Lecture Notes for Math 414: Linear Algebra II
Fall 2015, Michigan State University

Matthew Hirn

December 11, 2015

Beginning of Lecture 1

1 Vector Spaces

What is this course about?

1. Understanding the structural properties of a wide class of spaces which all share a similar additive and multiplicative structure
   structure = "vector addition and scalar multiplication" → vector spaces

2. The study of linear maps on finite dimensional vector spaces

We begin with vector spaces. First two examples:

1. $\mathbb{R}^n = n$-tuples of real numbers $x = (x_1, \ldots, x_n)$, $x_k \in \mathbb{R}$
   vector addition: $x+y = (x_1, \ldots, x_n)+(y_1, \ldots, y_n) = (x_1+y_1, \ldots, x_n+y_n)$
   scalar multiplication: $\lambda \in \mathbb{R}$, $\lambda x = \lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n)$

2. $\mathbb{C}^n$ [on your own: review 1.A on complex numbers]

1.B Definition of Vector Space

Scalars: Field $\mathbb{F}$ (assume $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ unless otherwise stated). So the previous two vector spaces can be written as $\mathbb{F}^n$ with scalars $\mathbb{F}$

Let V be a set (for now).

Definition 1 (Vector addition). $u, v \in V$, assigns an element $u + v \in V$

Definition 2 (Scalar multiplication). $\lambda \in \mathbb{F}$, $v \in V$, assigns an element $\lambda v \in V$

Definition 3 (Vector space). A set V is a vector space over the field $\mathbb{F}$ if vector addition and scalar multiplication are defined, and the following properties hold ($u, v, w \in V$, $a, b \in \mathbb{F}$):

1. Commutativity: $u + v = v + u$
2. Associativity: $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$
3. Additive Identity: $\exists\, 0 \in V$ such that $v + 0 = v$
4. Additive Inverse: for every v there exists w such that $v + w = 0$
5. Multiplicative Identity: $1v = v$
6. Distributive Properties: $a(u + v) = au + av$ and $(a + b)v = av + bv$

If $\mathbb{F} = \mathbb{R}$, "real vector space"
If $\mathbb{F} = \mathbb{C}$, "complex vector space"

From here on out V will always denote a vector space

Two more examples of vector spaces:

1. $\mathbb{F}^\infty$: $x = (x_1, x_2, \ldots)$ just like $\mathbb{F}^n$
2. $\mathbb{F}^S =$ the set of functions $f : S \to \mathbb{F}$ from S to $\mathbb{F}$ [check on your own]

Now for some important properties...

Proposition 1. *The additive identity is unique.*

Proof. Let 0_1 and 0_2 be any two additive identities. Then

$$0_1 = 0_1 + 0_2 = 0_2 + 0_1 = 0_2$$

□

Proposition 2. *The additive inverse is unique.*

Proof. Let w_1 and w_2 be two additive inverses of v. Then:

$$w_1 = w_1 + 0 = w_1 + (v + w_2) = (v + w_1) + w_2 = 0 + w_2 = w_2$$

□

Now we can write $-v$ as the additive inverse of v and define subtraction as $v - w = v + (-w)$. On the other hand, we still don't "know" that $-1v = -v$!

Notation: We have $0_{\mathbb{F}}$ and 0_V. In the previous two propositions we dealt with 0_V. Next we will handle $0_{\mathbb{F}}$. We just write 0 for either and use the context to determine the meaning.

Proposition 3. $0_{\mathbb{F}}v = 0_V$ *for every* $v \in V$

Proof.

$$0v = (0 + 0)v = 0v + 0v \Longrightarrow 0v = 0$$

□

Now the other way around...

Proposition 4. $\lambda 0 = 0$ *for every* $\lambda \in \mathbb{F}$

Proposition 5. $(-1)v = -v$ *for all* $v \in V$

Proof.

$$v + (-1)v = 1v + (-1)v = (1 + (-1))v = 0v = 0$$

Now use uniqueness of additive inverse. □

End of Lecture 1

Beginning of Lecture 2

Warmup: Is the empty set $\emptyset$ a vector space?
Answer: No since $0 \notin \emptyset$

1.C Subspaces

A great way to find "new" vector spaces is to identify subsets of an existing vector space which are closed under addition and multiplication.

Definition 4 (Subspace)**.** $U \subset V$ is a subspace of V if U is also a vector space (using the same vector addition and scalar multiplication as V).

Proposition 6. *$U \subset V$ is a subspace if and only if:*

1. $0 \in U$
2. $u, w \in U \Longrightarrow u + w \in U$
3. $\lambda \in \mathbb{F}$ *and* $u \in U \Longrightarrow \lambda u \in U$

Now we can introduce more interesting examples of vector spaces, many of which are subspaces of $\mathbb{F}^S$ for some set S [you should verify these are vector spaces]:

1. $\mathcal{P}(\mathbb{F}) = \{p : \mathbb{F} \to \mathbb{F} : p(z) = \underbrace{a_0 + a_1 z + \cdots a_m z^m}_{\deg(p)=m}, \;\; a_k \in \mathbb{F} \;\; \forall\, k, m \in \mathbb{N}\}$
2. $C(\mathbb{R};\mathbb{R})$ = real valued continuous functions
3. $C^m(\mathbb{R}^n;\mathbb{R})$ = real valued functions with continuous partial derivatives up to order m
4. $\mathcal{R}([0,1]) = \{f : [0,1] \to \mathbb{R} : \int_0^1 f(x)\,dx < \infty\}$.
5. $\mathbb{F}^{m,n}$ = the set of all $m \times n$ matrices with entries in $\mathbb{F}$
6. $\mathcal{S} = \{x : [0,1] \to \mathbb{R}^n : x'(t)$ is continuous and $x'(t) = Ax(t),$ where $A \in \mathbb{R}^{n,n}\}$

Another convenient way to get new vector spaces is to add subspaces together (this is like the union of two sets, but for vector spaces!).

Definition 5 (Sum of subsets)**.** Suppose $U_1, \ldots, U_m \subset V$. Then:

$$U_1 + \cdots + U_m := \{u_1 + \cdots + u_m : u_1 \in U_1, \ldots, u_m \in U_m\}.$$

Proposition 7. *Suppose $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is the smallest subspace of V containing $U_1, \ldots, U_m$.*

An example:

$$\begin{aligned}
U_1 &= \{x \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\} \\
U_2 &= \{x \in \mathbb{R}^3 : x_3 = 0\} \\
U_1 + U_2 &= \{x \in \mathbb{R}^3 : x = y + z,\ y_1 + y_2 + y_3 = 0 \text{ and } z_3 = 0\} \\
U_1 + U_2 &= \{x \in \mathbb{R}^3 : x = a(-1,0,1) + b(1,-1,0) + c(1,0,0) + d(0,1,0)\} \qquad (1) \\
U_1 + U_2 &= \mathbb{R}^3
\end{aligned}$$

Note there is redundancy in (1). We will be especially interested in situations that avoid this redundancy, i.e., subspace summations $U_1 + \cdots + U_m$ when the representation $u_1 + \cdots + u_m$ is unique.

Definition 6 (Direct sum)**.** Suppose that $U_1, \ldots, U_m$ are subspaces of V.

- $U_1 + \cdots + U_m$ is a direct sum if each element of $U_1 + \cdots + U_m$ can be written in only one way as $u_1 + \cdots + u_m$ where $u_k \in U_k$.
- If $U_1 + \cdots + U_m$ is a direct sum, then we denote it as $U_1 \oplus \cdots \oplus U_m$

Examples:

1. Let U_k be the subspace of $\mathbb{F}^n$ such that only the k^{th} coordinate is nonzero:
$$U_k = \{(\underbrace{0, \ldots, 0}_{k-1}, x, 0, \ldots, 0) : x \in \mathbb{F}\}$$
Then
$$\mathbb{R}^n = U_1 \oplus \cdots \oplus U_n$$

2. Recall the previous example with redundancy. That is not a direct sum. We can change U_2 though to get a direct sum:
$$\begin{aligned}
U_1 &= \{x \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\} \\
U_2 &= \{x \in \mathbb{R}^3 : x_1 = x_2 = x_3\} \\
\mathbb{R}^3 &= U_1 \oplus U_2
\end{aligned}$$

Notice in the second example that $U_1 \cap U_2 = \{0\}$. This leads us to the following proposition.

Proposition 8. *Let U, W be subspaces of V. Then,*

$$V = U \oplus W \iff U \cap W = \{0\}$$

The first example makes it tempting to propose the same pairwise intersection property for any number of subspaces, but this is not true! [try to come up with an example, then see the book] Instead we have the following proposition, which we can use to prove Proposition 8.

Proposition 9. *Suppose $U_1, \ldots, U_m$ are subspaces of V. Then*

$$\begin{aligned} &U_1 + \cdots + U_m \text{ is a direct sum} \iff \\ &0 = u_1 + \cdots + u_m,\ u_k \in U_k,\ \underline{\text{only}} \text{ when } u_k = 0\ \forall\, k \end{aligned}$$

Proof. The $\Rightarrow$ direction is clear.
For the $\Leftarrow$ direction, let $v \in U_1 + \cdots + U_m$ and suppose we have two representations:

$$v = u_1 + \cdots + u_m = w_1 + \cdots + w_m$$

Then

$$0 = (u_1 - w_1) + \cdots + (u_m - w_m)$$

Since $u_k - w_k \in U_k$, we must have $u_k = w_k$ for each k. □

[try to prove Proposition 8 on your own using Proposition 9, then see the book].

2 Finite Dimensional Vector Spaces

2.A Span and Linear Independence

We saw last time that summing subspaces gives rise to new vector spaces. Now we keep track of each of the vectors that generate these spaces.

Definition 7 (Linear combination)**.** w is a <u>linear combination</u> of the vectors $v_1, \ldots, v_m \in V$ if $\exists\, a_1, \ldots, a_m \in \mathbb{F}$ such that

$$w = a_1 v_1 + \cdots a_m v_m$$

Definition 8 (Span)**.** The span of $v_1, \ldots, v_m \in V$ is

$$\text{span}(v_1, \ldots, v_m) = \{a_1 v_1 + \cdots a_m v_m : a_k \in \mathbb{F} \;\; \forall\, k\}$$

Analogous to the sum of subspaces, we have the following result.

Proposition 10. span$(v_1, \ldots, v_m)$ *is the smallest subspace of* V *containing* $v_1, \ldots, v_m$.

Nomenclature: If span$(v_1, \ldots, v_m) = V$ then we say that $v_1, \ldots, v_m$ spans V.

Definition 9 (Finite dimensional vector space)**.** V is finite dimensional if there exists a finite number of vectors $v_1, \ldots, v_m$ (a list) such that span$(v_1, \ldots, v_m) = V$.

Definition 10 (Infinite dimensional vector space)**.** V is infinite dimensional if it is not finite dimensional.

End of Lecture 2

Beginning of Lecture 3

Warmup: Is this a vector space?

1. $\{f \in C((0,1);\mathbb{R}) : f(x) = x^{-p} \text{ for some } p > 0\}$
   Answer: No (all three properties fail)

2. $\{f \in C(\mathbb{R};\mathbb{R}) : f \text{ is periodic of period } \sigma\}$
   Answer: Yes (contains zero function, closed under addition and scalar multiplication)

Examples:

1. $\mathcal{P}(\mathbb{F})$ is infinite dimensional [see the proof in the book].

2. $\mathcal{P}_m(\mathbb{F}) = \{p \in \mathcal{P}(\mathbb{F}) : \deg(p) \leq m\}$ is finite dimensional:

$$\operatorname{span}(1, z, z^2, \ldots, z^m) = \mathcal{P}_m(\mathbb{F})$$

3. $U = \{f \in C(\mathbb{R};\mathbb{R}) : f \text{ is periodic of period } n \text{ for some } n \in \mathbb{N}\}$
   U is infinite dimensional

   Proof. Let $\mathcal{L} = v_1, \ldots, v_m$ be an arbitrary list from U, so that each v_k has period $n_k \in \mathbb{N}$. If $\ell = \operatorname{lcm}(n_1, \ldots, n_m)$, then any linear combination from $\mathcal{L}$ will have period which is at most ℓ. Therefore if p is a prime number such that $p > \ell$, $\sin(\frac{2\pi}{p}x) \notin \mathcal{L}$, but $\sin(\frac{2\pi}{p}x) \in U$, and thus $\operatorname{span}(\mathcal{L}) \neq U$. Since $\mathcal{L}$ was arbitrary we can conclude that no finite list will span U. □

It will be *very useful* to record if a list of vectors $v_1, \ldots, v_m$ has no redundancy in its span, just as we isolated sums of subspaces with no redundancy by defining the direct sum.

Definition 11 (Linear independence)**.** $v_1, \ldots, v_m \in V$ are linearly independent if whenever $0 = a_1 v_1 + \cdots + a_m v_m$, then necessarily $a_1 = \cdots = a_m = 0$.

Definition 12 (Linear dependence)**.** $v_1, \ldots, v_m \in V$ are linearly dependent if $\exists\, a_1, \ldots a_m$ with at least one $a_k \neq 0$ and $0 = a_1 v_1 + \cdots + a_m v_m$.

The notions of linear independence and linear dependence are extremely important!

Examples:

1. $(1,0,0),(0,1,0)$ are linearly independent in $\mathbb{F}^3$
2. $1, z, \ldots, z^m$ are linearly independent in $\mathcal{P}(\mathbb{F})$ [Why? Use the fact that a polynomial of degree m has at most m distinct zeros]
3. Recall example from sum of subspaces:
   - $(-1,0,1),(1,-1,0),(1,0,0),(0,1,0)$ are linearly dependent
   - $(-1,0,1),(1,-1,0),(1,1,1)$ are linearly independent

The following is a very useful lemma...

Lemma 1 (Linear Dependence Lemma, LDL)**.** *If $v_1, \ldots, v_m \in V$ are linearly dependent and $v_1 \neq 0$, then $\exists\, k \in \{2, \ldots, m\}$ such that*

1. $v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$

2. If the v_k is removed from $v_1, \ldots, v_m$ then the resulting span *is the same as the original.*

Proof. Let $\mathcal{L} = v_1, \ldots, v_m$. For #1, by definition of linear dependence $\exists\, a_1, \ldots, a_m$ not all zero such that $0 = a_1 v_1 + \cdots + a_m v_m$. Let $k \in \{2, \ldots, m\}$ be the largest index such that $a_k \neq 0$. Then:

$$v_k = -\frac{a_1}{a_k} v_1 - \cdots - \frac{a_{k-1}}{a_k} v_{k-1} \tag{2}$$

For #2, let $\mathcal{L}^* = \mathcal{L} \setminus \{v_k\}$. Since $\mathcal{L}^* \subset \mathcal{L}$, $\operatorname{span}(\mathcal{L}^*) \subset \operatorname{span}(\mathcal{L})$. Let $u \in \operatorname{span}(\mathcal{L})$. Then:

$$u = a_1 v_1 + a_{k-1} v_{k-1} + a_k v_k + a_{k+1} v_{k+1} + \cdots + a_m v_m$$

Substitute (2) in for v_k and the sum is now in terms of $\mathcal{L}^*$, i.e., $u \in \operatorname{span}(\mathcal{L}^*)$. Thus $\operatorname{span}(\mathcal{L}) \subset \operatorname{span}(\mathcal{L}^*)$. □

Now for our first theorem.

Theorem 1. *If $V = \operatorname{span}(v_1, \ldots, v_n)$ and $w_1, \ldots, w_m$ are linearly independent in V, then $m \leq n$.*

Proof. We will use the two lists and make successive reductions and additions using Lemma 1.
Note: $w_1, \ldots, w_m$ linearly indpendent $\Rightarrow w_k \neq 0 \ \forall\, k$ [why?]

Add & reduce: Since $V = \text{span}(v_1, \ldots, v_n)$ and $w_1 \in V$, then $w_1, v_1, \ldots, v_n$ are linearly dependent. So Lemma 1 says at least one of the v_k can be removed. Up to a relabeling, we may assume it is v_n. So $\text{span}(w_1, v_1, \ldots, v_{n-1})$ is the same as $\text{span}(v_1, \ldots, v_n)$.

Now we can repeat: $w_2 \in V = \text{span}(w_1, v_1, \ldots, v_{n-1})$ so $w_2, w_1, v_1, \ldots, v_{n-1}$ are linearly dependent. Use Lemma 1 again, which says that one of them can be removed. The question is which? If it is w_1, then $w_1 \in \text{span}(w_2)$, which is a contradiction; so it must be one of the $v_1, \ldots, v_{n-1}$. Without loss of generality (WLOG), we may assume it is v_{n-1} and so $\text{span}(w_2, w_1, v_1, \ldots, v_{n-2}) = \text{span}(w_2, v_1, \ldots, v_{n-1}) = V$.

Keep repeating. At each stage one of the v_k must be removed, else Lemma 1 implies that $w_j \in \text{span}(w_1, \ldots w_{j-1})$ which is a contradiction.

The process stops when either we run out of w's ($m \leq n$) or we run out of v's ($m > n$). If $m > n$, then $\text{span}(w_1, \ldots, w_n) = V$ and $m > n$. Thus $w_m \notin \text{span}(w_1, \ldots, w_n) = V$, but this is a contradiction since $w_k \in V \ \forall\, k$. □

Proposition 11. *If V is finite dimensional and U is a subspace of V, then U is finite dimensional.*

End of Lecture 3

Beginning of Lecture 4

2.B Bases

span+ linear independence = basis

Definition 13. $v_1, \dots, v_n \in V$ is a basis of V if $\text{span}(v_1, \dots, v_n) = V$ and $v_1, \dots, v_n$ are linearly independent.

Proposition 12. *$v_1, \dots, v_n \in V$ is a basis of V if and only if $\forall\ v \in V,\ \exists!\, a_1, \dots, a_n \in \mathbb{F}$ such that*

$$v = a_1 v_1 + \cdots a_n v_n$$

The notion of a basis is extremely important because it allows us to define a coordinate system for our vector spaces!

Examples:

1. $(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$ is the standard basis of $\mathbb{F}^n$.
2. $1, z, \dots, z^m$ is the standard basis for $\mathcal{P}_m(\mathbb{F})$
3. Let $\mathbb{Z}_N = \{0, 1, \dots, N-1\}$ (with addition mod N) and let $V = \{f : \mathbb{Z}_N \to \mathbb{C}\}$. The standard (time side) basis for V is $\delta_0, \dots, \delta_{N-1}$ where

   $$\delta_k(n) = \begin{cases} 1 & n = k \\ 0 & n \neq k \end{cases}$$

   Indeed,

   $$f(n) = \sum_{k=0}^{N-1} f(k)\delta_k(n)$$

   Fourier analysis tells us that another (frequency side) basis for V is $e_0, \dots, e_{N-1}$ where

   $$e_k(n) = \frac{1}{\sqrt{N}} e^{2\pi i k n / N}$$

   and

   $$f(n) = \sum_{k=0}^{N-1} a_k e_k(n)$$

with

$$a_k = \hat{f}(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f(n) e^{-2\pi i k n/N}$$

The coefficients a_k define the function $\hat{f}(k)$ which is the Fourier transform of f.

If $v_1, \ldots, v_n$ spans V, it should have enough vectors to make a basis. Indeed:

Proposition 13. *If $\mathcal{L} = v_1, \ldots, v_n$ spans V, then $\mathcal{L}$ can be reduced to a basis.*

Proof. If $\mathcal{L}$ is linearly independent, then we are done. So assume it is not. We will selectively throw away vectors using the LDL.

Step 1: If $v_1 = 0$ remove v_1
Step 2: If $v_2 \in \text{span}(v_1)$, remove v_2
Step k: If $v_k \in \text{span}(v_1, \ldots, v_{k-1})$, remove v_k

Stop at Step n, getting a new list $\mathcal{L}^* = w_1, \ldots, w_m$. We still have $\text{span}(\mathcal{L}^*) = V$ since we only discarded vectors that were in the span of other vectors. We also have the property:

$$w_k \notin \text{span}(w_1, \ldots, w_{k-1}), \quad \forall\, k > 1$$

Thus by the contrapositive of LDL, $\mathcal{L}^*$ is linearly independent, and hence a basis. □

Corollary 1. *If V is finite dimensional, it has a basis.*

We just removed stuff from a spanning set to get a basis. We can also add stuff to a linearly independent set to get a basis.

Proposition 14. *If $\mathcal{L} = u_1, \ldots, u_m \in V$ is linearly independent, then $\mathcal{L}$ can be extended to a basis.*

Proof. Let $w_1, \ldots, w_n$ be a basis of V. Thus

$$\mathcal{L}^* = u_1, \ldots, u_m, w_1, \ldots, w_n$$

spans V. Apply the procedure in the proof of Proposition 13, and note that none of the u's get deleted [why?]. □

Now we show that every subspace U has a complementary subspace W that together direct sum to V.

Proposition 15. *Suppose V is finite dimensional and that U is a subspace of V. Then there exists another subspace W such that*

$$V = U \oplus W$$

Proof. V finite dimensional $\Rightarrow$ U finite dimensional $\Rightarrow$ U has a basis $u_1, \dots, u_m$. By the previous proposition we can extend $u_1, \dots, u_m$ to a basis of V, say $\mathcal{L} = u_1, \dots, u_m, w_1, \dots, w_n$. We show that $W = \operatorname{span}(w_1, \dots, w_n)$ is the answer.

We need to show: (1) $V = U + W$, and (2) $U \cap W = \{0\}$. Since $\mathcal{L}$ is a basis, for any $v \in V$ we have:

$$v = \underbrace{a_1u_1 + \cdots + a_mu_m}_{u \in U} + \underbrace{b_1w_1 + \cdots + b_nw_n}_{w \in W} = u + w \in U + W$$

Now suppose that $v \in U \cap W$. Then

$$v = a_1u_1 + \cdots + a_mu_m = b_1w_1 + \cdots + b_nw_n$$

which implies

$$a_1u_1 + \cdots + a_mu_m - b_1w_1 - \cdots - b_nw_n = 0$$

But $\mathcal{L}$ is linearly independent so $a_1 = \cdots = a_m = b_1 = \cdots = b_n = 0$. □

2.C Dimension

Since a basis gives a unique representation of each $v \in V$, we should be able to say that the number of vectors in basis is the dimension of V. But to do so, we need to make sure every basis of V has the same number of vectors. Indeed:

Theorem 2. *Any two bases of a finite dimensional vector space have the same length.*

Proof. Let $\mathcal{B}_1 = v_1, \dots, v_m$ and $\mathcal{B}_2 = w_1, \dots, w_n$ be two bases of V. Since $\mathcal{B}_1$ is linearly independent and $\mathcal{B}_2$ spans V, $m \leq n$. Flipping the roles of $\mathcal{B}_1$ and $\mathcal{B}_2$, we get $n \leq m$. □

Definition 14. The dimension of V is the length of $\mathcal{B}$ for any basis $\mathcal{B}$.

Proposition 16. *If U is a subspace of V, then* $\dim U \leq \dim V$

Examples:

1. $\dim \mathbb{F}^n = n$
   Remark: $\dim \mathbb{R}^2 = 2$ and $\dim \mathbb{C} = 1$, even though $\mathbb{R}^2$ can be identified with $\mathbb{C}$. The scalar field $\mathbb{F}$ cannot be ignored when computing the dimension of V!

2. $\dim \mathcal{P}_m(\mathbb{F}) = m + 1$

Let $\mathcal{L} = v_1, \ldots, v_n$. If $\dim V = n$, then we need only check if $\mathcal{L}$ is linearly independent OR if $\operatorname{span}(\mathcal{L}) = V$ to conclude that $\mathcal{L}$ is a basis for V.

Proposition 17. *Suppose* $\dim V = n$ *and let* $\mathcal{L} = v_1, \ldots, v_n$.

1. *If $\mathcal{L}$ is linearly independent, then $\mathcal{L}$ is a basis*

2. *If* $\operatorname{span}(\mathcal{L}) = V$*, then $\mathcal{L}$ is a basis.*

Proof. Use Proposition 14 for (1) and Proposition 13 for (2). $\square$

End of Lecture 4

Beginning of Lecture 5

Theorem 3. *$\dim V < \infty$, U_1 and U_2 subspaces of V. Then*

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$$

Proof. Proof will use 3 objects:

1. $\mathcal{B} = u_1, \ldots, u_m =$ basis of $U_1 \cap U_2$
2. $\mathcal{L}_1 = v_1, \ldots, v_j =$ extension of $\mathcal{B}$ so that $\mathcal{B} \cup \mathcal{L}_1 =$ basis for U_1
3. $\mathcal{L}_2 = w_1, \ldots, w_k =$ extension of $\mathcal{B}$ so that $\mathcal{B} \cap \mathcal{L}_2 =$ basis for U_2.

We will show that $\mathcal{L} = \mathcal{B} \cup \mathcal{L}_1 \cup \mathcal{L}_2$ is a basis for $U_1 + U_2$. This will complete the proof since if it is true, then

$$\dim(U_1+U_2) = m+j+k = (m+j)+(m+k)-m = \dim U_1+\dim U_2-\dim(U_1 \cap U_2)$$

Clearly $\mathcal{L}$ spans $U_1 + U_2$ since $\operatorname{span}(\mathcal{L})$ contains both U_1 and U_2.

Now we show linear independence. Suppose:

$$\sum_i a_i u_i + \sum_l b_l v_l + \sum_p c_p w_p = 0 \tag{3}$$

Then:

$$\sum_p c_p w_p = -\sum_i a_i u_i - \sum_l b_l v_l \in U_1$$

But $w_p \in U_2$ by assumption, so

$$\sum_p c_p w_p \in U_1 \cap U_2 \Rightarrow \sum_p c_p w_p = \sum_q d_q u_q \text{ for some } d_q$$

Now, $(u_1, \ldots, u_m, w_1, \ldots, w_k)$ is a basis for U_2. Thus:

$$\sum_p c_p w_p - \sum_q d_q u_q = 0 \Rightarrow c_p = 0, d_q = 0, \;\; \forall\, p, q$$

Therefore (3) reduces to

$$\sum_i a_i u_i + \sum_l b_l v_l = 0$$

Repeat the previous argument. □

3 Linear Maps

V, W always vector spaces.

3.A The Vector Space of Linear Maps

Definition 15. Let V, W be vector spaces over the same field $\mathbb{F}$. A function $T : V \to W$ is a linear map if it has the following two properties:

1. additivity: $T(u + v) = Tu + Tv, \ \forall\, u, v \in V$
2. homogeneity: $T(\lambda v) = \lambda(Tv) \ \forall\, \lambda \in \mathbb{F}, v \in V$

The set of all linear maps from V to W is denoted $\mathcal{L}(V, W)$.

Note: You could say T is linear if it "preserves the vector space structures of V and W."

Examples (read the ones in the book too!):

- Fix a point $x_0 \in \mathbb{R}$. Evaluation at x_0 is a linear map:

$$\begin{aligned} &T : C(\mathbb{R}; \mathbb{R}) \to \mathbb{R} \\ &Tv = v(x_0) \end{aligned}$$

- The anti-derivative is a linear map:

$$\begin{aligned} &T : C(\mathbb{R}; \mathbb{R}) \to C^1(\mathbb{R}; \mathbb{R}) \\ &(Tv)(x) = \int_0^x v(y)\, dy \end{aligned}$$

- Fix $b \in \mathbb{F}$. Define the forward shift operator as:

$$\begin{aligned} &T : \mathbb{F}^\infty \to \mathbb{F}^\infty \\ &T(v_1, v_2, v_3, \ldots) = (b, v_1, v_2, v_3, \ldots) \end{aligned}$$

  T is a linear map if and only if $b = 0$ [why?].

Next we show that we can always find a linear map that takes whatever values we want on a basis, and furthermore, that it is completely determined by these values.

Theorem 4. *Let $v_1, \dots, v_n$ be a basis for V and let $w_1, \dots, w_n \in W$. Then there exists a unique linear map $T : V \to W$ such that*

$$Tv_k = w_k, \quad \forall\, k$$

Proof. Define $T : V \to W$ as

$$T(a_1 v_1 + \cdots a_n v_n) = a_1 w_1 + \cdots + a_n w_n$$

Clearly $Tv_k = w_k$ for all k. It is easy to see that T is linear as well [see the book].

For uniqueness, let $S : V \to W$ be another linear map such that $Sv_k = w_k$ for all k. Then:

$$S(a_1 v_1 + \cdots a_n v_n) = \sum_{k=1}^{n} S(a_k v_k) = \sum_{k=1}^{n} a_k S v_k = \sum_{k=1}^{n} a_k w_k = T(a_1 v_1 + \cdots + a_n v_n)$$

□

The previous theorem is elementary, but highlights the fact that amongst all the maps from V to W, linear maps are very special.

Theorem 5. *$\mathcal{L}(V, W)$ is a vector space with the following vector addition and scalar multiplication operations:*

- *vector addition:* $S, T \in \mathcal{L}(V, W)$, $(S + T)(v) = Sv + Tv \;\; \forall\, v \in V$
- *scalar mult.:* $T \in \mathcal{L}(V, W)$, $\lambda \in \mathbb{F}$, $(\lambda T)(v) = \lambda(Tv) \;\; \forall\, v \in V$

Theorem 6. *$\mathcal{L}(V, W)$ is finite dimensional and*

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$$

Proof. Suppose $\dim V = n$ and $\dim W = m$ and let

$$\mathcal{B}_V = v_1, \dots, v_n$$
$$\mathcal{B}_W = w_1, \dots, w_m$$

be bases for V and W respectively. Define the linear transform $E_{p,q} : V \to W$ as

$$E_{p,q}(v_k) = \begin{cases} 0 & k \neq q \\ w_p & k = q \end{cases}, \quad p = 1, \dots, m, \;\; q = 1, \dots, n$$

By Theorem 4, this uniquely defines each $E_{p,q}$. We are going to show that these mn transformations $\{E_{p,q}\}_{p,q}$ form a basis for $\mathcal{L}(V,W)$.

Let $T : V \to W$ be a linear map. For each $1 \leq k \leq n$, let $a_{1,k}, \ldots, a_{m,k}$ be the coordinates of Tv_k in the basis $\mathcal{B}_W$:

$$Tv_k = \sum_{p=1}^{m} a_{p,k} w_p$$

To prove spanning, we wish to show that:

$$T = \sum_{p=1}^{m} \sum_{q=1}^{n} a_{p,q} E_{p,q} \tag{4}$$

Let S be the linear map on the right hand side of (4). Then for each k,

$$\begin{aligned} Sv_k &= \sum_p \sum_q a_{p,q} E_{p,q} v_k \\ &= \sum_p a_{p,k} w_p \\ &= Tv_k \end{aligned}$$

So $S = T$, and since T was arbitrary, $\{E_{p,q}\}_{p,q}$ spans $\mathcal{L}(V,W)$.

To prove linear independence, suppose that

$$S = \sum_p \sum_q a_{p,q} E_{p,q} = 0$$

Then $Sv_k = 0$ for each k, so

$$\sum_p a_{p,k} w_p = 0, \quad \forall\, k$$

But $w_1, \ldots, w_m$ are linearly independent, so $a_{p,k} = 0$ for all p and k. $\square$

End of Lecture 5

Beginning of Lecture 6

Warmup: Let U, W be 5-dimensional subspaces of $\mathbb{R}^9$. Can $U \cap W = \{0\}$?
Answer: No. First note that $\dim\{0\} = 0$. Then, using Theorem 3 we have:

$$\begin{aligned} \dim \mathbb{R}^9 = 9 \geq \dim(U_1 + U_2) &= \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2) \\ &= 10 - \dim(U_1 \cap U_2) \\ &\Rightarrow \dim(U_1 \cap U_2) \geq 1 \end{aligned}$$

Proposition 18. *If $T : V \to W$ is a linear map, then $T(0) = 0$.*

Proof.

$$T(0) = T(0 + 0) = T(0) + T(0) \Rightarrow T(0) = 0$$

□

Usually the product of a vector from one vector space with a vector from another vector space is not well defined. However, for some pairs of linear maps, it is useful to define their product.

Definition 16. If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then the product $ST \in \mathcal{L}(U, W)$ is

$$(ST)(u) = S(Tu), \quad \forall\, u \in U$$

Note: You must make sure the range of T is in the domain of S!
Another note: Multiplication of linear maps is not commutative! In other words, in general $ST \neq TS$.

3.B Null Spaces and Ranges

For a linear map T, the collection of vectors that get mapped to zero and the collection of those that do not are very important.

Definition 17. For $T \in \mathcal{L}(V, W)$, the null space of T, $\operatorname{null} T$, is:

$$\operatorname{null} T = \{v \in V : Tv = 0\}$$

See examples in the book.

Proposition 19. *For $T \in \mathcal{L}(V, W)$, $\operatorname{null} T$ is a subspace of V.*

Proof. Check if it contains zero, closed under addition, closed under scalar multiplication:

- $T(0) = 0$ so $0 \in \operatorname{null} T$
- $u, v \in \operatorname{null} T$, then $T(u+v) = Tu + Tv = 0 + 0 = 0$
- $u \in \operatorname{null} T$, $\lambda \in \mathbb{F}$, then $T(\lambda u) = \lambda T u = \lambda 0 = 0$

□

Definition 18. A function $T : V \to W$ is injective if $Tu = Tv$ implies $u = v$.

Proposition 20. *Let $T \in \mathcal{L}(V, W)$. Then*

$$T \text{ is injective} \iff \operatorname{null} T = \{0\}$$

Proof. For the $\Rightarrow$ direction, we already know that $0 \in \operatorname{null} T$. Thus $T(v) = 0 = T(0)$, but since T is injective $v = 0$.

For the $\Leftarrow$ direction, we have:

$$Tu = Tv \Rightarrow T(u - v) = 0 \Rightarrow u - v = 0 \Rightarrow u = v$$

□

Definition 19. For $T : V \to W$, the range of T is:

$$\operatorname{range} T = \{Tv : v \in V\}$$

Proposition 21. *If $T \in \mathcal{L}(V, W)$, then $\operatorname{range} T$ is a subspace of W.*

Definition 20. A function $T : V \to W$ is surjective if $\operatorname{range} T = W$.

Theorem 7 (Rank-Nullity Theorem)**.** *Suppose V is finite dimensional and $T \in \mathcal{L}(V, W)$. Then $\operatorname{range} T$ is finite dimensional and*

$$\dim V = \dim(\operatorname{null} T) + \dim(\operatorname{range} T)$$

Proof. Let $u_1, \ldots, u_m$ be a basis for $\operatorname{null} T$, and extend it to a basis $u_1, \ldots, u_m, v_1, \ldots, v_n$ of V. So we need to show that $\dim \operatorname{range} T = n$. To do so we prove that $Tv_1, \ldots, Tv_n$ is a basis for $\operatorname{range} T$.

Let $v \in V$ and write:

$$\begin{aligned} v &= a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n \\ \Rightarrow Tv &= b_1Tv_1 + \cdots + b_nTv_n \end{aligned}$$

Thus $\operatorname{span}(Tv_1, \ldots, Tv_n) = \operatorname{range} T$

Now we show that $Tv_1, \ldots, Tv_n$ are linearly independent. Suppose

$$\begin{aligned} c_1Tv_1 + \cdots + c_nTv_n &= 0 \\ \Rightarrow T(c_1v_1 + \cdots + c_nv_n) &= 0 \\ \Rightarrow c_1v_1 + \cdots + c_nv_n &\in \operatorname{null} T \\ \Rightarrow c_1v_1 + \cdots + c_nv_n &= d_1u_1 + \cdots + d_mu_m \end{aligned}$$

But $v_1, \ldots, v_n, u_1, \ldots, u_m$ are linearly independent, so $c_j = d_k = 0$ for all j, k. Thus $Tv_1, \ldots, Tv_n$ are linearly independent. □

Corollary 2. *Suppose V, W are finite dimensional and let $T \in \mathcal{L}(V, W)$. Then:*

1. *If $\dim V > \dim W$ then T is not injective.*
2. *If $\dim V < \dim W$ then T is not surjective.*

Proof. Use the Rank-Nullity Theorem:

1. $\dim \operatorname{null} T = \dim V - \dim \operatorname{range} T \geq \dim V - \dim W > 0$
2. $\dim \operatorname{range} T = \dim V - \dim \operatorname{null} T \leq \dim V < \dim W$

□

End of Lecture 6

Beginning of Lecture 7

Very important applications:

- Homogeneous systems of equations
  m equations and n unknowns:

$$\begin{aligned} \sum_{k=1}^{n} a_{1,k} x_k &= 0 \\ &\vdots \\ \sum_{k=1}^{n} a_{m,k} x_k &= 0 \end{aligned} \tag{5}$$

  where $a_{j,k} \in \mathbb{F}$ and $x = (x_1, \ldots, x_n) \in \mathbb{F}^n$.
  Can you solve all m equations simultaneously? Clearly $x = 0$ is a solution. Are there any others? Define $T : \mathbb{F}^n \to \mathbb{F}^m$:

$$T(x_1, \ldots, x_n) = \left(\sum_{k=1}^{n} a_{1,k} x_k, \ldots, \sum_{k=1}^{n} a_{m,k} x_k \right) \tag{6}$$

  Note: $T(0) = 0$ is equivalent to saying 0 is a solution of (5). Furthermore,

$$\text{Nontrivial solutions exist for (5)} \iff \dim \operatorname{null} T > 0$$

  But by the Rank-Nullity Theorem:

$$\dim \operatorname{null} T > 0 \iff \dim \mathbb{F}^n - \dim \operatorname{range} T > 0$$

  Since $\dim \operatorname{range} T \leq m$,

$$\text{if } n > m \implies \text{Nontrivial solutions exist for (5)}$$

- Inhomogeneous systems of equations: Let $c_k \in \mathbb{F}$ and consider:

$$\begin{aligned} \sum_{k=1}^{n} a_{1,k} x_k &= c_1 \\ &\vdots \\ \sum_{k=1}^{n} a_{m,k} x_k &= c_m \end{aligned} \tag{7}$$

New question, can you say for all $c = (c_1, \ldots, c_m) \in \mathbb{F}^m$ there exists at least one solution to (7)? Using the same T as defined in (6), we have:

$$\begin{aligned} \text{A solution exists for (6)} &\iff \forall c \in \mathbb{F}^m, \exists\, x \in \mathbb{F}^n \text{ s.t. } T(x) = c \\ &\iff \operatorname{range} T = \mathbb{F}^m \\ &\iff \dim \operatorname{range} T = m \\ &\iff \dim \mathbb{F}^n - \dim \operatorname{null} T = m \\ &\iff \dim \operatorname{null} T = n - m \end{aligned}$$

Since $\dim \operatorname{null} T \geq 0$, if $n < m$ then certainly there exists $c \in \mathbb{F}^m$ such that no solution exists for (7).

3.C Matrices

Definition 21. Let $T \in \mathcal{L}(V, W)$ and let $\mathcal{B}_V = v_1, \ldots, v_n$ and $\mathcal{B}_W = w_1, \ldots, w_m$ be bases of V and W respectively. The matrix of T with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$ is the $m \times n$ matrix $\mathcal{M}(T; \mathcal{B}_V, \mathcal{B}_W)$ (or just $\mathcal{M}(T)$ when $\mathcal{B}_V$ and $\mathcal{B}_W$ are clear) with entries $A_{j,k}$ defined by:

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j, \quad \forall\, k = 1, \ldots, n$$

Note: Recall the proof of the fact that $\dim \mathcal{L}(V, W) = mn$. In that proof we were implicitly using the matrix representation of T.

Another note: Recall the idea that a basis $\mathcal{B}_V = v_1, \ldots, v_n$ for a vector space V gives *coordinates* for V. That is, for all $v \in V$, there exists $a_1, \ldots, a_n \in \mathbb{F}$ such that

$$v = a_1 v_1 + \cdots + a_n v_n$$

So the n-tuple $(a_1, \ldots, a_n) \in \mathbb{F}^n$ is a *coordinate representation* of the vector v in the basis $\mathcal{B}_V$. If we change the basis, say to $\mathcal{B}'_V$, we change the coordinate representation of v say to $(a'_1, \ldots, a'_n)$, but we *do not* change v.

Similarly, the matrix $\mathcal{M}(T; \mathcal{B}_V, \mathcal{B}_W)$ can be thought of as a coordinate representation of the linear map $T \in \mathcal{L}(V, W)$ with respect to the bases $\mathcal{B}_V$ and $\mathcal{B}_W$. If we change the bases, we get a *new* matrix representation of T, but we *do not* change T; it is still the same linear map. [we will come back to this with an example later]

Definition 22. $\mathbb{F}^{m,n}$ is the set of all $m \times n$ matrices with entries in $\mathbb{F}$.

Proposition 22. $\mathbb{F}^{m,m}$ *is a vector space with the standard matrix addition and scalar multiplication.*

Proposition 23. $\dim \mathbb{F}^{m,n} = mn$.

We will derive matrix multiplication from the desire that $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$ for all S, T for which ST makes sense. Suppose $T : U \to V$, $S : V \to W$, and that $\mathcal{B}_V = \{v_r\}_{r=1}^n$ is basis for V, $\mathcal{B}_W = \{w_j\}_{j=1}^m$ is a basis for W, and $\mathcal{B}_U = \{u_k\}_{k=1}^p$ is a basis for U. Let $\mathcal{M}(S) = A$ and $\mathcal{M}(T) = C$. Then for each $1 \leq k \leq p$:

$$\begin{aligned}
(ST)u_k &= S\left(\sum_{r=1} nC_{r,k}v_r\right) \\
&= \sum_{r=1}^{n} C_{r,k}Sv_r \\
&= \sum_{r=1}^{n} C_{r,k} \sum_{j=1}^{m} A_{j,r}w_j \\
&= \sum_{j=1}^{m} \left(\sum_{r=1}^{n} A_{j,r}C_{r,k}\right) w_j
\end{aligned}$$

Thus we define matrix multiplication as:

$$(AC)_{j,k} = \sum_{r=1}^{n} A_{j,r}C_{r,k}$$

[read the rest of 3.C on matrix multiplication on your own]

End of Lecture 7

Beginning of Lecture 8

3.D Invertibility and Isomorphic Vector Spaces

Definition 23. A linear map that is both injective and surjective is called bijective.

Definition 24. A linear map $T \in \mathcal{L}(V, W)$ is invertible if $\exists\, S \in \mathcal{L}(W, V)$ such that $ST = I_V$ and $TS = I_W$. Such a map S is an inverse of T.

Proposition 24. *An invertible linear map has a unique inverse.*

Proof. Let S_1 and S_2 be two inverses of $T \in \mathcal{L}(V, W)$. Then:

$$S_1 = S_1 I = S_1(TS_2) = (S_1 T)S_2 = IS_2 = S_2$$

□

Notation: Thus we can denote the inverse of T as $T^{-1} \in \mathcal{L}(W, V)$.

Theorem 8.

$$T \in \mathcal{L}(V, W) \text{ is invertible} \iff T \text{ is bijective}$$

Proof. For the $\Longrightarrow$ direction: Need to show T is injective and surjective.
Suppose:

$$Tv_1 = Tv_2 \Rightarrow T^{-1}Tv_1 = T^{-1}Tv_2 \Rightarrow v_1 = v_2$$

since $T^{-1}T = I$. Thus T is injective.
Now suppose $w \in W$. Then:

$$TT^{-1}w = w \Rightarrow T\underbrace{(T^{-1}w)}_{\in V} = w$$

and so T is surjective.

Now for the $\Longleftarrow$ direction: Need to show T is invertible. To do so we define a map $S \in \mathcal{L}(W, V)$ and show that $ST = I$ and $TS = I$.
Define $S : W \to V$ as:

$$Sw := \text{unique } v \in V \text{ s.t. } Tv = w \text{ (i.e., } Sw = v \Leftrightarrow Tv = w)$$

Note S is well defined only because T is bijective! By construction we have $TS = I$. To show that $ST = I$, let $v \in V$, then:

$$T(STv) = (TS)(Tv) = Tv \Rightarrow ST = I \text{ since } T \text{ is injective}$$

Now we need to show that $S \in \mathcal{L}(W, V)$. For additivity let $w_1, w_2 \in W$:

$$\begin{aligned} T(Sw_1 + Sw_2) &= TSw_1 + TSw_2 = w_1 + w_2 \\ &\Rightarrow S(w_1 + w_2) = Sw_1 + Sw_2 \text{ by definition of } S \end{aligned}$$

For homogeneity use a similar argument:

$$T(\lambda Sw) = \lambda T(Sw) = \lambda w \Rightarrow S(\lambda w) = \lambda Sw$$

□

We now want to formalize the notion of when two vector spaces are essentially the same.

Definition 25. Two parts:

- An isomorphism is an invertible linear map (i.e., a bijection)
- V, W are isomorphic if there exists $T \in \mathcal{L}(V, W)$ such that T is an isomorphism. We write $V \cong W$.

Theorem 9.

$$V \cong W \iff \dim V = \dim W$$

Proof. For the $\Longrightarrow$ direction, we know then there is a bijection $T \in \mathcal{L}(V, W)$. Thus $\operatorname{null} T = \{0\}$ and $\operatorname{range} T = W$, so by Rank-Nullity Theorem:

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T = 0 + \dim W = \dim W$$

For the $\Longleftarrow$ direction, let $v_1, \ldots, v_n$ be a basis for V and let $w_1, \ldots, w_n$ be a basis for W. Define $T : V \to W$ as:

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots c_n w_n$$

It is easy to see $T \in \mathcal{L}(V, W)$, T is injective, T is surjective. Thus T defines an isomorphism. □

Corollary 3. *If* $\dim V = n$, *then* $V \cong \mathbb{F}^n$.

Remark: This *proves* that we can think of the coordinates of any $v \in V$ in a basis $\mathcal{B}_V = v_1, \ldots, v_n$ as a unique representation in $\mathbb{F}^n$, with the vector space structure of V carried over to $\mathbb{F}^n$. Indeed, define the matrix of $v \in V$ with respect to the basis $\mathcal{B}_V$ as the $n \times 1$ matrix:

$$\mathcal{M}(v; \mathcal{B}_V) := \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

where

$$v = c_1 v_1 + \cdots + c_n v_n$$

The linear map $\mathcal{M}(\cdot, \mathcal{B}_V) : V \to \mathbb{F}^n$ (note $\mathbb{F}^{n,1} \cong \mathbb{F}^n$ trivially) is an isomorphism.

Corollary 4. *If* $\dim V = n$ *and* $\dim W = m$*, then* $\mathcal{L}(V, W) \cong \mathbb{F}^{m,n}$*.*

Proof. This follows easily since we already proved that $\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$. □

Proposition 25. *Let* $\mathcal{B}_V = v_1, \ldots, v_n$ *be a basis of* V *and let* $\mathcal{B}_W = w_1, \ldots, w_m$ *be a basis of* W*. Then* $\mathcal{M}(\cdot; \mathcal{B}_V, \mathcal{B}_W) : \mathcal{L}(V, W) \to \mathbb{F}^{m,n}$ *is an isomorphism.*

Proposition 26. *Let* $T \in \mathcal{L}(V, W)$*, let* $v \in V$*, and let* $\mathcal{B}_V$ *and* $\mathcal{B}_W$ *be bases of* V *and* W *respectively. Then:*

$$\mathcal{M}(Tv; \mathcal{B}_W) = \mathcal{M}(T; \mathcal{B}_V, \mathcal{B}_W)\mathcal{M}(v; \mathcal{B}_V)$$

[See the book for the proofs of the previous two propositions.]

Example: Let $D \in \mathcal{L}(\mathcal{P}_3(\mathbb{R}), \mathcal{P}_2(\mathbb{R}))$ be the differentiation operator, defined by $Dp = p'$. Let's compute the matrix $\mathcal{M}(D)$ of D with respect to the standard bases $\mathcal{B}_3 = 1, x, x^2, x^3$ of $\mathcal{P}_3(\mathbb{R})$ and $\mathcal{B}_2 = 1, x, x^2$ of $\mathcal{P}_2(\mathbb{R})$. Since $Dx^n = (x^n)' = nx^{n-1}$ we have:

$$\mathcal{M}(D; \mathcal{B}_3, \mathcal{B}_2) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

End of Lecture 8

Beginning of Lecture 9

Example: Let $D \in \mathcal{L}(\mathcal{P}_3(\mathbb{R}), \mathcal{P}_2(\mathbb{R}))$ be the differentiation operator, defined by $Dp = p'$. Let's compute the matrix $\mathcal{M}(D)$ of D with respect to the standard bases $\mathcal{B}_3 = 1, x, x^2, x^3$ of $\mathcal{P}_3(\mathbb{R})$ and $\mathcal{B}_2 = 1, x, x^2$ of $\mathcal{P}_2(\mathbb{R})$. Since $Dx^n = (x^n)' = nx^{n-1}$ we have:

$$\mathcal{M}(D; \mathcal{B}_3, \mathcal{B}_2) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Now lets consider a different basis for $\mathcal{P}_3(\mathbb{R})$, for example $\mathcal{B}_3' = 1 + x, x + x^2, x^2 + x^3, x^3$. Compute:

$$\begin{aligned} D(1+x) &= 1 \\ D(x+x^2) &= 1 + 2x \\ D(x^2+x^3) &= 2x + 3x^2 \\ D(x^3) &= 3x^2 \end{aligned}$$

Thus:

$$\mathcal{M}(D; \mathcal{B}_3', \mathcal{B}_2) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 3 & 3 \end{pmatrix}$$

Now consider the specific polynomial $p \in \mathcal{P}_3(\mathbb{R})$,

$$p(x) = 2 + x + 3x^2 + 5x^3 \Longrightarrow p'(x) = 1 + 6x + 15x^2$$

The coordinates of p in $\mathcal{B}_3$ and $\mathcal{B}_3'$, as well as p' in $\mathcal{B}_2$, are:

$$\mathcal{M}(p; \mathcal{B}_3) = \begin{pmatrix} 2 \\ 1 \\ 3 \\ 5 \end{pmatrix} \quad \mathcal{M}(p; \mathcal{B}_3') = \begin{pmatrix} 2 \\ -1 \\ 4 \\ 1 \end{pmatrix} \quad \mathcal{M}(p'; \mathcal{B}_2) = \begin{pmatrix} 1 \\ 6 \\ 15 \end{pmatrix}$$

Computing Dp in terms of matrix multiplication with respect to $\mathcal{B}_3$ and $\mathcal{B}_2$

we should get back $\mathcal{M}(p';\mathcal{B}_2)$; indeed:

$$\begin{aligned}\mathcal{M}(Dp;\mathcal{B}_2) &= \mathcal{M}(D;\mathcal{B}_3,\mathcal{B}_2)\mathcal{M}(p;\mathcal{B}_3)\\ &= \begin{pmatrix} 0 & 1 & 0 & 0\\ 0 & 0 & 2 & 0\\ 0 & 0 & 0 & 3\end{pmatrix}\begin{pmatrix}2\\1\\3\\5\end{pmatrix}\\ &= \begin{pmatrix}1\\6\\15\end{pmatrix}\\ &= \mathcal{M}(p';\mathcal{B}_2)\end{aligned}$$

We should also be able to compute Dp in terms of matrix multiplication but with respect to $\mathcal{B}_3'$ and $\mathcal{B}_2$ and still get back $\mathcal{M}(p';\mathcal{B}_2)$; indeed:

$$\begin{aligned}\mathcal{M}(Dp;\mathcal{B}_2) &= \mathcal{M}(D;\mathcal{B}_3',\mathcal{B}_2)\mathcal{M}(p;\mathcal{B}_3')\\ &= \begin{pmatrix} 1 & 1 & 0 & 0\\ 0 & 2 & 2 & 0\\ 0 & 0 & 3 & 3\end{pmatrix}\begin{pmatrix}2\\-1\\4\\1\end{pmatrix}\\ &= \begin{pmatrix}1\\6\\15\end{pmatrix}\\ &= \mathcal{M}(p';\mathcal{B}_2)\end{aligned}$$

Remark: As we said earlier, the choice of bases determines the matrix representation $\mathcal{M}(T;\mathcal{B}_V,\mathcal{B}_W)$ of the linear map $T\in\mathcal{L}(V,W)$. Later on we will prove important results about the choice of the bases the give the “nicest” possible matrix representation of T.

Definition 26. A linear map $T\in\mathcal{L}(V,V) =: \mathcal{L}(V)$ is an operator.

Remark: For the matrix of an operator $T\in\mathcal{L}(V)$, we assume that we take the same basis $\mathcal{B}_V$ for both the domain V and the range V, and thus write it as $\mathcal{M}(T;\mathcal{B}_V) := \mathcal{M}(T;\mathcal{B}_V,\mathcal{B}_V)$. Furthermore, $\mathcal{M}(T;\mathcal{B}_V)\in\mathbb{F}^{n,n}$, where $\dim V = n$, and so we see that $\mathcal{M}(T;\mathcal{B}_V)$ is a square matrix.

Theorem 10. *Suppose V is finite dimensional and $T\in\mathcal{L}(V)$. Then the following are equivalent:*

1. *T is bijective (i.e., invertible)*
2. *T is surjective*
3. *T is injective*

Remark: Not true if V is infinite dimensional!

Proof. We prove this by proving that $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$.
Clearly $1 \Rightarrow 2$ so that part is done.
Now suppose T is surjective, i.e., $\operatorname{range} T = V$. Then by the Rank-Nullity Theorem:

$$\begin{aligned} \dim V &= \dim \operatorname{null} T + \dim \operatorname{range} T \\ \Rightarrow \dim V &= \dim \operatorname{null} T + \dim V \\ \Rightarrow \dim \operatorname{null} T &= 0 \\ \Rightarrow \operatorname{null} T &= \{0\} \\ \Rightarrow T &\text{ is injective} \end{aligned}$$

So that takes care of $2 \Rightarrow 3$.
Now suppose T is injective. Then $\operatorname{null} T = \{0\}$ and $\dim \operatorname{null} T = 0$. Once again use the Rank-Nullity Theorem:

$$\begin{aligned} \dim V &= \dim \operatorname{null} T + \dim \operatorname{range} T \\ \Rightarrow \dim V &= 0 + \dim \operatorname{range} T \\ \Rightarrow \operatorname{range} T &= V \end{aligned}$$

Thus T is surjective. Since we assumed it was injective, this means T is bijective and so we have $3 \Rightarrow 1$ and we are done. □

4 Polynomials

Read on your own!

5 Eigenvalues, Eigenvectors, and Invariant Subspaces

Extremely important subject matter that is the heart of Linear Algebra and is used all over mathematics, applied mathematics, data science, and more.

For example, consider a graph $G = (\mathcal{V}, \mathcal{E})$ consisting of vertices $\mathcal{V}$ and edges $\mathcal{E}$; for example see Figure 1. You can encode this graph with a 6×6 matrix

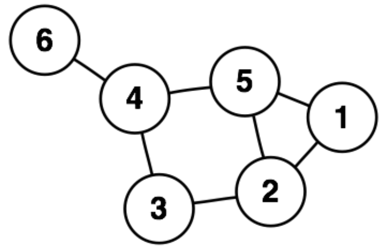

Figure 1: Graph with 6 vertices and 7 edges

L so that:

$$L_{j,k} = \begin{cases} \text{degree of vertex } k, & j = k \\ -1, & j \neq k \text{ and there is an edge between vertices } j \text{ and } k \\ 0, & \text{otherwise} \end{cases}$$

This matrix is called the graph Laplacian and it encodes connectivity properties of the graph through its eigenvalues and eigenvectors. If the nodes in the graph represent webpages, and the edges represent hyperlinks between the webpages, then a similar type of matrix represents the world wide web, and its eigenvectors and eigenvalues form the foundation of how Google computes search results!

5.A Invariant Subspaces

At the beginning of the course we defined a structure on sets V through the notion of a vector space. We then examined this structure further through subspaces, bases, and related notions. We then extended our study through linear maps between vector spaces, culminating in the Rank-Nullity Theorem and the notion of an isomorphism between two vector spaces with the same structure. Now we examine the structure of linear operators. The idea is that we will study the structure of $T \in \mathcal{L}(V)$ by finding nice structural decompositions of V relative to T.

Thought experiment: Let $T \in \mathcal{L}(V)$ and suppose

$$V = U_1 \oplus \cdots \oplus U_m$$

To understand T, we would need only understand $T_k = T|_{U_k}$ for each $k = 1, \ldots, m$. However, T_k may not be in $\mathcal{L}(U_k)$; indeed, T_k might map U_k to some other part of V. This is a problem, since we would like each restricted linear map T_k to be an operator itself on the subspace U_k. This leads us to the following definition.

Definition 27. Suppose $T \in \mathcal{L}(V)$. A subspace U of V is invariant under T if $Tu \in U$ for all $u \in U$, i.e., $T|_U \in \mathcal{L}(U)$.

Examples: $\{0\}$, V, $\operatorname{null} T$, $\operatorname{range} T$

Must an operator have any invariant subspaces other than $\{0\}$ and V? We will see... We begin with the study of one dimensional invariant subspaces.

End of Lecture 9

Beginning of Lecture 10

Definition 28. Suppose $T \in \mathcal{L}(V)$. A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of T if there exists $v \in V$, $v \neq 0$, such that

$$Tv = \lambda v$$

Such a v is called an eigenvector of T.

Proposition 27. *$T \in \mathcal{L}(V)$ has a one dimensional invariant subspace if and only if T has an eigenvalue.*

Proof. First suppose that T has a one dimensional invariant subspace, which we denote as U. Since $\dim U = 1$, U must be of the form:

$$U = \{\lambda v : \lambda \in \mathbb{F}\} = \text{span}(v)$$

for some $v \in V$, $v \neq 0$. Since T is invariant under U, $Tv \in U$. Thus there exists $\lambda \in \mathbb{F}$ such that $Tv = \lambda v$.

Now suppose that T has an eigenvalue $\lambda \in \mathbb{F}$. Then there exists $v \in V$, $v \neq 0$, such that $Tv = \lambda v$. Then $U = \text{span}(v)$ is an invariant subspace under T. □

Proposition 28. *Suppose V is finite dimensional, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{F}$. The following are equivalent:*

1. *λ is eigenvalue of T*
2. *$T - \lambda I$ is not injective*
3. *$T - \lambda I$ is not surjective*
4. *$T - \lambda I$ is not invertible*

Example: The Laplacian for $V = \{f \in C^\infty([-\pi, \pi]; \mathbb{C}) : f(-\pi) = f(\pi)\}$ is defined as:

$$\Delta f = \frac{d^2 f}{dx^2}$$

The eigenvalues and eigenvectors of Δ are:

$$\lambda = -k^2, k \in \mathbb{Z}, \quad v(x) = e^{ikx} = \cos kx + i \sin kx$$

Notice the similarity between the eigenvectors of Δ and the Fourier Transform defined earlier on $\mathbb{Z}_N$...

Theorem 11. *Let $T \in \mathcal{L}(V)$. If $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T and $v_1, \ldots, v_m$ are corresponding eigenvectors, then $v_1, \ldots, v_m$ are linearly independent.*

Proof. Proof by contradiction. Suppose $v_1, \ldots, v_m$ are linearly dependent. Using the LDL, let k be the smallest index such that

$$v_k \in \operatorname{span}(v_1, \ldots, v_{k-1}) \tag{8}$$

Thus

$$\begin{aligned} v_k &= a_1 v_1 + \cdots + a_{k-1} v_{k-1} \\ \Rightarrow T v_k &= a_1 T v_1 + \cdots + a_{k-1} T v_{k-1} \\ \Rightarrow \lambda_k v_k &= a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1} \end{aligned}$$

We also can conclude:

$$\begin{aligned} v_k &= a_1 v_1 + \cdots + a_{k-1} v_{k-1} \\ \Rightarrow \lambda_k v_k &= a_1 \lambda_k v_1 + \cdots + a_{k-1} \lambda_k v_{k-1} \end{aligned}$$

Combining the two expansions of $\lambda_k v_k$ yields:

$$0 = a_1(\lambda_k - \lambda_1) v_1 + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1}) v_{k-1}$$

Since k is the smallest index satisfying (8), $v_1, \ldots, v_{k-1}$ must be linearly independent. Thus $a_1 = \cdots = a_{k-1} = 0$ since $\lambda_k - \lambda_j \neq 0$ for all $k \neq j$. But then $v_k = 0$, which is a contradiction. □

Corollary 5. *Suppose V is finite dimensional. Then $T \in \mathcal{L}(V)$ has at most $\dim V$ distinct eigenvalues.*

End of Lecture 10

Beginning of Lecture 11

5.B Eigenvectors and Upper-Triangular Matrices

One of the main differences between operators and general linear maps is that we can take powers of operators! This will lead to many interesting results...

Definition 29. Let $T \in \mathcal{L}(V)$ and let $m \in \mathbb{Z}$, $m > 0$.

- $T^m = T \cdots T$ (composition m times)
- $T^0 = I$
- If T is invertible, then $T^{-m} = (T^{-1})^m$

Definition 30. Suppose $T \in \mathcal{L}(V)$ and let $p \in \mathcal{P}(\mathbb{F})$ be given by:

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

Then $p(T) \in \mathcal{L}(V)$ is defined as:

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m$$

Theorem 12. *Let $V \neq \{0\}$ be a finite dimensional vector space over $\mathbb{C}$. Then every $T \in \mathcal{L}(V)$ has an eigenvalue.*

Proof. Suppose $\dim V = n > 0$ and choose $v \in V$, $v \neq 0$. Then:

$$\mathcal{L} = v, Tv, T^2 v, \ldots, T^n v$$

is linearly dependent because the length of $\mathcal{L}$ is $n+1$. Thus there exists $a_0, \ldots, a_n \in \mathbb{C}$, not all zero, such that

$$0 = a_0 v + a_1 T v + a_2 T^2 v + \cdots + a_n T^n v$$

Consider the polynomial $p \in \mathcal{P}(\mathbb{C})$ with coefficients given by $a_0, \ldots, a_n$. By the Fundamental Theorem of Algebra,

$$p(z) = a_0 + a_1 z + \cdots + a_n z^n = c(z - \lambda_1) \cdots (z - \lambda_m), \quad \forall\, z \in \mathbb{C},$$

where $m \leq n$, $c \in \mathbb{C}$, $c \neq 0$, and $\lambda_k \in \mathbb{C}$. Thus:

$$\begin{aligned} 0 &= a_0 v + a_1 T v + \cdots + a_n T^n v \\ &= (a_0 I + a_1 T + \cdots + a_n T^n) v \\ &= c(T - \lambda_1 I) \cdots (T - \lambda_m I) v \end{aligned}$$

Thus $(T - \lambda_k I)v = 0$ for at least one k, which means $T - \lambda_k I$ is not injective, which implies that λ_k is eigenvalue of T. □

Example: Theorem 12 is not true for real vector spaces! Take for example the following operator $T \in \mathcal{L}(\mathbb{F}^2)$ defined as:

$$T(w, z) = (-z, w)$$

If $\mathbb{F} = \mathbb{R}$, then T is a counterclockwise rotation by 90 degrees. Since a 90 degree rotation of any nonzero $v \in \mathbb{R}^2$ will never equal a scalar multiple of itself, T has no eigenvalues!

On the other hand, if $\mathbb{F} = \mathbb{C}$, then by Theorem 12 T must have at least one eigenvalue. Indeed it has two, $\lambda = i$ and $\lambda = -i$ [see the book p. 135].

Recall we want a nice decomposition of V as $V = U_1 \oplus \cdots \oplus U_m$, where each U_k is an invariant subspace of T, so that to understand $T \in \mathcal{L}(V)$ we only need to understand $T|_{U_k}$. We will accomplish this by finding bases of V that yield matrices $\mathcal{M}(T)$ with lots of zeros.

As a first baby step, let V be a complex vector space. Then $T \in \mathcal{L}(V)$ must have at least one eigenvalue λ and a corresponding eigenvector v_*. Extend v_* to a basis of V:

$$\mathcal{B}_V = v_*, v_2, \ldots, v_n$$

Then:

$$\mathcal{M}(T; \mathcal{B}_V) = \begin{pmatrix} \lambda & & \\ 0 & & * \\ \vdots & & \\ 0 & & \end{pmatrix} \tag{9}$$

Furthermore, if we define $U_1 = \operatorname{span}(v_*)$ and $U_2 = \operatorname{span}(v_2, \ldots, v_n)$, then $V = U_1 \oplus U_2$. The subspace U_1 is a one dimensional invariant subspace of V under T, but U_2 is not necessarily. It is a start though! Now let's try to do better...

Definition 31. A matrix is upper triangular if all the entries below the diagonal equal 0:

$$\begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}$$

There is a useful connection between upper triangular matrices and invariant subspaces:

Proposition 29. *Suppose $T \in \mathcal{L}(V)$ and $\mathcal{B}_V = v_1, \ldots, v_n$ is a basis for V. Then the following are equivalent:*

1. $\mathcal{M}(T; \mathcal{B}_V)$ *is upper triangular*
2. $Tv_k \in \operatorname{span}(v_1, \ldots, v_k)$ *for each* $k = 1, \ldots, n$
3. $\operatorname{span}(v_1, \ldots, v_k)$ *is invariant under T for each* $k = 1, \ldots, n$

Proof. First we prove $1 \iff 2$. Let $A = \mathcal{M}(T; \mathcal{B}_V)$. Then by the definition of A we have:

$$Tv_k = \sum_{j=1}^{n} A_{j,k} v_j$$

But then

$$Tv_k \in \operatorname{span}(v_1, \ldots, v_k) \iff \underbrace{A_{j,k} = 0 \ \forall\, j > k}_{A \text{ is upper triangular}}$$

Clearly $3 \Longrightarrow 2$

We finish the proof by showing $2 \Longrightarrow 3$. Fix k. From 2 we have:

$$\begin{aligned}
&Tv_1 \in \operatorname{span}(v_1) \subset \operatorname{span}(v_1, \ldots, v_k) \\
&Tv_2 \in \operatorname{span}(v_1, v_2) \subset \operatorname{span}(v_1, \ldots, v_k) \\
&\quad \vdots \\
&Tv_k \in \operatorname{span}(v_1, \ldots, v_k)
\end{aligned}$$

Thus if $v \in \operatorname{span}(v_1, \ldots, v_k)$, then $Tv \in \operatorname{span}(v_1, \ldots, v_k)$ as well. □

Now can improve upon our "baby step" (9) above by showing that given an eigenvector v_* with eigenvalue λ, we can extend it to a basis $\mathcal{B}_V$ such that $\mathcal{M}(T; \mathcal{B}_V)$ is upper triangular.

Theorem 13. *Suppose V is a finite dimensional complex vector space and $T \in \mathcal{L}(V)$. Then there exists a basis $\mathcal{B}_V$ such that $\mathcal{M}(T; \mathcal{B}_V)$ is upper triangular.*

END OF LECTURE 11

Beginning of Lecture 12

Warmup: Suppose $T \in \mathcal{L}(V)$ and $6I - 5T + T^2 = 0$. What are the possible eigenvalues of T?

Answer: $6I - 5T + T^2 = 0$ implies that $(T - 2I)(T - 3I) = 0$. Now let $v \neq 0$ be an eigenvector of T with eigenvalue λ. Then $0 = (T - 2I)(T - 3I)v = (\lambda - 2)(\lambda - 3)v$, which implies that $\lambda = 2$ or $\lambda = 3$.

Theorem 14. *Suppose V is a finite dimensional complex vector space and $T \in \mathcal{L}(V)$. Then there exists a basis $\mathcal{B}_V$ such that $\mathcal{M}(T; \mathcal{B}_V)$ is upper triangular.*

Proof. Induction on $\dim V$. Clearly the result is true when $\dim V = 1$.

Now suppose the result is true for all complex vector spaces with dimension $n - 1$ or less, and let V be a complex vector space with $\dim V = n$. We know that V has one eigenvalue λ. Define:

$$U = \text{range}\,(T - \lambda I)$$

Since $T - \lambda I$ is not surjective, $\dim U < \dim V$. Furthermore, U is invariant under T; indeed, let $u \in U$:

$$Tu = \underbrace{(T - \lambda I)u}_{\in U} + \underbrace{\lambda u}_{\in U}$$

Thus $\widetilde{T} = T|_U \in \mathcal{L}(U)$, and we can apply the induction hypothesis to $\widetilde{T}$ and U. In particular, there exists a basis $\mathcal{B}_U = u_1, \ldots, u_m$ of U such that $\mathcal{M}(\widetilde{T}; \mathcal{B}_U)$ is upper triangular.

Extend $\mathcal{B}_U$ to a basis for V:

$$\mathcal{B}_V = u_1, \ldots, u_m, v_1, \ldots, v_\ell, \quad \ell + m = n$$

Since $\mathcal{M}(\widetilde{T}; \mathcal{B}_U)$ is upper triangular, by Proposition 29 we have:

$$Tu_k = \widetilde{T}u_k \in \text{span}(u_1, \ldots, u_k) \text{ for all } k = 1, \ldots, m.$$

Furthermore,

$$Tv_j = \underbrace{(T - \lambda I)v_j}_{\in U} + \underbrace{\lambda v_j}_{\in \text{span}(v_j)} \in \text{span}(u_1, \ldots, u_m, v_j) \subset \text{span}(u_1, \ldots, u_m, v_1, \ldots, v_j)$$

Thus T and $\mathcal{B}_V$ satisfy condition 2 of Proposition 29, and so $\mathcal{M}(T;\mathcal{B}_V)$ is upper triangular. □

Upper triangular matrices a very useful for determining if $T \in \mathcal{L}(V)$ is invertible...

Proposition 30. *Let $T \in \mathcal{L}(V)$ and let $\mathcal{B}$ be a basis for which $\mathcal{M}(T;\mathcal{B})$ is upper triangular. Then*

$$T \text{ is invertible} \iff \text{all diagonal entries of } \mathcal{M}(T;\mathcal{B}) \text{ are nonzero}$$

Proof. Let $\mathcal{B} = v_1, \ldots, v_n$ and let $A = \mathcal{M}(T;\mathcal{B})$. Easier to prove "not (a) $\iff$ not (b)".

First suppose T is not invertible; we want to show that some entry of $\mathcal{M}(T;\mathcal{B})$ is zero. T not invertible $\Rightarrow$ T not injective $\Rightarrow$ there exists $v \neq 0$ such that $Tv = 0$. Expand v in $\mathcal{B}$:

$$v = \sum_{j=1}^{n} c_j v_j$$

Let k be the index satisfying the following: $c_k \neq 0$ and $c_j = 0$ for all $j > k$ (note that possibly $k = n$). If $k = 1$, then $v = c_1 v_1 \Rightarrow Tv_1 = 0 \Rightarrow A_{1,1} = 0$. If $k > 1$ then:

$$\begin{aligned}
v &= \sum_{j=1}^{k} c_j v_j \\
Tv &= \sum_{j=1}^{k} c_j T v_j \\
0 &= \sum_{j=1}^{k-1} c_j T v_j + c_k T v_k \\
\Rightarrow T v_k &= -\sum_{j=1}^{k-1} \left(\frac{c_j}{c_k} \right) T v_j \in \operatorname{span}(v_1, \ldots, v_{k-1}),
\end{aligned}$$

where in the last line we used Proposition 29. But also by Proposition 29,

$$\sum_{j=1}^{k-1} b_j v_j = T v_k = \sum_{j=1}^{k} A_{j,k} v_j$$

and since $\mathcal{B}$ is a basis we must have $A_{k,k} = 0$.

Now suppose some entry on the diagonal of $\mathcal{M}(T; \mathcal{B})$ is zero. If $A_{1,1} = 0$ then $Tv_1 = 0$ and so T is not injective, and hence not invertible. If $A_{k,k} = 0$ for $k > 1$, then by Proposition 29 we have:

$$Tv_k = \sum_{j=1}^{k} A_{j,k} v_j = \sum_{j=1}^{k-1} A_{j,k} v_j \in \operatorname{span}(v_1, \ldots, v_{k-1}) \tag{10}$$

Consider now the linear map $\widetilde{T} = T|_{\operatorname{span}(v_1, \ldots v_k)}$. By (10),

$$\widetilde{T} \in \mathcal{L}(\operatorname{span}(v_1, \ldots, v_k), \operatorname{span}(v_1, \ldots, v_{k-1}))$$

Thus $\widetilde{T}$ cannot be injective since it maps a k-dimensional vector space to a $(k-1)$-dimensional vector space. In particular, there exists $v_* \in \operatorname{span}(v_1, \ldots, v_k)$ such that $\widetilde{T}v_* = 0$. But then $Tv_* = 0$, and so T is not injective, and hence not invertible. □

End of Lecture 12

Beginning of Lecture 13

Not only can upper triangular matrices tell us when $T \in \mathcal{L}(V)$ is invertible, they also tell us precisely what the eigenvalues of T are!

Proposition 31. *Let $T \in \mathcal{L}(V)$ and suppose $A = \mathcal{M}(T)$ is upper triangular. Then:*

$$\lambda \text{ is an eigenvalue of } T \iff \lambda = A_{k,k} \text{ for some } k$$

Proof. Let $A = \mathcal{M}(T)$ have diagonal entries given by $A_{k,k} = \lambda_k$:

$$A = \mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}$$

Let $\lambda \in \mathbb{F}$. Then

$$\mathcal{M}(T - \lambda I) = \begin{pmatrix} \lambda_1 - \lambda & & * \\ & \ddots & \\ 0 & & \lambda_m - \lambda \end{pmatrix}$$

Thus by Proposition 30 $T - \lambda I$ is not invertible (and hence λ is an eigenvalue) if and only if $\lambda = \lambda_k$ for some k. □

5.C Eigenspaces and Diagonal Matrices

Definition 32. A diagonal matrix is a square matrix that is 0 everywhere except possibly the diagonal:

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}$$

Note: If $\mathcal{M}(T; \mathcal{B})$ is upper triangular, then the diagonal entries are precisely the eigenvalues of T (since diagonal matrices are upper triangular).

Definition 33. Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The eigenspace of T corresponding to λ is:

$$E(\lambda, T) = \operatorname{null}(T - \lambda I)$$

Note: $T|_{E(\lambda,T)} = \lambda I$ (so eigenspaces are invariant subspaces)

Proposition 32. *Suppose V is finite dimensional and $T \in \mathcal{L}(V)$. Suppose also that $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T. Then:*

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T) \tag{11}$$

is a direct sum and furthermore

$$\dim E(\lambda_1, T) + \cdots + \dim(\lambda_m, T) \leq \dim V$$

Proof. Let $u_k \in E(\lambda_k, T)$ and suppose that

$$u_1 + \cdots + u_m = 0$$

Since eigenvectors corresponding to distinct eigenvalues are linearly independent, each $u_k = 0$ and so (11) is a direct sum.

Furthermore, by #16 of 2.C (HW1),

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) = \dim(E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)) \leq \dim V$$

□

End of Lecture 13

Beginning of Lecture 14

Definition 34. An operator $T \in \mathcal{L}(V)$ is diagonalizable if there exists a basis $\mathcal{B}$ such that $\mathcal{M}(T;\mathcal{B})$ is diagonal.

Proposition 33. *Suppose V is finite dimensional and $T \in \mathcal{L}(V)$. Then: T is diagonalizable $\Leftrightarrow$ V has a basis of eigenvectors of T.*

Proof. An operator $T \in \mathcal{L}(V)$ has a diagonal matrix with respect to a basis $\mathcal{B} = v_1, \ldots, v_n$ if and only if $Tv_k = \lambda_k v_k$ for each k. □

Example: Not every operator is diagonalizable, even over complex vector spaces! Consider $T \in \mathcal{L}(\mathbb{C}^2)$ defined as:

$$T(w, z) = (z, 0)$$

Then $T^2 = 0$. Now let $v \neq 0$ be an eigenvector with eigenvalue λ. Then $0 = T^2 v = T(Tv) = \lambda T v = \lambda^2$. Thus $\lambda = 0$. Even though $\dim E(0, T^2) = 2$, we see that

$$E(0, T) = \{(w, 0) : w \in \mathbb{C}\}$$

and so $\dim(0, T) = 1$. Therefore V does not have a basis of eigenvectors of T, and so T is not diagonalizable. We will address examples like this much later with the notion of generalized eigenvectors...

On the other hand, if we have enough distinct eigenvalues, we know that T is diagonalizable:

Proposition 34. *If $T \in \mathcal{L}(V)$ has $\dim V < \infty$ distinct eigenvalues, then T is diagonalizable.*

Proof. Let $\dim V = n$ and suppose $T \in \mathcal{L}(V)$ has distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $v_1, \ldots, v_n$. The eigenvectors are linearly independent because they correspond to distinct eigenvalues, and thus they form a basis for V. Thus T is diagonalizable. □

Note: The converse is not true! Take any diagonal matrix with non-unique entries on the diagonal.

Finally, our main result for this chapter. Namely, if T is diagonalizable, then we can achieve our stated goal of decomposing V as $V = U_1 \oplus \cdots \oplus U_n$, where each U_k is an invariant subspace of V under T and $\dim U_k = 1$.

Theorem 15. *Suppose V is finite dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote distinct eigenvalues of T. Then the following are equivalent:*

1. *T is diagonalizable*
2. *V has a basis consisting of eigenvectors of T*
3. *There exist one dimensional invariant subspaces $U_1, \ldots, U_n$ of V such that $V = U_1 \oplus \cdots \oplus U_n$*
4. *$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$*
5. *$\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$*

Proof. Many parts. The plan is:

$$1 \iff 2 \iff 3, \quad 2 \implies 4 \implies 5 \implies 2$$

- $1 \iff 2$: Simply Proposition 33.
- $2 \implies 3$: Let $\mathcal{B} = v_1, \ldots, v_n$ be basis of eigenvectors of V. Define $U_k = \operatorname{span}(v_k)$. Then each U_k is a 1-dimensional invariant subspace of V under T, and since $\mathcal{B}$ is a basis it is clear $V = U_1 \oplus \cdots \oplus U_n$.
- $3 \implies 2$: For each k, let $v_k \in U_k$, $v_k \neq 0$. Since U_k is a 1-dimensional invariant subspace under T, each v_k is an eigenvector of T. Furthermore each $v \in V$ can be written uniquely as:

  $$v = u_1 + \cdots + u_n,$$

  where $u_k \in U_k$ and therefore $u_k = a_k v_k$ for some $a_k \in \mathbb{F}$. Thus $v_1, \ldots, v_n$ is a basis for V.
- $2 \implies 4$: Let $v_1, \ldots, v_n$ be a basis of eigenvectors for V, and subdivide the list according to the unique eigenvalues of T, so that:

  $$v_1^{(\ell)}, \ldots, v_{k_\ell}^{(\ell)} \text{ corresponds to } \lambda_\ell, \quad \text{for } \ell = 1, \ldots, m$$

  and $k_1 + k_2 + \cdots + k_m = n$. Then any $v \in V$ can be written as:

  $$v = \sum_{\ell=1}^{m} \underbrace{\sum_{j=1}^{k_\ell} a_{j,\ell} \, v_j^{(\ell)}}_{\in E(\lambda_\ell, T)} \in E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$$

- 4 $\Longrightarrow$ 5: This is simply 2.C #16, which you did for homework!

- 5 $\Longrightarrow$ 2: Choose a basis for each $E(\lambda_\ell, T)$, say $v_1^{(\ell)}, \ldots, v_{k_\ell}^{(\ell)}$, where $k_1 + \cdots + k_m = n$ by assumption. Let $\mathcal{L}$ be the list of all of these vectors concatenated together. To show $\mathcal{L}$ is linearly independent, suppose:

$$\sum_{\ell=1}^{m} \underbrace{\sum_{j=1}^{k_\ell} a_{j,\ell}\, v_j^{(\ell)}}_{u_\ell \in E(\lambda_\ell, T)} = 0$$

$$\sum_{\ell=1}^{m} u_\ell = 0$$

  Each u_ℓ is eigenvector of T corresponding to a distinct eigenvalue λ_ℓ; thus $u_1, \ldots, u_m$ must be linearly independent and so $u_\ell = 0$ for all ℓ. But then $a_{j,\ell} = 0$ for all $j = 1, \ldots, k_\ell$ and for each ℓ, since $v_1^{(\ell)}, \ldots, v_{k_\ell}^{(\ell)}$ are linearly independent.

$\square$

End of Lecture 14

Beginning of Lecture 15

6 Inner Product Spaces

We now introduce geometrical aspects such as length and angle into the setting of abstract vector spaces.

6.A Inner Products and Norms

We begin by looking at $\mathbb{R}^n$.

Definition 35. The norm of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is:

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

Definition 36. For $x, y \in \mathbb{R}^n$, the dot product of x and y is:

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n.$$

Notice that $\|x\|^2 = x \cdot x$.

Example: In $\mathbb{R}^2$, $\|x\| = \sqrt{x_1^2 + x_2^2}$ which is just the length of x, and

$$x \cdot y = \|x\| \|y\| \cos\theta,$$

where θ is the angle between x and y.

Properties of the dot product:

- $x \cdot x \geq 0 \;\; \forall\, x \in \mathbb{R}^n$
- $x \cdot x = \|x\|^2 = 0 \iff x = 0$
- $x \cdot y = y \cdot x$
- Fix $y \in \mathbb{R}^n$. Then $T_y(x) = x \cdot y$ is a linear map, i.e., $T_y \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$.

Now we want to generalize the dot product to abstract vector spaces. First lets consider $\mathbb{C}^n$. Let $\lambda = a + ib \in \mathbb{C}$ be a complex scalar. Recall that:

- $|\lambda| = \sqrt{a^2 + b^2}$

- $|\lambda|^2 = \lambda\bar{\lambda}$

For $z \in \mathbb{C}^n$, the norm is defined as:

$$\|z\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}$$

Note that:

$$\|z\|^2 = z_1\bar{z}_1 + \cdots + z_n\bar{z}_n$$

If we want $z \cdot z = \|z\|^2$, then the previous line implies that we should define the dot product on $\mathbb{C}^n$ as:

$$w \cdot z = w_1\bar{z}_1 + \cdots + w_n\bar{z}_n$$

This leads us to the generalization of the dot product to abstract vector spaces:

Definition 37. An inner product on V is a function $\langle \cdot, \cdot \rangle : \mathbb{F}^2 \to \mathbb{F}$ that has the following properties:

1. Positive Definitness:
   $\langle v, v \rangle \geq 0 \;\; \forall\, v \in V$
   $\langle v, v \rangle = 0 \iff v = 0$

2. Linearity in the first argument:
   $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle \;\; \forall\, u, v, w \in V$
   $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle \;\; \forall \lambda \in \mathbb{F}, \;\; \forall\, u, v \in V$

3. Conjugate Symmetry:
   $\langle u, v \rangle = \overline{\langle v, u \rangle} \;\; \forall\, u, v \in V$

Examples:

1. Euclidean inner product on $\mathbb{F}^n$. Let $w = (w_1, \ldots, w_n, z = (z_1, \ldots, z_n) \in \mathbb{F}^n$:

$$\langle w, z \rangle = w_1\bar{z}_1 + \cdots + w_n\bar{z}_n$$

2. Weighted Euclidean inner product on $\mathbb{F}^n$. Fix $c = (c_1, \ldots, c_n) \in \mathbb{R}^n$ with $c_k \geq 0$. Then for $w, z \in \mathbb{F}^n$,

$$\langle w, z \rangle_c = c_1 w_1\bar{z}_1 + \cdots + c_n w_n\bar{z}_n$$

3. Define $V = L^2(\mathbb{R})$ as:

$$L^2(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} : \int_{-\infty}^{\infty} |f(x)|^2 \, dx < \infty\}$$

One can verify this is a real vector space. Since it is a subset of the vector space of all functions mapping $\mathbb{R}$ to $\mathbb{R}$, we need to show (1) it contains an additive identity (zero), (2) it is closed under addition, and (3) it is closed under scalar multiplication. Indeed, $f \equiv 0 \in L^2(\mathbb{R})$, and furthermore if $f \in L^2(\mathbb{R})$ then $\lambda f \in L^2(\mathbb{R})$ for any $\lambda \in \mathbb{R}$ since

$$\int_{-\infty}^{\infty} |\lambda f(x)|^2 \, dx = \lambda^2 \int_{-\infty}^{\infty} |f(x)|^2 \, dx < \infty$$

The trickiest part is that it is closed under addition; i.e., if $f, g \in L^2(\mathbb{R})$, then $f + g \in L^2(\mathbb{R})$. First note:

$$\int_{-\infty}^{\infty} |f(x)+g(x)|^2 \, dx = \underbrace{\int_{-\infty}^{\infty} |f(x)|^2 \, dx}_{\text{I}} + \underbrace{\int_{-\infty}^{\infty} |g(x)|^2 \, dx}_{\text{II}} + 2 \underbrace{\int_{-\infty}^{\infty} f(x)g(x) \, dx}_{\text{III}}$$

Since $f, g \in L^2(\mathbb{R})$, we know that the first two terms are finite. That leaves the third term. That this is finite follows from what's known in Real Analysis as Hölder's Inequality. However, we can in fact prove it with more elementary tools. First let $a, b \in \mathbb{R}$ and note that:

$$(a - b)^2 \geq 0 \Rightarrow a^2 - 2ab + b^2 \geq 0 \Rightarrow ab \leq \frac{a^2}{2} + \frac{b^2}{2}$$

Now let $f(x) = a$ and $g(x) = b$. Then:

$$\int_{-\infty}^{\infty} f(x)g(x) \, dx \leq \int_{-\infty}^{\infty} \frac{|f(x)|^2}{2} + \frac{|g(x)|^2}{2} \, dx < \infty \tag{12}$$

Thus $L^2(\mathbb{R})$ is a vector space! We can add an inner product to it by defining the inner product as:

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x) \, dx$$

By what we just showed in (12), the inner product is well defined. Furthermore, it is easy to verify that all of the properties of an inner product hold, except for "definiteness" property: $\langle f, f\rangle = 0 \Rightarrow f = 0$. This is a bit technical but follows from Real Analysis. Now $L^2(\mathbb{R})$ is what we call an inner product space. Any inner product can always be used to define the norm of a vector. In this case, we get the L^2-norm:

$$\|f\|_2 = \sqrt{\langle f, f\rangle} = \left(\int_{-\infty}^{\infty} |f(x)|^2\, dx\right)^{1/2}$$

In fact $L^2(\mathbb{R})$ is a special inner product space called a Hilbert space, but we leave that for more advanced math classes...

Definition 38. An inner product space is a vector space V along with an inner product on V.

Important Note: For the rest of chapter 6, we assume V is an inner product space.

Definition 39. For $v \in V$ an inner product space, the norm of v is:

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

END OF LECTURE 15

Beginning of Lecture 16

Proposition 35. *The following basic properties hold:*

1. *For each fixed $u \in V$, the function $T_u(v) = \langle v, u \rangle$ is linear, i.e., $T_u \in \mathcal{L}(V, \mathbb{F})$.*
2. $\langle 0, v \rangle = 0 \;\; \forall\, v \in V$
3. $\langle v, 0 \rangle = 0 \;\; \forall\, v \in V$
4. $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle \;\; \forall\, u, v, w \in V$
5. $\langle u, \lambda v \rangle = \bar{\lambda} \langle u, v \rangle \;\; \forall\, \lambda \in \mathbb{F}$ *and* $u, v \in V$
6. $\|v\| = 0 \iff v = 0$
7. $\|\lambda v\| = |\lambda| \|v\| \;\; \forall\, \lambda \in \mathbb{F}$

Proof. The proofs are all very simple and in the book. □

Definition 40. $u, v \in V$ are orthogonal if $\langle u, v \rangle = 0$.

In plane geometry, two vectors are orthogonal if they are perpendicular, see Figure 2.

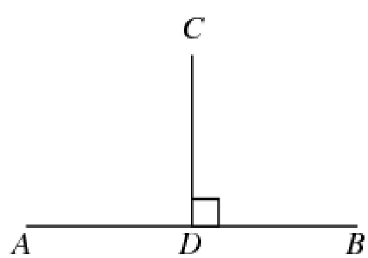

Figure 2: Orthogonal line segments

It is easy to see the following two basic facts:

- 0 is orthogonal to every $v \in V$
- 0 is the only vector in V orthogonal to itself

Theorem 16 (Pythagorean Theorem)**.** *Suppose u and v are orthogonal vectors in V. Then:*

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

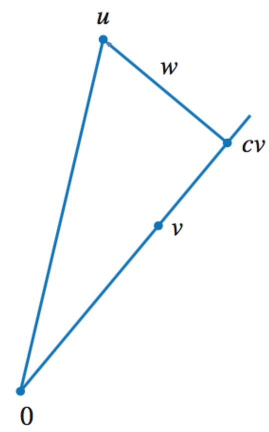

Figure 3: Orthogonal decomposition of u

Proof.

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

□

Now consider the following problem: Suppose $u, v \in V$ with $v \neq 0$. We want to write u as:

$$u = cv + w, \quad \langle v, w\rangle = 0$$

From the book, we have the picture in Figure 3. The question is, what are c and w?

First write u as:

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

We need to choose c such that:

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

We summarize this in the following proposition:

Proposition 36. *Suppose $u, v \in V$ with $v \neq 0$. Set:*

$$c = \frac{\langle u, v\rangle}{\|v\|^2} \quad \textit{and} \quad w = u - \frac{\langle u, v\rangle}{\|v\|^2} v.$$

Then:

$$\langle w, v\rangle = 0 \quad \textit{and} \quad u = cv + w$$

Theorem 17 (Cauchy-Schwarz Inequality)**.** *Suppose* $u, v \in V$*. Then:*

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

Furthermore,

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

Proof. If $v = 0$ then both sides are zero. Thus assume $v \neq 0$, and apply the orthogonal decomposition to u:

$$u = \frac{\langle u, v\rangle}{\|v\|^2} + w, \quad \langle v, w\rangle = 0$$

By the Pythagorean Theorem:

$$\begin{aligned}\|u\|^2 &= \left\|\frac{\langle u, v\rangle}{\|v\|^2}v\right\|^2 + \|w\|^2 \\ &= \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|w\|^2 \\ &\geq \frac{|\langle u, v\rangle|^2}{\|v\|^2}\end{aligned}$$

Now multiply both sides by $\|v\|^2$.

For the second part, we see from the above proof that equality holds if and only if $w = 0$. But then:

$$w = u - \frac{\langle u, v\rangle}{\|v\|^2}v = 0 \Longleftrightarrow u = \frac{\langle u, v\rangle}{\|v\|^2}v$$

□

The Cauchy-Schwarz Inequality is one of the most important, and most used, inequalities in all of mathematics! Lets now use it to prove the triangle inequality for general inner product spaces; Figure 6 gives the plane geometry intuition.

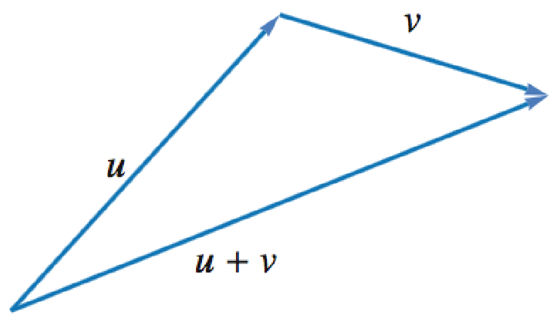

Figure 4: The triangle inequality for $\mathbb{R}^2$

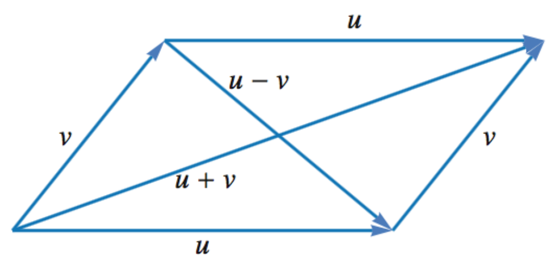

Figure 5: Parallelogram equality in $\mathbb{R}^2$

Theorem 18 (Triangle Inequality)**.** *Suppose $u, v \in V$. Then:*

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

with equality if and only if $u = cv$ for $c \geq 0$.

The next result is the Parallelogram Equality, which also has a geometric interpretation in $\mathbb{R}^2$; see Figure 7.

Proposition 37. *Suppose $u, v \in V$. Then:*

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

End of Lecture 16

Lecture 17: Midterm 1

Chapters 1-5

Lecture 18: Review of Midterm 1 Solutions

Beginning of Lecture 19

Theorem 19 (Triangle Inequality)**.** *Suppose $u, v \in V$. Then:*

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

with equality if and only if $u = cv$ for $c \geq 0$.

Proof. For the first part:

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v, u+v\rangle \\
&= \langle u,u\rangle + \langle v,v\rangle + \langle u,v\rangle + \langle v,u\rangle \\
&= \langle u,u\rangle + \langle v,v\rangle + \langle u,v\rangle + \overline{\langle u,v\rangle} \\
&= \|u\|^2 + \|v\|^2 + 2\text{Re}\langle u,v\rangle \\
&\leq \|u\|^2 + \|v\|^2 + 2|\langle u,v\rangle| \\
&\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| \qquad \text{[Cauchy-Schwarz]} \\
&= (\|u\| + \|v\|)^2
\end{aligned}$$

The proof above shows that equality holds if and only if:

1. $\text{Re}\langle u,v\rangle = |\langle u,v\rangle|$, and
2. $|\langle u,v\rangle| = \|u\|\|v\|$

From the Cauchy-Schwartz inequality, we know #2 holds if and only if $u = cv$ for some $c \in \mathbb{F}$. For #1, consider an arbitrary $\lambda = a + ib \in \mathbb{C}$, where $a, b \in \mathbb{R}$. Then $\text{Re}\lambda = a$ and $|\lambda| = \sqrt{a^2+b^2}$, so $\text{Re}\lambda = |\lambda|$ if and only if $\lambda = a \geq 0$. Thus #1 holds if and only if $\langle u,v\rangle \geq 0$, which combined with $u = cv$, implies that equality holds if and only if $c \geq 0$. □

The next result is the Parallelogram Equality, which also has a geometric interpretation in $\mathbb{R}^2$; see Figure 7.

Proposition 38. *Suppose $u, v \in V$. Then:*

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

Proof. Simply compute:

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

□

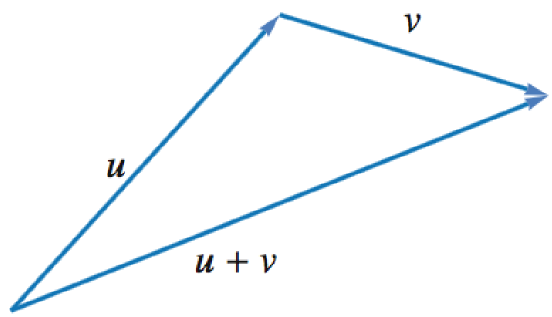

Figure 6: The triangle inequality for $\mathbb{R}^2$

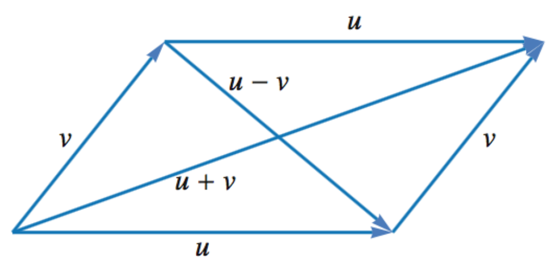

Figure 7: Parallelogram equality in $\mathbb{R}^2$

6.B Orthonormal Bases

Definition 41. A list of vectors $e_1, \ldots, e_m \in V$ is <u>orthonormal</u> if

$$\langle e_j, e_k \rangle = \left\{ \begin{array}{lll} 1 & \text{if } j = k & \text{[norm 1]} \\ 0 & \text{if } j \neq k & \text{[orthogonal]} \end{array} \right\} = \delta(j-k),$$

where

$$\delta : \mathbb{Z} \to \mathbb{C}, \quad \delta(0) = 1 \text{ and } \delta(n) = 0, \ \forall\, n \neq 0.$$

Examples:

1. The standard basis in $\mathbb{F}^n$

2. Recalls the vector space $V = \{f : \mathbb{Z}_N \to \mathbb{C}\}$, where $\mathbb{Z}_N = \{0, \ldots, N-1\}$, and the Fourier basis:

$$e_k : \mathbb{Z}_N \to \mathbb{C}, \quad e_k(n) = \frac{1}{\sqrt{N}} e^{2\pi i k n / N}.$$

Define an inner product on this vector space:

$$\langle f, g \rangle = \sum_{n=0}^{N-1} f(n)\overline{g(n)}$$

Now V is an inner product space and $e_0, \ldots, e_{N-1}$ is an orthonormal list. We can verify this:

$$\begin{aligned}
\langle e_j, e_k \rangle &= \sum_{n=0}^{N-1} e_j(n)\overline{e_k(n)} \\
&= \frac{1}{N}\sum_{n=0}^{N-1} e^{2\pi i j n/N} e^{-2\pi i k n/N} \\
&= \frac{1}{N}\sum_{n=0}^{N-1} e^{2\pi i (j-k) n/N} \\
&= \begin{cases} \frac{1}{N}\sum_{n=0}^{N-1} 1 = \frac{1}{N}\cdot N = 1 & \text{if } j = k \\ \frac{1}{N}\cdot\frac{1-(e^{2\pi i(j-k)/N})^N}{1-e^{2\pi i(j-k)/N}} = \frac{1}{N}\cdot\frac{1-e^{2\pi i(j-k)}}{1-e^{2\pi i(j-k)/N}} = \frac{1}{N}\cdot\frac{1-1}{1-e^{2\pi i(j-k)/N}} = 0 & \text{if } j \neq k \end{cases}
\end{aligned}$$

Since $e_0, \ldots, e_{N-1}$ is also a basis, we call it an orthonormal basis.

Definition 42. An orthonormal basis of V is an orthonormal list of vectors in V that is also a basis of V.

End of Lecture 19

Beginning of Lecture 20

Definition 43. An orthonormal basis of V is an orthonormal list of vectors in V that is also a basis of V.

Orthonormal lists and bases are very convenient! For example:

Proposition 39. *If $e_1, \ldots, e_m$ is an orthonormal list of vectors in V, then:*

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2, \quad \forall\, a_1, \ldots, a_m \in \mathbb{F}.$$

Proof. Expand the left hand side:

$$\begin{aligned} \left\| \sum_{k=1}^m a_k e_k \right\|^2 &= \sum_{j=1}^m \sum_{k=1}^m \langle a_j e_j, a_k e_k \rangle \\ &= \sum_{j=1}^m \sum_{k=1}^m a_j \overline{a}_k \langle e_j, e_k \rangle \\ &= \sum_{j=1}^m \sum_{k=1}^m a_j \overline{a}_k \delta(j-k) \\ &= \sum_{k=1}^m |a_k|^2 \end{aligned}$$

□

Corollary 6 (Important!). *Every orthonormal list of vectors is linearly independent.*

Proof. Suppose $e_1, \ldots, e_m$ is an orthonormal list of vectors in V and $a_1, \ldots, a_m \in \mathbb{F}$ are such that:

$$\sum_{k=1}^m a_k e_k = 0.$$

Then by the previous proposition, $|a_1|^2 + \cdots + |a_m|^2 = 0$, which means $a_k = 0$ for all k since $|a_k|^2 \geq 0$. Thus $e_1, \ldots, e_m$ are linearly independent. □

Proposition 40. *If $\dim V = n$ and $e_1, \ldots, e_n \in V$ is an orthonormal list of vectors, then $e_1, \ldots, e_m$ is an orthonormal basis.*

Proof. By the previous corollary such a list must be linearly independent, and since $n = \dim V$ it then must be a basis. □

In general, given a basis $v_1, \ldots, v_n$ of V and a vector $v \in V$, we know there exists $a_1, \ldots, a_n \in \mathbb{F}$ such that

$$v = a_1 v_1 + \cdots + a_n v_n$$

However, computing $a_1, \ldots, a_n$ can be difficult. If we use an orthonormal basis though, the calculation becomes very easy!

Theorem 20. *Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V and $v \in V$. Then:*

$$v = \sum_{k=1}^{n} \underbrace{\langle v, e_k \rangle}_{a_k} e_k$$

and

$$\|v\|^2 = \sum_{k=1}^{n} |\langle v, e_k \rangle|^2$$

Proof. Because $e_1, \ldots, e_n$ is a basis of V, there exists $a_1, \ldots, a_n \in \mathbb{F}$ such that

$$v = \sum_{k=1}^{n} a_k e_k$$

Now compute the inner product of both sides of the previous equation with e_j:

$$\langle v, e_j \rangle = \langle \sum_{k=1}^{n} a_k e_k, e_j \rangle = \sum_{k=1}^{n} a_k \langle e_k, e_j \rangle = \sum_{k=1}^{n} a_k \delta(j-k) = a_j$$

The second equation on $\|v\|^2$ now follows immediately from Proposition 39. $\square$

Since orthonormal bases are so useful, how do we go about finding them? The next algorithm shows how to turn any linearly independent list into an orthonormal list with the same span.

Theorem 21 (Gram-Schmidt)**.** *Suppose $v_1, \ldots, v_m \in V$ are linearly independent. Define:*

$$e_1 = \frac{v_1}{\|v_1\|},$$

and then for $k = 2, \ldots, m$, define e_k inductively by:

$$e_k = \frac{v_k - \sum_{j=1}^{k-1} \langle v_k, e_j \rangle e_j}{\left\| v_k - \sum_{j=1}^{k-1} \langle v_k, e_j \rangle e_j \right\|} \tag{13}$$

Then $e_1, \ldots, e_m \in V$ is an orthonormal list of vectors such that:

$$\mathrm{span}(v_1, \ldots, v_k) = \mathrm{span}(e_1, \ldots, e_k), \quad \forall\, k = 1, \ldots, m.$$

Remark: Step two of the Gram-Schmidt algorithm is:

$$e_2 = \frac{v_2 - \langle v_2, e_1 \rangle e_1}{\|v_2 - \langle v_2, e_1 \rangle e_1\|} = \frac{1}{\|v_2 - \langle v_2, e_1 \rangle e_1\|} \cdot \left(v_2 - \frac{\langle v_2, v_1 \rangle}{\|v_1\|^2} v_1 \right)$$

which looks very similar to our orthogonal decomposition theorem from 6.A. The only difference is that e_2 is normalized to have norm one. The idea of Gram-Schmidt is to iterate on this decomposition. Now for the proof:

Proof. Proof by induction on k. For $k = 1$, clearly $\mathrm{span}(v_1) = \mathrm{span}(e_1)$, and so our base case holds.

Now suppose that for $1 < k < m$ we have

$$\mathrm{span}(v_1, \ldots, v_{k-1}) = \mathrm{span}(e_1, \ldots, e_{k-1}),$$

and let us consider $e_1, \ldots, e_k$. First note that $v_k \notin \mathrm{span}(v_1, \ldots, v_{k-1})$ (because they are linearly independent) and thus $v_k \notin \mathrm{span}(e_1, \ldots, e_{k-1})$. Thus denominator of e_k in (13) is not zero and so it is well defined. Clearly it has norm one, i.e., $\|e_k\| = 1$.

Now let $1 \leq j < k$. Then:

$$\begin{aligned}
\langle e_k, e_j \rangle &= \left\langle \frac{v_k - \sum_{j=1}^{k-1} \langle v_k, e_j \rangle e_j}{\left\| v_k - \sum_{j=1}^{k-1} \langle v_k, e_j \rangle e_j \right\|}, e_k \right\rangle \\
&= \frac{\langle v_k, e_j \rangle - \langle v_k, e_j \rangle}{\left\| v_k - \sum_{j=1}^{k-1} \langle v_k, e_j \rangle e_j \right\|} \\
&= 0
\end{aligned}$$

Thus $e_1, \ldots, e_k$ is an orthonormal list.

By the definition of e_k, we have:

$$v_k = \left\| v_k - \sum_{j=1}^{k-1} \langle v_k, e_j \rangle e_j \right\| e_k + \sum_{j=1}^{k-1} \langle v_k, e_j \rangle e_j$$

and so $v_k \in \text{span}(e_1, \ldots, e_k)$. Thus by the inductive hypothesis,

$$\text{span}(v_1, \ldots, v_k) \subset \text{span}(e_1, \ldots, e_k).$$

But both lists $v_1, \ldots, v_k$ and $e_1, \ldots, e_k$ are linearly independent, and thus both subspaces have dimension k. Therefore they must be equal. □

END OF LECTURE 20

Beginning of Lecture 21

Example: Let's use Gram-Schmidt find an orthonormal basis of $\mathcal{P}_2([-1,1];\mathbb{R})$ with the inner product:

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

Let's start with the standard basis $1, x, x^2$ which is linearly independent but not orthonormal. We start by computing:

$$\|1\|^2 = \int_{-1}^{1} 1^2\,dx = 2$$

Thus:

$$e_1 = 1/\|1\| = 1/\sqrt{2}$$

Now we need to compute e_2. So we compute:

$$x - \langle x, e_1\rangle e_1 = x - \left(\int_{-1}^{1} x\frac{1}{\sqrt{2}}\,dx\right)\frac{1}{\sqrt{2}} = x$$

and also:

$$\|x - \langle x, e_1\rangle e_1\|^2 = \|x\|^2 = \int_{-1}^{1} x^2\,dx = \frac{2}{3}.$$

Thus:

$$e_2 = \frac{x - \langle x, e_1\rangle e_1}{\|x - \langle x, e_1\rangle e_1\|} = \sqrt{\frac{3}{2}}x$$

Now we need to compute e_3. We have:

$$\begin{aligned} x^2 - \langle x^2, e_1\rangle e_1 - \langle x^2, e_2\rangle e_2 &= x^2 - \left(\int_{-1}^{1} x^2\frac{1}{\sqrt{2}}\,dx\right)\frac{1}{\sqrt{2}} - \left(\int_{-1}^{1} x^2\sqrt{\frac{3}{2}}x\,dx\right)\sqrt{\frac{3}{2}}x \\ &= x^2 - \frac{1}{3} \end{aligned}$$

and also

$$\begin{aligned}\|x^2 - \langle x^2, e_1\rangle e_1 - \langle x^2, e_2\rangle e_2\|^2 &= \left\| x^2 - \frac{1}{3} \right\|^2 \\ &= \int_{-1}^{1} \left(x^2 - \frac{1}{3} \right)^2 \, dx \\ &= \int_{-1}^{1} \left(x^4 - \frac{2}{3}x^2 + \frac{1}{9} \right) \, dx = \frac{8}{45}.\end{aligned}$$

Hence:

$$e_3 = \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3} \right).$$

Thus

$$\mathcal{B} = \frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{45}{8}}(x^2 - 1/3)$$

is an orthonormal basis for $\mathcal{P}_2([-1, 1]; \mathbb{R})$.

The Gram-Schmidt algorithm can be used to prove several useful facts, which we do now.

Proposition 41. *Every finite dimensional inner product space has an orthonormal basis.*

Proof. Choose any basis of V and apply the Gram-Schmidt algorithm to it to get an orthonormal basis. □

Just as we can extend any linearly independent list to a basis, we can also extend any orthonormal list to an orthonormal basis.

Proposition 42. *If V is a finite dimensional inner product space, then every list of orthonormal vectors in V can be extended to an orthonormal basis of V.*

Proof. Let $e_1, \ldots, e_m \in V$ be an orthonormal list. Since they are linearly independent, we can extend them to a basis:

$$e_1, \ldots, e_m, v_1, \ldots, v_n$$

Now apply the Gram-Schmidt algorithm to this basis. Since $e_1, \ldots, e_m$ are orthonormal, as you can verify the Gram-Schmidt algorithm will leave them unchanged. Thus we get an orthonormal basis of the form:

$$e_1, \ldots, e_m, f_1, \ldots, f_n$$

□

Now we return to upper-triangular matrices. Recall that we previously showed that if V is a finite dimensional complex vector space, then for each $T \in \mathcal{L}(V)$ there is a basis $\mathcal{B}$ such that $\mathcal{M}(T; \mathcal{B})$ is upper triangular. When V is an inner product space, we would like to take $\mathcal{B}$ to be an orthonormal basis.

Proposition 43. *Suppose $T \in \mathcal{L}(V)$. If $\mathcal{M}(T; \mathcal{B})$ is upper triangular for some basis $\mathcal{B}$, then there exists an orthonormal basis $\mathcal{B}'$ such that $\mathcal{M}(T; \mathcal{B}')$ is upper triangular.*

Proof. Suppose $\mathcal{M}(T; \mathcal{B})$ is upper triangular and $\mathcal{B} = v_1, \ldots, v_n$. Then $U_k = \operatorname{span}(v_1, \ldots, v_k)$ is invariant under T for each $k = 1, \ldots, n$.

Apply the Gram-Schmidt algorithm to $\mathcal{B}$, producing an orthonormal basis $\mathcal{B}' = e_1, \ldots, e_n$. We claim $\mathcal{B}'$ is the desired basis. Indeed,

$$\operatorname{span}(e_1, \ldots, e_k) = \operatorname{span}(v_1, \ldots, v_k) = U_k, \quad \forall\, k = 1, \ldots, n.$$

Therefore $\operatorname{span}(e_1, \ldots, e_k)$ is invariant under T for each $k = 1, \ldots, n$. Thus $\mathcal{M}(T; \mathcal{B}')$ is upper triangular. □

Remark: The above proposition holds for *any* inner product space and operator T for which there exists some basis $\mathcal{B}$ such that $\mathcal{M}(T; \mathcal{B})$ is upper triangular. In particular, V can be a real vector space, if such a $\mathcal{B}$ exists. Of course when V is a complex vector space, we can guarantee the result...

Theorem 22 (Schur's Theorem)**.** *If V is a finite dimensional complex inner product space and $T \in \mathcal{L}(V)$, then there exists an orthonormal basis $\mathcal{B}'$ such that $\mathcal{M}(T; \mathcal{B}')$ is upper triangular.*

Proof. Since V is a finite dimensional complex vector space, there exists a basis $\mathcal{B}$ such that $\mathcal{M}(T; \mathcal{B})$ is upper triangular. Now apply the previous proposition. □

END OF LECTURE 21

Beginning of Lecture 22

Definition 44. A function φ is a linear functional on V if $\varphi \in \mathcal{L}(V, \mathbb{F})$.

Examples:

- *Fix* an arbitrary $u \in V$. Then:

$$\begin{aligned} \varphi : V &\to \mathbb{F} \\ v &\mapsto \varphi(v) = \langle v, u \rangle \end{aligned}$$

is a linear functional on V.

- *Fix* an arbitrary continuous function $f \in C([-1,1];\mathbb{R})$. Then:

$$\begin{aligned} \varphi : \mathcal{P}_2([-1,1];\mathbb{R}) &\to \mathbb{R} \\ p &\mapsto \varphi(p) = \int_{-1}^{1} p(x) f(x)\, dx \end{aligned}$$

is a linear functional on $\mathcal{P}([-1,1];\mathbb{R})$.

Remark: It is tempting to write $\varphi(p) = \langle p, f \rangle$, but we may not have $f \in \mathcal{P}_2([-1,1];\mathbb{R})$. For example, $f(x) = \cos(x)$ or $f(x) = e^x$, and so $\langle p, f \rangle$ does not necessarily make sense. Thus the next result is quite remarkable...

Theorem 23 (Riesz Representation Theorem)**.** *Suppose V is finite-dimensional and $\varphi \in \mathcal{L}(V, \mathbb{F})$. Then there is a unique vector $u \in V$ such that*

$$\varphi(v) = \langle v, u \rangle, \quad \forall v \in V$$

Proof. First we show that there exists a $u \in V$ such that $\varphi(v) = \langle v, u \rangle$, then we show that u is unique. Let $e_1, \ldots, e_n$ be an orthonormal basis of V. Then:

$$\begin{aligned} \varphi(v) &= \varphi\left(\sum_{k=1}^{n} \langle v, e_k \rangle e_k \right) \\ &= \sum_{k=1}^{n} \langle v, e_k \rangle \varphi(e_k) \\ &= \langle v, \sum_{k=1}^{n} \overline{\varphi(e_k)} e_k \rangle \end{aligned}$$

Thus setting:

$$u = \sum_{k=1}^{n} \overline{\varphi(e_k)} e_k$$

we have $\varphi(v) = \langle v, u \rangle$ for all $v \in V$.

Now we prove that u is unique. Suppose $u_1, u_2 \in V$ such that

$$\varphi(v) = \langle v, u_1 \rangle = \langle v, u_2 \rangle, \quad \forall\, v \in V$$

Then:

$$0 = \langle v, u_1 \rangle - \langle v, u_2 \rangle = \langle v, u_1 - u_2 \rangle, \quad \forall\, v \in V$$

Taking $v = u_1 - u_2$ implies $\|u_1 - u_2\|^2 = 0$ which implies that $u_1 - u_2 = 0$ and so $u_1 = u_2$. Therefore u is unique. □

Remark (con't): Returning to the example above, even if $f \in C([-1,1];\mathbb{R})$ and $f \notin \mathcal{P}_2([-1,1];\mathbb{R})$, there still exists a unique $q \in \mathcal{P}_2([-1,1];\mathbb{R})$ such that:

$$\varphi(p) = \int_{-1}^{1} p(x)f(x)\,dx = \int_{-1}^{1} p(x)q(x)\,dx = \langle p, q \rangle, \quad \forall\, p \in \mathcal{P}_2([-1,1];\mathbb{R}).$$

Furthermore, we can compute what q is by selecting an orthonormal basis e_1, e_2, e_3 for $\mathcal{P}_2([-1,1];\mathbb{R})$ (like the one we computed earlier) and using the formula:

$$q = \overline{\varphi(e_1)}e_1 + \overline{\varphi(e_2)}e_2 + \overline{\varphi(e_3)}e_3$$

This works in general.

End of Lecture 22

Beginning of Lecture 23

6.C Orthogonal Complements and Minimization Problems

Definition 45. If $U \subset V$, then the orthogonal complement of U is:

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

Geometrical Examples:

- If U is a line in $V = \mathbb{R}^2$, then $U^\perp$ is the line orthogonal U that passes through the origin.
- If U is a line in $V = \mathbb{R}^3$, then $U^\perp$ is the plane orthogonal to U that contains the origin.
- If U is a plane in $V = \mathbb{R}^3$, then $U^\perp$ is the line orthogonal to U that passes through the origin.

Proposition 44. *The following are basic properties of the orthogonal complement:*

1. *If $U \subset V$, then $U^\perp$ is a subspace of V*
2. $\{0\}^\perp = V$
3. $V^\perp = \{0\}$
4. *If $U \subset V$, then $U \cap U^\perp \subset \{0\}$. If U is a subspace of V, then $U \cap U^\perp = \{0\}$.*
5. *If $U \subset V$ and $W \subset V$ ad $U \subset W$, then $W^\perp \subset U^\perp$*

Proof. We go through the list:

1. We need to show $U^\perp$ contains 0, is closed under addition and scalar multiplication.
   - Clearly $\langle 0, u\rangle = 0 \ \forall\, u \in U$, thus $0 \in U^\perp$.
   - Now suppose $v, w \in U^\perp$. If $u \in U$, then:

   $$\langle v + w, u\rangle = \langle v, u\rangle + \langle w, u\rangle = 0 + 0 = 0 \Longrightarrow v + w \in U^\perp$$

- Suppose $\lambda \in \mathbb{F}$ and $v \in U^\perp$. If $u \in U$, then

$$\langle \lambda v, u\rangle = \lambda\langle v, u\rangle = \lambda \cdot 0 = 0 \Longrightarrow \lambda v \in U^\perp$$

2. $\langle v, 0\rangle = 0 \;\forall\, v \in V \Longrightarrow v \in \{0\}^\perp$, so $\{0\}^\perp = V$

3. Suppose $v \in V^\perp$. Then $\langle v, v\rangle = 0 \Longrightarrow v = 0$. Thus $V^\perp = \{0\}$.

4. Suppose $U \subset V$ and $v \in U \cap U^\perp$. Then we must have $\langle v, v\rangle = 0 \Longrightarrow v = 0$, and so $U \cap U^\perp \subset \{0\}$. If U is a subspace of V, then $0 \in U$ and by above $0 \in U^\perp$, so $U \cap U^\perp = \{0\}$.

5. This is clear.

□

Recall early on we proved that if U is a subspace of V, then there exists a second subspace W of V such that $V = U \oplus W$. We now show that we can take $W = U^\perp$.

Proposition 45. *If U is a finite dimensional subspace of V, then*

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

Proof. From the previous proposition we know that $U \cap U^\perp = \{0\}$, so we just need to show that $U + U^\perp = V$. Let $v \in V$ and let $e_1, \ldots, e_m$ be an ONB of U. Clearly then:

$$v = \underbrace{\sum_{k=1}^{m} \langle v, e_k\rangle e_k}_{u \in U} + \underbrace{v - \sum_{k=1}^{m} \langle v, e_k\rangle e_k}_{w}$$

We want to show $w \in U^\perp$. But this is clear since:

$$\forall\, k = 1, \ldots, m, \quad \langle w, e_k\rangle = \langle v, e_k\rangle - \langle v, e_k\rangle = 0 \Longrightarrow w \in U^\perp$$

□

Corollary 7. *If V is finite dimensional and U is a subspace of V, then:*

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

Proposition 46. *If U is a finite dimensional subspace of V, then*

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

Proof. We prove this in two parts:

- First we show that $U \subset (U^\perp)^\perp$. Suppose $u \in U$. Then by definition of $U^\perp$,
$$\langle v, u\rangle = 0 = \langle u, v\rangle, \quad \forall\, v \in U^\perp$$
But the above also implies that $u \in (U^\perp)^\perp$ since
$$(U^\perp)^\perp = \{w \in V : \langle w, v\rangle = 0, \ \forall\, v \in U^\perp\}$$

- Now we show that $(U^\perp)^\perp \subset U$. Suppose that $v \in (U^\perp)^\perp$. $v \in V$ so we can write it as:
$$v = u + w, \quad u \in U,\, w \in U^\perp \Longrightarrow v - u = w \in U^\perp$$
But by the above, we also have $u \in U \subset (U^\perp)^\perp$ and so:
$$v - u \in (U^\perp)^\perp \Rightarrow v - u \in U^\perp \cap (U^\perp)^\perp \Rightarrow v - u = 0 \Rightarrow v = u \Rightarrow v \in U$$

□

End of Lecture 23

Beginning of Lecture 24

Now we use the fact that $V = U \oplus U^\perp$ to define the orthogonal projection of V onto U.

Definition 46. Suppose U is a finite dimensional subspace of V.
The orthogonal projection of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined as:

$$P_U v = u, \quad \text{where } v = u + w,\ \ u \in U,\ w \in U^\perp$$

Remark: Since the decomposition $v = u + w \in U \oplus U^\perp$ is unique, the orthogonal projection P_U is well defined.

Example: Recall from earlier we have: If $u, v \in V$ and $u \neq 0$, then

$$v = cu + w, \quad \langle u, w \rangle = 0,\ \ c = \frac{\langle v, u \rangle}{\|u\|^2}$$

Thus if $U = \text{span}(u)$, then

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

More generally, if U is an arbitrary finite dimensional subspace of V and $e_1, \ldots, e_m$ is an ONB for U, then:

$$P_U v = \sum_{k=1}^{m} \langle v, e_k \rangle e_k \tag{14}$$

This is just one of many properties of P_U:

Proposition 47. *If U is a finite dimensional subspace of V and $v \in V$, then:*

1. $P_U \in \mathcal{L}(V)$
2. $P_U u = u \ \forall\, u \in U$
3. $P_U w = 0 \ \forall\, w \in U^\perp$
4. $\text{range}\, P_U = U$
5. $\text{null}\, P_U = U^\perp$

6. $v - P_U v \in U^{\perp}$

7. $P_U^2 = P_U$

8. $\|P_U v\| \leq \|v\|$

Proof. We prove each part:

1. This follows from (14) and the linearity of the inner product in the first argument.
2. If $u \in U$, then $u = u + 0 \in U \oplus U^{\perp}$, and thus $P_U u = u$.
3. If $w \in U^{\perp}$, then $w = 0 + w \in U \oplus U^{\perp}$, and thus $P_U w = 0$.
4. This is clear
5. Part 3 implies that $U^{\perp} \subset \operatorname{null} P_U$. Now suppose that $v \in \operatorname{null} P_U$, i.e., $P_U v = 0$. Then if $v = u + w \in U \oplus U^{\perp}$, we must have $P_U v = u = 0$, which implies that $v = 0 + w = w \in U^{\perp}$ and so $\operatorname{null} P_U \subset U^{\perp}$.
6. If $v = u + w \in U \oplus U^{\perp}$, then:
$$v - P_U v = (u + w) - u = w \in U^{\perp}$$
7. If $v = u + w \in U \oplus U^{\perp}$ then:
$$(P_U^2)v = P_U(P_U v) = P_U u = u = P_U v$$
8. If $v = u + w \in U \oplus U^{\perp}$ then:
$$\|P_U v\|^2 = \|u\|^2 \leq \|u\|^2 + \|w\|^2 = \|v\|^2$$

□

We now turn to a very important minimization problem: Given a subspace U of V and a point $v \in V$, find a point $u_0 \in U$ such that $\|v - u_0\|$ is as small as possible. In other words, find $u_0 \in U$ such that:

$$\|v - u_0\| = \min_{u \in U} \|v - u\| \iff \|v - u_0\| \leq \|v - u\|, \ \forall\, u \in U$$

In fact the orthogonal projection gives the solution!

Theorem 24. *Suppose U is a finite dimensional subspace of V, $v \in V$, and $u \in U$. Then:*

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

Furthermore,

$$\|v - P_U v\| = \|v - u\| \iff u = P_U v$$

Proof. We have:

$$\begin{aligned} \|v - P_U v\|^2 &\leq \|\underbrace{v - P_U v}_{\in U^\perp}\|^2 + \|\underbrace{P_U v - u}_{\in U}\|^2 \\ &= \|(v - P_U v) + (P_U v - u)\|^2 \\ &= \|v - u\|^2 \end{aligned}$$

The inequality is an equality if and only if:

$$\begin{aligned} \|v - P_U v\| = \|v - u\| &\Leftrightarrow \|v - P_U v\|^2 = \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ &\Leftrightarrow \|P_U v - u\|^2 = 0 \\ &\Leftrightarrow P_U v = u \end{aligned}$$

□

Please read the very interesting Example 6.58 in the book.

End of Lecture 24

Beginning of Lecture 25

7 Operators on Inner Product Spaces

We now explore the structure of operators on inner product spaces, which we have been building towards for quite a while now. This will lead to some of the most important results in all of Linear Algebra. In particular, we will completely characterize those operators that are diagonalizable, giving us a complete solution to the questions we asked at the beginning of Chapter 5.

Notation: V and W are inner product spaces over the same field $\mathbb{F}$. Sometimes we will write $\langle \cdot, \cdot \rangle_V$ for the inner product on V and $\langle \cdot, \cdot \rangle_W$ for the inner product on W.

7.A Self-Adjoint and Normal Operators

Definition 47. Suppose $T \in \mathcal{L}(V, W)$. The adjoint of T is the function $T^* : W \to V$ such that:

$$\langle Tv, w \rangle_W = \langle v, T^* w \rangle_V, \quad \forall\, v \in V, \ \forall\, w \in W$$

Warmup: Show the T^* is well-defined.

Solution: Fix $T \in \mathcal{L}(V, W)$ and $w \in W$ and consider the linear functional $\varphi \in \mathcal{L}(V, \mathbb{F})$ defined as:

$$\varphi(v) = \langle Tv, w \rangle$$

By the Riesz Representation Theorem, there exists a unique vector $u \in V$ such that:

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

Define $T^* w := u$.

Example: Consider the vector space:

$$V = \{ f \in C^\infty([0,1]; \mathbb{R}) : f^{(k)}(0) = f^{(k)}(1) \ \forall\, k = 0, 1, 2, \ldots \}$$

Define the inner product on V as:

$$\langle f, g \rangle = \int_0^1 f(x) g(x)\, dx, \quad f, g \in V$$

Let $T \in \mathcal{L}(V)$ be the differentiation operator, i.e., $T = D$ where

$$Df = f'$$

Let's compute the adjoint of D using integration by parts:

$$\begin{aligned}
\langle Df, g\rangle &= \int_0^1 Df(x)g(x)\,dx \\
&= \int_0^1 f'(x)g(x)\,dx \\
&= f(x)g(x)|_{x=0}^{x=1} - \int_0^1 f(x)g'(x)\,dx \\
&= 0 - \int_0^1 f(x)g'(x)\,dx \\
&= \int_0^1 f(x)\cdot(-Dg(x))\,dx \\
&= \langle f, -Dg\rangle
\end{aligned}$$

Thus, $D^* = -D$.

We have used this technique before, but it will be especially useful when dealing with adjoints and so we write it down here:

Lemma 2. *Let $u, w \in V$. If:*

$$\langle v, u\rangle = \langle v, w\rangle, \quad \forall\, v \in V,$$

then $u = w$.

Proof. Indeed,

$$\langle v, u\rangle = \langle v, w\rangle,\ \forall\, v \in V \Longrightarrow \langle v, u - w\rangle = 0,\ \forall\, v \in V$$

Since it is true for all $v \in V$, we can take $v = u - w$ and we then have:

$$\langle u - w, u - w\rangle = 0 \Longrightarrow \|u - w\|^2 = 0 \Longrightarrow u - w = 0 \Longrightarrow u = w$$

□

Proposition 48. *If $T \in \mathcal{L}(V, W)$, then $T^* \in \mathcal{L}(W, V)$*

Proof. Need to show additivity and homogeneity of T^*. Let $w_1, w_2 \in W$ and $v \in V$:

$$\begin{aligned}\langle v, T^*(w_1 + w_2)\rangle_V &= \langle Tv, w_1 + w_2\rangle_W \\ &= \langle Tv, w_1\rangle_W + \langle Tv, w_2\rangle_W \\ &= \langle v, T^*w_1\rangle_V + \langle v, T^*w_1\rangle_V \\ &= \langle v, T^*w_1 + T^*w_2\rangle_V\end{aligned}$$

Therefore $T^*(w_1 + w_2) = T^*w_1 + T^*w_2$.

Now let $w \in W$, $\lambda \in \mathbb{F}$, and $v \in V$:

$$\begin{aligned}\langle v, T^*(\lambda w)\rangle_V &= \langle Tv, \lambda w\rangle_W \\ &= \bar{\lambda}\langle Tv, w\rangle_W \\ &= \bar{\lambda}\langle v, T^*w\rangle_V \\ &= \langle v, \lambda T^*w\rangle_V\end{aligned}$$

and so $T^*(\lambda w) = \lambda T^*w$ □

Proposition 49. *The following properties of the adjoint hold:*

1. $(S + T)^* = S^* + T^*,\ \forall\, S, T \in \mathcal{L}(V, W)$
2. $(\lambda T)^* = \bar{\lambda}T^*,\ \forall\, \lambda \in \mathbb{F},\ \forall\, T \in \mathcal{L}(V, W)$
3. $(T^*)^* = T,\ \forall\, T \in \mathcal{L}(V, W)$
4. $I^* = I$, *where* $I \in \mathcal{L}(V)$ *is the identity operator, i.e.,* $Iv = v\ \ \forall v \in V$
5. $(ST)^* = T^*S^*,\ \forall\, T \in \mathcal{L}(V, W),\ \forall\, S \in \mathcal{L}(W, U)$

Proof. The proofs of #1 and #2 are very similar to the proof that T^* is linear. The proof of #4 is also quite easy.

To prove #3, let $v \in V$ and $w \in W$,

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

To prove #5, let $v \in V$ and $u \in U$,

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

□

End of Lecture 25

Beginning of Lecture 26

The null space and range of T are related to the null space and range of T^* through the orthogonal complement, as we now prove.

Proposition 50. *If $T \in \mathcal{L}(V, W)$, then:*

1. $\operatorname{null} T^* = (\operatorname{range} T)^{\perp}$
2. $\operatorname{range} T^* = (\operatorname{null} T)^{\perp}$
3. $\operatorname{null} T = (\operatorname{range} T^*)^{\perp}$
4. $\operatorname{range} T = (\operatorname{null} T^*)^{\perp}$

Proof. We prove #1 first:

$$\begin{aligned} w \in \operatorname{null} T^* &\iff T^* w = 0 \\ &\iff \langle v, T^* w \rangle = 0, \quad \forall\, v \in V \\ &\iff \langle Tv, w \rangle = 0, \quad \forall\, v \in V \\ &\iff w \in (\operatorname{range} T)^{\perp} \end{aligned}$$

Thus $\operatorname{null} T^* = (\operatorname{range} T)^{\perp}$.

The rest now follow easily. Indeed, taking the orthogonal complement of both sides of #1 gives #4. Replacing T with T^* in #1 gives #3, and in number #4 gives #2. □

We now relate the adjoint to matrices.

Definition 48. The conjugate transpose of an $m \times n$ matrix $A \in \mathbb{F}^{m,n}$ is the $n \times m$ matrix $A^{\dagger} \in \mathbb{F}^{n,m}$ defined as:

$$A^{\dagger}_{j,k} = \overline{A}_{k,j}, \quad \forall\, j = 1, \ldots, n, \;\; k = 1, \ldots, m$$

Proposition 51. *Let $T \in \mathcal{L}(V, W)$, $\mathcal{B}_V = e_1, \ldots, e_n$ be an ONB of V, and $\mathcal{B}_W = f_1, \ldots, f_m$ be an ONB of W (note: they must be orthonormal!!). Then:*

$$\mathcal{M}(T^*; \mathcal{B}_W, \mathcal{B}_V) = \mathcal{M}(T; \mathcal{B}_V, \mathcal{B}_W)^{\dagger}$$

Proof. Let $A = \mathcal{M}(T; \mathcal{B}_V, \mathcal{B}_W)$. Recall that $A_{j,k}$ is defined by writing Te_k as a linear combinations of $f_1, \ldots, f_m$:

$$Te_k = \sum_{j=1}^{m} A_{j,k} f_j = \sum_{j=1}^{m} \langle Te_k, f_j \rangle_W f_j \Longrightarrow A_{j,k} = \langle Te_k, f_j \rangle_W$$

where the second equality follows since $\mathcal{B}_W$ is an ONB.

Now let $B = \mathcal{M}(T^*; \mathcal{B}_W, \mathcal{B}_V)$. Then B is defined as:

$$T^* f_k = \sum_{j=1}^{n} B_{j,k} e_j = \sum_{j=1}^{n} \langle T^* f_k, e_j \rangle_V e_j \Longrightarrow B_{j,k} = \langle T^* f_k, e_j \rangle_V$$

But then:

$$B_{j,k} = \langle T^* f_k, e_j \rangle = \overline{\langle e_j, T^* f_k \rangle} = \overline{\langle Te_j, f_k \rangle} = \overline{A}_{k,j} = A^{\dagger}_{j,k}$$

□

Now we focus in on operators $T \in \mathcal{L}(V)$, where V is an inner product space. We shall be particularly interested in the following operators.

Definition 49. An operator $T \in \mathcal{L}(V)$ is self-adjoint if $T = T^*$, i.e.,

$$\langle Tv, w \rangle = \langle v, Tw \rangle, \quad \forall\, v, w \in V$$

Remark: The previous proposition shows that for a general $T \in \mathcal{L}(V)$, if $\mathcal{B}$ is an ONB for V, then $\mathcal{M}(T^*; \mathcal{B}) = \mathcal{M}(T; \mathcal{B})^{\dagger}$. But if T is self-adjoint, then $T = T^*$ and so $\mathcal{M}(T; \mathcal{B}) = \mathcal{M}(T^*; \mathcal{B}) = \mathcal{M}(T; \mathcal{B})^{\dagger}$, which implies that $\mathcal{M}(T; \mathcal{B})$ is symmetric and real valued.

Proposition 52. *The eigenvalues of self-adjoint operators are real valued (even when $\mathbb{F} = \mathbb{C}$).*

Proof. Let $T \in \mathcal{L}(V)$ be self-adjoint, $\lambda \in \mathbb{F}$ and eigenvalue of T, and $v \in V$ a corresponding nonzero eigenvector so that $Tv = \lambda v$. Then:

$$\lambda \|v\|^2 = \langle \lambda v, v \rangle = \langle Tv, v \rangle = \langle v, Tv \rangle = \langle v, \lambda v \rangle = \bar{\lambda} \|v\|^2 \Rightarrow \lambda = \bar{\lambda} \Rightarrow \lambda \in \mathbb{R}$$

□

Proposition 53. *Let V be a complex inner product space and $T \in \mathcal{L}(V)$.*

$$\text{If } \langle Tv, v\rangle = 0, \ \forall\, v \in V, \text{ then } T = 0$$

Proof. Suppose $\langle Tv, v\rangle = 0, \ \forall\, v \in V$. Let $u, w \in V$ and consider the clever rewriting of $\langle Tu, w\rangle$:

$$\begin{aligned}
\langle Tu, w\rangle &= \frac{1}{4}\langle T(u+w), \underbrace{u+w}_{v_1}\rangle - \frac{1}{4}\langle T(u-w), \underbrace{u-w}_{v_2}\rangle \\
&\quad + \frac{1}{4}\langle T(u+iw), \underbrace{u+iw}_{v_3}\rangle i - \frac{1}{4}\langle T(u-iw), \underbrace{u-iw}_{v_4}\rangle i \\
&= \frac{1}{4}\left(\langle Tv_1, v_1\rangle + \langle Tv_2, v_2\rangle + \langle Tv_3, v_3\rangle i + \langle Tv_4, v_4\rangle i\right) \\
&= 0
\end{aligned}$$

Thus $\langle Tu, w\rangle = 0, \ \forall\, u, w \in V$. Taking $w = Tu$, we get $\|Tu\|^2 = 0, \ \forall\, u \in V$, which implies that $Tu = 0$ for all $u \in V$, and so $T = 0$. □

Remark: False if $\mathbb{F} = \mathbb{R}$. Take $V = \mathbb{R}^2$ and T to be a 90-degree rotation.

Proposition 54. *Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Then:*

$$T \text{ is self-adjoint} \iff \langle Tv, v\rangle \in \mathbb{R}, \ \forall\, v \in V$$

Proof. Let $v \in V$, then:

$$\langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = \langle Tv, v\rangle - \langle v, Tv\rangle = \langle Tv, v - \langle T^*v, v\rangle = \langle (T-T^*)v, v\rangle \quad (15)$$

If $\langle Tv, v\rangle \in \mathbb{R}$, then by (15):

$$0 = \langle (T - T^*)v, v\rangle \Longrightarrow T - T^* = 0 \text{ [by previous Proposition] } \Longrightarrow T = T^*$$

Conversely, if T is self=adjoint then (15) also implies:

$$\langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = 0 \Longrightarrow \langle Tv, v\rangle = \overline{\langle Tv, v\rangle} \Longrightarrow \langle Tv, v\rangle \in \mathbb{R}$$

□

Remark: Also false if $\mathbb{F} = \mathbb{R}$ since $\langle Tv, v\rangle \in \mathbb{R}$ for all $T \in \mathcal{L}(V)$, including those that are not self-adjoint.

End of Lecture 26

Beginning of Lecture 27

Proposition 55. *If T is self-adjoint and $\langle Tv, v\rangle = 0$ for all $v \in V$, then $T = 0$ (even if $\mathbb{F} = \mathbb{R}$).*

Proof. If $\mathbb{F} = \mathbb{C}$ then we already proved this. So assume that $\mathbb{F} = \mathbb{R}$, and suppose that T is self-adjoint and $\langle Tv, v\rangle = 0$ for all $v \in V$. Let $u, w \in V$; then:

$$\begin{aligned}
&\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle = \\
&= \langle Tu, u\rangle + \langle Tu, w\rangle + \langle Tw, u\rangle + \langle Tw, w\rangle - \langle Tu, u\rangle + \langle Tu, w\rangle + \langle Tw, u\rangle - \langle Tw, w\rangle \\
&= 2\langle Tu, w\rangle + 2\langle Tw, u\rangle \\
&= 2\langle Tu, w\rangle + 2\langle u, Tw\rangle \ \ [\mathbb{F} = \mathbb{R}] \\
&= 4\langle Tu, w\rangle \ \ [T = T^*]
\end{aligned}$$

Thus, let $v_1 = u + w$ and $v_2 = u - w$:

$$\langle Tu, w\rangle = \frac{1}{4}(\langle Tv_1, v_1\rangle + \langle Tv_2, v_2\rangle) = 0, \ \ \forall\, u, w \in V \Longrightarrow T = 0$$

□

Self-adjoint operators are a subset of the following class of important operators.

Definition 50. $T \in \mathcal{L}(V)$ is normal if

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

Example: Recall the vector space:

$$V = \{f \in C^\infty([0,1]; \mathbb{R}) : f^{(k)}(0) = f^{(k)}(1), \ \ \forall\, k = 0, 1, 2, \ldots\}$$

and the differentiation operator $D \in \mathcal{L}(V)$,

$$Df = f'$$

We showed that $D^* = -D$. Thus D is not self adjoint, but

$$DD^* = D(-D) = -D^2 = (-D)D = D^*D$$

and so D is normal.

Proposition 56.

$$T \textit{ is normal } \iff \|Tv\| = \|T^*v\|, \ \forall\, v \in V$$

Proof. We have:

$$\begin{aligned} T \text{ is normal } &\iff T^*T - TT^* = 0 \\ &\iff \langle (T^*T - TT^*)v, v\rangle = 0, \ \forall\, v \in V \qquad (16) \\ &\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle, \ \forall\, v \in V \\ &\iff \|Tv\|^2 = \|T^*v\|^2, \ \forall\, v \in V \end{aligned}$$

where (16) follows from the previous Proposition since $T^*T - TT^*$ is self-adjoint. □

Corollary 8. *Suppose $T \in \mathcal{L}(V)$ is normal. If $v \in V$ is an eigenvector of T with eigenvalue λ, then v is an eigenvector of T^* with eigenvalue $\bar{\lambda}$.*

Proof. T normal implies that $T - \lambda I$ is normal since:

$$\begin{aligned} (T - \lambda I)(T - \lambda I)^* &= (T - \lambda I)(T^* - \bar{\lambda} I) \\ &= TT^* - \bar{\lambda}T - \lambda T^* + |\lambda|^2 I \\ &= T^*T - \lambda T^* - \bar{\lambda}T + |\lambda|^2 I \\ &= (T^* - \bar{\lambda} I)(T - \lambda I) \\ &= (T - \lambda I)^*(T - \lambda I) \end{aligned}$$

Then by the previous Proposition:

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

□

Proposition 57. *If T is normal, then the eigenvectors of T corresponding to distinct eigenvalues are orthogonal.*

Proof. Let α, β be distinct eigenvalues of T with corresponding eigenvectors u, v so that

$$Tu = \alpha u \qquad \text{and} \qquad Tv = \beta v$$

From the previous Corollary we have $T^*v = \bar{\beta} v$. Thus:

$$(\alpha - \beta)\langle u, v\rangle = \alpha\langle u, v\rangle - \beta\langle u, v\rangle = \langle \alpha u, v\rangle - \langle u, \bar{\beta} v\rangle = \langle Tu, v\rangle - \langle u, T^*v\rangle = 0$$

Since $\alpha \neq \beta$, we must have $\langle u, v\rangle = 0$. □

7.B The Spectral Theorem

Two flavors, real and complex. As is often the case, the complex version is in fact easier. So we start with that.

Complex Spectral Theorem

Theorem 25 (Complex Spectral Theorem)**.** *Suppose* $\mathbb{F} = \mathbb{C}$ *and* $T \in \mathcal{L}(V)$*. Then the following are equivalent:*

1. *T is normal*
2. *V has an ONB consisting of eigenvectors of T*
3. *T has a diagonal matrix with respect to some ONB of V*

END OF LECTURE 27

Beginning of Lecture 28

Theorem 26 (Complex Spectral Theorem). *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *T is normal*
2. *V has an ONB consisting of eigenvectors of T*
3. *T has a diagonal matrix with respect to some ONB of V*

Proof. We prove this in parts:

- (2) $\iff$ (3) follows from our work in Chapter 5.
- (3) $\implies$ (1): Let $\mathcal{B}$ be an ONB such that $\mathcal{M}(T;\mathcal{B})$ is diagonal. Then

  $$\mathcal{M}(T^*;\mathcal{B}) = \mathcal{M}(T;\mathcal{B})^\dagger \implies \mathcal{M}(T^*;\mathcal{B}) \text{ is diagonal}$$

  But then since diagonal matrices commute we have:

  $$\begin{aligned}\mathcal{M}(TT^*;\mathcal{B}) = \mathcal{M}(T;\mathcal{B})\mathcal{M}(T^*;\mathcal{B}) = \mathcal{M}(T^*;\mathcal{B})\mathcal{M}(T;\mathcal{B}) &= \mathcal{M}(T^*T;\mathcal{B}) \\ \implies TT^* &= T^*T\end{aligned}$$

  since $\mathcal{L}(V)$ and $\mathbb{C}^{n,n}$ are isomorphic under the linear map $\mathcal{M} : \mathcal{L}(V) \to \mathbb{C}^{n,n}$.

- Now suppose that T is normal. By Schur's Theorem there exists and ONB $\mathcal{B} = e_1, \ldots, e_n$ such that $\mathcal{M}(T;\mathcal{B})$ is upper triangular; write the matrix as:

  $$\mathcal{M}(T;\mathcal{B}) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{pmatrix} \implies \mathcal{M}(T^*;\mathcal{B}) = \mathcal{M}(T;\mathcal{B})^\dagger = \begin{pmatrix} \bar{a}_{1,1} & & 0 \\ \vdots & \ddots & \\ \bar{a}_{1,n} & \cdots & \bar{a}_{n,n} \end{pmatrix}$$

  We now show that $\mathcal{M}(T;\mathcal{B})$ is in fact a diagonal matrix. Indeed, because $\mathcal{B}$ is an ONB and T is normal:

  $$\begin{aligned}\|Te_1\|^2 &= |a_{1,1}|^2 \\ \|Te_1\|^2 &= \|T^*e_1\|^2 = |a_{1,1}|^2 + |a_{1,2}|^2 + \cdots + |a_{1,n}|^2 \\ &\implies a_{1,2} = \cdots = a_{1,n} = 0\end{aligned}$$

Now we also have:

$$\begin{aligned}\|Te_2\|^2 &= |a_{1,2}|^2 + |a_{2,2}|^2 = 0 + |a_{2,2}|^2 = |a_{2,2}|^2 \\ \|Te_2\|^2 &= \|T^*e_2\|^2 = |a_{2,2}|^2 + |a_{2,3}|^2 + \cdots |a_{2,n}|^2 \\ &\implies a_{2,3} = \cdots = a_{2,n} = 0\end{aligned}$$

Continuing in this fashion we see that all off-diagonal entries must be zero.

□

Next week we will prove the Real Spectral Theorem:

Theorem 27 (Real Spectral Theorem)**.** *Suppose* $\mathbb{F} = \mathbb{R}$ *and* $T \in \mathcal{L}(V)$*. Then the following are equivalent:*

1. *T is self-adjoint*
2. *V has an ONB consisting of eigenvectors of T*
3. *T has a diagonal matrix with respect to some ONB of V*

End of Lecture 28

Beginning of Lecture 29

Warmup: If we change $\mathbb{F} = \mathbb{R}$, where does the proof of the Complex Spectral Theorem fall apart?

Answer: To prove (1) $\implies$ (3) we used Schur's Theorem, which only applies to complex vector spaces.

Real Spectral Theorem

We now aim to prove the Real Spectral Theorem:

Theorem 28 (Real Spectral Theorem)**.** *Suppose $\mathbb{F} = \mathbb{R}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *T is self-adjoint*
2. *V has an ONB consisting of eigenvectors of T*
3. *T has a diagonal matrix with respect to some ONB of V*

The Real Spectral Theorem is harder to prove and as such we will first need some preliminary results.

Consider the quadratic polynomial $p \in \mathcal{P}_2(\mathbb{R})$:

$$p(x) = x^2 + bx + c, \quad x, b, c \in \mathbb{R}$$

Note the following:

$$\text{If } b^2 < 4c, \text{ then } x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right) > 0, \ \forall\, x \in \mathbb{R}$$

In particular $p(x) > 0$ so it has a multiplicative inverse for all $x \in \mathbb{R}$, namely $p(x) \cdot (1/p(x)) = 1$. A similar type of reasoning leads to the following result.

Proposition 58. *If $T \in \mathcal{L}(V)$ is self-adjoint and $b, c \in \mathbb{R}$ satisfy $b^2 < 4c$, then*

$$p(T) = T^2 + bT + cI$$

is invertible.

Proof. Let $v \in V$, $v \neq 0$. Then:

$$\begin{aligned}
\langle p(T)v, v\rangle &= \langle (T^2 + bT + cI)v, v\rangle \\
&= \langle T^2 v, v\rangle + b\langle Tv, v\rangle + c\langle v, v\rangle \\
&= \langle Tv, Tv\rangle + b\langle Tv, v\rangle + c\|v\|^2 \\
&\geq \|Tv\|^2 - |b|\|Tv\|\|v\| + c\|v\|^2 \text{ [Cauchy-Schwarz]} \\
&= \left(\|Tv\| - \frac{|b|\|v\|}{2}\right)^2 + \left(c - \frac{b^2}{4}\right)\|v\|^2 \\
&> 0
\end{aligned}$$

Thus $p(T)v \neq 0 \implies p(T)$ is injective, and hence invertible. □

END OF LECTURE 29

Beginning of Lecture 30

Proposition 59. *If $V \neq \{0\}$ and $T \in \mathcal{L}(V)$ is self-adjoint, then T has an eigenvalue (even if $\mathbb{F} = \mathbb{R}$).*

To prove this, we are going to need the following proposition from Chapter 4 on polynomials.

Proposition 60. *If $p \in \mathcal{P}(\mathbb{R})$ is a non-constant polynomial, then p has a unique factorization (except for re-ordering) of the form:*

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M),$$

where

$$\begin{aligned} m + M &\geq 1 \\ c &\in \mathbb{R}, c \neq 0 \\ \lambda_1, \ldots, \lambda_m &\in \mathbb{R} \\ b_1, \ldots, b_M &\in \mathbb{R} \\ c_1, \ldots, c_M &\in \mathbb{R} \\ b_j^2 &< 4c_j, \ \ \forall j \end{aligned}$$

Proof of Proposition 59. If $\mathbb{F} = \mathbb{C}$ then T has an eigenvalue even if it is not self-adjoint (recall this from Chapter 5), so we can assume that $\mathbb{F} = \mathbb{R}$.

Let $n = dim V$ and choose any $v \in V$ with $v \neq 0$; then:

$$v, Tv, T^2 v, \ldots, T^n v \text{ must be linearly dependent}$$

Thus there exists $a_0, \ldots, a_n \in \mathbb{R}$, not all zero, such that

$$0 = \sum_{k=0}^{n} a_k T^k v$$

Define $p(x) = a_0 + a_1 x + \cdots a_n x^n$. Then using Proposition 60 we have:

$$\begin{aligned} p(x) &= a_0 + a_1 x + \cdots + a_n x^n \\ &= c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M) \end{aligned}$$

Thus:

$$\begin{aligned} 0 &= a_0 v + a_1 Tv + \cdots a_n T^n v \\ &= (a_0 + a_1 T + \cdots a_n T^n) v \\ &= c(T - \lambda_1 I) \cdots (T - \lambda_m I)(T^2 + b_1 T + c_1 I) \cdots (T^2 + b_M T + c_M I) v \quad (17) \end{aligned}$$

By Proposition 58, each $T^2 + b_j T + c_j I$ is invertible. Also, $c \neq 0$, so $m > 0$ since otherwise the RHS of (17) would be an invertible (hence injective) operator acting a nonzero vector v, but the LHS is zero, and thus we would have a contradiction. Therefore,

$$\begin{aligned} 0 = (T - \lambda_1 I) \cdots (T - \lambda_m I) v &\implies T - \lambda_j I \text{ is not injective for some } j \\ &\implies \lambda_j \text{ is an eigenvalue of } T \end{aligned}$$

□

Proposition 61. *If $T \in \mathcal{L}(V)$ is self-adjoint and U is an invariant subspace of V under T, then:*

1. *$U^\perp$ is invariant under T*
2. *$T|_U \in \mathcal{L}(U)$ is self-adjoint*
3. *$T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint*

Proof. We prove each part:

1. Let $v \in U^\perp$ and $u \in U$; then:

$$\langle Tv, u \rangle = \langle v, \underbrace{Tu}_{\in U} \rangle = 0, \ \ \forall\, u \in U \implies Tv \in U^\perp$$

   Thus $U^\perp$ is invariant under T.

2. If $u, v \in U$, then:

$$\langle (T|_U) u, v \rangle = \langle Tu, v \rangle = \langle u, Tv \rangle = \langle u, (T|_U) v \rangle \implies (T|_U)^* = T|_U$$

3. Replace U with $U^\perp$ in the proof of #2. This is valid because we have already proved #1.

□

Theorem 29 (Real Spectral Theorem)**.** *Suppose $\mathbb{F} = \mathbb{R}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *T is self-adjoint*
2. *V has an ONB consisting of eigenvectors of T*
3. *T has a diagonal matrix with respect to some ONB of V*

Proof. We prove (1) $\implies$ (2) $\implies$ (3) $\implies$ (1) in parts:

- (2) $\implies$ (3) is clear
- (3) $\implies$ (1): Let $\mathcal{B}$ be an ONB such that $\mathcal{M}(T;\mathcal{B}) \in \mathbb{R}^{n,n}$ is diagonal. Then $\mathcal{M}(T^*;\mathcal{B}) = \mathcal{M}(T;\mathcal{B})^\dagger = \mathcal{M}(T;\mathcal{B})$, and so we must have $T^* = T$.
- (1) $\implies$ (2): Proof by induction on $\dim V$. For the base case, let $\dim V = 1$. Since T is guaranteed to have one eigenvalue, it has an eigenvector v and necessarily $\operatorname{span}(v) = V$.

  Now suppose that $\dim V = n > 1$ and that (1) $\implies$ (2) for all vector spaces U with $\dim U \leq n-1$ and all self-adjoint $S \in \mathcal{L}(U)$. Let $T \in \mathcal{L}(V)$ be self-adjoint. Let $u \in V$ be an eigenvector of T with $\|u\| = 1$, and let $U = \operatorname{span}(u)$. Then U is a 1-dimensional subspace of V that is invariant under T. Thus $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint.

  By the induction hypothesis, there exists an ONB $\mathcal{B}_\perp = u_1, \ldots, u_{n-1}$ of $U^\perp$ consisting of eigenvectors of $T|_{U^\perp}$. But then $\mathcal{B} = u_1, \ldots, u_{n-1}, u$ is an ONB of V consisting of eigenvectors of T.

□

End of Lecture 30

Beginning of Lecture 31

7.C Positive Operators and Isometries

Positive Operators

Definition 51. An operator $T \in \mathcal{L}(V)$ is positive if T is self-adjoint and

$$\forall\, v \in V,\ \ \langle Tv, v\rangle \geq 0.$$

Examples:

1. Orthogonal projections P_U (when U is a subspace of V)
2. T self-adjoint and $b, c \in \mathbb{R}$ such that $b^2 < 4c$, then $T^2 + bT + cI$ is a positive operator (see our proof proving that $T^2 + bT + cI$ is invertible)

Definition 52. An operator R is the square root of an operator T if $R^2 = T$.

Example: Suppose $T \in \mathcal{L}(\mathbb{R}^2)$ is a rotation by the angle $\theta \in [0, 2\pi)$, i.e.,

$$T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

If R is a rotation by $\theta/2$,

$$R = \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix}$$

then $R^2 = T$.

Positive operators mimic the numbers $[0, \infty)$. The next two theorems formalize this statement.

Theorem 30. *Let $T \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *T is positive*
2. *T is self-adjoint and all eigenvalues of T are nonnegative*
3. *T has a positive square root*

4. *T has a self-adjoint square root*

5. *There exists and operator $R \in \mathcal{L}(V)$ such that $T = R^*R$*

Proof. The plan is: (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5) $\Rightarrow$ (1).

- (1) $\Rightarrow$ (2): By definition T is self-adjoint. So let λ be an eigenvalue of T with eigenvector v (recall this means $v \neq 0$). Then:

$$0 \geq \langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle = \lambda\|v\|^2 \Rightarrow \lambda \geq 0$$

- (2) $\Rightarrow$ (3): Since T is self-adjoint, by The Spectral Theorem there is an ONB $e_1, \ldots, e_n$ of V consisting of eigenvectors of T; let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues. By assumption each $\lambda_k \geq 0$. Define $R \in \mathcal{L}(V)$ be defining it on $e_1, \ldots, e_n$:

$$Re_k = \sqrt{\lambda_k} e_k$$

We claim that R is a positive operator and that $R^2 = T$. The second point is clear since:

$$R^2 e_k = \lambda_k e_k = Te_k, \forall\, k = 1, \ldots, n$$

Thus R^2 and T agree on a basis and so they must be the same operator. Furthermore R is positive since:

$$\begin{aligned}
\langle Rv, v\rangle &= \left\langle R\left(\sum_{j=1}^n \langle v, e_j\rangle e_j\right), \sum_{k=1}^n \langle v, e_k\rangle e_k\right\rangle = \left\langle \sum_{j=1}^n \langle v, e_j\rangle Re_j, \sum_{k=1}^n \langle v, e_k\rangle e_k\right\rangle \\
&= \sum_{j=1}^n \sum_{k=1}^n \langle\langle v, e_j\rangle Re_j, \langle v, e_k\rangle e_k\rangle \\
&= \sum_{j=1}^n \sum_{k=1}^n \langle v, e_j\rangle \overline{\langle v, e_k\rangle} \langle Re_j, e_k\rangle \\
&= \sum_{j=1}^n \sum_{k=1}^n \langle v, e_j\rangle \overline{\langle v, e_k\rangle} \langle \sqrt{\lambda_j} e_j, e_k\rangle \\
&= \sum_{j=1}^n \sum_{k=1}^n \langle v, e_j\rangle \overline{\langle v, e_k\rangle} \sqrt{\lambda_j} \langle e_j, e_k\rangle \\
&= \sum_{j=1}^n \sqrt{\lambda_j} \cdot |\langle v, e_j\rangle|^2 \geq 0
\end{aligned}$$

- (3) $\Rightarrow$ (4): By definition
- (4) $\Rightarrow$ (5): (4) means that $T = R^2$ and $R = R^*$. Thus: $T = R^2 = RR = R^*R$.
- (5) $\Rightarrow$ (1): We need to show T is self-adjoint and $\langle Tv, v\rangle \geq 0$ for all $v \in V$. For the first part,

$$T^* = (R^*R)^* = R^*(R^*)^* = R^*R = T$$

For the second part,

$$\langle Tv, v\rangle = \langle R^*Rv, v\rangle = \langle Rv, Rv\rangle = \|Rv\|^2 \geq 0, \ \ \forall\, v \in V$$

□

Theorem 31. *Every positive operator has a unique positive square root.*

Proof. Suppose $T \in \mathcal{L}(V)$ is positive. Since T is self-adjoint, by The Spectral Theorem it has an ONB $\mathcal{B}$ of eigenvectors. Let $v \in \mathcal{B}$ be one of these eigenvectors, and let λ be its associated eigenvalue so that $Tv = \lambda v$. By the previous theorem $\lambda \geq 0$ and T has a positive square root, say R. We will prove that $Rv = \sqrt{\lambda}v$. Thus R will be uniquely determined on the basis $\mathcal{B}$, which means that it is the unique positive square root of T.

Now we prove that $Rv = \sqrt{\lambda}v$. Since R is positive, and hence self-adjoint, The Spectral Theorem impies that there exists an ONB $e_1, \ldots, e_n$ of V consisting of eigenvectors of R. Let $\eta_1, \ldots, \eta_n$ be the corresponding eigenvalues; because R is also positive, we know from the previous theorem that $\eta_k \geq 0$ for all k. Define $\lambda_k = \eta_k^2$; then $\sqrt{\lambda_k} = \eta_k$ and

$$Re_k = \sqrt{\lambda_k}e_k$$

Since $e_1, \ldots, e_n$ is an ONB, we can write

$$v = \sum_{k=1}^{n} \langle v, e_k\rangle e_k$$

Thus:

$$Rv = \sum_{k=1}^{n} \langle v, e_k\rangle \sqrt{\lambda_k}\, e_k \Longrightarrow R^2 v = \sum_{k=1}^{n} \langle v, e_k\rangle \lambda_k e_k$$

But $R^2 = T$ and $Tv = \lambda v$, so $R^2 v = Tv = \lambda v$ which implies:

$$\begin{aligned}\sum_{k=1}^{n}\langle v, e_k\rangle \lambda_k e_k = \sum_{k=1}^{n}\langle v, e_k\rangle \lambda e_k &\Longrightarrow \sum_{k=1}^{n}\langle v, e_k\rangle(\lambda - \lambda_k)e_k = 0 \\ &\Longrightarrow \langle v, e_k\rangle(\lambda - \lambda_k) = 0, \ \ \forall\, k = 1, \ldots, n\end{aligned}$$

Hence either $\langle v, e_k\rangle = 0$ or $\lambda = \lambda_k$ for each k; thus:

$$\begin{aligned}v = \sum_{\{k\,:\,\lambda_k = \lambda\}}\langle v, e_k\rangle e_k \Longrightarrow Rv &= \sum_{\{k\,:\,\lambda_k = \lambda\}}\langle v, e_k\rangle\sqrt{\lambda_k}e_k \\ &= \sqrt{\lambda}\sum_{\{k\,:\,\lambda_k = \lambda\}}\langle v, e_k\rangle e_k = \sqrt{\lambda}v\end{aligned}$$

□

End of Lecture 31

Beginning of Lecture 32

Isometries

Definition 53. An operator $S \in \mathcal{L}(V)$ is an isometry if

$$\forall\, v \in V, \quad \|Sv\| = \|v\|$$

Thus an isometry is an operator that preserves norms, or equivalently, preserves distances since the definition implies:

$$\forall\, u, w \in V, \quad \|S(u - w)\| = \|u - w\|$$

Example: Suppose $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ with $|\lambda_k| = 1$ for each k. Let $e_1, \ldots, e_n \in V$ be an ONB, and let $S \in \mathcal{L}(V)$ satisfy:

$$Se_k = \lambda_k e_k, \quad \forall\, k$$

Then we can show that S is an isometry.

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

$$v = \sum_{k=1}^{n} \langle v, e_k \rangle e_k$$
$$\|v\|^2 = \sum_{k=1}^{n} |\langle v, e_k \rangle|^2$$

Thus:

$$\begin{aligned} Sv &= \sum_{k=1}^{n} \langle v, e_k \rangle S e_k \\ &= \sum_{k=1}^{n} \lambda_k \langle v, e_k \rangle e_k \\ \Rightarrow \|Sv\|^2 &= \sum_{k=1}^{n} |\lambda_k|^2 |\langle v, e_k \rangle|^2 = \sum_{k=1}^{n} |\langle v, e_k \rangle|^2 = \|v\|^2 \end{aligned}$$

□

Theorem 32. *Suppose $S \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *S is an isometry*
2. $\langle Su, Sv \rangle = \langle u, v \rangle, \quad \forall\, u, v \in V$
3. *If $e_1, \ldots, e_m$ is orthonormal in V, then $Se_1, \ldots, Se_m$ is orthonormal.*
4. *There exists an ONB $e_1, \ldots, e_n$ such that $Se_1, \ldots, Se_n$ is orthonormal*
5. $S^*S = I$
6. $SS^* = I$
7. *S^* is an isometry*
8. *S is invertible and $S^{-1} = S^*$*

Proof. Proof in parts:

- (1) $\implies$ (2): If $\mathbb{F} = \mathbb{R}$ we use the formula:

  $$\langle u, v \rangle = \frac{1}{4}(\|u + v\|^2 - \|u - v\|^2)$$

  while if $\mathbb{F} = \mathbb{C}$ we use:

  $$\langle u, v \rangle = \frac{1}{4}(\|u + v\|^2 - \|u - v\|^2 + \|u + iv\|^2 i - \|u - iv\|^2 i)$$

  For example, if $\mathbb{F} = \mathbb{R}$ then:

  $$\begin{aligned}
  \langle Su, Sv \rangle &= \frac{1}{4}(\|Su + Sv\|^2 - \|Su - Sv\|^2) \\
  &= \frac{1}{4}(\|S(u + v)\|^2 - \|S(u - v)\|^2) \\
  &= \frac{1}{4}(\|u + v\|^2 - \|u - v\|^2) \\
  &= \langle u, v \rangle
  \end{aligned}$$

  $\mathbb{F} = \mathbb{C}$ is similar and you should verify it on your own.

- (2) $\implies$ (3): This one is easy since:

  $$\langle Se_j, Se_k \rangle = \langle e_j, e_k \rangle = \delta(j - k)$$

- (3) $\Longrightarrow$ (4): Obvious
- (4) $\Longrightarrow$ (5): First note we have:

  $$\langle e_j, e_k \rangle = \delta(j-k) = \langle Se_j, Se_k \rangle = \langle S^*Se_j, e_k \rangle, \quad \forall\, j,k$$

  Now:

  $$\begin{aligned}
  \langle S^*Su, v\rangle &= \left\langle S^*S\sum_{j=1}^{n} a_j e_j, \sum_{k=1}^{n} b_k e_k \right\rangle \\
  &= \left\langle \sum_{j=1}^{n} a_j S^*Se_j, \sum_{k=1}^{n} b_k e_k \right\rangle \\
  &= \sum_{j=1}^{n}\sum_{k=1}^{n} \langle a_j S^*Se_j, b_k e_k\rangle \\
  &= \sum_{j=1}^{n}\sum_{k=1}^{n} a_j \bar{b}_k \langle S^*Se_j, e_k\rangle \\
  &= \sum_{j=1}^{n}\sum_{k=1}^{n} a_j \bar{b}_k \langle e_j, e_k\rangle \\
  &= \langle u, v\rangle \quad \text{[just unwind what we did to get to previous line]}
  \end{aligned}$$

- (5) $\Longrightarrow$ (6): Since I is invertible, S^*S is invertible, which by 3.D #9 (HW 3) means that S^* and S are invertible. So we have:

  $$S^*S = I \Rightarrow S^*SS^{-1} = S^{-1} \Rightarrow SS^* = SS^{-1} = I$$

- (6) $\Longrightarrow$ (7): We compute:

  $$\|S^*v\|^2 = \langle S^*v, S^*v\rangle = \langle SS^*v, v\rangle = \langle v, v\rangle = \|v\|^2$$

- (7) $\Longrightarrow$ (8): We have proven already:

  $$S \text{ an isometry } \Longrightarrow S^*S = I \text{ and } SS^* = I$$

  Replacing S with S^* and using $(S^*)^* = S$ we obtain:

  $$S^* \text{ an isometry } \Longrightarrow SS^* = I \text{ and } S^*S = I$$

  Therefore, by definition S is invertible with $S^{-1} = S^*$.

- (8) $\implies$ (1): $S^{-1} = S^*$ implies $S^*S = I$. Thus:

$$\|Sv\|^2 = \langle Sv, Sv\rangle = \langle S^*Sv, v\rangle = \langle v, v\rangle = \|v\|^2$$

□

Theorem 33 (Spectral Theorem for Isometries when $\mathbb{F} = \mathbb{C}$)**.** *Suppose* $\mathbb{F} = \mathbb{C}$ *and* $S \in \mathcal{L}(V)$*. Then the following are equivalent:*

1. *S is an isometry*
2. *There is an ONB of V consisting of eigenvectors of S whose corresponding eigenvalues all have complex modulus equal to* 1*.*

Proof. The example earlier proved (2) $\implies$ (1). So we just have to prove (1) $\implies$ (2). Since S is an isometry, the previous theorem implies $S^*S = I = SS^*$ which means that S is normal. Therefore we can apply the Complex Spectral Theorem. Thus there exists an ONB $e_1, \ldots, e_n$ consisting of eigenvectors of S; let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues. Then:

$$|\lambda_k| = |\lambda_k|\|e_k\| = \|\lambda_k e_k\| = \|Se_k\| = \|e_k\| = 1$$

□

End of Lecture 32

Beginning of Lecture 33

Rigid Motions in $\mathbb{R}^n$

Definition 54. A rigid motion in an inner product space V is a transformation $f : V \to V$ that preserves distances, i.e.,

$$\forall\, u, v \in V, \quad \|f(u) - f(v)\| = \|u - v\|$$

Note, f is not assumed to be linear!

Examples:

- Any isometry $S \in \mathcal{L}(V)$ is a rigid motion.
- The translation map is a rigid motion, i.e., fix $w \in V$. Then the map:

  $$f(v) = v + w$$

  is a rigid motion, since

  $$\|f(u) - f(v)\| = \|u + w - (v + w)\| = \|u + w - v - w\| = \|u - v\|$$

  Note that translations are not linear unless $w = 0$. Indeed, $f(0) = w$, and we know that all linear maps take 0 to 0.

We are now going to prove that *every* rigid motion on a *real* inner product space is the composition of an isometry and a translation.

Theorem 34. *Let $f : V \to V$ be a rigid motion on a real inner product space V, and let $S(v) = f(v) - f(0)$. Then S is an isometry.*

Note the above theorem implies that $f(v) = S(v) + f(0)$, i.e., f can be decomposed as an isometry plus a translation.

To prove the theorem, we need the following lemma:

Lemma 3. *Let f be a rigid motion on a real inner product space V, and let $S(v) = f(v) - f(0)$. Then:*

1. $\|S(v)\| = \|v\|, \ \forall\, v \in V$
2. $\|S(u) - S(v)\| = \|u - v\|, \ \forall\, u, v \in V$

3. $\langle S(u), S(v)\rangle = \langle u, v\rangle, \ \ \forall\, u, v \in V$

Proof. We prove each statement individually:

1. Notice that:

$$\|S(v)\| = \|f(v) - f(0)\| = \|v - 0\| = \|v\|$$

2. A similar calculation yields:

$$\begin{aligned}\|S(u) - S(v)\| &= \|(f(u) - f(0)) - (f(v) - f(0))\| \\ &= \|f(u) - f(0) - f(v) + f(0)\| \\ &= \|f(u) - f(v)\| \\ &= \|u - v\|\end{aligned}$$

3. Using part 1 and the fact that we are working in a real inner product space we have:

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

On the other hand, using part 2 we also have:

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

Therefore,

$$\|u\|^2 + \|v\|^2 - 2\langle S(u), S(v)\rangle = \|u\|^2 + \|v\|^2 - 2\langle u, v\rangle \Rightarrow \langle S(u), S(v)\rangle = \langle u, v\rangle$$

□

Now we can prove our theorem:

Proof of Theorem 34. By Lemma 3, part 1 we know that $\|S(v)\| = \|v\|$ for all $v \in V$. So S preserves the norm, but now we need to show that S is linear.

Let $e_1, \ldots, e_n$ be an ONB of V, and define:

$$\forall\, k = 1, \ldots, n, \quad g_k = S(e_k)$$

Note that $g_1, \ldots, g_n$ is in fact an ONB too. Indeed, using Lemma 3, part 3,

$$\langle g_j, g_k \rangle = \langle S(e_j), S(e_k) \rangle = \langle e_j, e_k \rangle = \delta(j-k)$$

Let $v \in V$ and write v as a linear combination of the ONB $e_1, \ldots, e_n$:

$$v = \sum_{k=1}^{n} a_k e_k, \quad a_k = \langle v, e_k \rangle$$

Do the same for $S(v)$ in the ONB $g_1, \ldots, g_n$:

$$S(v) = \sum_{k=1}^{n} b_k g_k, \quad b_k = \langle S(v), g_k \rangle$$

Using Lemma 3, part 3 again:

$$b_k = \langle S(v), g_k \rangle = \langle S(v), S(e_k) \rangle = \langle v, e_k \rangle = a_k$$

Therefore,

$$S\left(\sum_{k=1}^{n} a_k e_k\right) = \sum_{k=1}^{n} a_k g_k = \sum_{k=1}^{n} a_k S(e_k)$$

Thus S is a linear map (recall the proof of 3.5 in the book). □

This concludes the material covered on the second midterm. In particular the second midterm will cover Chapter 6, sections 7.A, 7.B, 7.C from Chapter 7, and this material on rigid motions.

End of Lecture 33

Beginning of Lecture 34

7.D Polar Decomposition and Singular Value Decomposition

Polar Decomposition

First we recall the polar form of a complex number $z \in \mathbb{C}$. Let $z = x + iy$. Every complex number z can also be written in polar form as:

$$z = re^{i\theta}, \quad r \geq 0, \ \ \theta \in [0, 2\pi),$$

where

$$\begin{aligned} r &= x^2 + y^2 \\ x &= r\cos\theta \\ y &= r\sin\theta \end{aligned}$$

Note that in the polar formulation, r is nonnegative (and positive if $z \neq 0$), and $|e^{i\theta}| = 1$. We are going prove a polar decomposition for operators $T \in \mathcal{L}(V)$, using the analogy:

$$\begin{aligned} r \geq 0 &\longleftrightarrow \text{ positive operators} \\ e^{i\theta} &\longleftrightarrow \text{ isometries} \end{aligned}$$

First, recall that $R \in \mathcal{L}(V)$ is a square root of $T \in \mathcal{L}(V)$ if $R^2 = T$, and if T is a positive operator then T has a unique positive square root.

Notation: If T is a positive operator, let $\sqrt{T}$ denote the unique positive square root of T.

Second, recall from HW9, 7.C, #4, that for any $T \in \mathcal{L}(V, W)$, $T^*T \in \mathcal{L}(V)$ and $TT^* \in \mathcal{L}(W)$ are positive operators, and thus have unique positive square roots.

Theorem 35 (Polar Decomposition Theorem). *If $T \in \mathcal{L}(V)$, then there exists an isometry $S \in \mathcal{L}(V)$ such that*

$$T = S\sqrt{T^*T}$$

Proof. Define a function $S_1 : \operatorname{range}\sqrt{T^*T} \to \operatorname{range} T$ as:

$$S_1(\sqrt{T^*T}v) = Tv$$

An outline of the proof is:

1. Show S_1 is well defined and linear
2. Extend S_1 to an isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$.

To show S_1 is well defined, first we show:

$$\|Tv\| = \|\sqrt{T^*T}v\|, \quad \forall\, v \in V \tag{18}$$

Indeed,

$$\begin{aligned}
\|Tv\|^2 &= \langle Tv, Tv\rangle \\
&= \langle T^*Tv, v\rangle \\
&= \langle \sqrt{T^*T}\sqrt{T^*T}v, v\rangle \\
&= \langle \sqrt{T^*T}v, \sqrt{T^*T}v\rangle \ \text{[recall positive operators are self-adjoint]} \\
&= \|\sqrt{T^*T}v\|^2
\end{aligned}$$

Now to show S_1 is well defined, suppose that $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2$; we must show that $Tv_1 = Tv_2$. We have:

$$\begin{aligned}
\|Tv_1 - Tv_2\| &= \|T(v_1 - v_2)\| \\
&= \|\sqrt{T^*T}(v_1 - v_2)\| \ \text{[by (18)]} \\
&= \|\sqrt{T^*T}v_1 - \sqrt{T^*T}v_2\| \\
&= 0
\end{aligned}$$

You should verify on your own that S_1 is linear. That completes part 1.

Now we extend $S_1 \in \mathcal{L}(\operatorname{range}\sqrt{T^*T}, \operatorname{range} T)$ so an isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$. First note that by (18) and the definition of S_1, we have:

$$\|S_1u\| = \|u\|, \quad \forall\, u \in \operatorname{range}\sqrt{T^*T}$$

Note this implies that S_1 is injective since only 0 maps to 0. Thus:

$$\dim \operatorname{range}\sqrt{T^*T} = \dim \operatorname{range} T \Rightarrow \dim(\operatorname{range}\sqrt{T^*T})^\perp = \dim(\operatorname{range} T)^\perp$$

Let $e_1, \ldots, e_m$ be an ONB for $(\text{range}\sqrt{T^*T})^\perp$ and let $f_1, \ldots, f_m$ be an ONB for $(\text{range}\, T)^\perp$. Note, both ONBs have the same length! Define a second linear map $S_2 : (\text{range}\sqrt{T^*T})^\perp \to (\text{range}\, T)^\perp$ by

$$S_2\left(\underbrace{\sum_{k=1}^{m} a_k e_k}_{w}\right) = \underbrace{\sum_{k=1}^{m} a_k f_k}_{S_2 w}$$

Since $e_1, \ldots, e_m$ and $f_1, \ldots, f_m$ are ONBs, by the definition of S_2 it is clear that

$$\forall\, w \in (\text{range}\sqrt{T^*T})^\perp, \quad \|S_2 w\| = \|w\|$$

Since:

$$V = \text{range}\sqrt{T^*T} \oplus (\text{range}\sqrt{T^*T})^\perp$$

we have for each $v \in V$,

$$v = u + w, \quad u \in \text{range}\sqrt{T^*T}, \;\; w \in (\text{range}\sqrt{T^*T})^\perp \;\; [u, w \text{ are unique}]$$

Thus we can define $S \in \mathcal{L}(V)$ as:

$$S(v) = S(u + w) = S_1 u + S_2 w$$

Then for each $v \in V$,

$$S\sqrt{T^*T}v = S(\sqrt{T^*T}v) = S(\sqrt{T^*T}v + 0) = S_1(\sqrt{T^*T}v) + S_2(0) = Tv + 0 = Tv$$

and so $T = S\sqrt{T^*T}$.

Finally, we need to show that S is an isometry, i.e., it preserves norms:

$$\|Sv\|^2 = \|S_1 u + S_2 w\|^2 = \|S_1 u\|^2 + \|S_2 w\|^2 = \|u\|^2 + \|w\|^2 = \|v\|^2$$

$\square$

Thus we can decompose *any* operator T into two very nice operators: an isometry and a positive operator. Furthermore, when $\mathbb{F} = \mathbb{C}$, our Spectral Theory tells us that there exists an ONB $\mathcal{B}_1$ such that $\mathcal{M}(S; \mathcal{B}_1)$ is diagonal and another ONB $\mathcal{B}_2$ such that $\mathcal{M}(\sqrt{T^*T}; \mathcal{B}_2)$ is diagonal. Unfortunately in general $\mathcal{B}_1 \neq \mathcal{B}_2$!

End of Lecture 34

Lecture 35: Midterm 2

Lecture 36: Discussion of Rigid Motion Practice Problems

Beginning of Lecture 37

Singular Value Decomposition

Let V, W be finite dimensional inner product spaces over the field $\mathbb{F}$ with $\dim V = n$ and $\dim W = m$.

Definition 55. Suppose $T \in \mathcal{L}(V, W)$. The Hermitian square of T is $T^*T \in \mathcal{L}(V)$.

Proposition 62. *Suppose $T \in \mathcal{L}(V, W)$. Its Hermitian square $T^*T \in \mathcal{L}(V)$ is a positive operator.*

Proof. Need to show T^*T is self-adjoint and $\langle T^*Tv, v\rangle \geq 0$ for all $v \in V$.

- Self adjoint:
$$(T^*T)^* = T^*(T^*)^* = T^*T$$
- Nonnegative:
$$\langle T^*Tv, v\rangle = \langle Tv, Tv\rangle = \|Tv\|^2 \geq 0$$

□

Recall that for any $T \in \mathcal{L}(V, W)$, $T^*T \in \mathcal{L}(V)$ is a positive operator, and thus has a unique positive squareroot $\sqrt{T^*T}$. We are going to call

$$|T| := \sqrt{T^*T}$$

the modulus of T. The modulus of T shows how "big" the operator T is:

Proposition 63. *For any $T \in \mathcal{L}(V, W)$,*

$$\||T|v\|_V = \|Tv\|_W, \quad \forall\, v \in V$$

Proof. For any $v \in V$,

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

□

Recall our polar decomposition. In this language we have for any $T \in \mathcal{L}(V)$,

$$T = S\sqrt{T^*T} = S|T|, \quad S \text{ is an isometry}$$

We are now going to go down a different path with $|T|$.

Remark: Since T^*T is positive, the Spectral Theorem implies that V has an ONB $e_1, \ldots, e_n$ consisting of eigenvectors of T^*T with:

$$T^*Te_k = \lambda_k e_k, \quad \lambda_k \geq 0 \ \forall\, k$$

Then $e_1, \ldots, e_n$ are eigenvectors of $|T| = \sqrt{T^*T}$ with corresponding eigenvalues $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$.

Definition 56. Suppose $T \in \mathcal{L}(V, W)$. The singular values of T are the eigenvalues of $|T|$, i.e., if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of T^*T, then $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$ are the singular values of T.

Remark: Every $T \in \mathcal{L}(V, W)$ has $n = \dim V$ singular values.

Let $T \in \mathcal{L}(V, W)$ and let $\sigma_1, \ldots, \sigma_n$ be the singular values of T (counting multiplicities). Assume also that $\sigma_1, \ldots, \sigma_r$ are the non-zero singular values of T (also counting multiplicities), so that in particular $\sigma_k = 0$ for $k > r$.

By definition, the numbers $\sigma_1^2, \ldots, \sigma_n^2$ are eigenvalues of T^*T. Let $e_1, \ldots, e_n$ be an ONB of V consisting of eigenvectors of T^*T so that

$$T^*Te_k = \sigma_k^2 e_k, \quad \forall\, k = 1, \ldots, n \tag{19}$$

Proposition 64. *The system*

$$f_k := \frac{1}{\sigma_k} Te_k \in W, \quad k = 1, \ldots, r, \tag{20}$$

is an orthonormal system.

Proof. Observe:

$$\langle Te_j, Te_k \rangle = \langle T^*Te_j, e_k \rangle = \langle \sigma_k^2 e_j, e_k \rangle = \sigma_k^2 \langle e_j, e_k \rangle = \begin{cases} 0 & \text{if } j \neq k \\ \sigma_j^2 & \text{if } j = k \end{cases}$$

□

Theorem 36 (Singular Value Decomposition). *Suppose $T \in \mathcal{L}(V,W)$ and let $\sigma_1, \ldots, \sigma_n$ be the singular values of T. Let $e_1, \ldots, e_n \in V$ be an ONB consisting of eigenvectors of T^*T satisfying* (19)*, and let $f_1, \ldots, f_r \in W$ be the orthonormal system defined by* (20)*. Then:*

$$Tv = \sum_{k=1}^{r} \sigma_k \langle v, e_k \rangle f_k, \quad \forall\, v \in V$$

Proof. Define $S \in \mathcal{L}(V,W)$ as:

$$Sv := \sum_{k=1}^{r} \sigma_k \langle v, e_k \rangle f_k$$

We are going to show $S = T$ by showing that S and T agree on a basis of V. In particular, take $e_1, \ldots, e_n$ which is an ONB for V. Then:

- For $j = 1, \ldots, r$,

$$Se_j = \sum_{k=1}^{k} \sigma_k \langle e_j, e_k \rangle f_k = \sigma_j \langle e_j, e_j \rangle f_j = \sigma_j \|e_j\|^2 f_j = \sigma_j f_j = Te_j$$

- For $j > r$,

$$Se_j = \sum_{k=1}^{r} \sigma_k \langle e_j, e_k \rangle f_k = 0 = Te_j$$

  where the last inequality $Te_j = 0$ follows since:

$$\|Te_j\| = \||T|e_j\| = \|\sigma_j e_j\| = 0$$

□

Recall that for a general linear map $T \in \mathcal{L}(V,W)$, we used one basis $\mathcal{B}_V = v_1, \ldots, v_n$ for V and another basis $\mathcal{B}_W = w_1, \ldots, w_m$ for W to construct the matrix $A = \mathcal{M}(T; \mathcal{B}_V, \mathcal{B}_W)$ which has entries defined as:

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j$$

Take $\mathcal{B}_V = e_1, \ldots, e_n$ (as above), and extend $f_1, \ldots, f_r$ to an ONB $\mathcal{B}_W = f_1, \ldots, f_r, f_{r+1}, \ldots, f_m$ of W. The Singular Value Decomposition (SVD) says that $A = \mathcal{M}(T; \mathcal{B}_V, \mathcal{B}_W)$ has a "diagonal structure", i.e.,

$$\begin{aligned} Te_k &= \sum_{j=1}^{m} A_{j,k} f_k \\ &= \sum_{j=1}^{r} \sigma_j \langle e_k, e_j \rangle f_j = \sigma_k f_k \text{ [since } \sigma_k = 0 \text{ for } k > r] \\ &\implies A_{j,k} = \begin{cases} \sigma_k & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \end{aligned}$$

When $V = W$ this means $\mathcal{M}(T; (e_1, \ldots, e_n), (f_1, \ldots, f_n))$ is diagonal!

End of Lecture 37

Beginning of Lecture 38

Applications of the Singular Value Decomposition

Let $T \in \mathcal{L}(V, W)$ and recall the singular value decomposition of T:

- $\sigma_1, \ldots, \sigma_n$ the singular values of T with $\sigma_1, \ldots, \sigma_r$ the nonzero singular values.
- $e_1, \ldots, e_n$ an ONB of V consisting of eigenvectors of T^*T with $T^*Te_k = \sigma_k^2 e_k$
- $f_1, \ldots, f_r \in W$ orthonormal and defined as $f_k := (1/\sigma_k)Te_k$.

Then:

$$Tv = \sum_{k=1}^{r} \sigma_k \langle v, e_k \rangle f_k, \quad \forall\, v \in V$$

If $T \in \mathcal{L}(V)$ then $f_1, \ldots, f_r \in V$. In this case:

$$T^m v = \sum_{k=1}^{r} \sigma_k \langle v, e_k \rangle T^{m-1} f_k \neq \sum_{k=1}^{r} \sigma_k^m \langle v, e_k \rangle f_k$$

unlike an ONB $g_1, \ldots, g_n$ of eigenvectors of T in which

$$v = \sum_{k=1}^{n} \langle v, g_k \rangle g_k \Longrightarrow T^m v = \sum_{k=1}^{n} \langle v, g_k \rangle T^m g_k = \sum_{k=1}^{n} \lambda_k^m \langle v, g_k \rangle g_k$$

Nevertheless the SVD is still very useful! Indeed, the SVD tells us a lot about the “metric properties” of a linear transformation.

Computational Remark: The SVD requires finding the eigenvalues and eigenvectors of T^*T. In general computing eigenvalues of an operator (matrix) is hard, but for self-adjoint operators (like T^*T) there are algorithms that can do it very effectively. This will be good to keep in mind, even though we will not say any more on the subject.

Application 1: Image of the unit ball

Let $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and let

$$B = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$$

be the closed unit ball. We want to describe the shape of B after it is transformed by T, i.e., we want to know what $T(B)$ looks like.

Suppose first that $T \in \mathcal{L}(\mathbb{R}^n)$ and T takes a diagonal form, i.e., $\tilde{e}_1, \dots, \tilde{e}_n$ with

$$\tilde{e}_k = (0, \dots, 0, \underbrace{1}_{k}, 0, \dots, 0)$$

are eigenvectors of T with eigenvalues $\sigma_1, \dots, \sigma_n$, each $\sigma_k > 0$, so that in particular for any $v = (v_1, \dots, v_n) \in \mathbb{R}^n$,

$$Tx = T(x_1, \dots, x_n) = (\sigma_1 x_1, \dots, \sigma_n x_n)$$

Therefore, if $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, then

$$y = (y_1, \dots, y_n) = T(x_1, \dots, x_n) = Tx, \quad \text{for } x \in B$$

if and only if

$$\sum_{k=1}^{n} \frac{y_k^2}{\sigma_k^2} \leq 1 \tag{21}$$

Indeed, $T(x_1, \dots, x_n) = (\sigma_1 x_1, \dots, \sigma_n x_n) = (y_1, \dots, y_n)$ if and only if $y_k = \sigma_k x_k$, or equivalently $x_k = y_k/\sigma_k$. But then:

$$x \in B \Leftrightarrow \|x\| \leq 1 \Leftrightarrow \|x\|^2 \leq 1 \Leftrightarrow \sum_{k=1}^{n} x_k^2 \leq 1 \Leftrightarrow \sum_{k=1}^{n} \frac{y_k^2}{\sigma_k^2} \leq 1$$

The set of points satisfying (21) is called an ellipsoid. In $\mathbb{R}^2$ it is an ellipse (with its interior) with half-axes σ_1 and σ_2 [draw a picture]. The vectors $\tilde{e}_1, \dots, \tilde{e}_n$ defined above as the standard ONB of $\mathbb{R}^n$ are the principal axes.

Now consider $T \in \mathcal{L}(\mathbb{R}^n)$ with singular values $\sigma_1, \dots, \sigma_n$ and $\sigma_k > 0$ for each $k = 1, \dots, n$. Let $e_1, \dots, e_n \in \mathbb{R}^n$ and $f_1, \dots, f_n \in \mathbb{R}^n$ be the two ONBs associated to the SVD of T so that

$$Tx = \sum_{k=1}^{n} \sigma_k \langle x, e_k \rangle f_k, \quad \forall\, x \in \mathbb{R}^n$$

Note that:

$$x \in B \Leftrightarrow \|x\| \leq 1 \Leftrightarrow \|x\|^2 \leq 1 \Leftrightarrow \sum_{k=1}^{n} |\langle x, e_k \rangle|^2 \leq 1$$

We also have:

$$\begin{aligned} y = Tx, \text{ for } x \in B &\Leftrightarrow \sum_{k=1}^{n} \langle y, f_k \rangle f_k = \sum_{k=1}^{n} \sigma_k \langle x, e_k \rangle f_k, \text{ for } x \in B \\ &\Leftrightarrow \langle y, f_k \rangle = \sigma_k \langle x, e_k \rangle, \text{ for } x \in B, \ \forall\, k = 1, \ldots, n \\ &\Leftrightarrow \sum_{k=1}^{n} \frac{|\langle y, f_k \rangle|^2}{\sigma_k^2} \leq 1 \end{aligned}$$

But that is also an ellipsoid! It is just rotated so that its principal axes are $f_1, \ldots, f_n$.

Now consider the fully general case of $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with nonzero singular values $\sigma_1, \ldots, \sigma_r$. We have:

$$Tx = \sum_{k=1}^{r} \sigma_k \langle x, e_k \rangle f_k$$

Notice this implies that $\operatorname{range} T = \operatorname{span}(f_1, \ldots, f_r)$. In particular,

$$y = Tx \text{ for } x \in \mathbb{R}^n \Leftrightarrow y \in \operatorname{range} T \Leftrightarrow y \in \operatorname{span}(f_1, \ldots, f_r)$$

Now we can use essentially the same calculation as before to get:

$$\begin{aligned} y = Tx, \text{ for } x \in B &\Leftrightarrow \sum_{k=1}^{r} \langle y, f_k \rangle f_k = \sum_{k=1}^{r} \sigma_k \langle x, e_k \rangle f_k, \text{ for } x \in B \\ &\Leftrightarrow \langle y, f_k \rangle = \sigma_k \langle x, e_k \rangle, \text{ for } x \in B, \ \forall\, k = 1, \ldots, r \\ &\Leftrightarrow \sum_{k=1}^{r} \frac{|\langle y, f_k \rangle|^2}{\sigma_k^2} \leq 1 \end{aligned}$$

Thus we have shown:

Theorem 37. *The image $T(B)$ of the closed unit ball B is an ellipsoid in* $\operatorname{range} T$ *with half axes $\sigma_1, \ldots, \sigma_r$ along the principal axes $f_1, \ldots, f_r$, where $\sigma_1, \ldots, \sigma_r$ are the nonzero singular values of T and $f_k := (1/\sigma_k) T e_k$ with $e_1, \ldots, e_n$ an ONB of V consisting of eigenvectors of T^*T.*

End of Lecture 38

Beginning of Lecture 39

Application 2: Operator norm of a linear transformation

Let $T \in \mathcal{L}(V, W)$ and recall that $B := \{v \in V : \|v\| \leq 1\}$ is the closed unit ball. Consider the following optimization problem:

$$\max_{v \in B} \|Tv\|$$

Let's first consider a positive operator $T \in \mathcal{L}(V)$. Suppose $e_1, \ldots, e_n$ is an ONB for V consisting of eigenvectors of T with eigenvalues $\sigma_1, \ldots, \sigma_n \geq 0$. Note that since T is positive, T is self-adjoint, and so $|T| = \sqrt{T^*T} = \sqrt{T^2} = T$. Therefore the singular values of T are also $\sigma_1, \ldots, \sigma_n$. Let σ_1 the largest singular value. We are going to show that:

$$\sigma_1 = \max_{v \in B} \|Tv\| \tag{22}$$

Note that:

$$v = \sum_{k=1}^{n} \langle v, e_k \rangle e_k$$
$$\Rightarrow Tv = \sum_{k=1}^{n} \langle v, e_k \rangle T e_k = \sum_{k=1}^{n} \sigma_k \langle v, e_k \rangle e_k$$

Thus for any $v \in V$, and in particular $v \in B$,

$$\|Tv\|^2 = \sum_{k=1}^{n} \sigma_k^2 |\langle v, e_k \rangle|^2 \leq \sigma_1^2 \sum_{k=1}^{n} |\langle v, e_k \rangle|^2 = \sigma_1^2 \|v\|^2$$
$$\Rightarrow \|Tv\|^2 \leq \sigma_1^2 \|v\|^2$$

On the other hand:

$$\|Te_1\| = \|\sigma_1 e_1\| = \sigma_1 \|e_1\|$$

Thus we have shown (22).

Now let $T \in \mathcal{L}(V, W)$ and let $\sigma_1, \ldots, \sigma_n$ be the singular values of T, with $\sigma_1, \ldots, \sigma_r$ nonzero and σ_1 the largest singular value. We will use the singular value decomposition of T:

$$Tv = \sum_{k=1}^{r} \sigma_k \langle v, e_k \rangle f_k$$

For any $v \in V$, and in particular $v \in B$,

$$\|Tv\|^2 = \sum_{k=1}^{r} \sigma_k^2 |\langle v, e_k\rangle|^2 \leq \sigma_1^2 \sum_{k=1}^{n} |\langle v, e_k\rangle|^2 = \sigma_1^2 \|v\|^2$$

Additionally,

$$\|Te_1\| = \|\sigma_1 f_1\| = \sigma_1 \|f_1\| = \sigma_1 \|e_1\|$$

where the last inequality follows because $e_1, \dots, e_n$ and $f_1, \dots, f_r$ are both orthonormal, and hence $\|e_1\| = \|f_1\| = 1$. Therefore in the general case as well we see that

$$\sigma_1 = \max_{v \in B} \|Tv\| \tag{23}$$

Definition 57. The quantity

$$\|T\| := \max\{\|Tv\| : v \in V,\ \|v\| \leq 1\}$$

is the operator norm of T.

It is easy to see that $\|T\|$ is indeed a norm on $\mathcal{L}(V, W)$, meaning that:

- $\|\lambda T\| = |\lambda| \|T\|$ for all $\lambda \in \mathbb{F}$
- $\|T + S\| \leq \|T\| + \|S\|$ for all $S, T \in \mathcal{L}(V, W)$
- $\|T\| \geq 0$ for all $T \in \mathcal{L}(V, W)$
- $\|T\| = 0 \iff T = 0$

One of the main properties of the operator norm is:

$$\|Tv\| \leq \|T\| \|v\|$$

We have shown if σ_1 is the largest singular value of T, then $\|T\| = \sigma_1$. One can also show $\|T\| = C_0$, where $C_0 \in \mathbb{R}$ is defined as:

$$C_0 = \min\{C \in \mathbb{R} : \|Tv\| \leq C\|v\|,\ \ \forall\, v \in V\}$$

Other equivalent formulations of the operator norm are:

$$\|T\| = \max\{\|Tv\| : v \in V,\ \|v\| = 1\} = \max\left\{\frac{\|Tv\|}{\|v\|} : v \in V,\ v \neq 0\right\}$$

End of Lecture 39

Beginning of Lecture 40

Singular Value Decomposition of a matrix

Suppose $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is defined as:

$$Tx = Ax, \quad \forall\, x \in \mathbb{R}^n,$$

where $A \in \mathbb{R}^{m,n}$. So in particular, $\mathcal{M}(T) = A$ in the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, and we can identify T with the matrix A. We know that the singular value decomposition of T is:

$$Tx = \sum_{k=1}^{r} \sigma_k \langle x, e_k \rangle f_k$$

where $\sigma_1, \ldots, \sigma_1$ are the nonzero singular values of T, $e_1, \ldots, e_r \in \mathbb{R}^n$ are orthonormal, and $f_1, \ldots, f_r \in \mathbb{R}^m$ are orthonormal. We can rewrite the SVD in terms of matrices, which gives the SVD decomposition of the matrix A. Consider any $x \in \mathbb{R}^n$ as an $n \times 1$ vector (similarly $y \in \mathbb{R}^m$ is an $m \times 1$ vector). Define:

$$\begin{aligned}
\widetilde{\Sigma} &= \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r,r} \\
\widetilde{B} &= (e_1, \ldots, e_r) \in \mathbb{R}^{n,r} \\
\widetilde{C} &= (f_1, \ldots, f_r) \in \mathbb{R}^{m,r}
\end{aligned}$$

Then:

$$Ax = Tx = \sum_{k=1}^{r} \sigma_k \langle x, e_k \rangle f_k \iff A = \widetilde{C}\widetilde{\Sigma}\widetilde{B}^\dagger$$

The representation $A = \widetilde{C}\widetilde{\Sigma}\widetilde{B}^\dagger$ is called the <u>compact SVD</u> for A.

We can also compute a (standard) SVD representation of A as:

$$A = C\Sigma B^\dagger$$

where $\Sigma \in \mathbb{R}^{m,n}$, $B \in \mathbb{R}^{n,n}$ and $C \in \mathbb{R}^{m,m}$ with $S \in \mathcal{L}(\mathbb{R}^n)$, $Sx := Bx$ and $R \in \mathcal{L}(\mathbb{R}^m)$, $Ry := Cy$, both being isometries. The matrix Σ is simply the "diagonal" extension of $\widetilde{\Sigma}$:

$$\Sigma_{j,k} = \begin{cases} \sigma_k & j = k \leq r \\ 0 & \text{otherwise} \end{cases}$$

To define B, extend $e_1, \ldots, e_r$ to an ONB $e_1, \ldots, e_n$ for $\mathbb{R}^n$. Define B as:

$$B = (e_1, \ldots, e_n) \in \mathbb{R}^{n,n}$$

Similarly, extend $f_1, \ldots, f_r$ to an ONB $f_1, \ldots, f_m$ for $\mathbb{R}^m$ and define C as:

$$C = (f_1, \ldots, f_m) \in \mathbb{R}^{m,m}$$

Note that by definition the columns of B and C or orthonormal bases, and $\mathcal{M}(S) = B$ and $\mathcal{M}(R) = C$, so by Theorem 7.42 in the book, S and R are isometries.

Application 3: Condition number of a matrix

Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is defined as:

$$Tx = Ax, \quad \forall\, x \in \mathbb{R}^n,$$

where $A \in \mathbb{R}^{n,n}$. So in particular, $\mathcal{M}(T) = A$ in the standard basis, and we can identify T with the matrix A. Suppose additionally that A (and hence T) is invertible. Now suppose that we want to solve:

$$Ax = b$$

for some $b \in \mathbb{R}^n$. The solution is clearly:

$$x = A^{-1}b$$

However, as happens in "real life", we may only know the data approximately or round off errors during computations on a computer may occur, which distort the data. We consider a model in which b is only approximately known, so instead of having $Ax = b$ we are solving:

$$A\tilde{x} = b + \Delta b,$$

where Δb is a small perturbation of b, i.e.,

$$\|\Delta b\| < \epsilon \|b\|, \quad \epsilon \ll 1$$

The solution $\tilde{x}$ is approximately x; indeed:

$$\tilde{x} = A^{-1}b + A^{-1}\Delta b = x + \Delta x, \quad \text{where } \Delta x = A^{-1}\Delta b$$

We want to know how big the relative error $\|\Delta x\|/\|x\|$ in the solution $\tilde{x}$ is in comparison with the relative error $\|\Delta b\|/\|b\|$ of the initial data. Note that:

$$\begin{aligned}
\frac{\|\Delta x\|}{\|x\|} &= \frac{\|A^{-1}\Delta b\|}{\|x\|} \\
&= \frac{\|A^{-1}\Delta b\|}{\|b\|}\frac{\|b\|}{\|x\|} \\
&= \frac{\|A^{-1}\Delta b\|}{\|b\|}\frac{\|Ax\|}{\|x\|} \\
&\leq \frac{\|A^{-1}\|\|\Delta b\|}{\|b\|}\frac{\|A\|\|x\|}{\|x\|} \\
&\leq \|A\|\|A^{-1}\|\frac{\|\Delta b\|}{\|b\|}
\end{aligned}$$

The quantity $\|A\|\|A^{-1}\|$ is the condition number of A. It estimates how the relative error in the solution x depends on the relative error of the initial data b.

We can relate the condition number of A to its singular values. Let $\sigma_1, \ldots, \sigma_n$ be the singular values of A. Assume they are ordered so that:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$$

Note that $\sigma_n > 0$ since A is invertible and:

$$A = C\Sigma B^\dagger$$

where B and C are isometries and $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$. Thus B and C are invertible with

$$B^{-1} = B^\dagger, \;\; C^{-1} = C^\dagger \quad [\text{recall } \mathcal{M}(T^*) = \mathcal{M}(T)^\dagger]$$

Thus $\Sigma = C^{-1}A(B^\dagger)^{-1} = C^\dagger AB$ must also be invertible and:

$$A^{-1} = (C\Sigma B^\dagger)^{-1} = (B^\dagger)^{-1}\Sigma^{-1}C^{-1} = B\Sigma^{-1}C^\dagger$$

Note that $\Sigma^{-1} = \operatorname{diag}(1/\sigma_1, \ldots, 1/\sigma_n)$ and so the singular values of A^{-1} are

$$\frac{1}{\sigma_n} \geq \frac{1}{\sigma_{n-1}} \geq \cdots \geq \frac{1}{\sigma_1} > 0$$

We know that $\|A\| = \sigma_1$ and by the calculation we just completed $\|A^{-1}\| = 1/\sigma_n$. Therefore the condition number of A is:

$$\|A\|\|A^{-1}\| = \frac{\sigma_1}{\sigma_n}$$

A matrix is <u>well conditioned</u> if its condition number is not too large (the closer to one the better) or <u>ill conditioned</u> if its condition number is too large.

END OF LECTURE 40

Beginning of Lecture 41

Example: Let's do an example to show the problems that can occur with an ill conditioned matrix. Consider the system of equations:

$$\left.\begin{array}{rcl} x_1 + x_2 & = & 2 \\ x_1 + 1.001x_2 & = & 2 \end{array}\right\} \Longleftrightarrow \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 1.001 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 2 \\ 2 \end{pmatrix}}_{b}$$

The solution is:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} = A^{-1}b$$

Now let's consider the same system but with a small perturbation

$$\Delta b = \begin{pmatrix} 0 \\ 0.001 \end{pmatrix}$$

With the perturbation, the system now is:

$$\left.\begin{array}{rcl} \tilde{x}_1 + \tilde{x}_2 & = & 2 \\ \tilde{x}_1 + 1.001\tilde{x}_2 & = & 2.001 \end{array}\right\} \Longleftrightarrow \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 1.001 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix}}_{\tilde{x}} = \underbrace{\begin{pmatrix} 2 \\ 2 \end{pmatrix}}_{b} + \underbrace{\begin{pmatrix} 0 \\ 0.001 \end{pmatrix}}_{\Delta b}$$

The singular values of A are approximately $\sigma_1 \approx 2.0005$ and $\sigma_2 \approx 0.0005$. Thus the condition number of A is approximately (!):

$$\|A\|\|A^{-1}\| = \frac{\sigma_1}{\sigma_2} \approx \frac{2.0005}{0.0005} \approx 4000$$

The new solution is easily seen to be:

$$\tilde{x} = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 2 \\ 0 \end{pmatrix}}_{x=A^{-1}b} + \underbrace{\begin{pmatrix} -1 \\ 1 \end{pmatrix}}_{\Delta x = A^{-1}\Delta b}$$

which is completely different than x. Notice:

$$\text{Ratio of initial data perturbation} = \frac{\|\Delta b\|}{\|b\|} = \frac{\sqrt{8}}{0.001} \approx 0.00035$$

$$\text{Ratio of solution perturbation} = \frac{\|\Delta x\|}{\|x\|} = \frac{\sqrt{2}}{2} \approx 0.7$$

8 Bilinear and Quadratic Forms

Bilinear Forms

Definition 58. Let V and W be vector spaces over a field $\mathbb{F}$. The product $V \times W$ is defined as:

$$V \times W := \{(v, w) : v \in V,\, w \in W\}$$

Proposition 65. *Let V and W be vector spaces over a field $\mathbb{F}$. Then $V \times W$ is a vector space over $\mathbb{F}$ with vector addition and scalar multiplication defined as:*

- *Vector addition:* $(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2)$
- *Scalar multiplication:* $\lambda(v, w) = (\lambda v, \lambda w)$

Definition 59. Let V be a vector space over a field $\mathbb{F}$. A bilinear form on V is a function $L : V \times V \to \mathbb{F}$ that is linear in both arguments:

$$L(\alpha u + \beta v, w) = \alpha L(u, w) + \beta L(v, w), \quad \forall\, u, v, w \in V,\ \forall\, \alpha, \beta \in \mathbb{F}$$
$$L(u, \alpha v + \beta w) = \alpha L(u, v) + \beta L(u, w), \quad \forall\, u, v, w \in V,\ \forall\, \alpha, \beta \in \mathbb{F}$$

Examples:

1. Let $\varphi_1, \varphi_2 \in \mathcal{L}(V, \mathbb{F})$ be linear functionals on V. Define $L : V \times V \to \mathbb{F}$ as:
$$L(u, v) = \varphi_1(u)\varphi_2(v)$$
Then L is a bilinear form (as you can verify).

2. Let V be an inner product space over $\mathbb{R}$ and let $T \in \mathcal{L}(V)$. Then:
$$L(u, v) = \langle Tu, v\rangle$$
is a bilinear form. In fact every bilinear form on a real inner product space is of this form.

Theorem 38. *Let V be an inner product space over $\mathbb{R}$, and let $L : V \times V \to \mathbb{R}$ be a bilinear form on V. Then there exists a unique $T \in \mathcal{L}(V)$ such that*

$$L(u, v) = \langle Tu, v\rangle$$

Proof. Let $\mathcal{B} = e_1, \ldots, e_n$ be an ONB for V. Then:

$$u = \sum_{j=1}^{n} a_j e_j \quad \text{and} \quad v = \sum_{k=1}^{n} b_k e_k$$

Then:

$$\begin{aligned} L(u, v) &= L\left(\sum_{j=1}^{n} a_j e_j, v\right) \\ &= \sum_{j=1}^{n} a_j L(e_j, v) \\ &= \sum_{j=1}^{n} a_j L\left(e_j, \sum_{k=1}^{n} b_k e_k\right) \\ &= \sum_{j=1}^{n} a_j b_k L(e_j, e_k) \end{aligned}$$

Define $A \in \mathbb{R}^{n,n}$ as:

$$A_{k,j} = L(e_j, e_k)$$

Since $\mathcal{M}(\cdot; \mathcal{B}) : \mathcal{L}(V) \to \mathbb{R}^{n,n}$ is an isomorphism, there exists a unique $T \in \mathcal{L}(V)$ such that

$$\mathcal{M}(T; \mathcal{B}) = A$$

Note in particular, this means that

$$Te_j = \sum_{k=1}^{n} A_{k,j} e_k$$

We then have:

$$\begin{aligned}
\langle Tu, v\rangle &= \left\langle T\left(\sum_{j=1}^{n} a_j e_j\right), v\right\rangle \\
&= \sum_{j=1}^{n} a_j \langle Te_j, v\rangle \\
&= \sum_{j=1}^{n} a_j \Big\langle \sum_{k=1}^{n} A_{k,j} e_k, v\Big\rangle \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} a_j A_{k,j} \langle e_k, v\rangle \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} a_j A_{k,j} \Big\langle e_k, \sum_{l=1}^{n} b_l e_l\Big\rangle \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} a_j b_l A_{k,j} \langle e_k e_l\rangle \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} a_j b_k A_{k,j} \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} a_j b_k L(e_j, e_k) = L(u, v)
\end{aligned}$$

□

End of Lecture 41

Beginning of Lecture 42

Quadratic forms

Definition 60. Let V be a real inner product space. A quadratic form $Q : V \to \mathbb{R}$ is the "diagonal" of a bilinear form $L : V \times V \to \mathbb{R}$, i.e.,

$$Q(v) = L(v, v), \quad \text{for some bilinear form } L$$

Note by function of the form:

$$Q(v) = \langle Tv, v \rangle, \quad T \in \mathcal{L}(V)$$

Note by our previous theorem,

$$Q(v) = \langle Tv, v \rangle, \quad \text{for some } T \in \mathcal{L}(V)$$

Unlike bilinear forms, quadratic forms are not uniquely determined by $T \in \mathcal{L}(V)$ (we'll give an example in a bit). However, if restrict ourselves to self-adjoint T, then T is unique.

Proposition 66. *Let V be a finite dimensional real inner product space, and suppose $Q : V \to \mathbb{R}$ is a quadratic form on V. Then there exists a unique self-adjoint $T \in \mathcal{L}(V)$ such that*

$$Q(v) = \langle Tv, v \rangle, \quad T = T^*$$

Proof. We know that every quadratic form can be represented as $Q(v) = \langle \widetilde{T}v, v \rangle$ for some $\widetilde{T} \in \mathcal{L}(V)$. Define $T \in \mathcal{L}(V)$ as:

$$Tv = \frac{1}{2}(\widetilde{T} + \widetilde{T}^*)$$

Note that T is self-adjoint; indeed:

$$T^* = \left(\frac{1}{2}(\widetilde{T} + \widetilde{T}^*)\right)^* = \frac{1}{2}(\widetilde{T}^* + \widetilde{T}) = T$$

Additionally,

$$\begin{aligned}
\langle Tv, v\rangle &= \langle \frac{1}{2}(\widetilde{T} + \widetilde{T}^*)v, v\rangle \\
&= \frac{1}{2}\langle \widetilde{T}v + \widetilde{T}^*v, v\rangle \\
&= \frac{1}{2}\langle \widetilde{T}v, v\rangle + \frac{1}{2}\langle \widetilde{T}^*v, v\rangle \\
&= \frac{1}{2}\langle \widetilde{T}v, v\rangle + \frac{1}{2}\langle v, \widetilde{T}v\rangle \\
&= \frac{1}{2}\langle \widetilde{T}v, v\rangle + \frac{1}{2}\langle \widetilde{T}v, v\rangle \\
&= \langle \widetilde{T}v, v\rangle = Q(v)
\end{aligned}$$

Thus there exists a self-adjoint $T \in \mathcal{L}(V)$ such that $Q(v) = \langle Tv, v\rangle$.

We now show that T is unique. Suppose $S \in \mathcal{L}(V)$ is another self-adjoint operator such that $Q(v) = \langle Sv, v\rangle$. Then:

$$\begin{aligned}
\langle Tv, v\rangle &= \langle Sv, v\rangle, \quad \forall\, v \in V \\
\Rightarrow \langle (T - S)v, v\rangle &= 0, \qquad\quad \forall\, v \in V
\end{aligned}$$

But

$$(T - S)^* = T^* - S^* = T - S$$

so $T - S$ is self-adjoint. But by 7.16 in the book, this means $T - S = 0$, i.e., $T = S$. □

Quadratic Forms on $\mathbb{R}^n$

On $\mathbb{R}^n$, we can identify operators $T \in \mathcal{L}(\mathbb{R}^n)$ with their matrix $A = \mathcal{M}(T) \in \mathbb{R}^{n,n}$ in the standard basis. This means that quadratic forms $Q : \mathbb{R}^n \to \mathbb{R}$ can be written as:

$$Q(x) = \langle Ax, x\rangle = \sum_{j=1}^{n} \sum_{k=1}^{n} A_{j,k} x_j x_k, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Thus quadratic forms on $\mathbb{R}^n$ are homogeneous polynomials of degree 2, that is $Q \in \mathcal{P}_2(\mathbb{R}^n)$ and Q only contains terms of the form $a_{j,k} x_j x_k$.

Quadratic forms on $\mathbb{R}^n$ are uniquely represented by symmetric matrices, since $A = \mathcal{M}(T)$ is symmetric when $T = T^*$. There are though an infinite number of non-symmetric matrices that give the same quadratic form. For example, consider $Q : \mathbb{R}^2 \to \mathbb{R}$ defined as:

$$Q(x) = x_1^2 + x_2^2 - 4x_1x_2$$

Then

$$Q(x) = \langle Ax, x \rangle, \quad \text{for any } A = \begin{pmatrix} 1 & a-4 \\ -a & 1 \end{pmatrix}, \ a \in \mathbb{R}$$

Quadratic forms Q such that:

$$Q(x) = \sum_{k=1}^{n} a_k x_k^2$$

are particularly nice. Since $Q(x) = \langle Ax, x \rangle$, the nice form corresponds to A being a diagonal matrix wth $a_k = A_{k,k}$. In general though $A = \mathcal{M}(T)$ is not diagonal. Let $S \in \mathcal{L}(\mathbb{R}^n)$ be an isometry, and let $x = Sy$. Then:

$$Q(x) = Q(Sy) = \langle TSy, Sy \rangle = \langle S^*TSy, y \rangle$$

So we want an isometry S such that $\mathcal{M}(S^*TS)$ is diagonal. But since T is self-adjoint, by the Spectral Theorem there exists an ONB $e_1, \ldots, e_n$ of $\mathbb{R}^n$ consisting of eigenvectors of T. Define S so that:

$$\mathcal{M}(S) = \begin{pmatrix} | & & | \\ e_1 & \cdots & e_n \\ | & & | \end{pmatrix}$$

End of Lecture 42