# Notes on MAT247: Algebra 2

*Unversity of Toronto*

David Duan

Last Updated: September 27, 2023 (draft)

# Contents

# Chapter 1

# Determinants

## Section 1. Trace

**1.1. Notation:** Let $V, W$ be finite dimensional vector spaces over $F$, choose bases to identify $V \cong F^n$, $W \cong F^m$, we obtain

$$\mathcal{L}(V, W) \cong \mathcal{L}(F^n, F^m) \cong M_{m\times n}(F).$$

Here, $M_{m\times n}(F)$ is the vector space of $m \times n$-matrices with coefficients in $F$: matrices with $m$ rows and $n$ columns. Such a matrix can be viewed as a linear map $F^n \to F^m$, thus $M_{m\times n}(F) = \mathcal{L}(F^n, F^m)$. Thus, we associate each $T \in \mathcal{L}(V, W)$ to a matrix $A$ with respect to the choice of bases. A change of bases replaces $A$ with

$$A' = BAC^{-1}$$

where $B \in M_{m\times m}(F)$, $C \in M_{n\times n}(F)$ are the change-of-basis matrices. However, we are mainly interested in the case $W = V$, or linear operators. In this case, we can use the same basis for the source $V$ and target $W = V$. Thus $T$ is associated with a square matrix $A$, and a change of basis replaces $A$ with

$$A' = CAC^{-1}.$$

**1.2. Definition (Trace):** The trace of a square matrix is defined as the sum of diagoal entries.

$$tr(A) = \sum_{i=1}^{n} A_{ii}$$

**1.3. Proposition:** *Let $A, B$ be two square matrices, under matrix multiplication, $(AB)_{ij} = \sum_k A_{ik}B_{kj}$, hence*

$$tr(AB) = \sum_i^n \sum_k^n A_{ik}B_{ki} = \sum_k^n \sum_i^n A_{ki}B_{ik} = tr(BA)$$

*As a consequence, if $C$ is invertible*

$$tr(CAC^{-1}) = tr(C^{-1}(CA)) = tr(A)$$

*This property enabled the following to be well-defined.*

**1.4. Definition:** The trace of a transformation $T \in \mathcal{L}(V)$ is defined as

$$tr(T) = tr(A)$$

where $A$ is the matrix of $T$ with respect to some choice of basis of $V$

**1.5. Remark:**

- We also get $tr(TS) = tr(ST)$ for composition of transformations of $V$.
- if $F = \mathbb{C}$, and $\lambda_1, ..., \lambda_n$ are the eigenvalus (with multiplicities) of $T$, then

$$tr(T) = \lambda_1 + \cdots + \lambda_n$$

This follows since we may choose a basis in which the matrix $A$ is upper triangular, with $\lambda_1, ..., \lambda_n$ its diagonal entries.

## Section 2. Determinant

*The determiannt of $T$ is also an 'invariant' of a transformtion $T$. We will begin with some motivation.*

**2.1. Motivation (The inverse of a $2 \times 2$-matrix):** For a $2 \times 2$-matrix $A \in \mathcal{M}_{2\times 2}(F)$, given as

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

we define its *determinant* by the formula

$$det(A) = ad - bc$$

**2.2. Lemma:** *The $2 \times 2$-matrix $A$ is invertible if and only if $det(A) \neq 0$. In this case, the inverse is given by*

$$A^{-1} = \frac{1}{det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

*Proof.* Let

$$B = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Multiplying $A$ and $B$, we get

$$AB = det(A) \cdot I$$

where $I$ is the identity matrix. If $det(A) \neq 0$, this shows that $det(A)^{-1}B$ is a matrix inverse of $A$. If $det(A) = 0$, the identity becomes $AB = 0$, so if $A$ were invertible, then this would give $B = A^{-1}(AB) = A0 = 0$. Hence, all matrix entries $d, -b, -c, a$ of $B$ are zero which implies that $A = 0$, a contradiction. So $A$ cannot be invertible. □

**2.3. Example:** Solve the system of equations

$$2x_1 + 3x_2 = 4$$
$$2x_1 + x_2 = 3$$

**Solution:**

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

Invert the coefficient matrix, and apply to the column vector the right side:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 4 \\ 3 \end{pmatrix} = \frac{1}{-4} \begin{pmatrix} 1 & -3 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} 4 \\ 3 \end{pmatrix} = -\frac{1}{4} \begin{pmatrix} -5 \\ -2 \end{pmatrix} = \begin{pmatrix} 5/4 \\ 1/2 \end{pmatrix}$$

so $x_1 = \frac{5}{4}, x_2 = \frac{1}{2}$.

**2.4. Intuition:** So what is the meaning of this expression $det(A) = ad - bc$? Consider temporarily the case $F = \mathbb{R}$. Let $v_1, v_2 \in \mathbb{R}^2$ be vectors $v_1, v_2$ and

$$vol(v_1, v_2) \in \mathbb{R}$$

the *signed area* of the parallelogram spanned by the two vectors. The following facts are known from high school geometry.

(P1) $vol(av_1, v_2) = a\, vol(v_1, v_2) = vol(v_1, av_2)$ ,

(P2) $vol(v_1 + av_2) = vol(v_1, v_2)$,

for all vectors $v_1, v_2$ and scalars $a$.

**2.5. Definition (Bi-linear):** A bi-linear map is a map that is linear in each of its variables, that is, for a bi-linear map $f(x, y)$,

$$f(x + x', y) = f(x, y) + f(x', y)$$
$$f(x, y + y') = f(x, y) + f(x, y')$$
$$f(cx, y) = cf(x, y) = f(x, cy)$$

**2.6. Lemma:** *The map $vol : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is **bi-linear**.*

*Proof.* Let us first note two special cases of P1,P2. Taking $a = 0$ in P1, we have

$$vol(0, v_2) = 0 = vol(v_1, 0)$$

and taking $v_1 = 0, v_2 = v, a = 1$ in P2 we see

$$vol(v, v) = 0, \quad v \in \mathbb{R}^2$$

For linearity in the first argument we must show that

$$vol(v_1 + v_1', v_2) = vol(v_1, v_2) + vol(v_1', v_2)$$

for all vectors $v_1, v_1', v_2$. If $v_2 = 0$ then it is the fist special case, and if $v_1$ is a multiple of $v_2$ it follows from P2. Thus we can assume that $v_1, v_2$ are linearly independent and is a basis. Write $v_1' = \lambda v_1 + \mu v_2$, and simplify

$$\begin{aligned} vol(v_1 + v_1', v_2) &= vol((1+\lambda)v_1 + \mu v_2, v_2) \\ &= vol((1+\lambda)v_1, v_2) \\ &= (1+\lambda)vol(v_1, v_2) \\ &= vol(v_1, v_2) + vol(\lambda v_1 + \mu v_2, v_2) \\ &= vol(v_1, v_2) + vol(v_1', v_2). \end{aligned}$$

thus $vol$ is linear in the first argument, similarly it is also linear in the second argument. □

**2.7. Remark:**

- The same argument shows a more general statement: For any vector space $V$ over a field $F$, a map $f : V \times V \to V$ satisfies P1,P2 above if and only if it is bi-linear, with the property $f(v, v) = 0$ for all $v \in V$
- If a bilinear map $f : V \times V \to V$ satisfies $f(v, v) = 0$ for all $v$, then also

  $$f(v_1, v_2) = -f(v_2, v_1)$$

  for all $v_1, v_2$. This follows by expanding $0 = f(v_1 + v_2, v_1 + v_2)$ :

  $$\begin{aligned} 0 &= f(v_1 + v_2, v_1 + v_2) \\ &= f(v_1, v_1) + f(v_2, v_2) + f(v_1, v_2) + f(v_2, v_1) \\ &= f(v_1, v_2) + f(v_2, v_1) \end{aligned}$$

  The converse is also true if $2 \neq 0$ in $F$, because $f(v_1, v_2) + f(v_2, v_1) = 0$ implies that $2f(v, v) = 0$ (take $v = v_1 = v_2$).

*We can now calculate the volume of a parallelogram, using these formal properties of vol and the fact that the volume of a square is* $vol(e_1, e_2) = 1$ *for* $e_1, e_2$ *the standard basis of* $\mathbb{R}^2$

**2.8. Proposition:** *Let* $v_1, v_2 \in \mathbb{R}^2$ *be the column vectors of a* $2 \times 2$ *matrix* $A$. *Then*

$$vol(v_1, v_2) = det(A).$$

*Proof.* Write

$$v_1 = \begin{pmatrix} a \\ c \end{pmatrix} = ae_1 + ce_2, v_2 = \begin{pmatrix} b \\ d \end{pmatrix} = be_1 + de_2.$$

Using bilinearity to expand, we get

$$\begin{aligned} vol(v_1, v_2) &= a\, vol(e_1, v_2) + e\, vol(e_2, v_2) \\ &= ac\, vol(e_1, e_1) + ad\, vol(e_1, e_2) + cb\, vol(e_2, e_1) + cd\, vol(e_2, e_2) \\ &= ad - bc \\ &= det(A). \end{aligned}$$

□

**2.9. Note:** Altought this interpretation as an area only works for $F = \mathbb{R}$, we can generalized the definition of vol to arbitrary $F$, althought we can now rename it as det.
Namely, we currently have a unique bilinear functional

$$det: \ F^2 \times F^2 \to F, \quad (v_1, v_2) \mapsto det(v_1, v_2)$$

such that $det(v, v) = 0$ for all $v \in \mathbb{F}^2$, and with $det(e_1, e_2) = 1$ for the standard basis. In fact, we have showed that $det(v_1, v_2) = det(A) = ad - bc$.

**2.10. Motivation (Generalization to higher dimensions):** In $\mathbb{R}^n$, we consider the signed volume of the parallelpiped spanned by $v_1, ..., v_n$, denoted $vol(v_1, ..., v_n)$. This functional is uniquely determined by the properties:

$$vol(v_1, ..., \lambda v_1, ..., v_n) = \lambda vol(v_1, ..., v_i, ..., v_n),$$

$$vol(v_1, ...., v_i + \lambda v_j, ..., v_n) = vol(v_1, ..., v_i, ..., v_n), j \neq i$$

$$vol(e_1, ..., e_n) = 1$$

As above, the first two conditions ensure that vol is multi-linear and vanishes whenever two arguments coincide, and the third condition serves as a normalization property. Generalizing to arbitrary fields, we have

**2.11. Theorem:** *There exists a unique multi-linear functional*

$$det : F^n \times \cdots \times F^n \to F, (v_1, ..., v_n) \mapsto det(v_1, ..., v_n).$$

*with the property that* $det(v_1, ..., v_n) = 0$ *whenever* $v_i = v_j$ *for some* $i \neq j$*, and with*

$$det(e_1, ..., e_n) = 1$$

*for the standard ordered basis* $e_1, ..., e_n$ *of* $F^n$

*Before proving this theorem, lets remember a few facts about permutations.*

**2.12. Definition:** A permutation of a finite set $X$ is an invertible map $\sigma : X \to X$

**2.13. Note:** We will only consider permutations of the set $X = \{1, ..., n\}$. There are $n! = n(n-1)\cdots 1$ different permutations, and we can write a permutation $\sigma$ by listing the elemtns $\sigma(i)$ as

$$(\sigma(1), \sigma(2), ..., \sigma(n)).$$

For example, if $n = 4$, the permutation

$$\sigma(1) = 4, \sigma(2) = 3, \sigma(3) = 1, \sigma(4) = 2,$$

is depicted as

$$(4, 3, 1, 2)$$

**2.14. Definition:** A permutation $\sigma$ is called *even* if the 'number of pairs in the wrong order'

$$\#\{(\sigma(i), \sigma(j))|i < j, \sigma(i) > \sigma(j)\}$$

is even, and we write $sign(\sigma) = 1$. A permutation $\sigma$ is odd if it not even.

**2.15. Example:** For the permutation $(4, 3, 1, 2)$. There are five pairs of indices in wrong order,

$$(4, 3), (4, 1), (4, 2), (3, 1), (3, 2).$$

Hence, $sign(sigma) = -1$.

**2.16. Note:** Most of the times computing the signs by listing all pairs in the wrong order is too cumbersome. Fortunately, there are much simpler methods on finding the parity. First note that whenever one modifies a given permutation by interchanging two adjacent elements, then the sign of $\sigma$ changes. More specifically, that pair changes from 'correct order' to 'wrong order' or vice versa, whil preserving all other orderings.
By induction , we can conclude that for any permutation $\sigma$, we have that $sign(\sigma) = (-1)^N$ if one can put the elements back into their original order by $N$ transpositions of adjacent elements.

**2.17. Lemma:** *If $\sigma'$ is obtained from $\sigma$ by interchanging two elements(not necessarily adjacent), then $\sigma', \sigma$ have opposite signs.*

*Proof.* Since we know exchanging two adjacent element switches the parity, we can view swapping two elements as making a series of adjacent swaps. Let $a, b$ be the two elements we wish to switch. Say $a$ and $b$ are seperated by $k$ elements with $a$ in th elower slot. We first move $b$ below $a$ which means we need $k + 1$ adjacent swaps, then to move $a$ up to $b$'s original position it will take $k$ adjacent swaps. The total number of adjacent swaps required is $2k + 1$, an odd number, so the parity for $\sigma'$ will indeed be different from $\sigma$. $\square$

*Let's now return to the proof of the main theorem.*

*Proof.* We start with the uniqueness proof, assuming existence. Each $v_j \in F^n$ can be uniquely expressed in terms of the basis as

$$v_1 = \sum_{i_1=1}^{n} A_{i_1 1} e_{i_1}, ..., v_n = \sum_{i_n=1}^{n} A_{i_n n} e_{i_n}$$

By multi-linearity,

$$\begin{aligned} det(v_1, ..., v_n) &= det(\sum_{i_1=1}^{n} A_{i_1,1} e_{i_1}, ..., \sum_{i_n=1}^{n} A_{i_n,n} e_{i_n}) \\ &= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} det(A_{i_1,1} e_{i_1}, \cdots, A_{i_n,n} e_{i_n}) \\ &= \sum_{i_1 \cdots i_n}^{n} A_{i_1,1} \cdots A_{i_n,n}\ det(e_{i_1}, \cdots, e_{i_n}) \end{aligned}$$

(The notation $\sum_{i_1 \cdots i_n}$ just means the sum of all possible choices of the indices $i_1, ..., i_n$.)
By assumption, $det(e_{i_1}, ..., e_{i_n}) = 0$ whenever two of the indices coincide. So the only case giving a non-zero determinant is when

$$i_1 = \sigma(1), i_2 = \sigma(2), ..., i_n = \sigma(n)$$

for some permutation $\sigma$. This gives

$$det(v_1, ..., v_n) = \sum_{\sigma} A_{\sigma(1),1} \cdots A_{\sigma(n),n}\ det(e_{\sigma(1)}, ..., e_{\sigma(n)})$$

We can put $e_{\sigma(1)}, ..., e_{\sigma(n)}$ into the right order by a finite number of interchanges of indices, since swapping any two arguments of det gives a minus sign,

$$det(e_{\sigma(1)}, ..., e_{\sigma(n)}) = sign(\sigma) det(e_1, ..., e_n) = sign(\sigma).$$

we thus obtain the formula

$$det(v_1, ..., v_n) = \sum_{\sigma} sign(\sigma) A_{\sigma(1),1} \cdots A_{\sigma(n),n}.$$

This shows that the determinant is uniquely determined by its properties.
Now to show existence, we use the formula above as the *definition* of a multi-linear funcitonal. Clearly, $det(e_1, ..., e_n) = 1$, because in this case, $A_{ij} = \delta_{ij}$ (Kronecker delta) and only the trivial permutation $\sigma = I$ contributes the 1. So we only have to show that $det(v_1, ..., v_n)$ vanishes whenever $v_r = v_s$ for some $r < s$. In this case we have that $A_{ir} = A_{is}$ for all $i = 1, ..., n$. For every permutation $\sigma$, let $\sigma'$ be obtained by an additional permutation of the elements $\sigma(r)$ and $\sigma(s)$. Thus

$$\sigma'(r) = \sigma(s), \sigma'(s) = \sigma(r), \sigma'(j) = \sigma(j) \text{ for } j \neq r, s.$$

Since $\sigma'$ is obtained from $\sigma$ by a single transposition, we have that

$$sign(\sigma') = -sign(\sigma).$$

On the other hand, since $A_{ir} = A_{is}$ we have

$$A_{\sigma(r),r} A_{\sigma(s),s} = A_{\sigma(r),s} A_{\sigma(s),r} = A_{\sigma'(r),r} A_{\sigma'(s),s}$$

and thus

$$A_{\sigma(1),1}\cdots A_{\sigma(n),n} = A_{\sigma'(1),1}\cdots A_{\sigma'(s),s}.$$

From this, we can conclude that in the sum over all permutations, the terms corresponding to $\sigma, \sigma'$ cancel out, and so $det(v_1, ..., v_n) = 0$ whenever $v_r = v_s$ for $r < s$. □

**2.18. Definition:** The **determinant** of a square matrix $A \in M_{n\times n}(F)$ is defined as

$$det(A) = det(v_1, ..., v_n),$$

where $v_1, ..., v_n$ are the columns of $A$.

Which gives us a formula:

**2.19. Theorem:** *The determinant of an $n \times n$-matrix is given by the formula*

$$det(A) = \sum_{\sigma} sign(\sigma) A_{\sigma(1),1}\cdots A_{\sigma(n),n}.$$

**2.20. Example:** If $n = 3$ there are six permutations $(123), (132), (231), (213), (312), (321)$, of signs $+, -, +, -, +, -$, respectively. So we obtain

$$det(A) = A_{11}A_{22}A_{33} - A_{11}A_{32}A_{23} + A_{21}A_{32}A_{13} - A_{21}A_{12}A_{33} + A_{31}A_{12}A_{23} - A_{31}A_{22}A_{13}$$

**2.21. Remark:** In general, the number of terms in the formula is the numbers of permutations $n!$. This indicates that for large matrices, the formula is not efficient at all. But there are much simpler ways of computing determinants which we will discuss below.

**2.22. Theorem:** *If $A \in M_{n\times n}(F)$ is upper triangular (or lower triangular), then $det(A)$ is the product over diagonal entries.*

*Proof.* Upper triangular means that $A_{ij} = 0$ whenever $i > j$. Hence, the permutations $\sigma$ does not contribute to the sum unless $\sigma(1) \leq 1, \sigma(2) \leq 2$, ... Which means $\sigma(1) = 1$, $\sigma(2) = 2$, and so on. So $\sigma$ is just the identity permutation $\sigma = id$. Thus

$$Det(A) = A_{11}A_{22}\cdots A_{nn}$$

□

**2.23. Theorem:** *For every square matrix $A \in M_{n\times n}(F)$, we have*

$$det(A) = det(A^t)$$

*Proof.* Note that for any permutation $\sigma : \{1, ..., n\}$, we have the inverse permutation $\tau = \sigma^{-1} : \{1, ..., n\}$. If $i < j$ is an ordered pair such that $\sigma(i) > \sigma(j)$, then $k = \sigma(j), l = \sigma(i)$ is a pair such that $k < l$ but $\tau(k) > \tau(l)$. Hence $sign(\sigma) = sign(\tau)$. Furthermore,

$$A_{\sigma(1),1}\cdots A_{\sigma(n),n} = A_{1,\tau(1)}\cdots A_{n,\tau(n)}$$

(on the left we arrange the factors by their column index, and on the right we arrange them by row index). Hence

$$sign(\sigma)A_{\sigma(1),1}\cdots A_{\sigma(n),n} = sign(\tau)A^t_{\tau(1),1}\cdots A^t_{\tau(n),n}$$

Hence the sum the of all permutation $\sigma$ is the same as the sum of all permutation $\tau = \sigma^{-1}$, therefore $det(A) = det(A^t)$. □

**2.24. Theorem (Properties of determinant under row&column operations):** *Let $A \in M_{n\times n}(F)$.*

*(1). If $A'$ is obtained from $A$ by interchanging two columns, then $det(A') = -det(A)$.*

*(2). If $A'$ is obtained from $A$ by taking the c-th multiple of one column, then $det(A') = c\,det(A)$.*

*(3). If $A'$ is obtained from $A$ by additing a scalar multiple of one column to another column, then $det(A') = det(A)$.*

*Parallel statements hold for row operations.*

*Proof.* By construction, the determinant from $A \mapsto det(A)$ is linear in the columns of $A$, and vanishes whenever two columns coincide. This gives (2) and (3). As in the case for $n = 2$, the fact that $det(A)$ vanishes whenever the two of the columns columns are equal implies that it changes sign under the exchange of the two columns, which gives (1). Since $det(A) = det(A^t)$, we have the analogous statement for the row operations as well. □

**2.25. Theorem:** *For any $A \in M_{n\times n}(F)$,*

$$\begin{aligned} det(A) \neq 0 &\iff rank(A) = n \\ &\iff \text{ the columns of } A \text{ are linearly independent} \\ &\iff \text{the rows of } A \text{ are linearly independent} \\ &\iff A \text{ is invertible.} \end{aligned}$$

*Proof.* We only have to show the first equivalence as everything else has been proved in MAT240. Since rows operations and column operators do not change the rank of a matrix and change the determinant by a non-zero scalar. Using both row and column operations, any matrix can be bought into row echelon form, but for such matrix, the determinant is non-zero if and only if all diagonal entries are non-zero, which happens if and only if the rank is $n$. □

**2.26. Theorem:** *For $A, B \in M_{n\times n}(F)$,*

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

*In particular,* $\det(A^{-1}) = \det(A)^{-1}$.

*Proof.* If $A$ is not invertible, then $AB$ is also not invertible, and both sides are zero. Hence we may assume that $A$ is invertible, i.e. $\det(A) \neq 0$. The multi-linear functional

$$\phi : F^n \times \cdots \times F^n \to F,$$

$$\phi(w_1, ..., w_n) = \frac{\det(Aw_1, ..., Aw_n)}{\det(A)} \tag{1.1}$$

vanishes if any two of the $w_i$ coincide, and $\phi(e_1, ..., e_n) = 1$(since $Ae_1, ..., Ae_n$ are the columns of $A$, so the enumerator of (1.5) is $\det(A)$ in that case). Hence, by the uniqueness of the determinant function, $\phi(w_1, ..., w_n) = \det(w_1, ..., w_n)$ for all $w_j$'s.
Now take $w_j = Be_j$, the columns of $B$. Then the left hand side of (1.5) is

$$\phi(w_1, ..., w_n) = \det(w_1, ..., w_n) = \det(B),$$

while the enumerator on the right hand side is

$$\det(Aw_1, ..., Aw_n) = \det(AB(e_1), ..., AB(e_n)) = \det(AB)$$

Hence, (1.5) becomes

$$\det(B) = \frac{\det(AB)}{\det(A)}.$$

Multiply both side by $\det(A)$ and we get the required form. □

**2.27. Theorem:** *Suppose $A \in M_{n\times n}(F)$ has 'block upper triangular diagonal form'*

$$A = \begin{pmatrix} A' & * \\ 0 & A'' \end{pmatrix}$$

*where $A' \in M_{k\times k}(F)$ and $A'' \in M_{l\times l}(F)$. Then*

$$\det(A) = \det(A')\det(A'')$$

**2.28. Definition (Cofactor expansions):** Using linearity in the columns (or rows), we can expand and simplify our process of finding the determinant.

$$\det(A) =$$

$$A_{11}\det\begin{pmatrix} 1 & A_{12} & \cdots & A_{1n} \\ 0 & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & A_{n2} & \cdots & A_{nn} \end{pmatrix} + A_{21}\det\begin{pmatrix} 0 & A_{12} & \cdots & A_{1n} \\ 1 & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & A_{n2} & \cdots & A_{nn} \end{pmatrix} + \cdots + A_{n1}\det\begin{pmatrix} 0 & A_{12} & \cdots & A_{1n} \\ 0 & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ 1 & A_{n2} & \cdots & A_{nn} \end{pmatrix}$$

In the $i$-th determinant, we can use row operations to move th $i$-th row into the first position. This involves $i-1$ transpositions of rows, hence introduces a sign of $(-1)^{i-1}$. The resulting matrix is of block upper triangular form

$$\begin{pmatrix} 1 & * \\ 0 & \tilde{A}^{[i1]} \end{pmatrix}$$

where $\tilde{A}^{[i1]}$ is the matrix obtained from $A$ by removing the $i$-th row and $j$-th column. We therefore obtain

$$\det(A) = A_{11}\det(\tilde{A}^{[11]}) - A_{21}\det(\tilde{A}^{[21]}) + A_{31}\det(\tilde{A}^{[31]}) + \cdots$$

One can also apply the same technique to other columns or rows. We may reduce to the case just discussed by first moving the $j$-th column into 1st position by $j-1$ transpositions. This gives an extra sign $(-1)^{j-1}$, hence the total sign is $(-1)^{j-1}(-1)^{i-1} = (-1)^{i+j}$

**2.29. Theorem (Cofactor expansion):** *To expand across the $j$-th column, fix any $j$ (or fix any $i$ for expanding across $i$-th row),*

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

*where $\tilde{A}^{[ij]}$ denotes the matrix obtained from $A$ by removing the $i$-th row and $j$-th column.*

**2.30. Remark:** The determinant $\det(\tilde{A}^{[ij]})$ is known as the $(i,j)$ minor; the expression

$$C_{ij} = (-1)^{i+j}\det(\tilde{A}^{[ij]})$$

is called the $(i,j)$ cofactor. In practice, it is often a matter of finding a convinient row of column to do the expansion, i.e. a lot of zeros.

**2.31. Theorem (Cramer's rule):** *Let $A \in M_{n\times n}$ be an invertible matrix, with columns $v_1, ..., v_n$. Then the unique solution $x = (x_1, ..., x_n)^t$ to the equation $Ax = b$ is given by the formula*

$$x_i = \frac{1}{\det(A)}\det(v_1, ..., v_{i-1}, b, v_{i+1}, ..., v_n).$$

*Proof.* Suppose $x$ is a solution. By definition of matrix multiplication,

$$b = Ax = x_1v_1 + ... + x_nv_n.$$

Thus, expanding by linearity in the $i$th column,

$$\det(v_1, ..., v_{i-1}, b, v_{i+1}, ..., v_n) = \sum_{r=1}^{n} x_r \det(v_1, ..., v_{i-1}, v_r, v_{i+1}, ..., v_n).$$

But $\det(v_1, ..., v_{i-1}, b, v_{i+1}, ..., v_n) = 0$ unless $r = i$, in which case it is $\det(A)$. This shows

$$\det(v_1, ..., v_{i-1}, b, v_{i+1}, ..., v_n) = x_i\det(A).$$

□

**2.32. Theorem:** *Let $A \in M_{n\times n}(F)$ be a square matrix with $\det(A) \neq 0$. Then the inverse matrix $A^{-1}$ has entries*

$$(A^{-1})_{ij} = \frac{(-1)^{i+j}\det(\tilde{A}^{[ji]})}{\det(A)}.$$

*Proof.* Let $v_1, ..., v_n$ denote the columns of $A$, and $w_1, ..., w_n$ the columns of $A^{-1}$. Then $w_j = A^{-1}e_j$, i.e., $w_j$ is the solution to $Ax = e_j$, and the matrix entry $(A^{-1})_{ij}$ is the $i$-th component of this solution. Thus, by Cramer's rule

$$(A^{-1})_{ij} = \frac{1}{\det(A)}\det(v_1, ..., v_{i-1}, e_j, v_{i+1}, ..., v_n).$$

Since the $i$-th column only has one non-zero entry, given by a $'1'$ in the $j$-th row. We can use cofactor expansion in the $i$-th column and the only contribution to the summation comes from the $(j, i)$-entry, with a sign $(-1)^{i+j}$ The matrix obtained by removing the $j - th$ row and $i$-th column is just $\tilde{A}^{[ji]}$. Which gives us the desired equation. □

**2.33. Definition:** The determinant of a linear transformation $T \in \mathcal{L}(V)$ is defined as

$$\det(T) = \det(A),$$

where $A \in \mathcal{M}_{n\times n}(F)$ is the matrix of $T$ in any choice of basis of $V$.

**2.34. Note:** This is well defined only because the determinant of $A$ is invariant under change of basis.

(1). Under composition,

$$\det(T_1T_2) = \det(T_1)\det(T_2)$$

(2). $T$ is invertible if and only if $\det(T) \neq 0$, and in this case

$$\det(T^{-1}) = \det(T)^{-1}$$

(3). The dual map $T' \in \mathcal{L}(V')$ has determinant

$$\det(T') = \det(T).$$

(4). Suppose $W \subseteq V$ is a $T$-invariant subspace. Then $T$ induces a map on the quotient space $V/W$, given by

$$\tilde{T}(v + W) = (Tv) + W.$$

And its determinant with $T$ restricted to $W$, denotes $T|_W : W \to W$. is

$$\det(T) = \det(T|_W)\det(\tilde{T})$$

# Chapter 2

# Characteristic Polynomials

## Section 1. Characteristic Polynomials

**1.1. Theorem:** *A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of $T$ if and only if* $\det(\lambda I - t) = 0$

*Proof.*

$$\begin{aligned} \lambda \text{ is an eigenvalue of } T &\iff \exists v \in V, v \neq 0 : Tv = \lambda v \\ &\iff \ker(T - \lambda I) \neq 0 \\ &\iff (T - \lambda I) \text{ is not invertible} \\ &\iff \det(\lambda I - T) = 0. \end{aligned}$$

□

**1.2. Definition:** The polynomial

$$q(z) = \det(zI - T)$$

is called the characteristic polynomial of $T$. If $A \in \mathcal{M}_{n\times n}(F)$, we call $q(z) = \det(zI - A)$ the characteristic polynomial of the matrix $A$.

**1.3. Remark:** The previous theorem shows that $\lambda \in F$ is an eigenvalue of $T$ if and only if $\lambda$ is a root of the characteristic polynomial. If $F = \mathbb{C}$, we may use the fundamental theorem of linear algebra to factor the polynomial as

$$q(z) = (z - \lambda_1) \cdots (z - \lambda_n),$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues (repeated according to their multiplicity).

**1.4. Remark:** If $A$ is block upper triangular with $A', A''$, then the characteristic polynomials are

$$q_A(z) = q_{A'}(z) q_{A''}(z).$$

**1.5. Definition:** A vector $v \in V$ is called *cyclic* for $T \in \mathcal{L}(V)$ if $v$, $Tv$, $T^2v$,... span all of $v$

**1.6. Example:** Suppose $T \in \mathcal{L}(V)$ is a linear operator on a finite dimensional vector space, and $v \in V$ a non-zero vector. Consider the sequence of vectors

$$v_1 = v, v_2 = Tv, ..., v_i = T^{i-1}v, ...$$

Since $\dim V < \infty$, there is a unique smallest $k$ such that $v_1, ..., v_{k+1}$ are linearly dependent, hence

$$v_{k+1} \in W = \text{span}\{v_1, ..., v_k\}.$$

The subspace $W$ is $T$-invariant, and we call $W$ the $T$-cyclic subspace generated by $v$.

**1.7. Remark:** If $v$ is a cyclic, then the vectors

$$v_1 = v, v_2 = Tv, ..., v_n = T^{n-1}v$$

are a basis of $V$. In particular,

$$Tv_n = -a_0 v_1 - \cdots - a_{n-1} v_n.$$

for some $a_0, ..., a_n \in F$. Together with

$$Tv_1 = v_2, ..., Tv_{n-1} = v_n$$

this gives the matrix of $T$:

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}$$

such a matrix is called a companion matrix.

TBD

# Chapter 3

# Inner Product Spaces

## Section 1. Real Inner Product Spaces

**1.1. Definition:** Given vectors

$$v = a_1e_1 + \cdots + a_ne_n, w = b_1e_1 + \cdots + b_ne_n \in \mathbb{R}^n$$

their dot product is defined by

$$v \cdot w = a_1b_1 + a_2b_2 + \cdots + a_nb_n$$

**1.2. Remark:** The dot product can be thought of as matrix multiplication if the elements of $\mathbb{R}^n$ as column matrices:

$$v \cdot w = v^t w$$

**1.3. Definition:** The **norm** of $v = a_1e_1 + \cdots + a_ne_n \in \mathbb{R}^n$ is defined as

$$||v|| = \sqrt{v \cdot v} = \sqrt{a_1^2 + \cdots + a_n^2}$$

**1.4. Motivation:** These definitions are motivated by geometry. Consider the case of $\mathbb{R}^2$ and write $v, w$ in polar coordinates:

$$v = r\begin{pmatrix}\cos\alpha \\ \sin\alpha\end{pmatrix}, w = s\begin{pmatrix}\cos\beta \\ \sin\beta\end{pmatrix}$$

where $r, s \geq 0$. Then

$$||v|| = r, ||w|| = s,$$

are the lengths of the two vectors. From $v \cdot w = rs(\cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)) = rs\cos(\alpha - \beta)$. We see that

$$v \cdot w = ||v||\,||w||\cos(\theta)$$

where $\theta = \alpha - \beta$ is the angle between the two vectors. The same interpretation also holds for $n > 2$, or rather defines the lengths of vectors and the angle $\theta$ between vectors.

**1.5.** Generalizing from $\mathbb{R}^n$ to more general real vector spaces, we define

**1.6. Definition (Inner products, real case):** Let $V$ be a vector space (possibly infinite dimensional) over $F = \mathbb{R}$. An *inner product* on $V$ is a **positive definite symmetric bilinear form** on $V$. That is, it is given by a bilinear map:

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

with the following properties:

(1). Symmetry: $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$,

(2). Positivity: $\langle v, v \rangle \geq 0$ for all $v$,

(3). Definiteness: $\langle v, v \rangle = 0$ if and only if $v = 0$.

$V$ together with an inner product is called a *(real) inner product space.* The associated *norm* on $V$ is defined by

$$||v|| = \sqrt{\langle v, v \rangle}$$

**1.7. Example:** Let $\mathcal{P}_n(\mathbb{R})$ be the vector space of polynomials of degree $\leq n$. Let $[a, b] \subseteq \mathbb{R}$ be a closed interval with $a < b$. Then the formula

$$\langle p, q \rangle = \int_a^b p(x)q(x)dx$$

defines an inner product on $\mathcal{P}_n(\mathbb{R})$

**1.8. Example:** Let $\mathcal{M}_{n \times n}(\mathbb{R})$ be the space of real $n \times n$-matrices. Then

$$\langle A, B \rangle = \operatorname{tr}(AB^t)$$

defines an inner product. In fact,

$$\operatorname{tr}(AB^t) = \sum_{kl} A_{kl} B_{kl},$$

so this is really the standard inner product on $\mathbb{R}^{n^2} \cong \mathcal{M}_{n \times n}(\mathbb{R})$

## Section 2. Complex Inner Product Spaces

**2.1.** Recall that the absolute value of a complex number $x = a + bi$ is defined as $|x| = \sqrt{a^2 + b^2}$, consistent with the identification of $\mathbb{C} = \mathbb{R}^2$. For a complex vector

$$v = a_1 e_1 + \cdots + a_n e_n,$$

with $a_i \in \mathbb{C}$, we define

$$||v|| = \sqrt{|a_1|^2 + \cdots + |a_n|^2}$$

Note that this extends the definition of norm on $\mathbb{R}^n$, and it also agrees with the norm on $\mathbb{R}^{2n}$ (if we identify $\mathbb{C}^n = \mathbb{R}^{2n}$, by writing $a_k =\mathrm{Re}(a_k) + i\,\mathrm{Im}(a_k))$. The absolute value is needed here as we want $||v||$ to be a nonnegative number.

**2.2. Remark:** Since $||v|| = \sqrt{v \cdot v}$ is false, we instead define the inner product on $\mathbb{C}^n$ not as the dot product, but as

$$\langle v, w\rangle = a_1\overline{b_1} + \cdots + a_n\overline{b_n},$$

where the bars denote complex conjugate.

**2.3. Remark:**

(1). Letting $\overline{w} = \overline{b_1}e_1 + \cdots + \overline{b_n}e_n$, this can be written in terms of the dot product as $v \cdot \overline{w} = v^t\overline{w}$.

(2). The restriction of the inner product to $v, w \in \mathbb{R}^n \subseteq \mathbb{C}^n$ is the inner product on $\mathbb{R}^n$.

(3). Under the identification of

$$\mathbb{C}^n \cong \mathbb{R}^{2n}$$

$\mathrm{Re}\langle v, w\rangle \in \mathbb{R}$ of the complex inner product is the real inner product on $\mathbb{R}^2 n$. On the other hand, the imaginary part $\omega(v, w) = \mathrm{Im}\langle v, w\rangle$ is not an inner product because it is skew-symmetric: $\omega(v, w) = -\omega(w, v)$.

**2.4. Definition:** Let $V$ be a vector space (possibly infinite-dimensional) over $F = \mathbb{C}$. An *inner product* on $V$ is a bilinear map

$$\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$$

with the following properties:

(1). Linearity in the firs argument:

$$\langle v + v', w\rangle = \langle v, w\rangle + \langle v', w\rangle, \quad \langle\lambda v, w\rangle = \lambda\langle v, w\rangle.$$

(2). Symmetry: $\langle v, w\rangle = \overline{\langle w, v\rangle}$ for all $v, w \in V$,

(3). Positivity: $\langle v, v\rangle \geq 0$ for all $v$,

(4). Definiteness: $\langle v, v\rangle = 0$ if and only if $v = 0$.

$V$ together with an product is called a (complex) inner product space. The Associated norm on $V$ is defined by

$$||v|| = \sqrt{\langle v, v\rangle}.$$

**2.5. Note:** Notice that conjugate linearity in the second argument comes from the symmetry which involves a complex conjugation.

$$\langle v, \lambda w\rangle = \overline{\langle \lambda w, v\rangle} = \overline{\lambda\langle w, v\rangle} = \overline{\lambda}\,\overline{\langle w, v\rangle} = \overline{\lambda}\langle v, w\rangle$$

Let us also observe the scaling property of the norm:

$$||\lambda v|| = |\lambda|\ ||v||$$

**2.6. Example:** Let $P_n(\mathbb{C})$ be the vector space of complex polynomials of degree $\leq n$. We may restrict polynomials to complex valued function on $\mathbb{R} \subseteq \mathbb{C}$. Let $[a, b] \subseteq \mathbb{R}$ be a closed interval with $a < b$. Then the formula

$$\langle p, q\rangle = \int_a^b p(x)\overline{q(x)}dx$$

defines an inner product on $P_n(\mathbb{C})$.

**2.7. Example:** On the vector space of complex $n \times n$ matrices, $\mathcal{M}_{n\times n}(\mathbb{C})$, we have the operation of conjugate transpose (adjoint):

$$A^* = \overline{A}^t$$

The formula

$$\langle A, B\rangle = \operatorname{tr}(AB^*)$$

defines a complex inner product. In fact,

$$\operatorname{tr}(AB^*) = \sum_{kl} A_{kl}\overline{B_{kl}}$$

so this is really the standard inner product on $\mathbb{C}^{n^2} \cong \mathcal{M}_{n\times n}(\mathbb{C})$

**2.8. Proposition:** *Suppose $V$ is a inner product space, and $W \subseteq V$ a linear subspace. Then the restriction of the inner product to $W$ is an inner product on $W$.*

**2.9. Proposition:** *Inner products are non-degenerate: if $w \in V$ is such that $\langle v, w\rangle = 0$ for all $v \in V$, then $w = 0$. Similarly, if $v \in V$ such that $\langle v, w\rangle = 0$ for all $w \in V$, then $v = 0$.*

*Proof.* If $\langle v, w\rangle = 0$ for all $v \in V$, then in particular it holds for $v = w$. This gives $\langle w, w\rangle = 0$, hence $w = 0$ by definiteness. □

## Section 3. Important Properties and Theorems

**3.1. Remark (Basic properties of an inner product):**

(1). For each fixed $u \in V$, the function that takes $v$ to $\langle v, u\rangle$ is a linear map from $V$ to $F$.

(2). $\langle 0, u\rangle$ for every $u \in V$.

(3). $\langle u, 0\rangle = 0$ for every $u \in V$.

*Proof.* Part 1 follows from the linearity in the first slot. Part 2 follows from part 1 and the fact that every linear map takes 0 to 0. Part 3 follows from part 2 and the conjugate symmetry property of an inner product. □

**3.2. Definition:** Two vectors $u, v \in V$ are called **orthogonal** if $\langle u, v\rangle = 0$.

**3.3.** The order of the vectors does not matter because $\langle u, v\rangle = 0$ if and only if $\langle v, u\rangle = 0$. We usually say $u$ and $v$ are orthogonal or that $u$ is orthogonal to $v$.

**3.4. Motivation:** In the usual sense of plane geometry in $\mathbb{R}^2$, we can think of the word orthogonal as a fancy word meaning *perpendicular.* That is, in $\mathbb{R}^2$

$$\langle u, v\rangle = ||u||||v|| \cos\theta$$

so two vectors in $\mathbb{R}^2$ are orthogonal if and only if the cosine of the angle between them is 0, which happens if and only if the vectors are perpendicular.

**3.5. Proposition:**

*(a) 0 is orthogonal to every vector in $V$.*

*(b) 0 is the only vector in $V$ that is orthogonal to itself.*

**3.6. Theorem (Pythagorean Theorem):** *Suppose $u$ and $v$ are orthogonal vectors in $V$. Then*

$$||u+v||^2 = ||u||^2 + ||v||^2$$

*Proof.* We have

$$\begin{aligned} ||u+v||^2 &= \langle u+v, u+v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= ||u||^2 + ||v||^2 \end{aligned}$$

as desired. □

**3.7.** For vectors $u, v \in V$, with $v \neq 0$. We want to be able to write $u$ as a scalar multiple of $v$ plus a vector $w$ orthogonal to $v$

**3.8. Motivation:** To achieve this, let $c \in F$ denote a scalar, then

$$u = cv + (u - cv).$$

Thus we need to choose $c$ so that $v$ is orthogonal to $(u - cv)$. In other words, we want

$$0 = \langle u - cv, v\rangle = \langle u, v\rangle - c||v||^2.$$

So

$$c = \frac{\langle u, v\rangle}{||v||^2}$$

Making this choice of $c$, we can write

$$u = \frac{\langle u, v\rangle}{||v||^2}v + (u - \frac{\langle u, v\rangle}{||v||^2}v).$$

**3.9. Definition:** The vector

$$\operatorname{proj}_v(u) = \frac{\langle u, v\rangle}{||v||^2}v$$

is called the ***orthogonal projection*** of $u$ onto the 1-dimensional subspace $\operatorname{span}(v) \subseteq V$.

**3.10. Definition (Orthogonal Decomposition):** Suppose $u, v \in V$, with $v \neq 0$. Let $c = \frac{\langle u,v\rangle}{||v||^2}$ and $w = u - \frac{\langle u,v\rangle}{||v||^2}v$. Then

$$\langle w, v\rangle = 0 \text{ and } u = cv + w.$$

**3.11. Theorem (Cauchy-Schwarz inequality):** *Let $V$ be an inner product space. Suppose $u, v \in V$. Then*

$$|\langle u, v\rangle| \leq ||u||\,||v||.$$

*With equality if and only if $v, w$ are linearly dependent.*

*Proof.* If $v = 0$, then both side of the equation is 0, so we can assume $v \neq 0$. Consider the decomposition

$$u = \operatorname{proj}_v(u) + (u - \operatorname{proj}_v(u))$$

By Pythagorean theorem,

$$||u||^2 = ||\operatorname{proj}_v(u)||^2 + ||(u - \operatorname{proj}_v(u))||^2 \geq ||\operatorname{proj}_v(u)||^2 = \frac{|\langle v, w\rangle|^2}{||w||^2}$$

rearrange and the inequality follows. If the equality holds, then $v - \operatorname{proj}_v(u) = 0$, hence $v$ is a multiple of $w$. Conversely, if $v = aw$ then $|\langle v, w\rangle| = |a|\,||w||^2 = ||v||\,||w||$. ☐

**3.12. Theorem:** *suppose $u, v \in V$. Then*

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

*where equality holds if and only if one of $u, v$ is a nonnegative multiple of the other.*

*Proof.* We have

$$\begin{aligned}
||u+v||^2 &= \langle u+v, u+v\rangle \\
&= \langle u,u\rangle + \langle v,v\rangle + \langle u,v\rangle + \langle v,u\rangle \\
&= \langle u,u\rangle + \langle v,v\rangle + \langle u,v\rangle + \overline{\langle u,v\rangle} \\
&= ||u||^2 + ||v||^2 + 2\mathrm{Re}\langle u,v\rangle \\
&\leq ||u||^2 + ||v||^2 + 2|\langle u,v\rangle| \\
&\leq ||u||^2 + ||v||^2 + 2||u||\,||v|| \\
&= (||u|| + ||v||)^2
\end{aligned}$$

where the first inequality from the properties of complex numbers, and the second inequality follows from Cahchy-Schwarz. Taking the square roots of both sides of the inequality above gives the desired result.

The proof above also shows that the Triangle Inequality is an equality if and only if

$$\langle u,v\rangle = ||u||\,||v||$$

If one of $u, v$ is a nonnegative multiple of the other, then the above holds. Conversely, if the above equation is true, then Cauchy-Schwarz Inequality implies that one of $u, v$ must be a scalar multiple of the other, and the equation also forces the scalar in question to be nonnegative, as desired. □

**3.13. Theorem:** *Suppose $u, v \in V$. Then*

$$||u+v||^2 + ||u-v||^2 = 2(||u||^2 + ||v||^2).$$

*Proof.* We have

$$\begin{aligned}
||u+v||^2 + ||u-v||^2 &= \langle u+v, u+v\rangle + \langle u-v, u-v\rangle \\
&= ||u||^2 + ||v||^2 + \langle u,v\rangle + \langle v,u\rangle + ||u||^2 + ||v||^2 - \langle u,v\rangle - \langle v,u\rangle \\
&= 2(||u||^2 + ||v||^2),
\end{aligned}$$

as desired. □

## Section 4. Orthonormal bases

**4.1. Lemma:** *Suppose $v_1, ..., v_k \in V$ are pairwise orthogonal non-zero vectors. Then they are linearly independent.*

*Proof.* Suppose $a_1 v_1 + \cdots + a_k v_k = 0$. Taking the inner product with $v_j$:

$$0 = \langle a_1 v_1 + \cdots + a_k v_k . v_j \rangle = a_1 \langle v_1, v_j \rangle + \cdots + a_k \langle v_k, v_j \rangle = a_j \langle v_j, v_j \rangle = a_j ||v_j||^2.$$

Dividing by $||v_j||^2$, we obtain that $a_j = 0$. □

**4.2. Definition:** A list of vectors $v_1, ..., v_n$ of an inner product space $V$ is called *orthogonal* if it satisfies $\langle v_j, v_k \rangle = 0$ for $j \neq k$, and *orthonormal* if it satisfies

$$\langle v_j, v_k \rangle = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$$

**4.3. Proposition:** *If $e_1, ..., e_m$ is an orthonormal list of vectors in $V$, then*

$$||a_1 e_1 + \cdots + a_m e_m||^2 = |a_1|^2 + \cdots + |a_m|^2$$

*for all $a_1, ..., a_m \in F$.*

*Proof.* Because each $e_j$ has norm 1, this follows from repeated applications of the Pythagorean Theorem. □

**4.4. Proposition:** *Suppose $e_1, ...e_n$ is an orthonormal basis of $V$ and $v \in V$. Then*

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

*and*

$$||v||^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2.$$

*Proof.* Since $e_1, ..., e_n$ is a basis of $V$, there exist scalars $a_1, ..., a_n$ such that

$$v = a_1 e_1 + \cdots + a_n e_n.$$

Because $e_1, ..., e_n$ is orthonormal, taking the inner product of both sides of this equation with $e_j$ gives $\langle v, e_j = a_j$. Thus the first equation holds, the second equation follows immediately from the last proposition. □

**4.5. Motivation:** The formulas above shows the importance of orthonormal list of vectors. The following method which is an inductive method to turn any given lists into an orthogonal list with the same span as the original list is called the Gram-Schimdt Procedure.

**4.6. Definition (Gram-Schmidt Procedure):** Suppose $v_1, ..., v_m$ is a linearly independent list of vectors in $V$. Let $e_1 = \frac{v_1}{||v_1||}$. For $j = 2, ..., m$, define $e_j$ inductively by

$$e_j = \frac{v_j - \sum_{n=1}^{j-1} \text{proj}_{e_n}(v_j)}{||v_j - \sum_{n=1}^{j-1} \text{proj}_{e_n}(v_j)||}$$

Then $e_1, .., e_m$ is an orthonormal list of vectors in $V$ such that

$$\text{span}(v_1, ..., v_j) = \text{span}(e_1, ..., e_j)$$

for $j = 1, ..., m$.

*Proof.* We can show by induction on $j$ that the desired conclusion holds. When $j = 1$, span$(v_1)$ =span$(e_1)$ because $v_1$ is a positive multiple of $e_1$.
Suppose $1 < j < m$ and we have that

$$\text{span}(v_1, ..., v_{j-1}) = \text{span}(e_1, ..., e_{j-1}).$$

Since $v_1, ..., v_m$ is linearly independent, $v_j \notin$span$(e_1, ..., e_{j-1})$. Hence we are not dividing by 0 in our definition. Since dividing a vector by its norm produces a new vector with norm 1, thus $||e_j|| = 1$.
Let $1 \leq k < j$, then

$$\begin{aligned} \langle e_j, e_k \rangle &= \langle \frac{v_j - \sum_{n=1}^{j-1} \text{proj}_{e_n}(v_j)}{||v_j - \sum_{n=1}^{j-1} \text{proj}_{e_n}(v_j)||}, e_k \rangle \\ &= \frac{\langle v_j, e_k \rangle - \langle v_j, e_k \rangle}{||v_j - \sum_{n=1}^{j-1} \text{proj}_{e_n}(v_j)||} \\ &= 0. \end{aligned}$$

Thus $e_1, ..., e_j$ is an orthonormal list. From the definition of $e_j$ given above, we have $v_j \in$span$(e_1, ..., e_j)$, combining this with span$(v_1, ..., v_{j-1})$ = span$(e_1, ..., e_{j-1})$, we get

$$\text{span}(v_1, ..., v_j) \subset \text{span}(e_1, ..., e_j)$$

Since both lists are linearly independent, both spaces above have dimension $j$, and hence they must be equal, completing the proof. □

**4.7. Remark:** This orthonormal basis is also uniquely determined. Suppose

$$v_j + \sum_{i=1}^{j-1} a_i e_i$$

is orthogonal to $e_1, ..., e_{j-1}$. Then for $k = 1, ..., j-1$

$$\langle v_j, e_k \rangle + a_k \langle e_k, e_k \rangle = 0$$

which gives $a_k = -\frac{\langle v_j, e_k \rangle}{||e_k||^2}$ and so $a_k e_k = -\text{proj}_{e_k}(v_j)$.

**4.8. Remark:** The fact that span$(e_1, ..., e_k)$ =span$(v_1, ..., v_k)$ for all $k \leq j$ means that the change of basis matrix is upper triangular. In fact:

- The Gram Schimidt orthonormalization is the unique change to an orthonormal basis such that the change of basis matrix is upper triangular, with positive diagonal diagonal.

**4.9. Theorem:** *Suppose $V$ is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then $T$ has an upper-triangular matrix with respect to some orthonormal basis of $V$.*

*Proof.* Trivial. □

**4.10. Definition:** Suppose $V_1, V_2$ are two inner product spaces, with inner products denoted $\langle\cdot,\cdot\rangle_1, \langle\cdot,\cdot\rangle_2$. A linear isomorphism $T : V_1 \to V_2$ is called an *isometric isomorphism* if it satisfies

$$\langle Tx, Ty\rangle_2 = \langle x, y\rangle_1$$

for all $x, y \in V_1$. Thus, an isometric isomorphism specifies an identification of $V_1, V_2$ not only as vector spaces but also as inner product spaces.

**4.11. Proposition:** *For any two finite-dimensional inner product spaces $V, W$ of the same dimension, there exists an isometric isomorphism $V \to W$.*

*Proof.* Let $v_1, ..., v_n$ be an orthonormal basis for $V$ and $w_1, ..., w_n$ an orthonormal basis for $W$. Then the linear map $T : V \to W$ defined by $T(v_i) = w_i$ is an isometric isomorphism. Indeed we can check on the basis vectors:

$$\langle Tv_j, Tv_k\rangle = \langle w_j, w_k\rangle = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases} = \langle v_j, v_k\rangle$$

□

## Section 5. Linear Functionals on Inner Product Spaces

**5.1. Intuition:** Let $V$ be an inner product space over $F$, then every $w \in V$ defines a linear functional:

$$\varphi_w : V \to F, \quad v \mapsto \langle v, w\rangle.$$

This satisfies $\varphi_{w_1+w_2} = \varphi_{w_1} + \varphi_{w_2}$ and $\varphi_{\lambda w} = \overline{\lambda}\varphi_w$. Hence the map

$$V \to V', \quad w \mapsto \varphi_w$$

is conjugate linear. To make it linear, we define a new vector $\overline{V}$ as seen in the next proposition.

**5.2. Proposition:** *Suppose $V$ is finite dimensional, let $\overline{V}$ be the vector space which is equal to $V$ as a set, with the same addition, but with its scalar multiplication defined as:*

$$\lambda * v = \overline{\lambda} v.$$

*Then the map*

$$\overline{V} \to V', \quad w \mapsto \varphi_w$$

*is a linear isomorphism.*

*Proof.* Since both sides have the same dimension, it is enough to show that the map is injective. But $\varphi_w = 0$ (the zero linear functional) implies $0 = \varphi_w(w) = \langle w, w\rangle = ||w||^2$, thus $w = 0$. □

**5.3. Definition:** Suppose $V$ is finite-dimensional and $\varphi$ is a linear functional on $V$. Then there is a unique vector $u \in V$ such that

$$\varphi(v) = \langle v, u\rangle$$

for every $v \in V$.

*Proof.* By the previous proposition, every linear functional is of the form $\varphi_w$, for a unique $w$. □

*Proof 2.* First we show the existence of a vector $u \in V$ such that $\varphi(v) = \langle v, u\rangle$for every $v \in V$. Let $e_1, ..., e_n$ be an orthonormal basis of $V$, Then

$$\begin{aligned}\varphi(v) &= \varphi(\langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n) \\ &= \langle v, e_1\varphi(e_1)\rangle + \cdots + \langle v, e_n\varphi(e_n)\rangle \\ &= \langle v, \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n\rangle\end{aligned}$$

for every $v \in V$. Thus let $u = \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n$ gives us the desired result. Now to prove the uniqueness part, suppose $u_1, u_2 \in V$ are such that $\varphi(v) = \langle v, u_1\rangle = \langle v, u_2\rangle$ for every $v \in V$. Then

$$0 = \langle v, u_1\rangle - \langle v, u_2\rangle = \langle v, u_1 - u_2\rangle$$

for every $v \in V$. Taking $v = u_1 - u_2$ shows that $u_1 - u_2 = 0$. In other words, $u_1 = u_2$, completing the proof. □

**5.4. Proposition:** *Given any basis $v_1, ..., v_n$ of an inner product space $V$, there is a unique basis $w_1, ..., w_n$ of $V$ such that*

$$\langle v_j, w_k\rangle = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$$

*for all $j, k$.*

*Proof.* Let $\varphi_1, ..., \varphi_n \in V'$ be the dual basis of the dual space, i.e.

$$\varphi_k(v_j) = \begin{cases} 0 & j \neq k \\ 1 & j = k. \end{cases}$$

Since $\varphi_k$ is a linear functional on $V$, the Riesz representation theorem, gives $w_k \in V$ such that $\varphi_k = \varphi_{w_k} = \langle \cdot, w_k\rangle$. These then satisfy

$$\langle v_j, w_k\rangle = \varphi_k(v_j) = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$$

The vectors $w_1, ..., w_n \in V$ are again a basis: Indeed, suppose

$$\sum_j a_j w_j = 0.$$

Then $a_k = 0$ for all $k$, by the calculation

$$0 = \sum_j \langle v_k, a_j w_j \rangle = \sum_j \overline{a_j} \langle v_k, w_j \rangle = \sum_j \overline{a_j} \varphi_j(v_k) = \overline{a_k}.$$

□

## Section 6. Orthogonal Complements

**6.1. Definition:** Let $U$ be a subset of $V$, then the **orthogonal complement** of $U$, denoted $U^\perp$, is the set of all vectors in $V$ that are orthogonal to every vector in $U$:

$$U^\perp = \{v \in V : \langle v, u \rangle = 0 \text{ for every } u \in U\}$$

**6.2. Proposition:**

*(a) If $U$ is a subset of $V$, then $U^\perp$ is a subspace of $V$.*

*(b) $\{0\}^\perp = V$.*

*(c) $V^\perp = \{0\}$.*

*(d) If $U$ is a subset of $V$, then $U \cap U^\perp \subseteq \{0\}$*

*(e) If $U$ and $W$ are subsets of $V$ and $U \subseteq W$, then $W^\perp \subseteq U^\perp$.*

*Proof.* (a) Suppose $U$ is a subset of $V$, then $\langle 0, u \rangle = 0$ for every $u \in U$; thus $0 \in U^\perp$. Suppose $v, w \in U^\perp$, $\lambda \in F$. If $u \in U$, then

$$\langle \lambda v + w, u \rangle = \lambda \langle v, u \rangle + \langle w, u \rangle = 0 + 0 = 0.$$

Thus $U^\perp$ is closed under addition and scalar multiplication.

(b) Suppose $v \in V$. Then $\langle v, 0 \rangle = 0$, which implies that $v \in \{0\}^\perp$. Thus $\{0\}^\perp = V$.

(c) Suppose $v \in V^\perp$. Then $\langle v, v \rangle = 0$, which implies that $v = 0$, thus $V^\perp = \{0\}$.

(d) Suppose $U$ is a subset of $V$ and $v \in U \cap U^\perp$. Then $\langle v, v \rangle = 0$, which implies that $v = 0$. Thus $U \cap U^\perp \subseteq \{0\}$.

(e) Suppose $U$ and $W$ are subsets of $V$ and $U \subset W$. Suppose $v \in W^\perp$. Then $\langle v, u \rangle = 0$ for every $u \in W$, which implies that $\langle v, u \rangle = 0$ for every $u \in U$. Hence $v \in U^\perp$. Thus $W^\perp \subseteq U^\perp$.

□

**6.3. Proposition:** *Suppose $U$ is a finite-dimensional subspace of $V$. Then*

$$V = U \oplus U^\perp.$$

*Proof.* First we will show

$$V = U + U^\perp$$

Suppose $v \in V$. Let $e_1, ..., e_m$ be an orthonormal basis of $U$. Let $u = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$ and $w = v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m$, then

$$v = u + w$$

Clearly $u \in U$. Because $e_1, ..., e_m$ is an orthonormal list, for each $j = 1, ..., m$ we have

$$\langle w, e_j \rangle = \langle v, e_j \rangle - \langle v, e_j \rangle = 0.$$

Hence $w$ is orthogonal to every vector in $\text{span}(e_1, ..., e_m)$, so $w \in U^\perp$. Thus $V = U + U^\perp$. Then since $U \cap U^\perp = \{0\}$ this implies that $V = U \oplus U^\perp$. □

*Alternate proof.* Let $v_1, ..., v_m$ be an orthonormal basis of $U$, and define a map

$$T : V \to W, v \mapsto \sum_j \text{proj}_{v_j}(v) = \sum_{j=1}^{m} \langle v, v_j \rangle v_j.$$

Then $T(v_j) = v_j$ for $j = 1, ..., m$, hence $T$ is □

**6.4. Proposition:** *Suppose $V$ is finite-dimensional and $U$ is a subspace of $V$. Then*

$$\dim U^\perp = \dim V - \dim U$$

*Proof.* Trivial. □

**6.5. Proposition:** *Suppose $U$ is a finite-dimensional subspace of $V$. Then*

$$U = (U^\perp)^\perp$$

*Proof.* First we lets show $U \subset (U^\perp)^\perp$. Suppose $u \in U$. Then $\langle u, v \rangle = 0$ for every $v \in U^\perp$, thus $u$ is orthogonal to every vector in $U^\perp$, so we have $u \in (U^\perp)^\perp$. To prove the inclusion in the other direction, suppose $v \in (U^\perp)^\perp$. Then write $v = u + w$, where $u \in U$ and $w \in U^\perp$. We have $v - u = w \in U^\perp$. However $v$ and $u$ are both in $(U^\perp)^\perp$, which implies that $v - u \in (U^\perp)^\perp$. Thus $v - u \in U^\perp \cap (U^\perp)^\perp$, which implies that $v - u$ is orthogonal to itself, so $v - u = 0$, which implies that $v = 0$, which implies that $v \in U$. Thus $(U^\perp)^\perp \subset U$, completing the proof for the other direction. □

**6.6. Definition:** An operator $P \in \mathcal{L}(V)$ is called a *projection* if it satisfies

$$P^2 = P$$

This implies that $I - P$ is also a projection. Writing vectors as $v = Pv + (I - P)v$ determines a direct sum decomposition

$$V = V_1 \oplus V_2$$

where

$$V_1 = \text{range}(P), \ V_2 = \text{range}(I - P).$$

and also

$$V_1 = \text{null}(I - P), \ V_2 = \text{null}(P).$$

Conversely, every direct sum decomposition $V = V_1 \oplus V_2$ defines a projection: writing vectors in $V$ as $v = v_1 + v_2$, with $v_i \in V_i$ and $Pv = v_1$, then $P^2 = P$ and $\text{range}(P) = V_1$ and $\text{null}(P) = V_2$.

**6.7. Definition:** Suppose $U$ is a finite-dimension subspace of $V$. The **orthogonal projection** of $V$ onto $U$ is the operator $P_U \in \mathcal{L}(V)$ defined as follows: For $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Then $P_U v = u$.

**6.8. Proposition:** *Suppose $U$ is a finite-dimensional subspace of $V$ and $v \in V$. Then*

*(a)* $P_U \in \mathcal{L}(V)$;

*(b)* $v - P_U v \in U^\perp$;

*(c)* $||P_U v|| \leq ||v||$;

*Proof.* (a) Clearly $0 \in P_U$, suppose $v_1, v_2 \in V$, $\lambda \in F$. Write

$$v_1 = u_1 + w_1 \text{ and } v_2 = u_2 + w_2$$

with $u_1, u_2 \in U$ and $w_1, w_2 \in U^\perp$. Thus $P_U v_1 = u_1$ and $P_U v_2 = u_2$, so

$$P_U(\lambda v_1 + v_2) = \lambda u_1 + u_2 = \lambda P_U v_1 + P_U v_2.$$

hence $P_U$ is a linear map from $V$ to $V$.

(b) If $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$v - P_U v = v - u = w \in U^\perp.$$

(c) By orthogonal decomposition, $v = P_U v + (v - P_U v)$, by Pythagoras' theorem

$$||v||^2 = ||P_U v||^2 + ||v - P_U v||^2 \geq ||P_U v||^2$$

with equality if and only if $v - P_U v = 0$, which means $P_U v = v$, which implies $v \in U$.

□

**6.9. Theorem:** *Let $V$ be finite-dimensional. A projection $P \in \mathcal{L}(V)$ is an orthogonal projection if and only if it satisfies:*

$$||Pv|| \leq ||v||$$

*for all $v \in V$.*

*Proof.* We have already proved the forward direction. For the other direction, suppose $||Pv|| \leq ||v||$ for all $v \in V$, we must show that

$$V = \text{range}(P) \oplus \text{null}(P)$$

is an orthogonal decomposition, i.e. all vectors in null$(P)$ are orthogonal to all vectors in range$(P)$. This is equivalent to proving null$(P)^\perp \subseteq$ range$(P)$.
Let $v \in$null$(P)^\perp$. Then

$$\implies Pv = v - (I - P)v$$

is an orthogonal decomposition of $Pv$, as $v \in \text{null}(P)^\perp$ and $(I - P)v \in \text{range}(I - P) = \text{null}(P)$. By Pythagorean theorem, we have

$$||Pv||^2 = ||v||^2 + ||(I - P)v||^2 \geq ||v||^2$$

with equality if and only if $(I-P)v = 0$. But since we have $||Pv||^2 \leq ||v||^2$ by assumption, equality must hold. Hence, we have shown that $v \in \text{null}(P)^\perp$ implies $v \in \text{null}(I-P) = \text{range}(P)$, i.e.

$$\text{null}(P)^\perp \subseteq \text{range}(P)$$

□

**6.10.** Often times we are given a subspace $U$ of $V$ and a point $v \in V$, and we want to find a point $u \in U$ such that $||v-u||$ is as small as possible. The following result says that the desired point is the orthogonal projection $P_U v$.

**6.11. Proposition:** *Suppose $U$ is a finite-dimensional subspace of $V$, $v \in V$, and $u \in U$. Then*

$$||v - P_U v|| \leq ||v - u||.$$

*With equality if and only if $u = P_U v$.*

*Proof.* We have

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

Where the second line comes from the Pythagorean Theorem, as $v - P_U v \in U^\perp$ and $P_U v - u \in U$. The above inequality is an equality if and only if $||P_U v - u|| = 0$, which happens if and only if $u = P_U v$. □

# Chapter 4

# Adjoint Operators

## Section 1. Definitions and basic properties

**1.1. Definition:** Suppose $T \in \mathcal{L}(V, W)$. The **adjoint** of $T$ is the unique operator $T^* \in \mathcal{L}(W, V)$ such that

$$\langle Tv, w\rangle = \langle v, T^*w\rangle$$

for every $v \in V$, $w \in W$.

*Proof.* For fixed $w \in W$, the map $\varphi : v \mapsto \langle Tv, w\rangle$ is a linear functional on $V$. By the Riesz representation theorem, there is a unique vector $w' \in V$ such that $\varphi(v) = \langle v, w\rangle$ for all $v \in V$. Define $T^*w = w'$, then $T^* \in V$ is the unique vector such that

$$\langle Tv, w\rangle = \langle v, T^*w\rangle$$

for all $v \in V$. The proof that $T^*$ is a linear map is left as an exercise for the reader. □

**1.2. Proposition:**

*(1).* $(S + T)^* = S^* + T^*$ *for all* $S, T \in \mathcal{L}(V, W)$*;*

*(2).* $(\lambda T)^* = \overline{\lambda} T^*$ *for all* $\lambda \in \mathbb{F}$ *and* $T \in \mathcal{L}(V, W)$*;*

*(3).* $(T^*)^* = T$ *for all* $T \in \mathcal{L}(V, W)$*;*

*(4).* $I^* = I$*, where* $I$ *is the identity operator on* $V$*;*

*(5).* $(ST)^* = T^*S^*$ *for all* $T \in \mathcal{L}(V, W)$ *and* $S \in \mathcal{L}(W, U)$ *where* $U$ *is an inner product space of* $F$*.*

*(6).* *If* $T \in \mathcal{L}(V, W)$ *is invertible, then* $T^* \in \mathcal{L}(W, V)$ *is invertible, and* $(T^{-1})^* = (T^*)^{-1}$*.*

*Proof.* (1). Suppose $S, T \in \mathcal{L}(V, W)$. If $v \in V$, $w \in W$, then

$$\langle v, (S + T)^*w\rangle = \langle (S + T)v, w\rangle = \langle Sv, w\rangle + \langle Tv, w\rangle = \langle v, S^*w\rangle + \langle v, T^*w\rangle = \langle v, S^*w + T^*w\rangle,$$

as desired.

(2). Let $\lambda \in F$ and $T \in \mathcal{L}(V, W)$. If $v \in V$ and $w \in W$, then

$$\langle v, (\lambda T)^*w\rangle = \langle \lambda Tv, w\rangle = \lambda\langle Tv, w\rangle = \lambda\langle v, T^*w\rangle = \langle v, \overline{\lambda} T^*w\rangle,$$

as desired.

(3). Suppose $T \in \mathcal{L}(V, W)$. If $v \in V$ and $w \in W$, then

$$\langle w, (T^*)^*v\rangle = \langle T^*w, v\rangle = \overline{\langle v, T^*w\rangle} = \overline{\angle Tv, w\rangle} = \langle w, Tv\rangle.$$

Thus $(T^*)^*v = Tv$,as desired.

(4). If $v, u \in V$, then

$$\langle v, I^*u\rangle = \langle Iv, u\rangle = \langle v, u\rangle,$$

Thus $I^*u = u$, as desired.

(5). Suppose $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$. If $v \in V$ and $u \in U$, then

$$\langle v, (ST)^*u\rangle = \langle STv, u\rangle = \langle Tv, S^*u\rangle = \langle v, T^*(S^*u)\rangle,$$

as desired.

□

**1.3. Proposition:** *Suppose $T \in \mathcal{L}(V, W)$, then*

*(1). null $T^* =$ (range $T)^\perp$.*

*(2). range $T^* =$ (null $T)^\perp$.*

*(3). null $T =$ (range $T^*)^\perp$;*

*(4). range $T =$ (null $T^*)^\perp$.*

*Proof.* To prove (1), let $w \in W$, then

$$w \in \text{null } T^* \iff T^*w = 0 \iff \langle v, T^* \rangle = 0 \, \forall v \in V \iff \langle Tv, w \rangle = 0 \, \forall v \in V \iff w \in (\text{range } T)^\perp$$

Thus null $T^* =$ (range $T)^\perp$, proving (1). Taking the orthogonal complement of both side of (1), we get (4). Replacing $T$ by $T^*$ in (1) and (4) gives (2) and (3), since $(T^*)^* = T$. □

**1.4. Corollary:** *rank $(T^*) =$ rank $(T)$*

*Proof.*

$$\text{rank } (T^*) = \dim W - \dim \text{null } (T^*) = \dim W - \dim \text{range } (T)^\perp = \dim \text{range } (T) = \text{rank } (T)$$

In particular, $T$ is injective if and only if $T^*$ is surjective, and $T$ is surjective if and only if $T^*$ is injective.
http://mathonline.wikidot.com/injectivity-and-surjectivity-of-the-adjoint-of-a-linear-map □

**1.5. Definition:** The ***conjugate transpose*** of an $m$-by-$n$ matrix is the $n$-by-$m$ matrix obtained by taking the transpose and then takings the conjugate of each entry.

**1.6. Proposition:** *Let $T \in \mathcal{L}(V, W)$. Suppose $v_1, ..., v_n$ is an orthonormal basis of $V$ and $w_1, ..., w_m$ is an orthonormal basis of $W$. If the matrix of $T$ is given in these bases by $A \in \mathcal{M}_{m\times n}(F)$. Then the matrix of $T^* \in \mathcal{L}(W, V)$ is given by the conjugate transpose matrix, $\overline{A}^t$.*

*Proof.* Recall that we obtain the $k^{th}$ column of $A$ by writing $Av_k$ as a linear combination of $w_i$'s. Thus since

$$Tv_k = \langle Tv_k, w_1 \rangle w_1 + \cdots + \langle Tv_k, w_m \rangle w_m.$$

Thus $A_{jk} = \langle T_{v_k}, w_j \rangle$. Similarly, $B_{kj} = \langle T^*w_j, v_k \rangle$, but

$$B_{kj} = \langle T^*w_j, v_k \rangle = \overline{\langle v_k, T^*w_j \rangle} = \overline{\langle Tv_k, w_j \rangle} = \overline{A_{jk}},$$

as required. □

## 1.1 Adjoint and Dual Map

TBD

**1.7. Proposition:** *Let $V$ be a finite dimensional inner product space, and $P \in \mathcal{L}(V)$ be a projection, so that $P^2 = P$. Then $P^*$ is again a projection:*

$$(P^*)^2 = P^*P^* = (PP)^* = P^*$$

*Recall that $P$ is an orthogonal projection if and only if null $(P)^\perp$ =range $(P)$ are orthogonal. We can conclude:*

**1.8. Theorem:** *A projection $P \in \mathcal{L}(V)$ is an orthogonal projection if and only if $P = P^*$.*

**1.9. Proposition:** *The minimal polynomials of $T$ and $T^*$ are related by*

$$p_{T^*}(z) = \overline{p_T}(z).$$

*Proof.* By definition, $p_{T^*}$ is the monic polynomial of the smallest degree such that $p_{T^*}(T^*) = 0$. Let $p_{T^*}(z) = a_0 + a_1 z + \cdots + a_n z^n$ and $\overline{p}(z) = \overline{a_0} + \overline{a_1} z + \cdots + \overline{a_n} z^n$. Then $\overline{p(z)} = \overline{p}(\overline{z})$. For $T \in \mathcal{L}(V)$, we have

$$(p(T))^* = (a_0 + a_1 T + \cdots + a_n T^n)^* = \overline{a_0} + \overline{a_1} T^* + \cdots + \overline{a_n}(T^*)^n = \overline{p}(T^*)$$

And hence

$$\overline{p_{T^*}}(T) = 0$$

and so $\overline{p_{T^*}} = p_T$. □

**1.10. Proposition:** *For any $T \in \mathcal{L}(V)$,*

$$\det(T^*) = \overline{\det(T)}.$$

*Proof.* In terms of an orthonormal basis, if $A$ is the matrix of $T$, then the matrix of $T^*$ is $\overline{A}^t$. Which implies

$$\det(T*) = \det(\overline{A}^t) \ = \det(\overline{A}) = \overline{\det(A)} = \overline{\det(T)}$$

□

**1.11. Proposition:** *The characteristic polynomials of $T$ and $T^*$ are related by*

$$q_{T^*}(z) = \overline{q_T}(z).$$

*Proof.* $q_{T^*}(z) = \det(zI - T^*) = det((\overline{z}I - T)^*) = \overline{\det(\overline{z}I - T)} = \overline{q_T(\overline{z})} = \overline{q_T}(z)$. □

**1.12. Proposition:** *Let $J \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a JNF for $T \in \mathcal{L}(V)$. Then the complex conjugate $\overline{J} \in \mathcal{M}_{n\times n}(\mathbb{C})$ is a JNF for $T^*$*

*Proof.* Choose an orthonormal basis of $V$, and let $A$ be the matrix corresponding to $T$. Then

$$A = CJC^{-1}.$$

As we've shown, $A^* = \overline{A}^t$ is the matrix corresponding to $T^*$ as well as every complex matrix is similar to its transpose; hence the JNF of $\overline{A}^t$ coincides with $\overline{A}$, and since the JNF of $\overline{A}$ is $\overline{J}$, we get the desired result. $\square$

## Section 2. Self-Adjoint Operators

**2.1. Definition:** An operator $T \in \mathcal{L}(V)$ is ***self-adjoint*** if $T = T^*$, in other words,

$$\langle Tv, w\rangle = \langle v, Tw\rangle$$

for all $v, w \in V$. An operator is *skew-adjoint* if $T^* = -T$.

**2.2. Example:** For every $T \in \mathcal{L}(V)$, the operators

$$T + T^*, TT^*, T^*T$$

are self-adjoint, while $T - T^*$ is skew-adjoint.If $F = \mathbb{C}$, then multiplication by $i$ takes self-adjoint operators to skew-adjoint operators and vice versa.

**2.3. Example:** Let $v = F^n$, so that operators on $V$ are just matrices $A \in \mathcal{M}_{n\times n}(F)$. The matrix $A$ is self-adjoint if it is equal to its conjugate transpose. If $F = \mathbb{R}$, then it simply means that $A$ must be symmetric. for $F = \mathbb{C}$, we have that all diagonal entries must be real.

**2.4. Proposition:** *Every eigenvalue of a self-adjoint operator is real.*

*Proof.* Suppose $T$ is a self-adjoint operator on $V$. Let $\lambda$ be an eigenvalue of $T$, and let $v$ be a nonzero vector in $V$ such that $Tv = \lambda v$. Then

$$\lambda ||v||^2 = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}||v||^2$$

Since $\lambda = \overline{\lambda}$, it must be real. $\square$

**2.5. Proposition:** *Over $\mathbb{C}$, $Tv$ is orthogonal to $v$ for all $v$ only for the $0$ operator, i.e. suppose $V$ is a complex inner product space and $T \in \mathcal{L}(V)$. Suppose*

$$\langle Tv, v\rangle = 0$$

*for all $v \in V$. Then $T = 0$.*

*Proof.* We have

$$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4} + \frac{\langle T(u+iw), u+iw\rangle - \langle T(u-iw), u-iw\rangle}{4} i$$

for all $u, w \in V$, as can be verified by computing the right side. Note that each term on the right side is of the form $\langle Tv, v\rangle$ for appropriate $v \in V$. Thus our hypothesis implies that $\langle Tu, w\rangle = 0$ for all $u, w \in V$ if we take $w = Tu$. This implies that $T = 0$. $\square$

**2.6. Proposition:** *Suppose $V$ is a complex inner product space and $T \in \mathcal{L}(V)$. Then $T$ is self-adjoint if and only if*

$$\langle Tv, v\rangle \in \mathbb{R}$$

*for every $v \in V$.*

*Proof.* Let $v \in V$. Then

$$\langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = \langle Tv, v\rangle - \langle v, Tv\rangle = \langle Tv, v\rangle - \langle T * v, v\rangle = \langle (T - T^*)v, v\rangle.$$

If $\langle Tv, v \in \mathbb{R}$ for every $v \in V$, then the left side of the equation above equal 0, so $\langle (T - T^*)v, v\rangle = 0$ for every $v \in V$. This implies that $T - T^* = 0$. Hence $T$ is self-adjoint.
Conversely, if $T$ is self-adjoint, then the right side of the equation above equal 0, so $\langle Tv, v\rangle = \overline{Tv, v}$ for every $v \in V$. Thus $\langle Tv, v\rangle \in \mathbb{R}$ for every $v \in V$, as desired. □

**2.7. Proposition:** *Suppose $T$ is a self-adjoint operator on $V$ such that*

$$\langle Tv, v\rangle = 0$$

*for all $v \in V$. Then $T = 0$.*

*Proof.* We have already proved this when $V$ is a complex inner product space. Thus let $V$ be a real inner product space. If $u, w \in V$, then

$$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4};$$

this is proved by computing the right side using the equation

$$\langle Tw, u\rangle = \langle w, Tu\rangle = \langle Tu, w\rangle,$$

where the first equality holds because $T$ is self-adjoint and the second equality holds because we are in a real inner product space.
Each term on the right side of the big equation if of the form $\langle Tv, v\rangle$ for appropriate $v$. Thus $\langle Tu, w\rangle = 0$ for all $u, w \in V$, which implies that $T = 0$. □

*The proof below gives a more intuitive solution using theorem 3.6*

*Alternative proof.* If $S$ is self-adjoint, then there exist an orthonormal basis consisting of eigenvectors of $S$. Suppose $\langle Sv, v\rangle = 0$ for all $v \in V$. Let $v_i$ to be an eigenvector with eigenvalue $\lambda_i$, we get $\langle Sv_i, v_i\rangle = \langle \lambda v_i, v_i\rangle = 0$. This shows that $Sv_i = \lambda_i v_i = 0$ for all basis vectors and hence $S = 0$. □

## Section 3. Decomposition into eigenspaces

**3.1.** Throughout, we will take $V$ to be a finite-dimensional real or complex inner product space.

**3.2. Note:** Recall that in general, invariant subspaces need not admit invariant complements, for self-adjoint operators, this problem does not occur.

**3.3. Lemma:** *If $T = T^*$, then every $T$-invariant subspace $W$ has an invariant complement. In fact, $W^\perp$ is $T$-invariant*

*Proof.* Suppose $T = T^*$, and suppose $T(W) \subseteq W$.Let $v \in W^\perp$, we want to show that $Tv \in W^\perp$. This follows from the following: let $w \in W$,

$$\langle Tv, w\rangle = \langle v, T^*w\rangle = \langle v, Tw\rangle = 0.$$

Hence, since $Tv, w$ are orthogonal for all $w, v$ we get the desired result. □

**3.4. Lemma:** *If $T = T^*$, the eigenspaces $E(\lambda, T)$,$E(\mu, T)$ corresponding to distinct eigenvalues are orthogonal:*

$$E(\lambda, T) \subseteq E(\mu, T)^\perp$$

*Proof.* Let $\lambda, \mu$ be two distinct eigenvalues of $T$. Let $v \in E(\lambda, T)$ and $w \in E(\mu, T)$, we have

$$\lambda\langle v, w\rangle = \langle Tv, w\rangle = \langle v, Tw\rangle = \langle v, \mu w\rangle = \mu\langle v, w\rangle.$$

Since $\lambda \neq \mu$, this means $\langle v, w\rangle = 0$. □

**3.5. Lemma:** *Suppose $T \in \mathcal{L}(V)$ is self-adjoint, then $T$ has at least one eigenvector in $V$.*

*Proof.* If $F = \mathbb{C}$, we know every $T \in \mathcal{L}(V)$ has an eigenvector and the condition of self-adjoint isn't needed. For $F = \mathbb{R}$, it suffices to show that the set of eigenvalues of $T$ is non-empty. Let $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ be the matrix of $T$ in an orthonormal basis of $V$ and it suffices to show that $A$ has an eigenvector in $\mathbb{R}^n$. We may regard $A$ as a complex self-adjoint matrix whose entries happen to be real. As such it has eigenvectors $v \in \mathbb{C}^n$:

$$Av = \lambda v$$

By the previous lemma, $\lambda$ is real. Hence, taking the component wise real part of $v$, we find that $\text{Re}(v)$ is an eigenvector (unless it is zero, in which case take the imaginary prat.) □

**3.6. Theorem:** *suppose $T \in \mathcal{L}(V)$ is self-adjoint, where $n = \dim(V) < \infty$. Then $T$ is diagonalizable. In fact, $V$ admits an orthonormal basis $v_1, ..., v_n$ consisting of eigenvectors of $T$.*

*Proof.* Choose an eigenvector $v_1 \in V$ and normalize it to $||v_1||^2 = 1$. The orthogonal space $\text{span}\{v_1\}^\perp$ is $T$-invariant. Then the restriction of $T$ to the orthogonal complement has an eigenvector $v_2$, which we can normalize to $||v_2||^2 = 1$. Preceding by induction, we obtain the desired orthonormal basis consisting of eigenvectors.
Conversely, given $T$, suppose there exists an orthonormal basis $v_1, ..., v_n$ consisting of eignvectors of $T$ in which $T$ is diagonal, then

$$Tv_i = \lambda_i v_i$$

for some $\lambda_i \in C$. We claim that

$$T^* v_i = \overline{\lambda_i} v_i.$$

To see this, first note that $T^* v_i$ must be a scalar multiple of $v_i$, since its inner prducts with $v_j$ for $j \neq i$ are zero:

$$\langle T^* v_i, v_j \rangle = \langle v_i, T v_j \rangle = \lambda_j \langle v_i, v_j \rangle = 0.$$

Hence $v_i$ is an eigenvector for $T^*$. The eigenvalue can then be found:

$$\langle T^* v_i, v_i \rangle = \langle v_i, T v_i \rangle = \langle v_i, \lambda_i v_i \rangle = \overline{\lambda_i} \langle v_i, v_i \rangle = \overline{\lambda_i}.$$

Now, evaluating both sides of $T^* T = T T^*$ on basis vectors:

$$T^* T v_i = \lambda_i T^* v_i = |\lambda_i|^2 v_i$$

and similarly

$$T T^* v_i = |\lambda_i|^2 v_i.$$

Hence, $T$ is a normal operator $\square$

## Chapter 5

# Unitary Operators

## Section 1. Unitary Operators

**1.1. Definition:** Let $V$ be an inner product space, an operator $T \in \mathcal{L}(V)$ is called ***unitary*** if it is invertible and satisfies

$$\langle Tv, Tw \rangle = \langle v, w \rangle$$

for all $v, w \in V$. For $F = \mathbb{R}$, a unitary operator is also called an **orthogonal** operator. A matrix is called unitary if the associated linear operator is on $F^n$ is unitary.

**1.2. Proposition:** *If $T \in calL(V)$ is a unitary transformation, then so is $T^{-1}$. If $S, T \in \mathcal{L}(V)$ are unitary transformations, then so is $ST$.*

*Proof.* The first claim follows from:

$$\langle T^{-1}v, T^{-1}w \rangle = \langle TT^{-1}v, TT^{-1}w \rangle = \langle v, w \rangle$$

for all $v, w \in V$. The second claim is obtained similarly from

$$\langle STv, STw \rangle = \langle Tv, Tw \rangle = \langle v, w \rangle.$$

□

**1.3. Proposition:** *If $V$ is finite dimensional, an operato $T \in \mathcal{L}(V)$ satisfying*

$$\langle Tv, Tw \rangle = \langle v, w \rangle$$

*for all $v, w \in V$ is always invertible.*

*Proof.* Suppose $v \in V$ with $Tv = 0$. Then

$$||Tv||^2 = \langle Tv, Tv \rangle = \langle v, v \rangle = ||v||^2$$

shows that $v = 0$. □

**1.4. Proposition:** *Suppose $V$ is finite dimensional. Then the following are equivalent:*

*(1). $T \in \mathcal{L}(V)$ is unitary.*

*(2). There is an orthonormal basis $v_1, ..., v_n$ such that the images $Tv_1, ..., Tv_n$ are again an orthonormal basis.*

*(3). For every orthonormal basis $v_1, ..., v_n$, the images $Tv_1, ..., Tv_n$ are again an orthonormal basis.*

*Proof.* We have (3) implies (1) and (1) implies (3), since $\langle Tv_i, Tv_j \rangle = \langle v_i, v_j \rangle$. Suppose (2) hold, let $v_1, ..., v_n$ be an orthonormal basis such that $Tv_1, ..., Tv_n$ is on orthonormal basis. Given $v, w \in V$, we can write

$$v = \sum_i a_i v_i, w = \sum_i b_i v_i.$$

Then

$$\langle Tv, Tw\rangle = \langle \sum_i a_i T v_i, \sum_j b_j T v_j\rangle = \sum_{ij} a_i b_j \langle T v_i, T v_j\rangle = \sum_{ij} a_i b_j \langle v_i, v_j\rangle = \langle v, w\rangle.$$

Thus (2) implies (1). □

**1.5. Corollary:** *A matrix $A \in \mathcal{M}_{n\times n}(F)$ is unitary if and only if its columns are an orthonormal basis of $F^n$.*

*Proof.* The columns of $A$ are the image $Ae_i$ of the standard basis vector $e_i$. □

**1.6. Theorem:** *An operator $T$ is unitary if and only if*

$$TT^* = T^*T = I.$$

*Proof.* Suppose $T$ is unitary, then

$$\langle Tv, Tw\rangle = \langle v, w\rangle \implies \langle v, T^*Tw\rangle = \langle v, w\rangle$$

for all $v, w \in V$, hence $T^*Tw = w$ for all $w$, hence $T^*T = I$.
Since $T$ is invertible, we may multiply by $T^{-1}$ from the right and by $T$ from the left to obtain $TT^* = I$. This means, $T$ is unitary if and only if it is invertible, with inverse the adjoint:

$$T^{-1} = T^*.$$

In particular, $T$ is unitary implies $T^*$ is unitary. □

**1.7. Corollary:** *The determinant of a unitary operator satisfies*

$$|\det(T)| = 1.$$

*In particular, for $F = \mathbb{R}$ we have $\det(T) = 1$ or $\det(T) = -1$.*

*Proof.* Apply det to the equation $I = T^*T$:

$$1 = \det(I) = \det(T^*T) = \det(T^*)\det(T) = \overline{\det(T)}\det(T) = |\det(T)|^2.$$

□

**1.8. Corollary:** *If $A \in \mathcal{M}n \times n(F)$ is unitary then also $A^t$ and $\overline{A}$ are unitary.*

**1.9. Proposition:** *Suppose $T \in \mathcal{L}(V)$ is a unitary operator on a real or complex inner product space $V$, with $V$ finite dimensional. Then*

- *For every $T$-invariant subspace $W$ the orthogonal complement $W^\perp$is $T$-invariant*
- *All eigenvalues $\lambda$ of $T$ have absolute value equal to 1.*
- *The eigenspaces $E(\lambda, T)$,$E(\mu, T)$ corresponding to distinct eigenvalues $\lambda, \mu$ are orthogonal.*

*Proof.* (1). Note that since $T$ is invertible, the invariance property $TW \subseteq W$ implies $TW = W$. Applying $T^{-1}$ to both sides, $W = T^{-1}W$. Let $v \in W^{\perp}$, for all $w \in W$,

$$\langle Tv, w\rangle = \langle v, T^* w\rangle = \langle v, T^{-1}w\rangle = 0$$

Thus $Tv \in W^{\perp}$.

(2). Suppose $v \in V$ is an eigenvector with eigenvalue $\lambda$. Then

$$\langle v, v\rangle = \langle Tv, Tv\rangle = \langle \lambda v, \lambda v, \rangle = |\lambda|^2 \langle v, v\rangle$$

Hence $|\lambda|^2 = 1$.

(3). Suppose $v \in V$ is an eigenvector with eigenvalue $\lambda$, and $w \in V$ is an eigenvector eigenvalue $\mu$, where $\mu \neq \lambda$. Then

$$\langle v, w\rangle = \langle Tv, Tw\rangle = \langle \lambda v, \mu w\rangle = \lambda\overline{\mu}\langle v, w\rangle = \lambda\mu^{-1}\langle v, w\rangle.$$

Since $\lambda\mu^{-1} \neq 1$, hence $\langle v, w\rangle = 0$.

□

**1.10. Theorem:** *Suppose $T \in \mathcal{L}(V)$ is a unitary operator on a complex vector space of finite dimension. Then $T$ is diagonalizable. In fact, $T$ admits an orthonormal basis consisting of eigenvectors of $T$.*

*Proof.* Similar to the case of self-adjoint operators, proving by induction that for all $k \leq n$, there exists an orthonormal set of eigenvectors of $T$. The base case $k = 0$ is clearly true. For each $k < n$, given an orthonormal set of eigenvectors $v_1, ..., v_k$ of $T$, the space $\operatorname{span}\{v_1, ..., v_k\}$ is $T$-invariant, hence $\operatorname{span}\{v_1, ..., v_k\}^{\perp}$ is $T$-invariant, and $T$ acts as a unitary operator on this space, hence we can pick another unit length eigenvector $v_{k+1} \in \operatorname{span}\{v_1, ..., v_k\}^{\perp}$. □

# Chapter 6

# Normal Operators

## Section 1. Definitions and basic properties

**1.1. Definition:** An operator $T$ on an inner product space $V$ is called **normal** if it commutes with its adjoint:

$$TT^* = T^*T.$$

A matrix $A$ is normal if it satisfies $AA^* = A^*A$, where $A^* = \overline{A}^t$.

**1.2. Example:**

- Self-adjoint, skew-adjoint, and unitary operators are all normal.
- If $T$ is normal and $\lambda \in F$, then $\lambda T$ is normal:

$$(\lambda T)^*(\lambda T) = \overline{\lambda}T^*\lambda T = |\lambda|^2 T^*T = |\lambda|^2 TT^* = (\lambda T)(\lambda T)^*.$$

- If $T$ is normal and invertible, then $T^{-1}$ is normal:

$$T^{-1}(T^{-1})^* = T^{-1}(T^*)^{-1} = (TT^*)^{-1} = (T^*T)^{-1} = (T^*)^{-1}T^{-1} = (T^{-1})^*T^{-1}.$$

- If $T$ is normal and $p(z)$ is any polynomial, then $p(T)$ is normal. This is because all powers of $T$ commutes with all all powers of $T^*$:

$$(T^n)(T^*)^k = T\cdots TT^*\cdots T^* = T^*\cdots T^*T\cdots T = (T^*)^k(T^n).$$

  Along with $p(T)^* = \overline{p}(T^*)$, we get:

$$p(T)(p(T))^* = (a_0 + a_1T + \cdots + a_nT^n)(\overline{a_0} + \overline{a_1}T^* + \cdots + \overline{a_n}(T^*)^n),$$

  Hence we see they are equal since they commute term by term.

**1.3.** A simple fact which was used in the third bullet point above is that should be always be remembered is that the *adjoint of the inverse* is equal to *the inverse of the adjoint.*

**1.4. Fact:** *Let $T$ be an operator on an inner product space $V$. Then*

$$(T^*)^{-1} = (T^{-1})^*.$$

*Proof.* Fix $v \in V$, for any $w \in V$:

$$\langle T^*(T^{-1})^*v, w\rangle = \langle (T^{-1})^*v, Tw\rangle = \langle v, T^{-1}Tw\rangle = \langle v, w\rangle.$$

Hence $T^*(T^{-1})^* = I$ which implies $(T^{-1})^* = (T^*)^{-1}$. $\square$

**1.5. Proposition:** *Let $V$ be a finite-dimensional inner product space. Then $T \in \mathcal{L}(V)$ is normal if and only if it has the property*

$$||Tv||^2 = ||T^*v||^2$$

*for all $v \in V$.*

*Proof.*

$$\begin{aligned}
T \text{ is normal} &\iff T^*T - TT^* = 0 \\
&\iff \langle (T^*T - TT^*)v, v\rangle = 0 \text{ for all } v \in V \\
&\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle \text{ for all } v \in V \\
&\iff \langle Tv, Tv\rangle = \langle T^*v, T^*v\rangle \text{ for all } v \in V \\
&\iff ||Tv||^2 = ||T^*v||^2 \text{ for all } v \in V.
\end{aligned}$$

□

## Section 2. Eigenspace decomposition for normal operators

**2.1.** Our aim ultimately is to prove the complex spectral theorem (spectral theorem for normal operators). To begin, we will need to prove some smaller results.

**2.2. Lemma:** *If $T, S \in \mathcal{L}(V)$ are operators on a complex vector space, and $ST = TS$, then there exists a joint eigenvector $v \in V, v \neq 0$:*

$$Tv = \lambda v, Sv = \mu v.$$

*Proof.* Let $\lambda \in \mathbb{C}$ with $E(\lambda, T) \neq 0$. For $v \in E(\lambda, T)$, we have

$$TSv = STv = S\lambda v = \lambda Sv,$$

which implies $Sv \in E(\lambda, T)$. This shows that $E(\lambda, T)$ is $S$-invariant. Simply choose any $v \in E(\lambda, T)$ to be an eigenvector of $S|_{E(\lambda,T)}$ with eigenvalue $\mu$ gives us the desired result. □

**2.3. Remark:** This generalizes to any finite collection of operators $T_1, ..., T_k$ on a complex vector space, i.e. if the operators commute pairwise then there exists a joint eigenvector for all of these vectors.

**2.4. Lemma:** *Suppose $T \in \mathcal{L}(V)$ is any linear operator on a finite-dimensional inner product space. If $W \subseteq V$ is $T$-invariant then $W^\perp$ is $T^*$-invariant. Hence, if $W$ is invariant under both $T$ and $T^*$, then so is $W^\perp$.*

*Proof.* Suppose first $W \subseteq V$ is $T$-invariant. Let $v \in W^\perp$,then for all $w \in W$ we have

$$\langle T^* v, w \rangle = \langle v, Tw \rangle = 0,$$

this means all vectors in $W$ is orthogonal to $T^* v$, and hence

$$T^* W^\perp \subseteq W^\perp.$$

Similarly, if we have $T^* W \subseteq W$, then

$$TW^\perp = (T^*)^* W^\perp \subseteq W^\perp.$$

□

**2.5. Lemma:** *Let $T \in \mathcal{L}(V)$ be an operator on a finite dimensional vector space and $W$ an $T$-invariant subspace, then $(T|_W)^* = (T^*|_W)$.*

*Proof.* By properties of adjoint, fix $x \in W$, then for any $y \in W$.

$$\langle (T|_W)^* x, y \rangle = \langle x, T|_W y \rangle = \langle x, Ty \rangle = \langle T^* x, y \rangle$$

Hence,

$$\langle (T|_W)^* x - T^* x, y \rangle = 0.$$

Since $y$ can be anything, let $y = (T|_W)^* x - T^* x$, and by non-degeneracy of inner product

$$\langle (T|_W)^* x - T^* x, (T|_W)^* x - T^* x \rangle = 0 \implies (T|_W)^* x = T^* x.$$

□

**2.6. Theorem (Complex spectral theorem):** *Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional complex inner product space $V$. Then*

$$T \text{ is normal} \iff \text{There exists an orthonormal basis of } V \text{ consisting of eigenvectors of } T.$$

*Proof.* Suppose $T$ is normal. By induction we may construct an orthonormal set $\{v_1, ..., v_k\}$ consisting of joint eigenvectors for $T, T^*$. The induction start with $k = 0$ which by lemma 2.2, there exists a joint eigenvector $v$. For the inductive step, suppose $\{v_1, ..., v_k\}$ is an orthonormal set consisting of joint eigenvectors for $T, T^*$, then $\operatorname{span}\{v_1, ..., v_k\}$ is invariant under both $T$ and $T^*$, by 2.4, then $\operatorname{span}\{v_1, ..., v_k\}^\perp$ is then invariant under both $T$ and $T^*$ as well. Then using 2.5,

$$T|_W \circ T^*|_W = (T \circ T^*)|_W = (T^* \circ T)|_W = T^*|_W \circ T|_W,$$

and so the restrictions of $T, T^*$ to $\operatorname{span}\{v_1, ..., v_k\}^\perp$ commute, hence they must again admit a joint unit eigenvector $v_{k+1}$.
Conversely, given $T$, suppose there exists an orthonormal basis $v_1, ..., v_n$ in which $T$ is diagonal. Then

$$Tv_i = \lambda_i v_i \text{ and } T^* v_i = \overline{\lambda_i} v_i.$$

Now evaluating both sides of $TT^*$ and $T^*T$ on the basis vectors, we see

$$T^* T v_i = \lambda_i T^* v_i = |\lambda_i|^2 v_i,$$

and similarly

$$TT^*v_i = \overline{\lambda_i}Tv_i = |\lambda_i|^2 v_i,$$

and so $TT^* = T^*T$, as desired. ☐

**2.7.** Two more important propositions which we've seen before are as follows:

**2.8. Proposition:** *Suppose $T$ is normal and $v$ is an eigenvector of $T$ with eigenvalue $\lambda$, then $v$ is also an eigenvector of $T^*$ with eigenvalue $\overline{\lambda}$.*

*Proof.* Since $T$ is normal, $T - \lambda I$ can be easily verify to be normal as well. Using 1.5, we have

$$0 = ||(T - \lambda I)v|| = ||(T - \lambda I)^* v|| = (T^* - \overline{\lambda} I)v||.$$

Hence $v$ is an eigenvector of $T^*$ with eigenvalue $\overline{\lambda}$, as desired. ☐

**2.9. Proposition:** *Suppose $T$ is normal, then eigenvectors of $T$ corresponding to distinct eigenvalues are orthogonal.*

*Proof.* Suppose $\alpha, \beta$ are distinct eigenvalue of $T$, with corresponding eigenvectors $u, v$. Thus $Tu = \alpha u$ and $Tv = \beta v$. Thus,

$$\begin{aligned}(\alpha - \beta)\langle u, v\rangle &= \langle \alpha u, v\rangle - \langle u, \overline{\beta} v\rangle \\ &= \langle Tu, v\rangle = \langle u, T^* v\rangle \\ &= 0.\end{aligned}$$

Because $\alpha \neq \beta$, it implies that $\langle u, v\rangle = 0$, hence they are orthogonal. ☐

## Section 3. Spectral resolution

**3.1.** An important consequence of the spectral theorem is as follows:

**3.2. Theorem (Spectral resolution):** *Let $T$ be a normal operator on a finite dimensional complex inner product space $V$. Let $P_\lambda \in \mathcal{L}(V)$ be the orthogonal projection to the eigenspace $E(\lambda, T)$. Then*

$$P_\lambda P_\mu = \begin{cases} 0 & \lambda \neq \mu \\ , P_\lambda & \lambda = \mu \end{cases}$$

$$\sum_\lambda P_\lambda = I$$

*and the operator $T$ as a sum is equal to*

$$T = \sum_\lambda \lambda P_\lambda.$$

*This formula is most often referred to as the spectral resolution of $T$.*

*Proof.* The first identity simply states that the eigenspaces for different eigenvalues are orthogonal i.e. $P_\lambda$ vanishes on $\operatorname{ran}(P_\mu) = E(\mu, T)$. The second identity holds true on any eigenvector $v \in E(\mu, T)$, since $P_\lambda(v) = 0$ for all $\lambda \neq \mu$ and $P_\mu(v) = v$. Likewise, this implies the spectral resolution on any eigenvectors in $E(\mu, T)$, since

$$\sum_\lambda \lambda P_\lambda(v) = \mu P_\mu(v) = \mu v = Tv.$$

□

**3.3. Proposition (Orthogonal projection formula):** *If $W \subset F^n$ is a subspace, and $v_1, ..., v_k$ is an orthonormal basis of $W$, then the matrix of orthogonal projection to $W$ is given by*

$$P = \sum_{i=1}^{k} v_i v_i^*$$

*where $v_1 \in F^n$ is an element of $M_{n\times 1}(F)$, and $v_i^* \in M_{1\times n}(F)$ is the conjugate transpose matrix.*

*Proof.* Suppose $v \in W^\perp$, then it clearly holds as $v_i^* v = \langle v, v_i \rangle = 0$, and for the basis vectors $v_j \in W$, the properties of orthonormal basis gives us that $v_i^* v_j = \langle v_j, v_i \rangle$ vanishes when $i \neq j$, and is equal to 1 is $i = j$. □

**3.4. Remark:** The equation $v_i^* v = \langle v, v_i$ comes from the fact that inner product on complex vector space is defined as

$$\langle v, w \rangle = a_1 \overline{b_1} + \cdots + a_n \overline{b_2}$$

and the fact that an inner product of 2 vectors can be thought of as the dot product of 2 column vectors. Hence we would want to take the complex conjugate of $v_i^*$ to get back to $v_i$.

**3.5. Example:** Consider the matrix

$$A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

with complex numbers $a, b \in \mathbb{C}$. This is a normal opreator as it is the sum of two normal operators. The characteristic polynomial can be calculator to be $(z-a)^2 + b^2 = a^2 - 2az + (a^2 + b^2)$ with roots

$$\lambda_1 = a + ib, \quad \lambda_2 = a - ib.$$

Note that $a, b$ need not be real. The normalized eigenvector for $A$ is

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

By the above proposition, we can find that te corresponding projections are:

$$P_{\lambda_1} = v_1 v_1^8 = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}, \quad P_{\lambda_2} = v_2 v_2^* = \frac{1}{2}\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}.$$

Hence, the spectral resolution of our operator is

$$A = (a+ib)\frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} + (a-ib)\frac{1}{2}\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$$

**3.6. Remark:** Some neat applications of spectral resolution are:

- The spectral resolution of the adjoint is simply:

$$T^* = \sum_{\lambda} \overline{\lambda} P_{\lambda}.$$

- Taking powers of $T$, ans using the fact that $P_\lambda P_\mu = 0$ for $\mu \neq \lambda$:

$$T^k = (\sum_{\lambda} \lambda P_\lambda)^k = \sum_{\lambda_1} \cdots \sum_{\lambda_k} \lambda_1 \cdots \lambda_k P_{\lambda_1} \cdots P_{\lambda_k} = \sum_{\lambda} \lambda^k P_\lambda.$$

  Where the last equalities comes from the fact the only non-zero summation terms is when $\lambda_1 = \cdots = \lambda_k$ and hence $P_\lambda^k = P$ by properties of orthogonal projection.

More generally, for any polynomial $q(z)$, we can plug in and see

$$q(T) = \sum_{\lambda} q(\lambda) P_\lambda.$$

Hence, to take a polynomial function of $T$, we keep the same eigenvectors but get new eigenvalues. Note that $q(T)$ depends only on the value of $q$ on the eigenvalues of $T$, with this, an interesting possibility arises. Let

$$Spec(T) \subset \mathbb{C},$$

to be the spectrum of $T$, i.e. the set of eigenvalues of $T$, which is a finite subset of the complex plane. Then for any complex-valued function:

$$f : Spec(T) \to \mathbb{C}$$

on this finite set, we define

$$f(T) = \sum_{\lambda} f(\lambda) P_\lambda.$$

This implies that $f$ only has to be defined on the eigenvalues of $T$ for it to be defined for $f(T)$.

**3.7. Proposition:** *For any normal operator $T$, and any $f$, the operator $f(T)$ has the same eigenvectors as $T$, with eigenvalues $f(\lambda)$.*

*Proof.* We can verify by applying $f(T)$ to an eigenvector $v \in E(\mu, T)$. □

**3.8. (Cont'd):** $f(T)$ is also normal, since $V$ has an orthonormal basis consisting of eigenvectors of $f(T)$, and

$$Spec(f(T)) = f(Spec(T)).$$

Some other easily verifiable properties are:

- If $f$ is the restriction of some polynomial $q(z)$, then $f(T) = q(T)$.
- Special cases:

$$f(\lambda) = \lambda \implies f(T) = T, f(\lambda) = 1 \implies f(T) = I.$$

-

$$(f+g)(T) = f(T) + g(T), (\lambda f)(T) = \lambda f(T)$$

  for $a \in \mathbb{C}$, and

$$(fg)(T) = f(T)g(T)$$

-

$$(f \circ g)(T) = f(g(T))$$

  where $f$ is a function on the image $g(Spec(T)) = Spec(g(T))$.
- Let $\overline{f}$ be the complex conjugate of $f$ defined by $\overline{f}(\lambda) = \overline{f(\lambda)}$, then

$$f(T)^* = \overline{f}(T).$$

  Special cases: $f(\lambda) = \overline{\lambda} \implies f(T) = T^*$. We hence can also deduce that $T^*$ has the same eigenvectors as $T$, we complex conjugate eigenvalues $\overline{\lambda}$

All of this allows us to take fancy function of $T$ like the 'absolute value'

$$|T| \in \mathcal{L}(V)$$

which is a self-adjoint operator. However, always remember that this very much depends on $T$ being normal.

# Chapter 7

# Positive Operators

## Section 1. Definitions and basic properties

**1.1.** We continue with the assumption of $V$ being a complex inner product space of finite dimension. We have sen that self-adjoint operators are like real numbers, skew-adjoint operators like imaginary numbers, and unitary operators are like complex numbers of absolute value 1. The following definition generalizes positive real numbers:

**1.2. Proposition:** *For a self-adjoint operator $T \in \mathcal{L}(V)$, the following conditions are equivalent:*

$$\textit{The eigenvalues of } T \textit{ are non-negative} \iff \textit{For all } v \in V, \langle Tv, v\rangle \geq 0.$$

*Proof.* Suppose for all $\in V, \langle Tv, v\rangle \geq 0$, then in particular this must hold true for all eigenvectors. Let $v$ be an eigenvector for the eigenvalue $\lambda$, then $\langle Tv, v\rangle = \lambda\langle v, v\rangle \geq 0$. Hence $\lambda \geq 0$.
Now suppose that all eigenvalues of $T$ are non-negative. Let $T = \sum_\lambda \lambda P_\lambda$ be the spectral decomposition of $T$. FOr an orthogonal projection $P$, we have $\langle Pv, v\rangle = \langle P^2 v, v\rangle = \langle Pv, Pv\rangle = ||Pv||^2 \geq 0$. Hence

$$\langle Tv, v\rangle = \sum_\lambda \lambda ||P_\lambda v||^2 \geq 0.$$

□

**1.3. Definition:** An operator $T \in \mathcal{L}(V)$ is *positive* if $T$ is self-adjoint and

$$\langle Tv, v\rangle \geq 0$$

for all $v \in V$.

**1.4. Remark:** In this terminology, 0 is considered positive, however one could require a stronger condition that $\langle Tv, v\rangle > 0$ for all nonzero $v \in V$, which would be equivalent to all eigenvalues of $T$ being strictly positve. This stronger notion is called *strictly positive*.

**1.5. Example:**

(1). The zero operator and the identity operators are positive,

(2). For any $T \in calL(V)$, the operators $TT^*$ and $T^*T$ are positive:

$$\langle T^*Tv, v\rangle = \langle Tv, Tv\rangle = ||Tv||^2 \geq 0.$$

(3). All orthogonal projections $P = P^* = P^2$ are positive.

(4). If $T \in calL(V)$ is normal, and any function $f : Spec(T) \to \mathbb{C}$ taking values in $[0, \infty)$, then operator $f(T)$ is positive. In particular, $|T|$ is positive.

**1.6. Proposition:** *If $T_1, T_2 \in \mathcal{L}(V)$ are positive operators, and $a_1, a_2 \geq 0$, then $a_1T_1 + a_2T_2$ is positive.*

*Proof.* For all $v \in V$,

$$\langle (a_1T_1 + a_2T_2)v, v\rangle = a_1\langle T_1v, v\rangle + a_2\langle T_2v, v\rangle \geq 0.$$

□

**1.7. Definition:** An operator $R$ is called a **square root** of an operator $T$ if $R^2 = T$.

**1.8. Proposition (Characterization of positive operators):** *If $T \in \mathcal{L}(V)$ is positive, then $T$ admits a unique positive square root.*

*Proof.* To prove existence, since $\operatorname{Spec}(T) \subseteq [0,\infty)$, we may apply functional calculus to restriction of the positive square root $\sqrt{\cdot} : [0,\infty) \to [0,\infty)$ to define an operator $S = \sqrt{T}$ with $S^2 = T$, which is

$$\sqrt{T} = \sum_{\lambda} \sqrt{\lambda} P_\lambda.$$

To prove uniqueness, if $S$ is any square root, and $v \in E(\mu, S)$ is an eigenvector, then $v \in E(\mu^2, T)$, as $Tv = SSv = \mu Sv = \mu^2 v$. Hence $v \in E(\lambda, T)$ for some eigenvalue $\lambda$ of $T$ where $\mu = \sqrt{\lambda}$. In other words, $Sv = \sqrt{T}v$ for all eigenvectors of $S$ and so $S = \sqrt{T}$ (since there exists an orthonormal basis consisting of eigenvalues). □

## Section 2. Polar decomposition and singular values

**2.1. Motivation:** Any complex number $z$ may be written as a product of a positive number $r$ and a complex number $u$ of absolute value 1, formally:

$$z = |z| \cdot \frac{z}{|z|^{-1}} = ru$$

The decomposition is unique when $z \neq 0$, and with our analogy between complex numbers and operator, we hope to extend this to operators.

**2.2. Theorem:** *For any invertible $T \in \mathcal{L}(V)$, there are unique operators $U \in \mathcal{L}(V)$ and $R \in \mathcal{L}(V)$ such that $U$ is unitary, $R$ is positive, and*

$$T = UR.$$

*Proof.* Suppose $T = UR$, then $T^*T = R^*U^*UR = RU^{-1}UR = R^2$, hence $R = \sqrt{T^*T}$. If $T$ is invertible, then $T^*T$ is also invertible, and hence $R$ is invertible. For if not, then there would exist a nonzero vector $x$ such that $Rx = 0 \implies R^2x = 0 \implies T^*Tx = 0$, which contradicts the invertibility of $T^*T$. In this case $U = TR^{-1}$, and we can check that it is indeed unitary using $R^* = R$:

$$U^*U = (R^{-1})^*T^*TR^{-1} = R^{-1}R^2R^{-1} = I.$$

□

**2.3. Remark:** The unitary operator above is written to the left, there is another polar decomposition $T = RU$ where the unitary operator is written to the right, where we would have to taken $R = \sqrt{TT^*}$ instead.

*The polar decomposition is a bit more tricky for $T$ not invertible, we will need to first prove some lemmas:*

**2.4. Lemma:** *For finite-dimensional inner product spaces $V, W$, and any $T \in \mathcal{L}(V, W)$, there are orthogonal decomposition.*

$$V = null(T) \oplus ran(T^*), \quad W = null(T^*) \oplus ran(T).$$

*The operator $T$ restricts to an isomorphism $T_1 : ran(T^*) \to ran(T)$.*

*Proof.* We have already proved that the respective subspaces are indeed orthogonal. Since the dimension of a subspace and its orthogonal complement adds up to the total dimension, we have that they must be direct sums. An analogous argument can be made for the second formula. □

**2.5. Lemma:** *For finite-dimensional inner product spaces $V, W$ and any $T \in \mathcal{L}(V, W)$,*

$$null(T) = null(T^*T), \quad ran(T^*) = ran(T^*T).$$

*Proof.* For the first equality, let $v \in V$ such that $T^*Tv = 0$, then $0 = \langle T^*Tv, v\rangle = \langle Tv, Tv\rangle = ||Tv||^2$, hence $Tv = 0$; conversely, if $Tv = 0$, then $T^*Tv = 0$ as well. For the second equality, note that $\mathrm{ran}(T^*T)$ is the orthogonal complement of $\mathrm{null}(T^*T) = \mathrm{null}(T)$, using the previous lemma, $\mathrm{ran}(T^*) = \mathrm{ran}(T^*T)$. □

**2.6. Theorem (Polar decomposition of operators):** *For any $T \in calL(V)$, there are operators $U \in \mathcal{L}(V)$ and $R \in \mathcal{L}(V)$ where $U$ is unitary and $R$ is positive and*

$$T = UR.$$

*Here $R$ is unique and equal to $\sqrt{T^*T}$.*

*Proof.* First, if such a decomposition exists, then $T^*T = R^*U^*UR = R^2$, hence $R = \sqrt{T^*T}$. To prove the existence of such a decomposition, we shall define $U$ as a sum of two isometric isomorphisms

$$U_1 : ran(T^*) \to ran(T), U_2 : null(T) \to null(T^*).$$

Actually, we may take $U_2$ as any isometric isomorphisms(for example, the map from a given orthonormal basis of $\mathrm{null}(T)$ to an orthonormal basis of $\mathrm{null}(T^*)$). For the definition of $U_1$, let

$$T_1 : ran(T^*) \to ran(T)$$

be the restriction of $T$ by lemma above, and

$$R_1 = \sqrt{T_1^*T_1} : ran(T^*) \to ran(T).$$

Since $T_1^*T_1 = (T^*T)|_{ran(T^*)}$, $R_1$ then coincides with the restriction of $R = \sqrt{T^*T}$. The null space of $R$ (a normal operator) coincides with the null space of its square, $T^*T$, hence with $null(T) = ran(T^*)^\perp$, we have

$$null(R) = null(R^2) = null(T^*T) = null(T) = ran(T^*)^\perp.$$

Hence, $R_1$ is invertible, as all vectors in $null(R)$ are in $ran(T^*)^\perp$. We may put

$$U_1 = T_1 R_1^{-1}.$$

Note this is how we defined $U$ in the case where $T$ is invertible, and by the same reasoning it is unitary:

$$U_1^* U_1 = (R_1^{-1})^* T_1 T^* T_1 R_1^{-1} = (R_1^{-1}) R_1^2 R_1^{-1} = I.$$

Finally, let $U = U_1 + U_2$, so

$$Uv = U_1 v_1 + U_2 v_2$$

where we write $v = v_1 + v_2$ with $v_1 \in ran(T^*), v_2 \in null(T)$. Using pythagoras' theorem, we see that $U$ is indeed unitary(isometry):

$$||U_v||^2 = ||U_1 v_1||^2 + ||U_2 v_2||^2 = ||v_1||^2 + ||v_2||^2 = ||v||^2.$$

Therefore, we have $T = UR$ where $Tv = Tv_1$ and similarly

$$UR_v = URv_1 = U_1 R_1 v_1 = T_1 v_1 = Tv_1.$$

$\square$

**2.7. Remark:** The restriction of $U$ to ran$(T^*)$ is a uniquely determined isometric isomorphism $ran(T^*) \to ran(T)$, hence the ambiguity in the choice of $U$

## Section 3. Singular Values

**3.1. Definition:** Let $T \in \mathcal{L}(V, W)$ be a linear map between finite-dimensional inner product spaces. The **singular values** of $T$ are the eigenvalues of $\sqrt{T^*T}$, where the multiplicity of a singular value is its multiplicity as an eigenvalue of $\sqrt{T^*T}$.

**3.2. Remark:** In order to compute the singular values of $T$, it is not necessary to actually find the square root of $T^*T$, as they are simply the square roots of the eigenvalues of $T^*T$

**3.3.** Given $T \in calL(V, W)$, recall that $null(T)^\perp = ran(T^*) = ran(T^*T)$. Let

$$v_1, ..., v_k \in ran(T^*)$$

be an orthonormal basis consisting of eigenvectors of $T^*T$ with corresponding eigenvalues $s_1^2, ..., s_k^2$, the squares of the non-zero singular values. If we restrict $T$ to an isomorphism from $ran(T^*) \to ran(T)$, the basis $v_i$ determines a basis of $ran(T)$:

**3.4. Lemma:** *The vectors $w_1, ..., w_k$ given by*

$$w_i = \frac{1}{s_i} T(v_i)$$

*are an orthonormal basis of $ran(T)$. In fact, they are eigenvectors of $TT^*$, with eigenvalues $s_i^2$.*

*Proof.* We check:

$$\langle w_i, w_j \rangle = \frac{1}{s_i s_j} \langle T v_i, T v_j \rangle = \frac{1}{s_i s_j} \langle T^* T v_i, v_j \rangle = \frac{s_i}{s_j} \langle v_i, v_j \rangle = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

and

$$TT^* w_i = \frac{1}{s_i} TT^* T(v_i) = s_i T v_i = s_1^2 w_i.$$

□

**3.5.** The formula rearranged to

$$T(v_i) = s_i w_i$$

describes $T|_{ran(T^*)}$ in terms of a basis, but since $T$ vanishes on $ran(T^*)^\perp = null(T)$, it describes $T$ itself! Explicitly, we have

$$T(v) = \sum_{i=1}^{k} s_i \langle v, v_i \rangle w_i$$

Indeed, if $v \in ran(T^*)^\perp = null(T)$, both sides are 0, and for $v = v_k$, it reduces to the expression above

**3.6. Theorem (Singular value decomposition):** *Every operator $T \in \mathcal{L}(V, W)$ may be written in the form*

$$T(v) = \sum_{i=1}^{k} s_1 \langle v, v_i \rangle w_i$$

*where $s_1, .., w_k > 0$, and $v_1, ..., v_k$ and $w_1, ..., w_k$ are orthonormal sets of vectors. In this expression, the $s_i s$ are the strictly positive singular values, the $v_i$ are unit eigenvectors of $T^*T$, and the $w_i$ are unit eigenvector of $TT^*$, with*

$$T v_i = s_i w_i, T^* w_i = s_i v_i.$$

*Proof.* We have proved most of this above, it only remains to show that if $T$ is given by this formula, then $s_i$ are the strictly positive singular values. The formula gives that $T(v_j) = s_j w_j$, and the dual map is computed as

$$\langle T^* w, v \rangle = \langle w, Tv \rangle = \sum_{i=1}^{k} s_i \langle w, w_i \rangle \langle v_i, v \rangle,$$

for $w \in W, v \in V$. Hence

$$T^*(w) = \sum_{i=1}^{k} s_i \langle w, w_i \rangle v_i.$$

Putting $w = w_j$ this shows $T^*(w_j) = s_j v_j$. Together this gives

$$T^*Tv_i = s_i T^* w_i = s_i^2 v_i,$$

and similarly, $TT^*w_i = s_i^2 w_i$. These are all the eigenvector for non-zero eigenvalues for $T^*T$ and $TT^*$, since the rank of these operators cannot be larger than $k$. □

**3.7. Remark:** In the case of a normal operator, we can take $v_i$ to be eigenvectors of $T$, with eigenvalues $\lambda_i$, and so $s_i = |\lambda_i|$, and $w_i = \frac{\lambda_i}{|\lambda_i|} v_i$. Which we see that the spectral resolution $T = \sum_i \lambda_i v_i v_i^*$ is related to the singular value decomposition.

**3.8.** Singular value decomposition is also very useful for operators thought of as matrices. Suppose $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is invertible, then all of its singular values are strictly positive. Let $v_1, ..., v_n$ and $w_1, ..., w_n$ be constructed as above. Denoted by

$$U_1 = (v_1, ..., v_n), U_2 = (w_1, ..., w_n),$$

which are unitary matrices having these bases as their columns and

$$D = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & s_n \end{pmatrix}$$

the diagonal matrix having the singular values as its entries. Then $Av_i = s_i w_i$ means

$$AU_1 = U_2 D,$$

or

$$A = U_2 D U_1^{-1}.$$

This gives us the following version of SVD.

**3.9. Theorem:** *Every invertible $n \times n$ matrix $A$ can be written in the form*

$$A = U_2 D U_1^{-1}$$

*where $U_1, U_2$ are unitary, and $D$ is a diagonal matrix with strictly positive entries.*

**3.10. Remark:** This is also related to the polar decomposition $A = UR$ if we just let $U = U_2 U_1^{-1}$ and $R = U_1 D U_1^{-1}$. SO once we have the singular decomposition we also get the polar decomposition.

**3.11. Note:** For matrices $A \in \mathcal{M}_{m\times n}(\mathbb{C})$ that are not square, (or have a null space), the singular value decomposition is more messy. As usual, we can think of $A$ as a linear map $\mathbb{C}^n \to \mathbb{C}^m$. Construct an orthonormal basis for $ran(A^*)$ $v_1, ..., v_k$, and an orthonormal basis of $ran(A)$ $w_1, ..., w_k$ as before. Then extend to orthonormal basis $v_1, ..., v_n$ of $\mathbb{C}^n$ and $w_1, ..., w_m$ of $\mathbb{C}^m$, respectively. Let $U_1$ be the unitary matrix having $v_1, .., v_n$ as its columns, and $U_2$ be the matrix having $w_1, ..., w_n$ as its columns, and let $D \in \mathcal{M}_{m\times n}(\mathbb{C})$ be the matrix having $s_i$ as its $(i,i)$ entry for $i \leq k$, and all other entries equal to zero. Then

$$A = U_2 D U_1^{-1}.$$

In summary:

**3.12. Theorem:** *Every matrix $A \in \mathcal{M}_{m\times n}$ can be written in the form*

$$A = U_2 D U_1^{-1}$$

*where $U_1 \in \mathcal{M}_{n\times n}(\mathbb{C})$, $U_2 \in \mathcal{M}_{m\times m}$ are unitary, and $D \in \mathcal{M}_{m\times n}$ has strictly positive $(i,i)$ entries for $i \leq k$ and all other entries equal to zero.*

**3.13. Remark:** Note that

$$A^* = U_1 D^* U_2^{-1}$$

is the singular value decomposition for $A^*$. Hence, the calculation for $A, A^*$ are essentially the same. In practice, one would start with $AA^*$ or $A^*A$, depending on which of the matrices is smaller.

# Chapter 8

# More Decompositions

## Section 1. Schur's Theorem

**1.1. Theorem:** *Let $V$ be a finite-dimensional complex inner product space, and $T \in \mathcal{L}(V)$. Then there exists an orthonormal basis $v_1, ..., v_n$ such that the matrix of $T$ in this basis is upper triangular.*

*Proof.* Pick any basis $w_1, ..., w_n$ in which $T$ is upper triangular, which means

$$Tw_i = \text{span}\{w_1, ..., w_i\}$$

for all $i = 1, ..., n$. Such a basis always exist by the Jordan Normal Form. Let

$$W_i = \text{span}\{w_1, ..., w_i\} \subseteq V.$$

Then

$$0 = W_0 \subseteq W_1 \subseteq W_2 \subseteq \cdots \subseteq W_n = V$$

are all $T$-invariant subspaces with dim $W_i = i$, hence $\dim W_i - \dim W_{i-1} = 1$ and so $W_i \cap W_{i-1}$ is 1-dimensional. Let each $v_i$ be a unit vector in $W_i \cap W_{i-1}^{\perp}$, then each $v_i \in W_i$ while $v_{i+1} \in W_i^{\perp}$. Hence set of unit vectors $v_1, ..., v_n$ forms an orthonormal basis of $V$. Since

$$Tv_i \in W_i = \text{span}\{v_1, ..., v_n\}$$

we have $T$ is upper-triangular with respect to this basis. □

**1.2. Remark:** We can use Gram-Schmidt to show this as well.

**1.3.** Begin proving the following theorems, lets first prove an important lemma which will be quite useful.

**1.4. Lemma:** *A matrix $T$ is unitary, if and only if the columns of $T$ form an orthonormal basis.*

*Proof.* By definition, $T$ is unitary if and only if $TT^* = I$. The $(i, j)$-th entry of $TT^*$, by definition of matrix multiplication, is

$$[TT^*]_{i,j} = \sum_k t_{ik} t^*_{kj}$$

since $T^* = \overline{T}^t$, we have $t^*_{kj} = \bar{t}_{jk}$, hence we are really just taking the inner product of the $i$-th column and $j$-th columns. However $T$ is unitary if and only if

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

hence we can see that $T$ unitary if and only if the columns of $T$ are an orthonormal basis. □

**1.5. Theorem:** *The Schur's theorem for matrices state every complex matrix $A$ may be written where*

$$A = UBU^{-1}$$

*where $U$ is unitary and $B$ is upper triangular.*

*Proof.* Let $v_1, ..., v_n$ be the orthonormal basis from Schur's theorem, and let $B$ be the upper-triangular matrix with respect to this basis. Then $Av_i = Bv_i = \sum_{j \leq i} B_{ji} v_j$. Taking $v_1, ..., v_n$ as the columns of $U$, by the above lemma, $U$ is unitary, hence letting $B$ be the upper triangular matrix having $B_{ji}$ as its non-zero entries, we get $AU = UB$ which implies $A = UBU^{-1}$. $\square$

**1.6.** Note that in the above proof, we used the fact that given two matrices $A$ and $U$ having $v_1, ..., v_n$ as its columns.

$$A \left(v_1, ..., v_n\right) = \left(Av_1, ..., Av_n\right)$$

In general, we cannot bring $A$ into Jordan Normal Form through unitary transformations, but we can at least bring it into upper triangular form.

**1.7. Theorem:** *Every complex matrix $A \in \mathcal{M}n \times n(\mathbb{C})$ may be written in the form*

$$A = UB$$

*where $U$ is unitary, and $B$ is upper triangular with positive diagonal entries. If $A$ is invertible then this decomposition is unique with strictly positive diagonal entries.*

*Proof.* Let $w_1, ..., w_n$ be the columns of $A$. Then $A = UB$ says

$$w_i = \sum_j B_{ji}$$

where $v_1, ..., v_n$ is the orthonormal basis given by the columns of $U$.So given a matrix $A$, we want to construct the $v_i$'s such that $B_{ii} \geq 0$ and $B_{ji} = 0$ for all $j > i$. Equivalently, this is saying that we want

$$w_i \in \text{span}(v_1, ..., v_n)$$

for all $i = 0, ..., n$ and the $i$-th coefficient of each $w_i$ to be $\geq 0$.
If $A$ is invertible, then by Gram-Schmidt, there exists a unique orthonormal basis $w_1, ..., w_n$ which satisfies our requirement, as Gram-Schmidt produces an orthonormal basis in which the change of basis matrix is upper triangular with positive diagonal entries. If $A$ is not invertible, we use a slightly modified version of Gram Schmidt. First suppose by induction that we have an orthonormal set of vectors $v_1, ..., v_k$ satisfied our conditions. To construct $v_{k+1}$, we consider two cases:
**Case 1:** If $w_{k+1} \notin \text{span}\{v_1, ..., v_k\}$, simply apply Gram-Schmidt to $w_{k+1}$ to get $v_{k+1}$.
**Case 2:** If $w_{k+1} \in \text{span}\{v_1, ..., v_k\}$, then take any unit vector $v_{k+1}$ that is orthogonal to $v_1, ..., v_k$. Then we have $w_{k+1} \in \text{span}\{v_1, ..., v_{k+1}\}$ and the $k+1$-th coefficient is 0.
Since $U$ is unitary and $B$ is upper triangular, this concludes the proof. □

**1.8. Remark:** For $A$ invertible, we can go one step further and decompose $B = DN$ where $D$ is a diagonal matrix and $N$ is upper triangular with 1's on the diagonal with row reduction. The resulting decomposition

$$A = UDN$$

is called the Iwasawa decomposition of an invertible matrix.

**1.9.** One might try to relate the decomposition $A = UB$ to the polar decomposition $A = U'P$, of course, the unitary matrices are not the same, but we can relate $B$ and $P$ as follows:

$$P^2 = P^*P = A^*A = B^*B,$$

so $P = \sqrt{B^*B}$.

**1.10. Theorem (Cholesky decomposition):** *Every positive definite matrix $A$ can be written as*

$$A = B^*B$$

*where $B$ is upper triangular, with positive diagonal entries. If $A$ is strictly positive definite then* *this decomposition is unique.*

*Proof.* since $A$ is positive, it admits a unique positive square root $\sqrt{A}$, write $\sqrt{A} = UB$ where $B$ is upper triangular with positive diagonal entries. Then

$$A = (\sqrt{A})^2 = (\sqrt{A})^* \sqrt{A} = B^* B.$$

□

# Chapter 9

# Operators on real inner product spaces

## Section 1. Complexification

**1.1.** Most of our discussion of inner product spaces has been focused on the field of complex vectors, and that is because there are some important differences in the real case. In particular, the complex spectral theorem for normal operators becomes false in $\mathbb{R}$ unless the operator is self-adjoint. To make up for this, one of the main technique is to replace the real vector space with a complex one, through a process of 'complexification'.

**1.2. Note:** For a matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R})$, we can view it as a complex matrix who entries happen to be real. Lets denote it by $A_{\mathbb{C}} \in \mathcal{M}_{m\times n}(\mathbb{C})$ for now. As linear maps,

$$A : \mathbb{R}^n \to \mathbb{R}^m \quad A_{\mathbb{C}} : \mathbb{C}^n \to \mathbb{C}^m.$$

We call this extended map the *complexification* of $A$. If $m = n$, then

$$\det(A_{\mathbb{C}}) = \det(A)$$

since both are given by the same formula as a sum over permutations. Similarly,

$$tr(A) = tr(A_{\mathbb{C}}).$$

**1.3. Remark:** For a real polynomial

$$p(t) = a_0 + \cdots + a_m t^m$$

let

$$p_{\mathbb{C}}(z) = a_0 + \cdots + a_m z^m$$

be the corresponding complex polynomial (with real coefficients). This is the complexification of $p$.

**1.4. Proposition:** *Let $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ be a real $n \times n$-matrix, with characteristic polynomial $q$ and minimal polynomial $p$. Then $A_{\mathbb{C}}$ has characteristic polynomial $q_{\mathbb{C}}$ and minimal polynomial $p_{\mathbb{C}}$.*

*Proof.* The characteristic polynomial of $A$ is

$$q(t) = \det(tI - A) = \sum_{\sigma} \operatorname{sign}(\sigma)(tI - A)_{\sigma(1)1} \cdots (tI - A)_{\sigma(n)n}$$

where the characteristic polynomial of $A_{\mathbb{C}}$ is

$$q(t) = \det(tI - A_{\mathbb{C}}) = \sum_{\sigma} \operatorname{sign}(\sigma)(tI - A_{\mathbb{C}})_{\sigma(1)1} \cdots (tI - A_{\mathbb{C}})_{\sigma(n)n} = \sum_{\sigma} \operatorname{sign}(\sigma)(tI - A)_{\sigma(1)1} \cdots (tI - A)_{\sigma(n)n}.$$

Hence we get that the char polynomial of $A_{\mathbb{C}}$ can be obtained by replacing the real variable $t$ with the complex variable $z$, hence it is indeed $q_{\mathbb{C}}$.
The minimal polynomial of $A$ is the unique monic polynomial $p(t) = a_0 + a_1 t + \cdots + a_{m-1} t^{m-1} + t^m$ of smallest degree $m$ with $p(A) = 0$, then $P_{\mathbb{C}}(A_{\mathbb{C}}) = 0$, which means $p_{\mathbb{C}}$ is divisible by the minimal polynomial of $A_{\mathbb{C}}$. Now to show that the minimal polynomial is indeed $p_{\mathbb{C}}$, suppose the minimal polynomial of $A_{\mathbb{C}}$ is $c_0 + c_1 z + \cdots + c_{k-1} z^{k-1} + z^k$ with complex numbers $c_i$. Taking the real part

of each copmmlex coefficients and using the fact that entries of $A_{\mathbb{C}}$ and all its powers are real, we obtain

$$Re(c_0) + Re(c_1)A + \cdots + A^k = 0$$

Since $p$ is divisible by $A_{\mathbb{C}}$, this means that $k = m$. The uniqueness of the minimal polynomial of $A_{\mathbb{C}}$ shows that $c_i = a_i$. □

**1.5. Proposition:** *Let $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ be a real $n \times n$-matrix. In the Jordan Normal Form of $A_{\mathbb{C}}$, the number of Jordan blocks of size $k$ for $\lambda$ equals the number of Jordan blocks of size $k$ for $\overline{\lambda}$. In particular, $\lambda$ is an eigenvalue if and only if $\overline{\lambda}$ is an eigenvalue.*

*Proof.* Let $C \in \mathcal{M}_{n\times n}(\mathbb{C})$ be an invertible complex matrix with

$$CA_{\mathbb{C}}C^{-1} = J.$$

Taking the complex conjugate of this equation, $\overline{A_{\mathbb{C}}} = A_{\mathbb{C}}$, we have

$$\overline{C}A_{\mathbb{C}}\overline{C}^{-1} = \overline{J}.$$

clearly $\overline{J}$ is also a JNF. By uniqueness of JNF, $J$ and $\overline{J}$ must be the same matrix up to rearrangement of the Jordan blocks. Hence the desired result, in particular, the complex eigenvalues of a real matrix come matrix come in complex conjugate pairs, and its complex conjugate have the same geometric and algebraic multiplicities. Moreover, since

$$(\lambda I - A_{\mathbb{C}})v = 0 \iff (\overline{\lambda}I - A_{\mathbb{C}})\overline{v} = 0$$

complex conjugation, we see: □

**1.6. Lemma:** *For $A \in \mathcal{M}_{n\times n}(\mathbb{R})$, complex conjugation takes eigenvectors for the eigenvalues $\lambda$ of $A_{\mathbb{C}}$ to eigenvectors for the eigenvalue $\overline{\lambda}$.*

**1.7. Remark:** More generally,

$$(\lambda I - A_{\mathbb{C}})^k v = 0 \iff (\overline{\lambda}I - A_{\mathbb{C}})^k\overline{v} = 0$$

so we have a similar argument for generalized eigenvectors.

**1.8. Definition:** Let $V$ be a real vector space. The **complexification** of $V$, denoted $V_{\mathbb{C}} = V \times V$. An element of $V_{\mathbb{C}}$ is an ordered pair $(u, v)$ written as $u + iv$, where $u, v \in V$. Addition on $V_{\mathbb{C}}$ is defined by

$$(u_1 + iw_1) + (u_2 + iw_2) = (u_1 + u_2) + i(w_1 + w_2)$$

for $u_1, u_2, w_1, w_2 \in V$ and scalar multiplication by a complex number $a + bi$ is defined by

$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$

for $a, b \in \mathbb{R}$ and $u, v \in V$.

**1.9. Lemma:** *$V_c$ with the operations of addition and scalar multiplication defined as above is a complex vector space.*

**1.10. Remark:** We think of $V$ as a subset of $V_{\mathbb{C}}$ consisting of vectors $u + i0$, where we usually just write $u$ in place of $u + i0$.

**1.11. Example:**

- $(\mathbb{R}^n)_{\mathbb{C}} = \mathbb{C}^n$
- $\mathcal{M}_{m\times n}(\mathbb{R})_{\mathbb{C}} = \mathcal{M}_{m\times n}(\mathbb{C})$
- $\mathcal{P}(\mathbb{R})_{\mathbb{C}} = \mathcal{P}(\mathbb{C})$

**1.12. Note:** For $u, w \in V$, we call

$$\mathrm{Re}(u + iw) = u, \ \mathrm{Im}(u + iw) = w$$

the real and imaginary of $u + iw$.
On $V_{\mathbb{C}}$, we have an additional operation that is not present on general complex vector spaces: *complex conjugate*

$$V_{\mathbb{C}} \to V_{\mathbb{C}}, v = \mathrm{Re}(v) + i\mathrm{Im}(v) \mapsto \overline{v} = \mathrm{Re}(v) - i\mathrm{Im}(v).$$

This has the same properties as complex conjugation on $\mathbb{C}^n$:

$$\overline{\overline{v}} = v, \overline{v_1 + v_2} = \overline{v_1} + \overline{v_2}, \overline{\lambda v} = \overline{\lambda}\overline{v}$$

In particular, this operations defines a conjugate linear map, and elements of $V$ is fixed under complex conjugation.

**1.13. Lemma:** *A complex subspace $U \subseteq V_{\mathbb{C}}$ is of the form $U = W_{\mathbb{C}}$ for a subspace $W \subseteq V$ if and only if it is invariant under complex conjugation, i.e.:*

$$v \in U \implies \overline{v} \in U.$$

*Proof.* Let $u_1, ..., u_k$ be a basis of $U$, then

$$a_1u_1 + \cdots + a_nu_n = 0$$

implies $a_1 = ... = a_n = 0$, taking the complex conjugation of both side, since $\overline{0} = 0$, we have that $\overline{u_1}, ..., \overline{u_n}$ is basis since its linear independent of right dimension. Then $U$ is spanned by $u_1, ..., u_k$ and $\overline{u_1}, \overline{u_k}$, hence it is also spanned by Re($v_1$),...,Re($v_k$),Im($v_1$),...,Im($v_n$). Letting $W =\subsetneq$ be the real subspace spanned by these vectors, it follows that $U = W_{\mathbb{C}}$. □

**1.14. Proposition:** *If $v_1, ..., v_n$ is a basis of a real vector space $V$, then $v_1, ..., v_n$ is a basis of $V_{\mathbb{C}}$, and $\dim V = \dim V_{\mathbb{C}}$.*

*Proof.* Let $v_1, ..., v_n$ be a basis of $V$, given $v \in V_{\mathbb{C}}$, since $\mathrm{Re}(v) \in V$ and $\mathrm{Im}(v) \in V$, we can write $\mathrm{Re}(v) = \sum_k a_k v_k$ and $\mathrm{Im}(v) = \sum_k b_k v_k$, for $a_k, b_k \in \mathbb{R}$, then

$$v = \sum_k (a_k + ib_k) v_k,$$

hence $v_1, ..., v_n$ spans $V_{\mathbb{C}}$ as a complex vector space. To show that the set of vectors is linear independent, if $\sum_k (a_k + ib_k) v_k = 0$, we can take the real and imaginary parts to obtain $\sum_k a_k v_k = 0$ and $\sum b_k v_k = 0$, and hence they are indeed independent. The second statements follows immediately. □

**1.15.** Now we can define the complexification of an operator:

**1.16. Definition:** Given a linear operator $T \in \mathcal{L}(V, W)$, we obtain a complex-linear operator $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}}, W_{\mathbb{C}})$ defined by

$$T_{\mathbb{C}} v = T\mathrm{Re}(v) + iT\mathrm{Im}(v).$$

**1.17. Proposition:**

- *Complexification is linear in the sense that*

$$(T_1 + T_2)_{\mathbb{C}} = (T_1)_{\mathbb{C}} + (T_2)_{\mathbb{C}}, (aT)_{\mathbb{C}} = aT_{\mathbb{C}} \quad \text{for } a \in \mathbb{R}.$$

- *Complexification of the identity operator in on $V$ is $I_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$.*
- *Under composition of operators,*

$$(ST)_{\mathbb{C}} = S_{\mathbb{C}} T_{\mathbb{C}}.$$

*In particular, $T_{\mathbb{C}}$ is an isomorphism if and only if $T$ is an isomorphism, and in this case, $(T_{\mathbb{C}})^{-1} = (T^{-1})_{\mathbb{C}}$.*

**1.18. Proposition:** *Suppose $v_1, ..., v_n$ is a basis of $V$, and $w_1, ..., w_n$ is a basis of $W$. If $T$ has matrix $A \in \mathcal{M}_{m \times n}(\mathbb{R})$ in this basis. Then the matrix of $T_{\mathbb{C}}$ is $A_{\mathbb{C}}$.*

*Proof.* Trivial. □

**1.19. Remark:**

- For $T \in \mathcal{L}(V)$, the trace and determinant of $T_{\mathbb{C}}$ are equal to those of $T$.
- For $T \in \mathcal{L}(V)$, the characteristic and minimal polynomial of $T_{\mathbb{C}}$ are the complexification of those of $T$.
- For $T \in \mathcal{L}(V)$,

$$\dim null((\lambda I - T_{\mathbb{C}})^k) = \dim null((\overline{\lambda} I - T_{\mathbb{C}})^k)$$

for all $\lambda \in \mathbb{C}, k \in \mathbb{N}$.
- In particular, eigenvalues come in complex conjugate pairs, $\lambda, \overline{\lambda}$, where they also have the

same geometric and algebraic multiplicity.

**1.20. Proposition:** *Suppose $V$ is a real vector space and $T \in \mathcal{L}(V)$, $\lambda \in \mathbb{R}$, then $\lambda$ is an eigenvalue of $T_c$ if and only if $\lambda$ is an eigenvalue of $T$.*