations. The first 2 approaches also make use of the result the matrix of composite linear maps is the product of their matrices. Comparing **1** with **2**, they both make use of the same ideas, but the first is shorter.

1 Axler

$$M(T, y) = M(I, z, y)M(T, z)M(I, y, z) = M(P^{-1}, e)M(T, z)M(P, e)$$

The last equality is from the use of the lemma above.

2 MA2101 Method 1 Note that $M(T, z)_{i,j} = \zeta_i T z_j = \epsilon_i z^{-1} T z e_j = M(z^{-1} T z, e)$. We can similarly replace z with y .

$$M(T, y) = M(y^{-1} T y, e) = M((zP)^{-1} T (zP), e) = M(P^{-1}(z^{-1} T z)P, e) = M(P^{-1}, e)M(T, z)M(P, e)$$

3 MA2101 Method 2 See below. This will be presented separately.

5.8.2 Vectors

Let $v \in V$. Define the matrix of v with respect to a basis z as $M(v, z)$. Clearly, $M(v, z) = M(z^{-1}v, e)$

For simplicity/laziness, I will identify the column matrix $M(z^{-1}v, e) \in \mathbb{F}^{n,1}$ as just the column vector $z^{-1}v \in \mathbb{F}^n$. Technically they are different mathematical objects.

I also implicitly make use of the result that $M(Tu, basis) = M(T, basis)M(u, basis)$

Then $M(v, y) = y^{-1}v = (z \circ P)^{-1}v = P^{-1}z^{-1}v = M(P^{-1}, e)M(v, z)$

5.8.3 Dual vectors

Let $\alpha \in \hat{V}$.

Denote $e = \{1\}$ as the basis of \mathbb{F} (when viewed as a vector space).

The dual basis of e is $\epsilon = \{1\}$.

Denote e' as the basis of \mathbb{F}^n .

Then $M(\alpha, y, e) = M(\alpha, z, e)M(I, y, z) = M(\alpha, z, e)M(P, e')$

Alternative derivation:

Notice that $1\alpha y_j = 1(\alpha y)(e'_j)$ for each $1 \leq j \leq n$.

Hence, $M(\alpha, y, e) = M(\alpha y, e', e) = M(\alpha z P, e', e) = M(\alpha z, e', e)M(P, e', e')$

We can observe that the alternative way is rather analogous to how Prof Brett does it in **MA2101 Method 1** as stated above.

5.8.4 MA2101 Method 2

Lemma 2 $\epsilon_j \circ P = \sum_{1 \leq i \leq n} P_{ji} \epsilon_i$

Proof Clearly, $\epsilon_j \circ P \in \hat{\mathbb{F}}^n$.

For each $1 \leq k \leq n$, $\epsilon_j \circ P e_k = \epsilon_j \sum_i P_{ik} e_i = P_{jk}$. Also, $\sum_{1 \leq i \leq n} P_{ji} \epsilon_i(e_k) = P_{jk}$. And we are done.

We first summarise how the basis vectors of y, z, γ, ζ are represented in terms of each other.

- $y_j = y(e_j) = (zP)e_j = z \sum_i P_{ij} e_i = \sum_i P_{ij} z_i$
- Similarly, $z_j = \sum_i P_{ij}^{-1} y_i$

- Using Lemma 2, $\gamma_j = \epsilon_j \circ y^{-1} = \epsilon_j \circ (zP)^{-1} = (\epsilon_j P^{-1})z^{-1} = (\sum_i P_{j,i}^{-1} \epsilon_i)z^{-1} = \sum_i P_{j,i}^{-1} \zeta_i$
- Similarly, $\zeta_j = \sum_i P_{j,i} \gamma_i$

Vectors

Let v be a vector in V . Note that $z = y \circ P^{-1}$.

Denote $a_j := M(v, z)_{j,1}$

$$v = \sum_j a_j z_j = \sum_j a_j \sum_i P_{i,j}^{-1} y_i = \sum_i (\sum_j P_{i,j}^{-1} a_j) y_i$$

Hence, $M(v, y)_{i,1} = \sum_j P_{i,j}^{-1} M(v, z)_{j,1}$ for each $1 \leq i \leq n$. Hence, $M(v, y) = M(P^{-1}, e)M(v, z)$

We deal with dual vectors and linear maps in a similar fashion.

6 Eigenthings

Definition Given a linear operator T on a vector space V , a subspace $U \subseteq V$ is an eigenray if it is invariant with respect to T and is one-dimensional.

Proposition The eigenvalue of an eigenray is unique.

This is a consequence of the linearity of a linear operator.

Proposition Given n eigenrays R_1, \dots, R_n of a vector space V with pairwise different eigenvalues, and any choice of n eigenvectors $v_i \in R_i$ (one from each eigenray), then the set (v_1, \dots, v_n) is linearly independent.

Proof By induction on n .

A special case for $n = 2$ can be proven as follows. If v_1, v_2 are linearly dependent, then $v_2 = kv_1$ for some scalar k . But this says that v_2 lies in $\text{span}(v_1) = R_1$. It then follows directly that $R_1 = R_2$. But we know that the an eigenray only has one eigenvalue. Contradiction.

Corollary There are at most $\dim(V)$ distinct eigenrays in a finite-dimensional vector space V .

This is a consequence of the fact that a linearly independent set of vectors in a finite-dimensional vector space V can have at most $\dim(V)$ vectors. The proof is by Steinitz exchange/replacement lemma.

6.1 Fundamental Theorem of Algebra

We now assume the fundamental theorem of algebra without proof.

Proposition Every linear operator over a complex vector space V has an eigenvalue.

Consider $v, Tv, \dots, T^n v$, where $n = \dim(V)$. Since there are more than n vectors, they cannot be linearly independent. That is, there exists scalars, $\lambda_i, i \in \{0, \dots, n\}$ such that $\sum_{0 \leq i \leq n} \lambda_i T^i v = 0$.

$$\begin{aligned} \sum_{0 \leq i \leq n} \lambda_i T^i v &= \left(\sum_{0 \leq i \leq n} \lambda_i T^i \right) v \\ &= a(T - a_1 I) \dots (T - a_n I) v \quad \text{by FTA and linearity of } T \end{aligned}$$

Note 1: Recall that for all functions, linear or not, we have $(f + g)h = fh + gh$. However, additionally linear maps are bilinear under composition, that is, $f(g + h) = fg + fh$. This is what allows the second equality to hold.

Note 2: It is also important that V is a vector space over \mathbb{C} , otherwise, we would not be able to do the factorisation $a(T - a_1 I) \dots (T - a_n I)$, where a_i are possibly complex scalars. If V was a vector

space over \mathbb{R} , then $(T - iI)$ for instance would not be an operator over a real vector space, since $(T - iI)v = Tv - iv$ and we are trying to multiply v by a complex scalar. Additionally, if V was a vector space over \mathbb{R} , all the λ_i must also be real.

This lemma will allow us to "kickstart" our proof of the next theorem.

6.2 Upper triangular and diagonal matrices

Definition Given matrices A, B , say $A \equiv B$ or A is similar to B if there exists invertible matrix P such that $A = P^{-1}BP$.

Proposition Given a linear operator $T \in L(V)$, where V is a vector space over \mathbb{C} , there exists a choice of basis y of V such that $M(T, y)$ is upper triangular.

Note that this is equivalent to stating that given any basis z of V , there exists P such that $M(T, y) = M(P^{-1}, e)M(T, z)M(P, e)$. i.e. $M(T, z)$ is similar to some upper triangular matrix $M(T, y)$.

(For simplicity, we will denote $M(T, y, y)$ with $M(T, y)$.)

Before proceeding with the proof, we need to recall a few things.

- By choosing any 2 bases y, z of a vector space V , $P := z^{-1} \circ y$ is a change of basis map. P is the composition of 2 isomorphisms, so it is itself an isomorphism.
- We choose one basis z of V and some isomorphism $P \in L(\mathbb{F}^n)$. Then $y := z \circ P$ is another basis for V .
- As a consequence, we have the following results:
- Given 2 bases y, z and $P := z^{-1} \circ y$, we have $M(T, y) = M(P^{-1}, e)M(T, z)M(P, e)$.
- Given a basis z , isomorphism $P \in L(\mathbb{F}^n)$ and $y := z \circ P$, we have $M(T, y) = M(P^{-1}, e)M(T, z)M(P, e)$

My point here is that we can either go from z, P to y , or from z, y to P .

Proof The proof is by induction on dimension of V , n . Suppose $P(n)$ true. That is, for any n -dimensional vector space V , $T \in L(V)$, there exists a choice of basis y such that $M(T, y)$ is upper triangular.

Let V be a $n + 1$ -dimensional complex vector space. Let $T \in L(V)$. We define the basis z as follows. We know that T must have at least 1 eigenvalue, and hence some eigenvector. So we let z_1 be an eigenvector of T , and extend z_1 arbitrarily to a basis z for V .

Consider $M(T, z)$. The first column of $M(T, z)$ is $\begin{pmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ by our choice of z_1 . Hence $M(T, z)$ has the form

$$\begin{pmatrix} \lambda_1 & \mathbf{r} \\ \mathbf{0} & Q \end{pmatrix}$$

where $\mathbf{r} \in \mathbb{C}^{1,n}$, $\mathbf{0} \in \mathbb{C}^{n,1}$, $Q \in \mathbb{C}^{n,n}$

Note that Q is an $n * n$ matrix, and hence can be viewed as a linear operator on \mathbb{C}^n . So by induction hypothesis, there exists a matrix $M(P_1, e)$ such that $P_1^{-1}QP_1$ is upper triangular.

Now define $M(P, e) = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & M(P_1, e) \end{pmatrix}$ and we have the desired P .

See block matrix multiplication for more details.

Theorem The eigenvalues of a linear map T are precisely the diagonal entries of an upper triangular matrix A of T .

Proof First, we note that for any scalar λ and any basis z ,

$$\begin{aligned} M(T - \lambda I, z) &= M(T, z) - \lambda M(I, z) \\ &= A - \lambda I. \end{aligned}$$

Note that I in the first line is the identity map and I in the second line is the identity matrix.

Hence, λ is a diagonal entry of A

$\iff A - \lambda I$ is a non-invertible/singular matrix

$\iff T - \lambda I$ is a non-invertible/singular linear operator

$\iff \lambda$ is an eigenvalue of T .

6.3 Jordan Canonical Form

See lecture notes for definition of Jordan block and Jordan form.

As with any upper triangular matrix of a linear operator T , the diagonal entries of the Jordan canonical form of T are precisely the eigenvalues of T .

Definition The multiplicity of an eigenvalue λ is the sum of the sizes of Jordan blocks corresponding to that eigenvalue.

Remark It is incorrect to say the multiplicity of an eigenvalue λ is simply the number of occurrences of λ along any upper triangular matrix. This is because the Jordan canonical form has certain desired characteristics that makes this definition of multiplicity useful in the Cayley-Hamilton Theorem.

Remark 2 It turns out that the multiplicity of an eigenvalue λ can be the number of occurrences of λ along any upper triangular matrix (same as in the Jordan matrix form). The proof of this, however, is non-trivial and requires the notion of generalised eigenvalues.

Definition Given that $\lambda_i, 1 \leq i \leq k$ are precisely the eigenvalues of a linear operator T , and that m_i is the multiplicity of eigenvalue λ_i , the characteristic polynomial of T is

$$\chi_T(x) = \prod_{i=1}^k (x - \lambda_i)^{m_i}$$

Remark A reminder that eigenvalues and eigenvectors are properties of a linear operator, independent of choice of basis. Hence the characteristic polynomial is also a property of an operator, independent of basis.

6.3.1 Cayley-Hamilton Theorem

$$\chi_T(T) = 0_{L(V)}$$

To prove this, let y be a Jordan basis of T , such that $M(T, y)$ is in Jordan canonical form. We can show via block matrix multiplication that $\chi_T(M(T, y)) = 0_{\mathbb{C}^{n,n}}$.

Since $M(\chi_T(T), y) = \chi_T(M(T, y)) = 0_{\mathbb{C}^{n,n}}$, we have $\chi_T(T) = 0_{L(V)}$.

Proposition Let z be any basis of V . Then

$$\chi_T(M(T, z)) = 0_{\mathbb{C}^{n,n}}$$

Proof Since $\chi_T(T) = 0_{L(V)}$, i.e. $\chi_T(T)$ is the zero map, clearly any matrix of $\chi_T(T)$ is the zero matrix regardless of basis. (since $M(\chi_T(T), z) = \chi_T(M(T, z))$) Hence, we have the desired conclusion.

Additionally, let A be any matrix that is similar to $J = M(T, y)$. Then there exists invertible matrix P such that $A = P^{-1}JP$. Let $Q \in L(\mathbb{C}^n)$, $P = M(Q, e)$. Define $z = y \circ Q$. Then Q is a change of basis isomorphism. By our proposition above,

$$\chi_T(A) = \chi_T(M(T, z)) = 0_{\mathbb{C}^{n,n}}$$

An application of the Cayley-Hamilton Theorem is that when taking powers of linear operators, higher powers can be expressed in terms of lower powers.

Corollary Suppose the characteristic polynomial of linear operator T has degree k . Then all powers of $T^j, j \geq k$ can be expressed as a linear combination of I, T, \dots, T^{k-1} .

Similarly, if T is invertible, then T^{-1} can be expressed as a linear combination of I, T, \dots, T^{k-1} .

Stuff: To be integrated into the inner products section Given linear operator $T \in L(V)$, where V is a complex vsp. Let z be *any* basis of V . Let y be an orthonormal eigen-basis.

Let $M(P, e, e)$ be the matrix $M(I, y, z)$. Then $y_j = \sum_i P_{i,j} z_i$.

Prof's Claim: The columns of $M(P)$, $c_i = \sum_j P_{ij} e_j$ are orthonormal.

It suffices to show $\text{dot}(c_i, c_j) = \delta_{ij}$

$$M(P) = (c_1 \dots c_n)$$

$$\begin{aligned} \text{dot}(c_i, c_j) &= \text{dot}\left(\sum_k P_{ki} e_k, \sum_l P_{lj} e_l\right) \\ &= \sum_{k,l} P_{ki} P_{lj} \text{dot}(e_k, e_l) \\ &= \sum_k P_{ki} P_{kj} \end{aligned}$$

$$\begin{aligned} \delta_{ij} &= g(y_i, y_j) \\ &= g(zc_i, zc_j) \\ &= \sum_{k,l} P_{ki} P_{lj} g(z_k, z_l) \end{aligned}$$

Counterexample:

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

Hence, this is a slight inaccuracy in the lecture notes.

We need z to be orthonormal for this proof to follow through.

7 Inner Product Spaces

7.1 Bilinear forms

Given a vector space V of dimension n , and $B(V)$, the set of all bilinear forms over V (which is also a vector space), the basis of $B(V)$ is given by

$$\{\zeta_i \otimes \zeta_j : 1 \leq i, j \leq n\}$$

where ζ is a dual basis of V .

Denote the matrix of a bilinear form g with respect to basis z of V as $M(g, z)$ (the z here refers to both inputs to g). For each $1 \leq i, j \leq n$, $M(g, z)_{i,j} = g(z_i, z_j)$.

7.1.1 Change of basis

Let z, y be 2 bases. Let $P = z^{-1} \circ y$ be the change of basis from z to y .

Recall from the previous chapter, we can express $z_i, y_i, \zeta_i, \gamma_i$ in terms of each other as follows: TODO

7.2 Real inner products

A real inner product g is a bilinear form that has the following properties:

1. Positivity: $\forall v \in V, g(v, v) \geq 0$
2. Definiteness: $g(v, v) = 0 \iff v = 0$
3. Symmetry: $\forall u, v \in V, g(u, v) = g(v, u)$

7.2.1 Matrix of a real inner product

Let $M(g, z)$ be the matrix of real inner product g with respect to basis z of V . Denote $g_{i,j} = M(g, z)_{i,j}$. Let $u, v \in V$. Then $u = \sum_i a_i z_i, v = \sum_i b_i z_i$ for some scalars $a_i, b_i \in \mathbb{R}$.

The matrix of a real inner product has these properties:

1. Positivity: $g(u, u) = \sum_{i,j} a_i a_j g_{i,j} \geq 0$
2. Definiteness: $\sum_{i,j} a_i a_j g_{i,j} = 0 \iff \forall 1 \leq i \leq n, a_i = 0$
3. Symmetry: $g_{i,j} = g_{j,i}$

Remark $g_{i,j} = g_{j,i}$ is a sufficient and necessary condition for g to be symmetric.

Proposition Computing inner products by matrix multiplication

Use the same definitions of $g_{i,j}, u, v$ as above. Then the 1×1 matrix of $g(u, v)$ is given by

$$[g(u, v)] = M(u, z)^T M(g, z) M(v, z)$$

Proof Observe that

$$[g(u, v)]_{1,1} = \sum_{i,j} a_i b_j g_{i,j} = \sum_{i,j} a_i g_{i,j} b_j = \sum_{i,j} (M(u, z)^T)_{1,i} M(g, z)_{i,j} M(v, z)_{j,1}$$

and the result follows.

Remark Note that this is for a **real** inner product. For complex inner product, the transpose is replaced by conjugate transpose.

Proposition Matrix of an inner product cannot have non-positive eigenvalues.

Proof Let A be a matrix of real inner product g . Suppose A has a non-positive eigenvalue $\lambda \leq 0$. Then there exists eigenvector $v \in \mathbb{R}^n, v \neq 0$, such that $Av = \lambda v$. Note that v corresponds to some vector $u \in V$, i.e. $M(u, z) = v$.

Hence, $g(u, u) = v^T Av = v^T \lambda v = \lambda v^T v = \lambda |v|^2 \leq 0$. (Note that the norm here is the Euclidean norm since $v \in \mathbb{R}^n$).

Then g is not positive definite. Hence g cannot be an inner product.

7.3 Orthogonality

Definition z is an orthonormal basis of real inner product space (V, g) if for each $1 \leq i, j \leq n$, $g(z_i, z_j) = \delta_{i,j}$.

Hence $M(g, z)$ is the identity matrix.

An orthonormal basis can also be defined in the "mapping" way.

Definition z is an orthonormal basis of real inner product space (V, g) if $\forall a, b \in \mathbb{R}^n$,

$$g(za, zb) = \text{dot}(a, b)$$

(Note: dot here is the dot product) We can prove that the 2 definitions are indeed equivalent.

Proposition Pythagorean Theorem

Let $(y_i)_{1 \leq i \leq k}$ be an orthonormal list of vectors. Then

$$\left| \sum_{1 \leq i \leq k} a_i y_i \right|^2 = \sum_{1 \leq i \leq k} a_i^2$$

Corollary Orthonormal list of vectors are linearly independent.

7.4 Riesz Representation Theorem (Real)

Given inner product space (V, g) , define the linear map

$$\begin{aligned} \Gamma : V &\rightarrow \hat{V} \\ u &\mapsto \Gamma_u \end{aligned}$$

where $\forall v \in V, \Gamma_u(v) = g(u, v)$.

Since $\dim(V) = \dim(\hat{V})$, to show that Γ is an isomorphism, it suffices to show that this is a injection. And indeed, $\Gamma(u) = 0_{\hat{V}} \implies \Gamma(u)(u) = g(u, u) = 0 \implies u = 0$. Since kernel of Γ is trivial, it is injective.

In particular, there being a bijection between V and \hat{V} means that

$$\forall \alpha \in \hat{V}, \exists! u \in V, \Gamma(u) = \alpha$$

Proposition Cauchy Schwarz inequality

$$|g(u, v)| \leq |u||v|$$

The equality cases occurs iff $u = cv, c \in \mathbb{R}$.

Proof Write $u = \frac{g(u,v)}{|v|^2}v + (v - \frac{g(u,v)}{|v|^2}v)$.

Denote $p = \frac{g(u,v)}{|v|^2}v$, $n = u - p$.

Then

$$\begin{aligned}
|u|^2 &= |p|^2 + |n|^2 && \text{by orthogonality} \\
&\geq |p|^2 \\
&= \left| \frac{g(u,v)}{|v|^2}v \right|^2 \\
&= \frac{g(u,v)^2}{|v|^4} |v|^2 \\
&= \frac{g(u,v)^2}{|v|^2}
\end{aligned}$$

Multiplying $|v|^2$ across gives the desired inequality.

The equality case occurs iff

$$|n|^2 = 0 \implies n = 0 \implies u = p \implies u = \frac{g(u,v)}{|v|^2}v$$

Setting $c = \frac{g(u,v)}{|v|^2}$, we conclude the proof.

Proposition Triangle Inequality

$$|u + v| \leq |u| + |v|$$

If $v \neq 0$, the equality case occurs iff $u = cv$, $c \in \mathbb{R}_0^+$

Proof

$$\begin{aligned}
|u + v|^2 &= |u|^2 + g(u,v) + g(v,u) + |v|^2 \\
&= |u|^2 + 2g(u,v) + |v|^2 \\
&\leq |u|^2 + 2|u||v| + |v|^2 \\
&\leq (|u| + |v|)^2
\end{aligned}$$

The equality case occurs iff $g(u,v) = |u||v|$. Cauchy Schwarz equality case tells us that $u = cv$ for some $c \in \mathbb{R}$. Substituting in gives $c|v|^2 = |c||v|^2 \implies c = |c|$ assuming v non-zero. Hence, c must be non-negative.

7.4.1 Component form of Riesz Representation Theorem

Given a basis z of V . Let $\alpha = \sum_i p_i \zeta_i$. Then the Riesz theorem asserts that there exists a unique $u = \sum_i a_i z_i \in V$ such that for all $v = \sum_i b_i z_i \in V$,

$$\begin{aligned}
\alpha(v) &= g(u,v) \\
\sum_{i,j} p_i b_j \zeta_i(z_j) &= \sum_{i,j} a_i b_j g_{i,j} \\
\sum_j p_j b_j &= \sum_j \sum_i a_i g_{i,j} b_j
\end{aligned}$$

Since the scalars b_i are arbitrary, by setting $b_j = 1, b_k = 0 \forall k \neq j$, we obtain the equalities

$$p_j = \sum_i a_i g_{i,j}$$

If we view $(p_1 \ p_2 \ \dots \ p_n), (a_1 \ a_2 \ \dots \ a_n)$ as a row vectors, our equality becomes

$$p_{1,j} = \sum_i a_{1,i} g_{i,j}$$

It is immediately apparent the the row vector $(p_1 \ p_2 \ \dots \ p_n)$ is the product of $(a_1 \ a_2 \ \dots \ a_n)$ and the matrix $M(g, z)$.

As proven previously, the matrix of an inner product cannot have non-positive eigenvalues. In particular, 0 is not an eigenvalue. And this says that the matrix $M(g, z)$ is invertible.

Hence,

$$(a_1 \ a_2 \ \dots \ a_n) = (p_1 \ p_2 \ \dots \ p_n) (M(g, z))^{-1}$$

Hence,

$$a_j = a_{1,j} = \sum_i p_{1,i} (M(g, z)^{-1})_{i,j} = \sum_i p_i (M(g, z)^{-1})_{i,j} = \sum_i p_i g_{i,j}^{-1}$$

where $g^{-1} := M(g, z)^{-1}$

Since g is symmetric, $M(g, z)^{-1}$ is a symmetric matrix. Hence,

$$a_j = \sum_i g_{j,i}^{-1} p_i$$

Here, we view $\begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix}$ and $\begin{pmatrix} p_1 \\ p_2 \\ \dots \\ p_n \end{pmatrix}$ as column vectors.

7.5 Extension to complex vector spaces

Define sesquilinear forms as in lecture notes. (Linear in first component, conjugate linear in second component)

Changes to the following results:

Computing inner products by matrix multiplication The 1×1 matrix of $g(u, v)$ is given by

$$[g(u, v)] = M(u, z)^T M(g, z) \overline{M(v, z)}$$

Proof Observe that

$$[g(u, v)]_{1,1} = \sum_{i,j} a_i \overline{b_j} g_{i,j} = \sum_{i,j} a_i g_{i,j} \overline{b_j} = \sum_{i,j} (M(u, z)^T)_{1,i} M(g, z)_{i,j} \overline{M(v, z)_{j,1}}$$

and the result follows.

Eigenvalues of a matrix of a complex inner product Let A be a matrix of a complex inner product g of \mathbb{C}^n . Then A only has positive eigenvalues.

Proof Since A is a matrix of a complex inner product g , A must be Hermitian (see Spectral Theorem section for definition) since g is conjugate-symmetric. And it is also proven in a later section that a Hermitian matrix only has real eigenvalues.

Suppose $\lambda \in \mathbb{R}$ is a real eigenvalue of A . Suppose $\lambda \leq 0$. By a similar argument as in the real inner product case, we first find an eigenvector $v \in \mathbb{C}^n$ associated with λ . Then,

$$g(\bar{v}, \bar{v}) = \bar{v}^T A \bar{v} = \bar{v}^T A v = \bar{v}^T \lambda v = \lambda |v|^2 \leq 0$$

which contradicts the positive definiteness of g .

Remark If we replace \mathbb{C}^n with a generic inner product space V with basis z , then we need to do a bit more housekeeping. For e.g. we will find $u, u' \in V$ such that $v = M(u, z), \bar{v} = M(u', z)$. Since v has non-zero components, so does its conjugate \bar{v} , which says that $u' \neq 0$.

Remark This result applies to all Hermitian positive-definite matrices.

Remark For a non-symmetric/Hermitian matrix, the converse is not true. A matrix can have positive eigenvalues, but still be non-positive definite. However, for a symmetric matrix, the converse is true. See link below.

[math.stackexchange](https://math.stackexchange.com/questions/1111111/positive-definite-matrix-with-positive-eigenvalues)

This be extended to Hermitian matrices as well, as we see below.

A Hermitian matrix with positive eigenvalues Let A be a Hermitian matrix. Suppose all eigenvalues of A are positive. Then A is positive definite. (i.e. A is then a matrix of a complex inner product)

Proof For a matrix M , denote $M^* = \overline{M^T}$ (conjugate transpose). By the Spectral Theorem for complex matrices, there exists a unitary matrix U such that $U^* A U = D$, where D diagonal. Let $P = U^*$. Then $A = P^* D P$.

Let g be the conjugate-symmetric sesquilinear form in $S(V)$ represented by A (for some vector space V).

Let $u \in V, v \in \mathbb{C}^n$, where $v = M(u)$.

Then

$$\begin{aligned} g(u, u) &= v^T A \bar{v} \\ &= v^T (P^* D P) \bar{v} \\ &= (P \bar{v})^* D (P \bar{v}) \\ &= y^* D y \\ &= \sum_i \bar{y}_i y_i D_{i,i} \\ &= \sum_i |y_i|^2 D_{i,i} \end{aligned}$$

Since $|y_i|^2 \geq 0$, and $D_{i,i} > 0$ since $D_{i,i}$ is an eigenvalue of A , $g(u, u) \geq 0$. In particular, if $u \neq 0$, then not all $y_i = 0$, such that $|y_i|^2 > 0$ for some i . Then, $g(u, u) > 0$.

Hence g is positive definite, and A is positive definite.

Cauchy Schwarz inequality It suffices to adjust one line in the proof

$$\begin{aligned} & \left| \frac{g(u, v)}{|v|^2} v \right|^2 \\ &= \frac{|g(u, v)|^2}{|v|^4} |v|^2 \\ &= \frac{|g(u, v)|^2}{|v|^2} \end{aligned}$$

Triangle inequality Again, we just need to adjust one line in the proof

$$g(u, v) + g(v, u) = g(u, v) + \overline{g(u, v)} = 2\operatorname{Re}(g(u, v)) \leq 2|g(u, v)| \leq 2|u||v|$$

Assuming $v \neq 0$, the equality case also occurs then $u = cv, c \in \mathbb{R}_0^+$. The proof is as follows:

First we note that $|g(u, v)| = |u||v| \iff u = cv, c \in \mathbb{C}$

Then equality occurs iff

$$\operatorname{Re}(g(cv, v) = |c||v|^2 \implies \operatorname{Re}(c|v|^2) = |c||v|^2 \implies \operatorname{Re}(c) = |c|$$

The last equality says that c must be non-negative.

Riesz Representation Theorem Since the complex inner product g is a sesquilinear form, we will need to adjust our definition of Γ . Define

$$\begin{aligned} \Gamma : V &\rightarrow \hat{V} \\ u &\mapsto \Gamma_u \end{aligned}$$

where $\forall v \in V, \Gamma_u(v) = g(v, u)$

Note that Γ is no longer a linear map, since $\Gamma(\lambda u) = \bar{\lambda}\Gamma(u)$. However, we can still show that Γ is a conjugate isomorphism.

In particular, Γ is a bijection between V and \hat{V} .

Component form

$$\begin{aligned} \alpha(v) &= g(v, u) \\ \sum_{i,j} p_j b_j \zeta_i(z_j) &= \sum_{j,i} b_j \bar{a}_i g(z_j, z_i) = \sum_{j,i} b_j \bar{a}_i g_{j,i} \\ \sum_j p_j b_j &= \sum_j \sum_i \bar{a}_i g_{j,i} b_j \\ p_j &= \sum_i g_{j,i} \bar{a}_i \end{aligned}$$

Writing the last eqn by interpreting p_j, a_i as column vectors, we get,

$$p_{j,1} = \sum_i g_{j,i} \bar{a}_{i,1}$$

Hence multiplying the inverse of $M(g, z)$ over gives

$$\bar{a}_j = \sum_i g_{j,i}^{-1} p_i$$

Here we denote $M(g, z)^{-1}$ as g^{-1} . To obtain a_j , simply apply conjugation on both sides of the equality.

Schur's Theorem Given a linear operator T on a finite dimensional complex inner product space, there exists an orthonormal basis z such that $M(T, z)$ is upper-triangular.

Proof First choose a basis y of V such that $M(T, y)$ is upper triangular. Then apply Gram-Schmidt to y to obtain orthonormal basis z . Then the change of basis matrix $M(P)$ from y to z is upper-triangular.

Also, the inverse of an upper-triangular matrix $M(P)$ is upper-triangular (Proven by the corollary to Cayley-Hamilton theorem). Since the product of upper triangular matrices is upper triangular, $M(T, z) = M(P)^{-1}M(T, y)M(P)$ is also upper triangular.

7.6 Spectral Theorem

Lemma Given inner product space (V, g) , there is a bijection between the set of sesquilinear forms $S(V)$ and the set of linear operators $L(V)$ defined by the relation

$$\forall u, v \in V, \tau(u, v) = g(u, Tv)$$

We now consider the above relation in component form. Let z be a basis of V . Letting u, v be the basis vectors of z , we have

$$\tau_{i,j} = \tau(z_i, z_j) = g(z_i, Tz_j) = g(z_i, \sum_k T_{k,j}z_k) = \sum_k \overline{T_{k,j}}g(z_i, z_k) = \sum_k \overline{T_{k,j}}g_{i,k}$$

Hence,

$$\tau_{i,j} = \sum_k g_{i,k} \overline{T_{k,j}}$$

Definition A sesquilinear form τ is hermitian if $\forall u, v \in V, \tau(u, v) = \overline{\tau(v, u)}$.

Hermitian is essentially the same as conjugate symmetric.

Definition A linear operator T over (V, g) is Hermitian iff its corresponding sesquilinear form τ is Hermitian.

Note that we can only talk about Hermitian operators when V is equipped with an inner product.

Lemma If T is Hermitian, then $\forall u, v \in V, g(u, Tv) = g(Tu, v)$.

Corollary If T is Hermitian, T has only real eigenvalues.

Proof Let $\lambda \in \mathbb{C}$ be an eigenvalue of T . Let v be an eigenvector of T associated with λ . Then, setting $u = v$,

$$g(v, Tv) = g(Tv, v) \implies \lambda|v|^2 = \bar{\lambda}|v|^2 \implies \lambda = \bar{\lambda}$$

Hence, $\lambda \in \mathbb{R}$.

Lemma With respect to an orthonormal basis, the matrix of T is the conjugate of the matrix of τ .

Proof Recall that

$$\tau_{i,j} = \sum_k g_{i,k} \overline{T_{k,j}}$$

Since we are given that z is orthonormal, $g_{i,k} = \delta_{i,k}$ and

$$\tau_{i,j} = \overline{T_{i,j}}$$

Corollary If T is Hermitian, that with respect to an orthonormal basis z , the matrix of T is also Hermitian.

Proof The result follows by:

$$\begin{aligned} T_{i,j} &= \overline{\tau_{i,j}} \\ &= \tau_{j,i} \quad \text{since } \tau \text{ hermitian} \\ &= \overline{T_{j,i}} \end{aligned}$$

Remark Note that if z is not orthonormal, then the matrix of a Hermitian operator T , $M(T, z)$, is not necessarily a Hermitian matrix.

Spectral Theorem (Complex) Given a Hermitian linear operator T , there exists an orthonormal basis z such that $M(T, z)$ is diagonal.

Remark Since $M(T, z)$ is diagonal, z is made up of eigenvectors.

Proposition Suppose y is an orthonormal basis. Let T be Hermitian. Let z be the orthonormal basis where $M(T, z)$ is diagonal. Then the change of basis matrix from y to z is unitary.

Remark The condition that $M(T, z)$ is diagonal is not necessary. In fact, given any two orthonormal bases y, z , the change of basis matrix between these is orthogonal.

The matrix form of a spectral theorem does not immediately follow from the linear operator form of a spectral theorem. This is because a Hermitian matrix A may not be a Hermitian operator (i.e. its associated sesquilinear form may not be Hermitian). Hence, we cannot use one of the above lemmas that states that the eigenvalues of a Hermitian operator are real.

Lemma The eigenvalues of a Hermitian matrix are real.

<https://yutsumura.com/eigenvalues-of-a-hermitian-matrix-are-real-numbers/>

Matrix form of Spectral Theorem (Complex) Given a Hermitian matrix A , there exists a unitary matrix U such that $\overline{U}^T A U$ is diagonal.

Proof A can be interpreted as the matrix of a linear operator T in $L(\mathbb{C}^n)$ with respect to the standard basis e , which is orthonormal. By Schur's Theorem, there exists orthonormal basis z such that $M(T, z)$ is upper triangular. By the above lemma, the eigenvalues along the diagonal of $M(T, z)$ are real (since A and $M(T, z)$ share eigenvalues). Also, the matrix U of the change of basis transformation from e to z is unitary.

Note that

$$M(T, z) = \overline{U}^T A U = \overline{U}^T M(T, e) U$$

And $\overline{M(T, z)}^T = M(T, z)$, since A is Hermitian. Since $M(T, z)$ is Hermitian and upper triangular, it is diagonal.

Remark Notice that this is slightly different from the spectral theorem. In the operator form of the spectral theorem, we did not claim that the matrix of the change of basis transformation is unitary. That is because we did not necessarily start out with an orthonormal basis y .

In MA2101, the real forms of the spectral theorem (and its matrix variation) are stated without proof.

Spectral Theorem (Real) Given a symmetric linear operator T over real inner product space V , there exists an orthonormal basis z such that $M(T, z)$ is diagonal.

Matrix form of Spectral Theorem (Real) Given a symmetric matrix A , there exists an orthogonal matrix O such that $O^T A O$ is diagonal.

7.7 Determinants

Define multilinear forms as in lecture notes. Define n -forms as in lecture notes. In short, n -forms are multilinear forms that are antisymmetric with respect to any pair of components.

Properties of 2-forms Let f be a 2-form. Then by definition, $f(u, v) = -f(v, u)$. Hence, $f(u, u) = -f(u, u) \implies f(u, u) = 0$.

Define the wedge product

$$\zeta_1 \wedge \dots \wedge \zeta_k := \frac{1}{k!} (\zeta_1 \otimes \zeta_2 \otimes \dots \otimes \zeta_k - \zeta_2 \otimes \zeta_1 \otimes \dots \otimes \zeta_k + \dots)$$

For 2 dual vectors,

$$\zeta_1 \wedge \zeta_2 := \frac{1}{2!} (\zeta_1 \otimes \zeta_2 - \zeta_2 \otimes \zeta_1)$$

In other words, we enumerate all permutations of ζ_1, \dots, ζ_k and each "swap" we make, we multiply a factor of -1 .

Let V be a vector space of dimension n , and let z be a basis of V , ζ its dual basis.

We will prove that $\{\zeta_i \wedge \zeta_j : 1 \leq i < j \leq n\}$ is a basis for the set of 2-forms on V .

Let f be a 2-form, then f is a bilinear form, and hence

$$\begin{aligned} f &= \sum_{1 \leq i, j \leq n} f(z_i, z_j) \zeta_i \otimes \zeta_j \\ &= \sum_{1 \leq i \leq n} f(z_i, z_i) \zeta_i \otimes \zeta_i + \sum_{1 \leq i < j \leq n} f(z_i, z_j) \zeta_i \otimes \zeta_j + \sum_{1 \leq j < i \leq n} f(z_i, z_j) \zeta_i \otimes \zeta_j \\ &= 0 + \sum_{1 \leq i < j \leq n} f(z_i, z_j) \zeta_i \otimes \zeta_j + \sum_{1 \leq j < i \leq n} -f(z_j, z_i) \zeta_i \otimes \zeta_j \\ &= \sum_{1 \leq i < j \leq n} f(z_i, z_j) \zeta_i \otimes \zeta_j + \sum_{1 \leq i < j \leq n} -f(z_i, z_j) \zeta_j \otimes \zeta_i \\ &= \sum_{1 \leq i < j \leq n} (2!) f(z_i, z_j) \zeta_i \wedge \zeta_j \end{aligned}$$

which completes the proof.

The dimension of the set of 2-forms is $nC2$. In general, the dimension of the set of m -forms is nCm . The dimension of the set of n -forms is $nCn = 1$.

Let T be a linear operator on V . The transpose of T with respect to an m -form Ω is defined as follows: $\hat{T}\Omega(v_1, \dots, v_m) = \Omega(Tv_1, \dots, Tv_m)$.

Hence, $\hat{T}\Omega$ is another m -form.

If $m = n$, then $\hat{T}\Omega \in \text{span}(\Omega)$. Hence, we can associate a number $\Delta_\Omega(T)$, such that

$$\hat{T}\Omega = \Delta_\Omega(T)\Omega$$

Let z be a Jordan basis of V with respect to T , and $z_i, 1 \leq i \leq n$ are the Jordan basis vectors. Then letting $\hat{T}\Omega$ act on z , we have

$$\begin{aligned}\hat{T}\Omega(z_1, \dots, z_n) &= \Omega(Tz_1, \dots, Tz_n) \\ &= \prod_{i=1}^n T_{i,i} \Omega(z_1, \dots, z_n)\end{aligned}$$

Hence, $\Delta_\Omega(T) = \prod_i T_{i,i}$, the product of the diagonal entries of the matrix of T .

Now, notice that this holds for any choice of Ω . Hence we can ditch the Ω subscript, and write

$$\hat{T}\Omega = \Delta(T)\Omega$$

In other words, this is a property of T , independent of choice of Ω .

Remark If we are to view $J = M(T, z)$, the matrix of T w.r.t. the Jordan basis as a linear operator over \mathbb{F}^n , then it is clear that $\Delta(J) = \Delta(T)$, i.e. the product of the diagonal entries of J . This trivial observation that tells us that the determinant of a linear operator T is the same as the determinant of any matrix of T under any choice of basis.

This allows us to derive other properties of Δ by choosing any arbitrary Ω . Some properties derived in lecture are:

- If S, T are linear operators, then $\Delta(ST) = \Delta(S)\Delta(T)$
- An immediate consequence is that $\Delta(P)\Delta(P^{-1}) = 1$
- Let A, B be similar matrices. Then $B = P^{-1}AP$ for some invertible matrix P . We can view B, A, P, P^{-1} as linear operators in \mathbb{F}^n , hence $\Delta(B) = \Delta(P^{-1})\Delta(A)\Delta(P) = \Delta(A)$

We now refer to Δ as \det .

7.7.1 More properties of \det

By default, we will let A denote a square matrix, and T denote a linear operator over n -dimensional vector space V .

Transpose of a matrix preserves determinants We first prove that a Jordan matrix has the same determinant as its transpose. If J is a Jordan matrix, then its determinant is just the product of the diagonal entries, as we have proven above in $\Delta_\Omega(T) = \prod_i T_{i,i}$. Since J^T is a lower triangular matrix, the antisymmetric property of an n -form means that its determinant is also the product of the diagonal entries. Hence $\det(J) = \det(J^T)$.

If A is a complex square matrix, then it is similar to a Jordan matrix J . Also, A^T is similar to J^T . To see this, write

$$\begin{aligned}P^{-1}AP &= J \\ P^T A^T (P^{-1})^T &= J^T \\ P^T A^T (P^T)^{-1} &= J^T\end{aligned}$$

Hence

$$\det(A^T) = \det(J^T) = \det(J) = \det(A)$$

Conjugate of a matrix conjugates determinant As with the previous section, consider Jordan matrix J first.

$$\begin{aligned}\det(\overline{J}) &= \prod_i \overline{J_{i,i}} \\ &= \prod_i \overline{J_{i,i}} \\ &= \overline{\prod_i J_{i,i}} \\ &= \overline{\det(J)}\end{aligned}$$

Then for a complex matrix A , it is similar to some Jordan matrix J . Also, \overline{A} is similar to \overline{J} . To see this, write

$$\begin{aligned}P^{-1}AP &= J \\ \overline{P^{-1}AP} &= \overline{J} \\ \overline{P^{-1}} \overline{A} \overline{P} &= \overline{J} \\ (\overline{P})^{-1} \overline{A} \overline{P} &= \overline{J}\end{aligned}$$

Hence,

$$\det(\overline{A}) = \det(\overline{J}) = \overline{\det(J)} = \overline{\det(A)}$$

Remark In the above, we also used some minor propositions along the way.

- Transpose and inverse can be interchanged
- Conjugate and inverse can be interchanged
- Transpose and conjugate can be interchanged (this is not mentioned above, but this is easy to see)

8 Miscellaneous Lemmas

Linear isometries in \mathbb{R}^n are precisely orthogonal matrices (Technically an isometry is a linear map in $L(\mathbb{R}^n)$ and is not quite the same as $\mathbb{R}^{n,n}$, but we'll be a bit sloppy here.)

Parallelogram law Let g be a real inner product, and $|\cdot|$ is a norm induced by g . Then $g(u, v) = \frac{1}{4}(|u+v|^2 - |u-v|^2)$.

Remark Notice that this allows us to express the real inner product in terms of the induced norm.

Product of matrices

- Product of orthogonal matrices is orthogonal