## Contents

**Introduction**  
0.1. Goals and course plan (Lecture 1)  
0.2. Review  
0.3. Euler’s Theorem  

Chapter 1. Constructions  
1.1. Direct sum, direct product (Lectures 2-4)  
1.2. Quotients (Lecture 5)  
1.3. Hom spaces and duality (Lectures 6-8)  
1.4. Multilinear algebra and tensor products  

Chapter 2. Structure Theory  
2.1. Introduction (Lecture 15)  
2.2. Jordan Canonical Form  

Chapter 3. Vector and matrix norms  
3.1. Review of metric spaces  
3.2. Norms on vector spaces  
3.3. Norms on matrices  
3.4. Example: eigenvalues and the power method (Lecture, 17/)  
3.5. Sequences and series of vectors and matrices  

Chapter 4. The Holomorphic Calculus  
4.1. The exponential series (24/3/2014)  
4.2. 26/3/2014  
4.3. Invertibility and the resolvent (31/3/2014)  

Chapter 5. Vignettes  
5.1. The exponential map and structure theory for GLₙ(ℝ) (2/4/2014)  
5.2. Representation Theory of Groups  

Bibliography
Introduction

Lior Silberman, lior@Math.UBC.CA, http://www.math.ubc.ca/~lior/
Office: Math Building 229B
Phone: 604-827-3031

For administrative details see the syllabus.

0.1. Goals and course plan (Lecture 1)

<table>
<thead>
<tr>
<th>Math</th>
<th>Metamath</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reinforce basics</td>
<td>Abstraction</td>
<td>Hard problems, Multiple ideas</td>
</tr>
<tr>
<td>• Vector space, subspace</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Linear independence, basis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Linear map</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Eigenvalue, eigenvector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New ideas</td>
<td>Constructions, Universal properties</td>
<td>Abstract examples</td>
</tr>
<tr>
<td>• Direct sum, product</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Hom($U, V$) and duality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Quotients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Tensor products</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure theory for linear maps</td>
<td>Canonical forms, Algorithms</td>
<td></td>
</tr>
<tr>
<td>• Matrix decompositions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• $LU, LL^\dagger$ and Computation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Minimal poly, Cayley–Hamilton</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Jordan canonical form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Combining fields</td>
<td></td>
</tr>
<tr>
<td>• Norms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Holomorphic calculus ($e^X$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
0.2. Review

0.2.1. Basic definitions. We want to give ourselves the freedom to have scalars other than real or complex.

**Definition 1 (Fields).** A field is a quintuple \( (F, 0, 1, +, \cdot) \) such that \( (F, 0, +) \) and \( (F \setminus \{0\}, 1, \cdot) \) are abelian groups, and the distributive law \( \forall x, y, z \in F : x(y+z) = xy + xz \) holds.

**Example 2.** \( \mathbb{R}, \mathbb{C}, \mathbb{Q}, \mathbb{F}_2 \) (via addition and multiplication tables; ex: show this is a field), \( \mathbb{F}_p \).

**Exercise 3.** Every finite field has \( p^r \) elements for some prime \( p \) and some integer \( r \geq 1 \). Fact: there is one such field for every prime power.

**Definition 4.** A vector space over a field \( F \) is a quadruple \( (V, 0, +, \cdot) \) where \( (V, 0, +) \) is an abelian group, and \( \cdot : F \times V \to V \) is a map such that:

1. \( 1_F \cdot v = v \).
2. \( \alpha (\beta v) = (\alpha \beta) v \).
3. \( (\alpha + \beta)(v+w) = \alpha v + \beta v + \alpha w + \beta w \).

**Lemma 5.** \( 0_F \cdot v = 0 \) for all \( v \).

**Proof.** \( 0_F = (0+0) v = 0 v + 0 v \). Now subtract \( 0 v \) from both sides. \( \square \)

0.2.2. Bases and dimension. Fix a vector space \( V \).

**Definition 6.** Let \( S \subset V \).

- \( v \in V \) depends on \( S \) if there are \( \{v_i\}_{i=1}^r \subset S \) and \( \{a_i\}_{i=1}^r \subset F \) such that \( v = \sum_{i=1}^r a_i v_i \) [empty sum is \( 0 \)]
- Write \( \text{Span}_F(S) \subset V \) for the set of vectors that depend on \( S \).
- Call \( S \) linearly dependent if some \( v \in S \) depends on \( S \setminus \{v\} \), equivalently if there are distinct \( \{v_i\}_{i=1}^r \subset S \) and \( \{a_i\}_{i=1}^r \subset F \) not all zero such that \( \sum_{i=1}^r a_i v_i = 0 \).
- Call \( S \) linearly independent if it is not linearly dependent.

**Axiom 7 (Axiom of choice).** Every vector space has a basis.

0.2.3. Examples. \( \{0\}, \mathbb{R}^n, F^X \).
0.3. Euler’s Theorem

Let $G = (V, E)$ be a connected planar graph. A face of $G$ is a finite connected component of $\mathbb{R}^2 \setminus G$.

**Theorem 8 (Euler).** $v - e + f = 1$.

**Proof.** Arbitrarily orient the edges. Let $\partial_E : \mathbb{R}^E \rightarrow \mathbb{R}^V$ be defined by $f((u, v)) = 1_v - 1_u$, $\partial_F : \mathbb{R}^F \rightarrow \mathbb{R}^E$ be given by the sum of edges around the face.

**Lemma 9.** $\partial_F$ is injective.

**Proof.** Faces containing boundary edges are independent. Remove them and repeat. \[\square\]

**Lemma 10.** $\text{Ker} \partial_E = \text{Im} \partial_F$.

**Proof.** Suppose a combo of edges is in the kernel. Following a sequence with non-zero coefficients gives a closed loop, which can be expressed as a sum of faces. Now subtract a multiple to reduce the number of edges with non-zero coefficients. \[\square\]

**Lemma 11.** $\text{Im}(\partial_E)$ is the set of functions with total weight zero.

**Proof.** Clearly the image is contained there. Conversely, given $f$ of total weight zero move the weight to a single vertex using elements of the image. [remark: quotient vector spaces] \[\square\]

Now $\dim \mathbb{R}^E = \dim \text{Ker} \partial_E + \dim \text{Im} \partial_E = \dim \text{Im} \partial_F + \dim \text{Im} \partial_E$ so $e = f + (v - 1)$.

**Remark 12.** Using $\mathbb{F}_2$ coefficients is even simpler.
Fix a field $F$.

1.1. Direct sum, direct product (Lectures 2-4)

1.1.1. Simplest case (Lecture 2). 

**Construction 13** (External direct sum). Let $U, V$ be vector spaces. Their direct sum, denoted $U \oplus V$, is the vector space whose underlying set is $U \times V$, with coordinate-wise addition and scalar multiplication.

**Lemma 14.** This really is a vector space.

**Remark 15.** The Lemma serves to review the definition of vector space.

**Proof.** Every property follows from the respective properties of $U, V$. □

**Remark 16.** Direct products of groups are discussed in 322.

**Lemma 17.** $\dim_F (U \oplus V) = \dim_F U + \dim_F V$.

**Remark 18.** This Lemma serves to review the notion of basis.

**Proof.** Let $B_U, B_V$ be bases of $U, V$ respectively. Then $\{ (u, 0) \mid u \in B_U \} \cup \{ (0, v) \mid v \in B_V \}$ is a basis of $U \oplus V$. □

**Example 19.** $\mathbb{R}^n \oplus \mathbb{R}^m \cong \mathbb{R}^{n+m}$.

1.1.2. Internal sum and direct sum (Lecture 3). A key situation is when $U, V$ are subspaces of an “ambient” vector space $W$.

**Lemma 20.** Let $W$ be a vector space, $U, V \subset W$. Then $\text{Span}_F (U \cup V) = \{ u + v \mid u \in U, v \in V \}$.

**Proof.** RHS contained in the span by definition. It is a subspace (non-empty, closed under addition and scalar multiplication) which contains $U, V$ hence contains the span. □

**Definition 21.** The space in the previous lemma is called the sum of $U, V$ and denoted $U + V$.

**Lemma 22.** Let $U, V \subset W$. There is a unique homomorphism $U \oplus V \to U + V$ which is the identity on $U, V$.

**Proof.** Define $f((u, v)) = u + v$. Check that this is a linear map. □

**Proposition 23** (Dimension of sums). $\dim_F (U + V) = \dim_F U + \dim_F V - \dim_F (U \cap V)$.

**Proof.** Consider the map $f$ of Lemma 22. It is surjective by Lemma 20. Moreover $\ker f = \{ (u, v) \in U \oplus V \mid u + v = 0 \}$, that is

$$\ker f = \{ (w, -w) \mid w \in U \cap V \} \cong U \cap V.$$  

Since $\dim_F \ker f + \dim_F \text{Im} f = \dim (U \oplus V)$ the claim now follows from Lemma 17. □
REMARK 24. This was a review of that formula. Alternative proof by starting from a basis of
$U \cap V$ and extending to bases of $U, V$, which is basically revisiting the proof of the formula.

DEFINITION 25 (Internal direct sum). We say the sum is direct if $f$ is an isomorphism.

THEOREM 26. For subspaces $U, V \subset W$ TFAE

(1) The sum $U + V$ is direct and equals $W$;
(2) $U + V = W$ and $U \cap V = \{0\}$
(3) Every vector $w \in W$ can be uniquely written in the form $w = u + vv$.

PROOF. (1) ⇒ (2): $U + V = W$ by assumption, $U \cap V = \text{Ker } f$.
(2) ⇒ (3): the first assumption gives existence, the second uniqueness.
(3) ⇒ (1): existence says $f$ is surjective, uniqueness says $f$ is injective. □

1.1.3. Finite direct sums (Lecture 4). Three possible notions: $(U \oplus V) \oplus W, U \oplus (V \oplus W)$,
vector space structure on $U \times V \times W$. These are all the same. Not just isomorphic (that is, not
just same dimension), but also isomorphic when considering the extra structure of the copies of
$U, V, W$. How do we express this?

DEFINITION 27. $W$ is the internal direct sum of its subspaces $\{V_i\}_{i \in I}$ if it spanned by them
and each vector has a unique representation as a sum of elements of $V_i$ (either as a finite sum of
non-zero vectors or as a zero-extended sum).

REMARK 28. This generalizes the notion of “linear independence” from vectors to subspaces.

LEMMA 29. Each of the three candidates contains an embedded copy of $U, V, W$ and is the
internal direct sum of the three images.

PROOF. Easy. □

PROPOSITION 30. Let $A, B$ each be the internal direct sum of embedded copies of $U, V, W$.
Then there is a unique isomorphism $A \to B$ respecting this structure.

PROOF. Construct. □

REMARK 31. (1) Proof only used result of Lemma, not specific structure; but (2) proof implicitly
relies on isomorphism to $U \times V \times W$; (3) We used the fact that a map can be defined using
values on copies of $U, V, W$ (4) Exactly same proof as the facts that a function on 3d space can be
defined on bases, and that that all 3d spaces are isomorphic.

\begin{itemize}
  \item Dimension by induction.
\end{itemize}

DEFINITION 32. Abstract arbitrary direct sum.

\begin{itemize}
  \item Block diagonality.
  \item Block upper-triangularity. [first structural result].
\end{itemize}

1.2. Quotients (Lecture 5)

Recall that for a group $G$ and a normal subgroup $N$, we can endow the quotient $G/N$ with group
structure $(gN)(hN) = (gh)N$.

\begin{itemize}
  \item This is well-defined, gives group.
  \item Have quotient map $q$: $G \to G/N$ given by $g \mapsto gN$.
\end{itemize}
• Homomorphism theorem: any \( f : G \to H \) factors as \( G \to G/\ker(f) \) follows by isomorphism.
• If \( N < M < G \) with both \( N,M \) normal then \( q(M) \simeq M/N \) is normal in \( G/N \) and \( (G/N)/\langle M/N \rangle \simeq (G/M) \).

Now do the same for vector spaces.

**Lemma 33.** Let \( V \) be a vector space, \( W \) a subspace. Let \( \pi : V \to V/W \) be the quotient as abelian groups. Then there is a unique vector space structure on \( V/W \) making \( \pi \) a surjective linear map.

**Proof.** We must set \( \alpha (v + W) = \alpha v + W \). This is well-defined and gives the isomorphism. Use quotient map to verify vector space axioms. \( \square \)

**Example 34.** \( V = C^1(0,1) \), space of continuously differentiable functions. Then \( \frac{d}{dx} : V \to C(0,1) \) vanishes on \( \mathbb{R}1 \) and hence induces a map \( \frac{d}{dx} : (C^1(0,1)/\mathbb{R}1) \to C(0,1) \). Note that the inverse of this map is what we call “indefinite integral” – whose images is exactly an equivalence class “function+c”.

**Example 35.** The definite integral is a linear function which vanishes on functions which are non-zero at countably many points. More generally, integration works on functions modulo functions which are zero a.e.

**Fact 36.** The properties above persist for vector spaces.

• How to use quotients: “kill off” part of the vector spaces that is irrelevant (linear maps vanish there) or already understood.

### 1.3. Hom spaces and duality (Lectures 6-8)

#### 1.3.1. Hom spaces (Lecture 5 continued or start of lecture 6).

**Definition 37.** \( \text{Hom}_F(U,V) \) will denote the space of \( F \)-linear maps \( U \to V \).

**Lemma 38.** \( \text{Hom}_F(U,V) \subset V^U \) is a subspace, hence a vector space.

**Definition 39.** \( V' = \text{Hom}_F(V,F) \) is called the dual space.

Motivation 1: in PDE. Want solutions in some function space \( V \). Use that \( V' \) is much bigger to find solutions in \( V' \), then show they are represented by functions.

#### 1.3.2. The dual space, finite dimensions.

**Note 40.** In lecture ignore infinite dimensions (but make statements which are correct in general).

**Construction 41 (Dual basis).** Let \( B = \{ b_i \}_{i \in I} \subset V \) be a basis. Write \( v \in V \) uniquely as \( v = \sum_{i \in I} a_i b_i \) (almost all \( a_i = 0 \)) and set \( \varphi_i(v) = a_i \).

**Lemma 42.** These are linear functionals.

**Proof.** Represent \( \alpha v + v' \) in the basis. \( \square \)

**Example 43.** \( V = \mathbb{F}^n \) with standard basis, get \( \varphi_i(\vec{x}) = x_i \). Note every functional has the form \( \varphi(\vec{x}) = \sum_{i=1}^n \varphi(\vec{e}_i) \varphi_i(\vec{x}) \).
Remark 44. Alternative construction: $\varphi_i$ is the unique linear map to $F$ satisfying $\varphi_i(b_j) = \delta_{i,j}$.

Lemma 45. The dual basis is linearly independent. It is spanning iff $\dim_F V < \infty$.

Proof. Evaluate a linear combination at $b_j$.
If $V$ is finite-dimensional, enumerate the basis as $\{b_i\}_{i=1}^n$. Then for any $\varphi \in V'$ and any $v \in V$ write $v = \sum_i a_i b_i$ and then
$$\varphi(v) = \sum_i a_i \varphi(b_i) = \sum_i (\varphi(b_i)) \varphi_i(v)$$
so
$$\varphi = \sum_i (\varphi(b_i)) \varphi_i \in \text{Span}_F \{\varphi_i\}.$$  

In the infinite-dimensional case let $\phi = \sum_{i \in I} \varphi_i$. Then $\phi$ is a well-defined linear functional which depends on every coordinate hence not in the span of the $\{\varphi_i\}$. □

Remark 46. This isomorphism $V \rightarrow V'$ is not canonical: the functional $\varphi_i$ depends on the whole basis $B$ and not only on $b_i$, and the dual basis transforms differently from the original basis under change-of-basis.

The argument above used evaluation – let’s investigate that more.

Proposition 47 (Double dual). Given $v \in V$ consider the evaluation map $e_v : V' \rightarrow F$ given by $e_v(\varphi) = \varphi(v)$. Then $v \mapsto e_v$ is a linear injection $V \hookrightarrow V''$, an isomorphism iff $V$ is finite-dimensional.

Proof. The vector space structure on $V'$ (and on $F^V$ in general) is such that $e_v$ is linear. That the map $v \mapsto e_v$ is linear follows from the linearity of the elements of $V'$. For injectivity let $v \in V$ be non-zero. Extending $v$ to a basis, let $\varphi_v$ be the element of the dual such that $\varphi_v(v) = 1$. Then $e_v(\varphi_v) \neq 0$ so $e_v \neq 0$. If $\dim_F V = n$ then $\dim_F V' = n$ and thus $\dim_F V'' = n$ and we have an isomorphism. □

The map $V \hookrightarrow V''$ is natural: the image $e_v$ of $v$ is intrinsic and does not depend on a choice of basis.

1.3.3. The dual space, infinite dimensions (Lecture 7).

Lemma 48 (Interaction with past constructions). We have

1. $(V/U)' \hookrightarrow V'$ as $\{\varphi \in V' \mid \varphi(U) = \{0\}\}$.
2. $(U \oplus V)' \simeq U' \oplus V'$.

Proof. Universal property. □

Corollary 49. Since $(F)' \simeq F$, it follows by induction that $(F^n)' \simeq F^n$.

What about infinite sums?

- The universal property gives a bijection $(\bigoplus_{i \in I} V_i)' \leftrightarrow \bigotimes_{i \in I} V_i'$, more generally
$$\text{Hom}_F \left( \bigoplus_{i \in I} V_i, Z \right) \leftrightarrow \bigotimes_{i \in I} \text{Hom}_i(V_i, Z).$$
  - LHS has a vector space structure – should get one on the right.
  - Which leads us to observe that:
Any Cartesian product $\times_{i \in I} W_i$ has a natural vector space structure, coming from pointwise addition and scalar multiplication.

- Note that the underlying set is

$$\times_{i \in I} W_i = \{ f \mid f \text{ is a function with domain } I \text{ and } \forall i \in I : f(i) \in W_i \}$$

$$= \left\{ f : I \to \bigcup_{i \in I} W_i \mid f(i) \in W_i \right\}.$$  

* RMK: AC means Cartesian products nonempty, but our sets have a distinguished element so this is not an issue.

- Define $\alpha(w_i)_{i \in I} \defeq (\alpha w_i + w_i')_{i \in I}$. This gives a vector space structure.

- Denote the resulting vector space $\prod_{i \in I} W_i$ and called it the direct product of the $W_i$.

- The bijection $(\bigoplus_{i \in I} V_i') \leftrightarrow \prod_{i \in I} V_i'$ is now a linear isomorphism [in fact, the vector space structure on the right is the one transported by the isomorphism].

We now investigate $\prod_i W_i$ in general.

- Note that it contains a copy of each $W_i$ (map $w \in W_i$ to the sequence which has $w$ in the $i$th position, and 0 at every other position).

- And these copies are linearly independent: if a sum of such vectors from distinct $W_i$ is zero, then every coordinate was zero.

- Thus $\prod_i W_i$ contains $\bigoplus_{i \in I} W_i$ as an internal direct sum.

- This subspace is exactly the subset $\{ w \in \prod_i W_i \mid \text{supp}(w) \text{ is finite} \}$.

- And in fact, that subspace proves that $\bigoplus_{i \in I} W_i$ exists.

- But $\prod_i W_i$ contains many other vectors – it is much bigger.

**Example 50.** $\mathbb{R}^{\oplus \mathbb{N}} \subset \mathbb{R}^{\mathbb{N}}$ – and the latter is the dual!

**Corollary 51.** The dual of an infinite-dimensional space is much bigger than the sum of the duals, and the double dual is bigger yet.

**1.3.4. Question:** so we only have finite sums in linear algebra. What about infinite sums?

Answer: no infinite sums in algebra. Definition of $\sum_{n=1}^{\infty} a_n = A$ from real analysis relies on analytic properties of $A$ (close to partial sums), not algebraic properties.

But, calculating sums can be understood in terms of linear functionals.

**Lemma 52 (Results from Calc II, reinterpreted).** Let $S \subset \mathbb{R}^N$ denote the set of sequences $a$ such that $\sum_{n=1}^{\infty} a_n$ converges.

1. $\mathbb{R}^{\oplus \mathbb{N}} \subset S \subset \mathbb{R}^{\mathbb{N}}$ is a linear subspace.

2. $\Sigma : S \to \mathbb{R}$ given by $\Sigma(a) = \sum_{n=1}^{\infty} a_n$ is a linear functionals.

**Philosophy:** Calc I,II made element-by-element statements, but using linear algebra we can express them as statements on the whole space.

Now questions about summing are questions about intelligently extending the linear functional $\Sigma$ to a bigger subspace. BUT: if an extension is to satisfy every property of summing series, it is actually the trivial (no) extension.

For more information let’s talk about limits of sequences instead (once we can generalize limits just apply that to partial sums of a series).
Definition 53. Let $c \subseteq \ell^\infty \subset \mathbb{R}^\mathbb{N}$ be the sets of convergent, respectively bounded sequences.

Lemma 54. $c \subseteq \ell^\infty$ are subspaces, and $\lim_{n \to \infty} : c \to \mathbb{R}$ is a linear functional.

Example 55. Let $C : \mathbb{R}^\mathbb{N} \to \mathbb{R}^\mathbb{N}$ be the Cesàro map $(Ca)_N = \frac{1}{N} \sum_{n=1}^N a_n$. This is clearly linear. Let $CS = C^{-1}(c)$ be the set of sequences which are Cesàro-convergent, and set $L \in CS'$ by $L(a) = \lim_{n \to \infty}(Ca)$. This is clearly linear (composition of linear maps). For example, the sequence $(0, 1, 0, 1, \cdots)$ now has the limit $\frac{1}{2}$.

Lemma 56. If $a \in c$ then $Ca \in c$ and they have the same limit. Thus $L$ above is an extension of $\lim_{n \to \infty}$.

Theorem 57. There are two functionals $\text{LIM, lim}_\omega \in (\ell^\infty)'$ (“Banach limit”, “limit along ultrafilter”, respectively) such that:

1. They are positive (map non-negative sequences to non-negative sequences);
2. Agree with $\lim_{n \to \infty}$ on $c$;
3. And, in addition
   - $\text{LIM} \circ S = \text{LIM}$ where $S : \ell^\infty \to \ell^\infty$ is the shift.
   - $\lim_\omega (a_n b_n) = (\lim_\omega a_n) (\lim_\omega b_n)$.

1.3.5. Pairings and bilinear forms (Lecture 8). Goal: identify the dual of a vector space in concrete terms. For this we need an abstract notion of dual not tied to the particular realization $V'$ (similar to how we have an abstract notion of direct sum as “space generated by independent copies of $V_i$” which is not tied to the concrete realization as the subspace of tuples of finite support in $\prod_i V_i$).

Observation 58. The evaluation map $V \times V' \to F$ given by
\[
\langle v, \varphi \rangle = \varphi(v)
\]
is bilinear (=linear in each variable).

Note that linearity in the first variable is equivalent to the linearity of $\varphi$, while linearity in the second variable is equivalent to the definition of the vector space structure on $V'$.

Definition 59 (Pairings / bilinear maps). For any two vector spaces $U, V$ a (bilinear) pairing between $U, V$ is a map
\[
\langle \cdot, \cdot \rangle : U \times V \to F
\]
which is linear in each variable separately. Similarly we define a bilinear map $U \times V \to Z$.

Remark 60. Note that such a map (if non-zero) is never linear for as map $U \oplus V \to F$. For example $\langle \alpha u, \alpha v \rangle = \alpha^2 \langle u, v \rangle$.

Example 61. The standard inner product on $F^n$: $\langle u, v \rangle = \sum_i u_i v_i$. More generally, given $B \in M_{m \times n}(F)$ have a pairing on $F^m \times F^n$ given by
\[
\langle u, v \rangle = \sum_i u_i B_{ij} v_j .
\]

More generally, given any bilinear pairing of $U, V$ choose bases $\{u_i\}_{i \in I} \subseteq U$, $\{v_j\}_{j \in J} \subseteq V$ and define the Gram matrix by
\[
B_{ij} = \langle u_i, v_j \rangle .
\]
We can then compute the pairing of any two vectors: by the distributive law ("FOIL")

\[
\left\langle \sum_i a_i u_i, \sum_j b_j v_j \right\rangle = \sum_{i,j} a_j b_j B_{i,j}.
\]

Conversely, any matrix \( B \) defines a bilinear pairing (aside: this is a linear bijection if you give pairings the obvious vector space structure).

### 1.3.6. Pairings: Duality and degeneracy.

Fix a bilinear form \( \langle \cdot, \cdot \rangle : U \times V \to F \). Then for any \( u \in U \) we get a map \( \phi_u : V \to F \) by \( \phi_u(v) = \langle u, v \rangle \).

1. \( \phi_u \) is linear \((\in V')\) iff the pairing is linear in the second variable.

2. The map \( U \to V' \) given by \( u \to \phi_u \) is linear iff the pairing is linear in the first variable.

We conclude that every pairing gives a map \( U \to V' \), and equivalently also a map \( V \to U' \).

**Lemma 62.** We have a linear bijection \{pairings on \( U \times V \}\) \( \longleftrightarrow \) \( \text{Hom}_F(U,V') \)

**Proof.** The inverse map associates to each \( f \in \text{Hom}_F(U,V') \) the bilinear form

\[
\langle u, v \rangle_f = (f(u))(v).
\]

**Definition 63.** Call the bilinear map non-degenerate if both maps \( U \to V' \), \( V \to U' \) are embeddings.

**Lemma 64.** A pairing is non-degenerate iff for every non-zero \( u \in U \) there is \( v \in V \) such that \( \langle u, v \rangle \neq 0 \) and conversely.

Key idea: if the map \( V \to U' \) associated to a pairing is bijective, then we can use \( V \) as a model for \( U' \) via the pairing.

**Example 65.** The dot product is a non-degenerate pairing \( F^n \times F^n \) hence identifies \((F^n)'\) with \( F^n \).

Two further examples from functional analysis:

First we fix a compact topological space \( X \). Then for any finite Borel measure \( \mu \) on \( X \) and any continuous \( f \in C(X) \) we have the integral

\[
\int f \, d\mu.
\]

This is a bilinear pairing \( C(X) \times \{\text{finite signed measures on } X\} \) which is non-degenerate.

**Theorem 66 (Riesz representation theorem).** Let \( X \) be compact. Then every continuous linear functional on \( C(X) \) is given by a finite measure (that is, the continuous dual \( C(X)' \) can be represented by the space of measures).

Second, the inner product on Hilbert space is a non-degenerate pairing!

**Theorem 67 (Riesz representation theorem).** Let \( \mathcal{H} \) be a Hilbert space. Then every continuous linear functional on \( \mathcal{H} \) is of the form \( \langle u, \cdot \rangle \).
1.3.7. The dual of a linear map (Lecture 8, continued).

**Construction 68.** Let $T \in \text{Hom}(U, V)$. Set $T' \in \text{Hom}(V', U')$ by $(T' \varphi)(v) = \varphi(Tv)$.

**Lemma 69.** This is a linear map $\text{Hom}(U, V) \rightarrow \text{Hom}(V', U')$. An isomorphism if $U, V$ finite-dimensional.

**Lemma 70.** $(TS)' = S'T'$

**Proof.** PS3

\[\square\]

1.4. Multilinear algebra and tensor products

1.4.1. Multilinear forms (Lecture 9).

**Definition 71.** Let $\{V_i\}_{i \in I}$ be vector spaces, $W$ another vector space. A function $f: \times_{i \in I} V_i \rightarrow W$ is said to be multilinear if it is linear in each variable.

**Example 72 (Bilinear maps).** (1) $f(x, y) = xy$ is bilinear $F^2 \rightarrow F$.
(2) The map $(T, \psi) \mapsto T\psi$ is a multilinear map $\text{Hom}(V, W) \times V \rightarrow W$.
(3) For a matrix $A \in M_{n,m}(F)$ have $(x, y) \mapsto \sum_{t=0}^n A_{xy}$ on $F^n \times F^m$.
(4) For $\varphi \in U'$, $\psi \in V'$ have $(u, v) \mapsto \varphi(u)\psi(v)$, and finite combinations of those.

**Remark 73.** A multilinear function on $U \times V$ is not the same as a linear function on $U \oplus V$. For example: is $f(au, av)$ equal to $af(u, v)$ or to $a^2 f(u, v)$? That said, $\oplus V_i$ was universal for maps from $V_j$. It would be nice to have a space which is universal for multilinear maps. We only discuss the finite case.

**Example 74.** A multilinear function $B: U \times \{0\} \rightarrow F$ has $B(u, 0) = B(0, 0) = 0$. A multilinear function $B: U \times F \rightarrow F$ has $B(u, x) = B(u, x(1)) = xB(u, 1) = x\varphi(u)$ where $\varphi(u) = B(u, 1) \in U'$.

We can reduce everything to Example 72(3): Fix bases $\{u_j\}, \{v_j\}$. Then

$$B \left( \sum_i x_iu_i, \sum_i y_jv_j \right) = \sum_i x_iB(u_i, v_j) = xB_y$$

where $B_{ij} = B(u_i, v_j)$. Note: $x_i = \varphi_i(u)$ where $\{\varphi_i\}$ is the dual basis. Conclude that

\[1.4.1\]

$$B = \sum_{i,j} B(u_i, v_j) \varphi_i\psi_j.$$

Easy to check that this is an expansion in a basis (check against $(u_i, v_j)$). We have shown:

**Proposition 75.** The set $\{\varphi_i\psi_j\}_{i,j}$ is a basis of the space of bilinear forms $U \times V \rightarrow F$.

**Corollary 76.** The space of bilinear forms on $U \times V$ has dimension $\dim F U \cdot \dim F V$.

**Remark 77.** Also works in infinite dimensions, since can have the sum \[1.4.1\] be infinite – every pair of vectors only has finite support in the respective bases.
1.4.2. The tensor product (Lecture 10-11). Now let’s fix $U, V$ and try to construct a space that will classify bilinear maps on $U \times V$.

- Our space will be generated by terms $u \otimes v$ on which we can evaluate $f$ to get $f(u, v)$.
- Since $f$ is multilinear, $f(au, bv) = abf(u, v)$ so need $(au) \otimes (bv) = ab(u \otimes v)$.
- Similarly, since $f(u_1 + u_2, v) = f(u_1, v) + f(u_2, v)$ want $(u_1 + u_2) \otimes (v_1 + v_2) = u_1 \otimes v_1 + u_2 \otimes v_1 + u_1 \otimes v_2 + u_2 \otimes v_2$.

**Construction 78** (Tensor product). Let $U, V$ be spaces. Let $X = F^\otimes(U \times V)$ be the formal span of all expressions of the form $\{u \otimes v\}_{(u, v) \in U \times V}$. Let $Y \subset X$ be the subspace spanned by

$$\{(au) \otimes (bv) - ab(u \otimes v) \mid a, b \in F, (u, v) \in U \times V\}$$

and

$$\{(u_1 + u_2) \otimes (v_1 + v_2) - (u_1 \otimes v_1 + u_2 \otimes v_1 + u_1 \otimes v_2 + u_2 \otimes v_2) \mid \ast\ast\}.$$

Then set $U \otimes V = X/Y$ and let $\iota: U \times V \rightarrow U \otimes V$ be the map $\iota(u, v) = (u \otimes v) + Y$.

**Theorem 79.** $\iota$ is a bilinear map. For any space $W$ any any bilinear map $f: U \times V \rightarrow W$, there is a unique linear map $\tilde{f}: U \otimes V \rightarrow W$ such that $f = \tilde{f} \circ \iota$.

**Proof.** Uniqueness is clear, since $\tilde{f}(u \otimes v) = f(u, v)$ fixes $\tilde{f}$ on a generating set. For existence we need to show that if $\tilde{f}: X \rightarrow W$ is defined by $\tilde{f}(u \otimes v) = f(u, v)$ then $\tilde{f}$ vanishes on $Y$ and hence descends to $U \otimes V$.

**Proposition 80.** Let $B_U, B_V$ be bases for $U, V$ respectively. Then $\{u \otimes v \mid u \in B_U, v \in B_V\}$ is a basis for $U \otimes V$.

**Proof.** Spanning: use bilinearity of $\iota$. Independence: for $u \in B_U$ let $\{\varphi_u\}_{u \in B_U} \subset U'$, $\{\psi_v\}_{v \in B_V} \subset V'$ be the dual bases. Then $\varphi_u \psi_v$ is a bilinear map $U \times V \rightarrow F$, and the sets $\{u \otimes v\}_{(u, v) \in B_U \times B_V}$ and $\{\varphi_u \psi_v\}_{(u, v) \in B_U \times B_V}$ are dual bases.

**Corollary 81.** $\dim_F(U \otimes V) = \dim_F U \cdot \dim_F V$.

1.4.3. Symmetric and antisymmetric tensor products (Lecture 13, 3/2/2014). In this section we assume $\text{char}(F) = 0$.

Let $(12) \in S_2$ act on $V \otimes V$ by exchanging the factors (why is this well-defined?).

**Lemma 82.** Let $T \in \text{End}_F(U)$ satisfy $T^2 = \text{Id}$. Then $U$ is the direct sum of the two eigenspaces.

**Definition 83.** $\text{Sym}^2 V$ and $\wedge^2 V$ are the eigenspaces.

**Proposition 84.** Generating sets and bases.

In general, let $S_k$ act on $V^\otimes k$.

- What do we mean by that? Well, this classifies $n$-linear maps $V \times \cdots \times V \rightarrow Z$. Universal property gives isom of $(U \otimes V) \otimes W, U \otimes (V \otimes W)$.
- Why action well-defined? After all, the set of pure tensors is nonlinear. So see first as multilinear map $V^n \rightarrow V^\otimes n$.
- Single out $\text{Sym}^k V, \wedge^k V$. Note that there are other representations.
- Claim: bases
1.4.4. Bases of $\text{Sym}^k$, $\wedge^k$, determinants (Lecture 14, 5/2/2014).

**Proposition 85.** Symmetric/antisymmetric tensors are generating sets; bases coming from subsets of basis.

Tool: the maps $P^\pm_k : V \otimes^k V \to V \otimes^k V$ given by

$$P^\pm_k (v_1 \otimes \cdots \otimes v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} (\pm)^\sigma (v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(k)}).$$

**Lemma 86.** These are well defined (extensions of linear maps). Fix elements of $\text{Sym}^k V$, $\wedge^k V$ respectively, images are in those subspaces (check $\tau \circ P^\pm_k = (\pm) P^\pm_k$). Conclude that image is spanned by image of basis.

- Exterior forms of top degree and determinants.
CHAPTER 2

Structure Theory

2.1. Introduction (Lecture 15)

2.1.1. The two paradigmatic problems. Fix a vector space \( V \) of dimension \( n < \infty \) (in this chapter, all spaces are finite-dimensional unless stated otherwise), and a map \( T \in \text{End}(V) \). We will try for two kinds of structural results:

1. [“decomposition”] \( T = RS \) where \( R, S \in \text{End}(V) \) are “simple”
2. [“form”] There is a basis \( \{v_j\}_{j=1}^n \subset V \) in which the matrix of \( T \) is “simple”.

Example 87. (From 1st course)

1. (Gaussian elimination) Every matrix \( A \in M_n(F) \) can be written in the form \( A = E_1 \cdots E_k \cdot A_{\text{rr}} \) where \( E_i \) are “elementary” (row operations or rescaling) and \( A_{\text{rr}} \) is row-reduced.
2. (Spectral theory) Suppose \( T \) is diagonable. Then there is a basis in which \( T \) is diagonal.

As an example of how to use (1), suppose \( \det(A) \) is defined for matrices by column expansion. Then can show (Lemma 1) that \( \det(EX) = \det(X) \) whenever \( E \) is elementary and that (Lemma 2) \( \det(AX) = \det(X) \) whenever \( A \) is row-reduced. One can then prove

**Theorem 88.** For all \( A, B \), \( \det(AB) = \det(A) \det(B) \).

**Proof.** Let \( D = \{A \mid \forall X : \det(AX) = \det(A) \det(X)\} \). Then we know that \( A_{\text{rr}} \in cD \) and that if \( A \in D \) then for any elementary \( E \), \( \det((EA)X) = \det(EAX) = \det(E) \det(A) \det(X) = \det(E) \det(A) \det(X) = \det(EA) \det(X) \) so \( EA \in D \) as well. It now follows from Gauss’s Theorem that \( D \) is the set of all matrices. \( \square \)

2.1.2. Triangular matrices.

**Definition 89.** \( A \in M_n(F) \) is upper (lower) triangular if ...

Significance: these are very good for computation. For example:

**Lemma 90.** The upper-triangular matrix \( U \) is invertible iff its diagonal entries are non-zero.

**Algorithm 91** (Back-substitution). Suppose upper-triangular \( U \) is invertible. Then the solution to \( UX = b \) is given by setting \( x_i = \frac{b_i - \sum_{j=k+1}^n u_{ij}x_j}{u_{ii}} \) for \( i = n, n-1, n-2, \ldots, 1 \).

**Remark 92.** Note that the algorithm does exactly as many multiplications as non-zero entries in \( U \). Hence better than Gaussian elimination for general matrix (\( O(n^3) \)), really good for sparse matrix, and doesn’t require storing the matrix entries only the way to calculate \( u_{ij} \) (in particular no need to find inverse).

**Exercise 93.** (1) Give formula for inverse of upper-triangular matrix (2) Develop forward-substitution algorithm for lower-triangular matrices.
Corollary 94. If $A = LU$ we can efficiently solve $Ax = b$.

Note that we don’t like to store inverses. For example, because they are generally dense matrices even if $L,U$ are sparse.

We now try to look for a vector-space interpretation of being triangular. For this note that if $U \in M_n(F)$ is triangular then

\[
Ue_1 = u_{11}e_1 \in \text{Span}\{e_1\} \\
Ue_2 = u_{12}e_1 + u_{22}e_2 \in \text{Span}\{e_1, e_2\} \\
\vdots \\
Ue_k \in \text{Span}\{e_1, \ldots, e_k\} \\
\vdots 
\]

In particular, we found a family of subspaces $V_i = \text{Span}\{e_1, \ldots, e_i\}$ such that $U(V_i) \subset V_i$, such that $\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n = F^n$ and such that dim$V_i = i$.

Theorem 95. $T \in \text{End}(V)$ has an upper-triangular matrix wrt some basis iff there are $T$-invariant subspaces $\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n = F^n$ with dim$V_i = i$.

Proof. We just saw necessity. For sufficiency, given $V_i$ choose for $1 \leq i \leq n$, $v_i \in V_i \setminus V_{i-1}$. These exist (the dimension increases by 1), are a linearly independent set (each vector is independent of its predecessors) and the first $i$ span $V_i$ (by dimension count). Finally for each $i$, $Tv_i \in T(V_i) \subset V_i = \text{Span}\{v_1, \ldots, v_i\}$ so the matrix of $T$ in this basis is upper triangular. \hfill \Box

2.2. Jordan Canonical Form

2.2.1. The minimal polynomial. Recall we have an $n$-dimensional $F$-vector space $V$.

- A key tool for studying linear maps is studying polynomials in the maps (we saw how to analyze maps satisfying $T^2 = \text{Id}$, for example).
- We will construct a gadget (the “minimal polynomial”) attached to every linear map on $V$. It is a polynomial, and will tell us a lot about the map.
- Computationally speaking, this polynomial cannot be found efficiently. It is a tool of theorem-proving in abstract algebra.

Definition 96. Given a polynomial $f \in F[x]$, say $f = \sum_{i=0}^d a_i x^i$ and a map $T \in \text{End}(V)$ set (with $T^0 = \text{Id}$)

\[
f(T) = \sum_{i=0}^d a_i T^i.
\]

Lemma 97. Let $f, g \in F[x]$. Then $(f + g)(T) = f(T) + g(T)$ and $(fg)(T) = f(T)g(T)$. In other words, the map $f \mapsto f(T)$ is a linear map $F[x] \to \text{End}(V)$, also respecting multiplication (“a map of $F$-algebras”, but this is beyond our scope).

Proof. Do it yourself. \hfill \Box

- Given a linear map our first instinct is to study the kernel and the image. [Aside: the kernel is an ideal in the algebra].
- We’ll examine the kernel and leave the image for later.
**Lemma 98.** There is a non-zero polynomial \( f \in F[x] \) such that \( f(T) = 0 \). In fact, there is such \( f \) with \( \deg f \leq n^2 \).

**Proof.** \( F[x] \) is infinite-dimensional while \( \text{End}_F(V) \) is finite-dimensional. Specifically, \( \dim_F F[x] \leq n^2 = n^2 + 1 \) while \( \dim_F \text{End}_F(V) = n^2 \).

**Remark 99.** We will later show (Theorem of Cayley–Hamilton) that the characteristic polynomial \( P_T(x) = \det(x \text{Id} - T) \) from basic linear algebra has this property.

- Warning: we are about to divide polynomials with remainder.

**Proposition 100.** Let \( I = \{ f \in F(x) \mid f(T) = 0 \} \). Then \( I \) contains a unique non-zero monic polynomial of least degree, say \( m(x) \), and \( I = \{ g(x)m(x) \mid g \in F[x] \} \) is the set of multiples of \( m \).

**Proof.** Let \( m \in I \) be a non-zero member of least degree. Dividing by the leading coefficient we may assume \( m \) monic. Now suppose \( m' \) is another such. Then \( m - m' \in I \) (this is a subspace) is of strictly smaller degree. It must therefore be the zero polynomial, and \( m \) is unique. Clearly if \( g \in F[x] \) then \((gm)(T) = g(T)m(T) = 0 \). Conversely, given any \( f \in I \) we can divide with remainder and write \( f = qm + r \) for some \( q, r \in F[x] \) with \( \deg r < \deg m \). Evaluating at \( T \) we find \( r(T) = 0 \), so \( r = 0 \) and \( f = qm \).

**Definition 101.** Call \( m(x) = m_T(x) \) the minimal polynomial of \( T \).

**Remark 102.** We will later prove directly that \( \deg m_T(x) \leq n \).

**Example 103.** (Minimal polynomials)

1. \( T = \text{Id}, \ m(x) = x - 1 \).
2. \( T = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \ T^2 = 0 \) but \( T \neq 0 \) so \( m_T(x) = x^2 \).
3. \( T = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \ T^2 = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \) so \( (T^2 - \text{Id}) = 2(T - \text{Id}) \) so \( T^2 - 2T + \text{Id} = 0 \) so \( m_T(x)(x - 1)^2 \). But \( T - \text{Id} \neq 0 \) so \( m_T(x) = (x - 1)^2 \).
4. \( T = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ T^2 = \text{Id} \) so \( m_T(x) = x^2 - 1 = (x - 1)(x + 1) \).
   - In the eigenbasis \( \left\{ \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix} \right\} \) the matrix is \( \begin{pmatrix} 1 \\ -1 \end{pmatrix} \) — we saw this in a previous class.
5. \( T = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \ T^2 = -\text{Id} \) so \( m_T(x) = x^2 + 1 \).
   - (a) If \( F = \mathbb{Q} \) or \( F = \mathbb{R} \) this is irreducible. No better basis.
   - (b) If \( F = \mathbb{C} \) (or \( \mathbb{Q}(i) \)) then factor \( m_T(x) = (x - i)(x + i) \) and in the eigenbasis \( \left\{ \begin{pmatrix} 1 \\ i \end{pmatrix} \right\} \) the matrix has the form \( \begin{pmatrix} -i \\ i \end{pmatrix} \).
6. \( V = F[x]^{\leq n} \) (polynomials of degree less than \( n \)), \( T = \frac{d}{dx} \). Then \( T^n = 0 \) but \( T^{n-1} \neq 0 \) (why?) so \( m_T(x) = x^n \).
7. [To be proved in problem set] Let \( D = \text{diag}(a_1, \ldots, a_n) \) be diagonal, its entries being the distinct numbers \( \{b_1, \ldots, b_r\} \) (perhaps with repetition). Then its minimal polynomial is \( \prod_{i=1}^{r} (x - b_r) \) [cf (1),(4),(5)].
We now connect the minimal polynomial with the spectrum.

**Lemma 104 (Spectral calculus).** Suppose that \( T \mathbf{v} = \lambda \mathbf{v} \). Then \( f(T) \mathbf{v} = f(\lambda) \mathbf{v} \).

**Proof.** Work it out at home. \( \square \)

**Remark 105.** The same proof shows that if the subspace \( W \) is \( T \)-invariant \((T(W) \subset W)\) then \( W \) is \( f(T) \)-invariant for all polynomials \( f \).

**Corollary 106.** If \( \lambda \) is an eigenvalue of \( T \) then \( m_T(\lambda) = 0 \). In particular, if \( m_T(0) \neq 0 \) then \( T \) is invertible \((0 \text{ is cannot be eigenvalue})\).

We now use the *minimality* of the minimal polynomial.

**Theorem 107.** \( T \) is invertible iff \( m_T(0) \neq 0 \).

**Proof.** Suppose that \( T \) is invertible and that \( \sum_{i=1}^{d} a_i T^i = 0 \) [note \( a_0 = 0 \) here]. Then this is not the minimal polynomial since multiplying by \( T^{-1} \) also gives
\[
\sum_{i=0}^{d-1} a_{i+1} T^i = 0.
\]

\( \square \)

**Corollary 108.** \( \lambda \in F \) is an eigenvalue of \( T \) iff \( \lambda \) is a root of \( m_T(x) \).

**Proof.** Let \( S = T - \lambda \text{Id} \). Then \( m_S(x) = m_T(x+\lambda) \). Then \( \lambda \in \text{Spec}_F(T) \iff S \text{ not invertible} \iff m_S(0) = 0 \iff m_T(\lambda) = 0 \). \( \square \)

**Remark 109.** The characteristic polynomial \( P_T(x) \) also has this property – this is how eigenvalues are found in basic linear algebra.

### 2.2.2. Generalized eigenspaces.

Continue with \( T \in \text{End}_F(V), \dim_F(V) = n \). Recall that \( T \) is diagonalizable iff \( V \) is the direct sum of the eigenspace. For non-diagonalizable maps we need something more sophisticated.

**Problem 110.** Find a matrix \( A \in M_2(F) \) which only has a 1-d eigenspace.

**Definition 111.** Call \( \mathbf{v} \in V \) a generalized eigenvector of \( T \) if for some \( \lambda \in F \) and \( k \geq 1 \), \( (T - \lambda)^k \mathbf{v} = 0 \). Let \( V_\lambda \subset V \) denote the set of generalized \( \lambda \)-eigenvectors and \( 0 \). Call \( \lambda \) a generalized eigenvalue of \( T \) if \( V_\lambda \neq \{0\} \).

In particular, if \( T \mathbf{v} = \lambda \mathbf{v} \) then \( \mathbf{v} \in V_\lambda \).

**Proposition 112 (Generalized eigenspaces).**

1. Each \( V_\lambda \) is a \( T \)-invariant subspace.
2. Let \( \lambda \neq \mu \). Then \( (T - \mu) \) is invertible on \( V_\lambda \).
3. \( V_\lambda \neq \{0\} \) iff \( \lambda \in \text{Spec}_F(T) \).

**Proof.** Let \( \mathbf{v}, \mathbf{v}' \in V_\lambda \) be killed by \( (T - \lambda)^k, (T - \lambda)^{k'} \) respectively. Then \( \alpha \mathbf{v} + \beta \mathbf{v}' \) is killed by \( (T - \lambda)^{\max\{k,k'\}} \). Also, \( (T - \lambda)^k T \mathbf{v} = T (T - \lambda)^k \mathbf{v} = 0 \) so \( T \mathbf{v} \in V_\lambda \) as well.

Let \( \mathbf{v} \in \text{Ker}(T - \mu) \) be non-zero. By Lemma 104 for any \( k \) we have \( (T - \lambda)^k \mathbf{v} = (\mu - \lambda)^k \mathbf{v} \neq 0 \) so \( \mathbf{v} \notin V_\lambda \).

Finally, given \( \lambda \) and non-zero \( \mathbf{v} \in V_\lambda \) let \( k \) be minimal such that \( (T - \lambda)^k \mathbf{v} = 0 \). Then \( (T - \lambda)^{k-1} \mathbf{v} \) is non-zero and is an eigenvector of eigenvalue \( \lambda \). \( \square \)
THEOREM 113. The sum $\bigoplus_{\lambda \in \text{Spec}(F)} V_{\lambda} \subset V$ is direct.

PROOF. Let $\sum_{i=1}^{r} v_i = 0$ be a minimal dependence with $v_i \in V_{\lambda_i}$ for distinct $\lambda_i$. Applying $(T - \lambda_i)^k$ for $k$ large enough to kill $v_j$ we get the dependence.

$$\sum_{i=1}^{r-1} (T - \lambda_i)^k v_j = 0.$$ 

Now $(T - \lambda_i)^k v_j \in V_{\lambda_i}$ since these are $T$-invariant subspaces, and for $1 \leq i \leq r-1$ is non-zero since $T - \lambda_r$ is invertible there. This shorter dependence contradicts the minimality. □

REMARK 114. The sum may very well be empty – there are non-trivial maps without eigenvalues (for example $\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$).

2.2.3. Algebraically closed field. We all know that sometimes linear maps fail to have eigenvalues, even though they “should”. In this course we’ll blame the field, not the map, for this deficiency.

DEFINITION 115. Call the field $F$ algebraically closed if every non-constant polynomial $f \in F[x]$ has a root in $F$. Equivalently, if every non-constant polynomial can be written as a product of linear factors.

FACT 116 (Fundamental theorem of algebra). $\mathbb{C}$ is algebraically closed.

REMARK 117. Despite the title, this is a theorem of analysis.

Discussion. The goal is to create enough eigenvalues so that the generalized eigenspaces explain all of $V$. The first point of view is that we can simple “define the problem away” by restricting to the case of algebraically closed fields. But this isn’t enough, since sometimes we are given maps over other fields. This already appears in the diagonable case, dealt with in 223: we can view

$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \in M_2(\mathbb{R})$ instead as $\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \in M_2(\mathbb{C})$, at which point it becomes diagonable. In other words, we can take a constructive point of view:

- Starting with any field $F$ we can “close it” by repeatedly adding roots to polynomial equations until we can’t, obtaining an “algebraic closure” $\overline{F}$ [the difficulty is in showing the process eventually stops].
  - This explains the “closed” part of the name – it’s closure under an operation.
  - [Q: do you need the full thing? A: In fact, it’s enough to pass to the splitting field of the minimal polynomial]

- We now make this work for linear maps, with three points of view:
  (1) (matrices) Given $A \in M_n(F)$ view it as $A \in M_n(\overline{F})$, and apply the theory there.
  (2) (linear maps) Given $T \in \text{End}_F(V)$, fix a basis $\{v_i\}_{i=1}^{n} \subset V$, make the formal span $\overline{V} = \bigoplus_{i=1}^{n} \overline{F}v_i$ and extends $T$ to $\overline{V}$ by the property of having the same matrix.
  (3) (coordinate free) Given $V$ over $F$ set $\overline{V} = \overline{F} \otimes_F V$ (considering $\overline{F}$ as an $F$-vectorspace), and extend $T$ (by $\overline{T} = \text{Id}_K \otimes_F T$).
Back to the stream of the course.

**Lemma 118.** Suppose $F$ is algebraically closed and that $\dim_F V \geq 1$. Then every $T \in \text{End}_F(V)$ has an eigenvector.

**Proof.** $m_T(x)$ has roots. \hfill \Box

We suppose now that $F$ is algebraically closed, in other words that every linear map has an eigenvalue. The following is the key structure theorem for linear maps:

**Theorem 119.** (with $F$ algebraically closed) We have $V = \bigoplus_{\lambda \in \text{Spec}_F(T)} V_\lambda$.

**Proof.** Let $m_T(x) = \prod_{i=1}^r (x - \lambda_i)^{k_i}$ and let $W = \bigoplus_{i=1}^r V_{\lambda_i}$. Supposing that $W \neq V$, let $\bar{V} = V/W$ and consider the quotient map $\bar{T} \in \text{End}_F(\bar{V})$ defined by $\bar{T}(\bar{v} + W) = T\bar{v} + W$. Since $\dim_F \bar{V} \geq 1$, $\bar{T}$ has an eigenvalue there. We first check that this eigenvalue is one of the $\lambda_i$. Indeed, for any polynomial $f \in F[x]$, $f(\bar{T})(\bar{v} + W) = (f(T)v) + W$, and in particular $m_T(\bar{T}) = 0$ and hence $m_T | m_T$. 

Renumbering the eigenvalues, we may assume $V_{\lambda_r} \neq \{0\}$, and let $v \in V$ be such that $v + W \in V_{\lambda_r}$ is non-zero, that is $v \notin W$. Since $\prod_{i=1}^{r-1} (\bar{T} - \lambda_i)^{k_i}$ is invertible on $V_{\lambda_r}$, $u = \prod_{i=1}^{r-1} (T - \lambda_i)^{k_i} v \notin W$. But $(T - \lambda_r)^{k_r} u = m_T(T)v = 0$ means that $u \in V_{\lambda_r} \subset W$, a contradiction. \hfill \Box

**Proposition 120.** In $m_T(x) = \prod_{i=1}^r (x - \lambda_i)^{k_i}$, the number $k_i$ is the minimal $k$ such that $(T - \lambda_i)^k = 0$ on $V_{\lambda_i}$.

**Proof.** Let $T_i$ be the restriction of $T$ to $V_{\lambda_i}$. Then $(T_i - \lambda_i)^k$ is the minimal polynomial by assumption. But $m_T(T_i) = 0$. It follows that $(x - \lambda_i)^k | m_T$ and hence that $k \leq k_i$. Conversely, since $\prod_{j \neq i} (T - \lambda_j)^{k_j}$ is invertible on $V_{\lambda_i}$, we see that $(T - \lambda_i)^{k_i} = 0$ there, so $k_i \geq k$. \hfill \Box

Summary of the construction so far:

- $F$ algebraically closed field, $\dim_F V = n$, $T \in \text{End}_F(V)$.
- $m_T(x) = \prod_{i=1}^r (x - \lambda_i)^{k_i}$ the minimal polynomial.
- Then $V = \bigoplus_{i=1}^r V_{\lambda_i}$ where on $V_{\lambda_i}$ we have $(T - \lambda_i)^{k_i} = 0$ but $(T - \lambda_i)^{k_i - 1} \neq 0$.

We now study the restriction of $T$ to each $V_{\lambda_i}$, via the map $N = T - \lambda_i$, which is nilpotent of degree $k_i$.

### 2.2.4. Nilpotent maps.

We return to the case of a general field $F$.

**Definition 121.** A map $N \in \text{End}_F(V)$ such that $N^k = 0$ for some $k$ is called nilpotent. The smallest such $k$ is called its degree of nilpotence.

**Lemma 122.** Let $N \in \text{End}_F(V)$ be nilpotent. Then its degree of nilpotence is at most $\dim_F V$.

**Proof.** Exercise. \hfill \Box

**Proof of Corollary.** Define subspaces $V_k$ by $V_0 = V$ and $V_{i+1} = N(V_i)$. Then $V = V_0 \supset V_1 \cdots \supset V_i \supset \cdots$. If at any stage $V_i = V_{i+1}$ then $V_{i+j} = V_i$ for all $j \geq 1$, and in particular $V_i = \{0\}$ (since $V_k = 0$). It follows that for $i < k$, $\dim V_{i+1} < \dim V_i$ and the claim follows. \hfill \Box

**Corollary 123** (Cayley–Hamilton Theorem). Suppose $F$ is algebraically closed. Then $m_T(x) | p_T(x)$ and, equivalently, $p_T(T) = 0$. In particular, $\deg m_T \leq \dim_F V$. 

22
Recall that the characteristic polynomial of $T$ is the polynomial $p_T(x) = \det(x \text{Id} - T)$ of degree $\dim_F V$, and that is also has the property that $\lambda \in \text{Spec}_F (T)$ iff $p_T(\lambda) = 0$.

**Proof.** The linear map $x \text{Id} - T$ respects the decomposition $V = \bigoplus_{i=1}^r V_{\lambda_i}$. We thus have $p_T(x) = \prod_{i=1}^r p_{T|V_{\lambda_i}}(x)$. Since $p_{T|V_{\lambda_i}}(x)$ has the unique root $\lambda$, it is the polynomial $(x - \lambda)^{\dim_F V_{\lambda_i}}$, so

$$p_T(x) = \prod_{i=1}^r (x - \lambda_i)^{\dim F V_{\lambda_i}}.$$  

Finally, $k_i$ is the degree of nilpotence of $(T - \lambda_i)$ on $V_{\lambda_i}$. Thus $k_i \leq \dim F V_{\lambda_i}$.

We now resolve a lingering issue:

**Lemma 124.** The minimal polynomial is independent of the choice of the field. In particular, the Cayley–Hamilton Theorem holds over any field.

**Proof.** Whether $\{1, T, \ldots, T^{d-1}\} \subset \text{End}_F (V)$ are linearly dependent or not does not depend on the field. \hfill \Box

**Theorem 125 (Cayley–Hamilton).** Over any field we have $m_T(x)|p_T(x)$ and, equivalently, $p_T(T) = 0$.

**Proof.** Extend scalars to an algebraic closure. This does not change either of the polynomials $m_T, p_T$. \hfill \Box

We finally turn to the problem of finding good bases for linear maps, starting with the nilpotent case. Here $F$ can be an arbitrary field.

**Lemma 126.** Let $N \in \text{End}(V)$ be nilpotent. Let $B \subset V$ be a set of vectors such that $N(B) \subset B \cup \{0\}$. Then $B$ is linearly independent iff $B \cap \text{Ker}(N)$ is.

**Proof.** One direction is clear. For the converse, let $\sum_{i=1}^r a_i v_i = 0$ be a minimal dependence in $B$. Applying $N$ we obtain the dependence

$$\sum_{i=1}^r a_i Nv_i = 0.$$

If all $Nv_i = 0$ then we had a dependence in $B \cap \text{Ker}(N)$. Otherwise, no $Nv_i = 0$ (this would shorten the dependence), and by uniqueness it follows that, up to rescaling, $N$ permutes the $v_i$. But then the same is true for any power of $N$, contradicting the nilpotence. \hfill \Box

**Corollary 127.** Let $N \in \text{End}(V)$ and let $v \in V$ be non-zero such that $N^k v = 0$ for some $k$ (wlog minimal). Then $\{N^i v\}_{i=0}^{k-1}$ is linearly independent.

**Proof.** $N$ is nilpotent on $\text{Span}\{N^i v\}_{i=0}^{k-1}$, this set is invariant, and its intersection with $\text{Ker} N$ is exactly $\{N^{k-1} v\} \neq \{0\}$. \hfill \Box

Our goal is now to decompose $V$ as a direct sum of $N$ subspaces (“Jordan blocks”) each of which has a basis as in the Corollary.

**Theorem 128 (Jordan form for nilpotent maps).** Let $N \in \text{End}_F (V)$ be nilpotent. We then have a decomposition $V = \bigoplus_{j=1}^r V_j$ where each $V_j$ is an $N$-invariant Jordan block.
Example 129. \( A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ -1 & -2 & -3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} (1 \\ 2 \\ 3). \)

- \( A^2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} (1 \\ 2 \\ 3) \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} (1 \\ 2 \\ 3) = 0 \), so \( A \) is nilpotent. The characteristic polynomial must be \( x^3 \).

- The image of \( A \) is \( \text{Span} \left\{ \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \right\} \). Since \( A \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \), \( \text{Span} \left\{ \begin{pmatrix} 3 \\ -1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 0 \end{pmatrix} \right\} \) is a block.

- Taking any other vector in the kernel (say, \( \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} \)) we get the basis \( \left\{ \begin{pmatrix} 3 \\ -1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} \right\} \) in which \( A \) has the matrix

\[
\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.
\]

**Proof.** Let \( N \) have degree of nilpotence \( d \) and kernel \( W \). For \( 1 \leq k \leq d \) define \( W_k = \text{Im}(N^k) \cap W \), so that \( W_0 = W \supset W_1, W_d = \{0\} \). Now choose a basis of \( W \) compatible with this decomposition – in other words choose subsets \( B_k \subset W_k \) such that \( \bigcup_{k \geq k'} B_k \) is a basis for \( W_{k'} \). Let \( B = \bigcup_{k=0}^{d-1} B_k = \{v_j\}_{i \in I} \) and for each \( i \) define \( k_i \) by \( v_i \in B_{k_i} \). Choose \( u_i \) such that \( N^{k_i}u_i = v_i \), and for \( 1 \leq j \leq k_i \) set \( v_{i,1} = N^{k_i-j}u_i \) so that \( v_{i,1} = u_i \) and in general \( N^{k_i-j}u_i = \begin{pmatrix} v_{i,j-1} \\ 0 \end{pmatrix} \) for \( j \geq 1 \). It is clear that \( \text{Span}_F \{ v_{i,j} \}_{j=1}^{k_i} \) is a Jordan block, and that \( C = \{ v_{i,j} \}_{i,j} \) is a union of Jordan blocks. \( \square \)

- **The set \( C \) is linearly independent:** by construction, \( N(C) \subset C \cup \{0\} \) and \( C \cap W = B \) is independent.

- **The set \( C \) is spanning:** We prove by induction on \( k \leq d \) that \( \text{Span}_F(C) \supset \text{Ker}(N^k) \). This is clear for \( k = 0 \); suppose the result for \( 0 \leq k < d \), and let \( v \in \text{Ker}(N^{k+1}) \). Then \( N^{k+1}v \in W_k \), so we can write

\[
N^{k}v = \sum_{i:k_i \geq k} a_i v_i = \sum_{i:k_i \geq k} a_i N^k(v_{i,k+1}) = 0.
\]

It follows that

\[
N^{k} \left( v - \sum_{i:k_i \geq k} a_i v_{i,k+1} \right) = 0.
\]

By induction, \( v - \sum_{i:k_i \geq k} a_i v_{i,k+1} \in \text{Span}_F(C) \), and it follows that \( v \in \text{Span}_F(C) \).

**Definition 130.** A Jordan basis is a basis as in the Theorem.
**Lemma 131.** Any Jordan basis for $N$ has exactly $\dim_F W_k - \dim_F W_{k-1}$ blocks of length $k$. Equivalently, up to permuting the blocks, $N$ has a unique matrix in Jordan form.

**Proof.** Let $\{v_{i,j}\}$ be a Jordan basis. Then $\text{Ker}N = \text{Span}\{v_{i,1}\}$, while $\{v_{i,j} \mid j \leq k - k_i\}$ is a basis for $\text{Im}(N^{k_i})$. Clearly $\{v_{i,1} \mid j \geq k\}$ then spans $W_k$ and the claim follows. \hfill \Box

### 2.2.5 The Jordan canonical form.

**Theorem 132 (Jordan canonical form).** Let $T \in \text{End}_F(V)$ and suppose that $m_T$ splits into linear factors in $F$ (for example, that $F$ is algebraically closed). Then there is a basis $\{v_{\lambda,i,j}\}_{\lambda,i,j}$ of $V$ such that $\{v_{\lambda,i,j}\}_{i,j} \subset V_{\lambda}$ is a basis, and such that $(T - \lambda)v_{\lambda,i,j} = \begin{cases} v_{\lambda,i,j-1} & j \geq 1 \\ 0 & j = 1 \end{cases}$. Furthermore, writing $W_{\lambda} = \text{Ker}(T - \lambda)$ for the eigenspace, we have for each $\lambda$, that $1 \leq i \leq \dim_F W_{\lambda}$ and that the number of $i$ such that $1 \leq j \leq k$ is exactly $\dim_F ((T - \lambda)^{k-1}V_\lambda \cap W_{\lambda}) - \dim_F ((T - \lambda)^{k}V_\lambda \cap W_{\lambda})$. Equivalently, $T$ has a unique matrix in Jordan canonical form up to permuting the blocks.

**Corollary 133.** The algebraic multiplicity of $\lambda$ is $\dim_F V_{\lambda}$. The geometric multiplicity is the number of blocks.

**Example 134 (Jordan forms).** (1) $A_1 = \begin{pmatrix} 2 & 2 & 3 \\ 1 & 3 & 3 \\ -1 & -2 & -2 \end{pmatrix} = I + A$. This has characteristic polynomial $(x - 1)^3$, $A_1 - I = A$ and we are back in example [129].

(2) (taken from Wikibooks:Linear Algebra) $p_B(x) = (x - 6)^4$. Let $B = \begin{pmatrix} 7 & 1 & 2 & 2 \\ 1 & 4 & -1 & -1 \\ -2 & 1 & 5 & -1 \\ 1 & 1 & 2 & 8 \end{pmatrix}$, $B' = B - 6I = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & -2 & -1 & -1 \\ -2 & 1 & -1 & -1 \\ 1 & 1 & 2 & 2 \end{pmatrix}$. Gaussian elimination shows $B' = E \begin{pmatrix} 3 & -3 & 0 & 0 \\ 1 & -2 & -1 & -1 \\ -3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, $B'^2 = \begin{pmatrix} 0 & 3 & 3 & 3 \\ 0 & 3 & 3 & 3 \\ 0 & -6 & -6 & -6 \\ 0 & 3 & 3 & 3 \end{pmatrix}$ and $B'^3 = 0$. Thus $\text{Ker}B' = \{ (x,y,z,w)^t \mid x = y = -(z+w) \}$ is two-dimensional. We see that the image of $B'^2$ is spanned by $(3,3,-6,3)^t$, which is (say) $B'(2,-1,-1,2)^t$ which (being the last column) was $B'(0,0,0,1)^t$. Another vector in the kernel is $(-1,-1,1,0)^t$, and we get the Jordan basis $\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}$. 

25
(3) \( C = \begin{pmatrix} 4 & 0 & 1 & 0 \\ 2 & 2 & 3 & 0 \\ -1 & 0 & 2 & 0 \\ 4 & 0 & 1 & 2 \end{pmatrix} \) acting on \( V = \mathbb{R}^4 \) with \( p_C(x) = (x - 2)^2(x - 3)^2 \). Then \( C - 2I = \begin{pmatrix} 2 & 0 & 1 & 0 \\ 2 & 0 & 3 & 0 \\ -1 & 0 & 0 & 0 \\ 4 & 0 & 1 & 0 \end{pmatrix} \), \( C - 3I = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 2 & -1 & 3 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \), \( (C - 3I)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -3 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 1 \end{pmatrix} \). Thus \( \text{Ker}(C - 2I) = \text{Span} \{ e_2, e_4 \} \), which must be the 2d generalized eigenspace \( V_2 \) giving two \( 1 \times 1 \) blocks. For \( \lambda = 3 \), \( \text{Ker}(C - 3I) = \{ (x, y, z, w)^t : z = y = -x, w = 3x \} = \text{Span} \{ (1, -1, -1, 3)^t \} \). This isn’t the whole generalized eigenspace, and \( \text{Ker}(C - 3I)^2 = \{ (x, y, z, w)^t : y = 3x + 4z, w = x - 2z \} = \text{Span} \{ (1, -1, -1, 3)^t, (1, 3, 0, 1)^t \} \).

This must be the generalized eigenspace \( V_3 \), since it’s 2d. We need to find the image of \( (C - 3I)[V_3] \). One vector is in the kernel, so we try the other one, and indeed \( (C - 3I)(1, 3, 0, 1)^t = (1, -1, -1, 3) \). This gives us a 2x2 block, so in the basis \( \left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} \right\} \), the matrix would have the form \( \begin{pmatrix} (2) \\ (2) \end{pmatrix} \).

Note how the image of \( (C - 3I)^2 \) is exactly \( V_2 \) (why?)

(4) \( V = \mathbb{R}^6 \), \( p_D(x) = t^6 + 3t^5 - 10t^3 - 15t^2 - 9t - 2 = (t + 1)^3(t - 2) \):

\[
D = \begin{pmatrix}
0 & 0 & 0 & 0 & -1 & -1 \\
0 & -8 & 4 & -3 & 1 & -3 \\
-3 & 13 & -8 & 6 & 2 & 9 \\
-2 & 14 & -7 & 4 & 2 & 10 \\
1 & -18 & 11 & -11 & 2 & -6 \\
-1 & 19 & -11 & 10 & -2 & 7
\end{pmatrix},
\]

\( D + I = \begin{pmatrix}
1 & 0 & 0 & 0 & -1 & -1 \\
0 & -7 & 4 & -3 & 1 & -3 \\
-3 & 13 & -7 & 6 & 2 & 9 \\
-2 & 14 & -7 & 5 & 2 & 10 \\
1 & -18 & 11 & -11 & 3 & -6 \\
-1 & 19 & -11 & 10 & -2 & 8
\end{pmatrix}, \quad (D + I)^2 = \begin{pmatrix}
1 & -1 & 0 & 1 & -2 & -3 \\
-2 & 16 & 9 & -11 & 4 & -3 \\
-1 & 37 & -18 & 17 & 2 & 21 \\
1 & 35 & -18 & 19 & -2 & 15 \\
-1 & -53 & 27 & -28 & 2 & -24 \\
2 & 52 & -27 & 29 & -4 & 21
\end{pmatrix}, \quad (D + I)^3 = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & -54 & 27 & -27 & 0 & -27 \\
0 & 108 & -54 & 54 & 0 & 54 \\
0 & 108 & -54 & 54 & 0 & 54 \\
0 & -162 & 81 & -81 & 0 & -81 \\
0 & 162 & -81 & 81 & 0 & 81
\end{pmatrix}.

(5) First, \( V_2 \) must be a 1-dimensional eigenspace. Gaussian elimination finds the eigenvector \( (0, 1, -2, -2, 3, -3)^t \). Next, \( V_{-1} \) must be 5-dimensional. Row-reduction gives: \( D + I \rightarrow \begin{pmatrix}
1 & 0 & 0 & 0 & -1 & -1 \\
0 & 0 & 0 & 1 & -1/2 \\
0 & 1 & 0 & 0 & 1 & 3/2 \\
0 & 0 & 1 & 0 & 2 & 3/2 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}, \quad (D + I)^2 \rightarrow \begin{pmatrix}
2 & 0 & -1 & 3 & -4 & -5 \\
0 & 2 & -1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} \). So the \( \text{Ker}(D + I) \) is two-dimensional (since \( (D + I)^2 \neq 0 \) there will be a block of size at least 3; since \( (D + I)^3 \) has rank one, it has the 5d kernel \( V_{-1} = \{ x \mid x_1 - 2x_2 + x_4 + x_6 \} \) so the largest
block is 3, and so the other block must have size 2. We need a vector from the generalized eigenspace in the image of \((D + I)^2\). Since \((D + I)^3 \vec{e}_1 = 0\) but the first column of \((D + I)^2\) is non-zero, we see that \((D + I)^2 \vec{e}_1 = (1, -2, -1, 1, -1, 2)^t\) has preimage \((D + I) \vec{e}_1 = (1, 0, -3, -2, 1, -1)^t\), and we obtain our first block. Next, we need an eigenvector in the kernel and image of \(D + I\), but any vector in the kernel is also in the image (no blocks of size 1), so we can take any vector in \(\text{Ker}(D + I)\) independent of the one we already have. Using the row-reduced form we see that \((D + I)^2 \vec{e}_1 = (1, -1, -2, 0, 1, 0)^t\) is such a vector. Then we solve

\[
\begin{pmatrix}
1 & 0 & 0 & 0 & -1 & -1 \\
0 & -7 & 4 & -3 & 1 & -3 \\
-3 & 13 & -7 & 6 & 2 & 9 \\
-2 & 14 & -7 & 5 & 2 & 10 \\
1 & -18 & 11 & -11 & 3 & -6 \\
-1 & 19 & -11 & 10 & -2 & 8
\end{pmatrix}
\begin{pmatrix}
1 \\
1 \\
-1 \\
-2 \\
0 \\
0
\end{pmatrix}
= \begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
x_5 \\
x_6
\end{pmatrix},
\]

finding for example the vector \((1, 0, -1, -1, 0, 0)^t\) and our second block. We conclude that in the basis \(
\begin{pmatrix}
0 \\
1 \\
-2 \\
-2 \\
3 \\
-3
\end{pmatrix},
\begin{pmatrix}
1 \\
-2 \\
1 \\
1 \\
-1 \\
1
\end{pmatrix},
\begin{pmatrix}
1 \\
0 \\
-3 \\
-3 \\
1 \\
2
\end{pmatrix},
\begin{pmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{pmatrix},
\begin{pmatrix}
1 \\
1 \\
-1 \\
-1 \\
0 \\
0
\end{pmatrix},
\begin{pmatrix}
1 \\
1 \\
1 \\
1 \\
0 \\
0
\end{pmatrix}
\) the matrix has the form

\[
\begin{pmatrix}
(2) \\
-1 & 1 \\
-1 & 1 \\
-1 & -1 \\
(1 & 0 \\
-1 & 1
\end{pmatrix}
\]

27
CHAPTER 3

Vector and matrix norms

For the rest of the course our field of scalars is either \( \mathbb{R} \) or \( \mathbb{C} \).

3.1. Review of metric spaces

**Definition 135.** A **metric space** is a pair \( (X,d_X) \) where \( X \) is a set, and \( d_X: X \times X \rightarrow \mathbb{R}_{\geq 0} \) is a function such that for all \( x,y,z \in X \), \( d_X(x,y) = 0 \) iff \( x = y \), \( d_X(x,y) = d_X(y,z) \) and (the triangle inequality) \( d_X(x,z) \leq d_X(x,y) + d_X(y,z) \).

**Notation 136.** For \( x \in X \) and \( r \geq 0 \) we write \( B_X(x,r) = \{ y \in X \mid d_X(x,y) \leq r \} \) for the closed ball of radius \( r \) around \( x \), \( B_X^r(x,r) = \{ y \in X \mid d_X(x,y) < r \} \) for the open ball.

**Definition 137.** Let \( (X,d_X),(Y,d_Y) \) be metric spaces and let \( f: X \rightarrow Y \) be a function.

1. We say \( f \) is **continuous** if \( \forall x \in X : \forall \varepsilon > 0 : \exists \delta > 0 : \forall x \in X \mid d_X(x,\delta) \subset B_Y(f(x),\varepsilon) \).
2. We say \( f \) is **uniformly continuous** \( \forall \varepsilon > 0 : \exists \delta > 0 : \forall x \in X : d_X(x,\delta) \subset B_Y(f(x),\varepsilon) \).
3. We say \( f \) is **Lipschitz continuous** if in (2) we can take \( \delta = L\varepsilon \), in other words if for all \( x \neq x' \in X \),

\[
\frac{d_Y(f(x),f(x'))}{d_X(x,x')} \leq L
\]

In that case we let \( \|f\|_{\text{Lip}} \) denote the smallest \( L \) for which this holds.

Clearly (3) \( \Rightarrow \) (2) \( \Rightarrow \) (1).

**Lemma 138.** The composition of two functions of type (1),(2),(3) is again a function of that type. In particular, \( \|f \circ g\|_{\text{Lip}} \leq \|f\|_{\text{Lip}} \|g\|_{\text{Lip}} \).

**Definition 139.** We call the metric space \( (X,d_X) \) **complete** if every Cauchy sequence converges.

3.2. Norms on vector spaces

Fix a vector space \( V \).

**Definition 140.** A **norm** on \( V \) is a function \( \| \cdot \| : V \rightarrow \mathbb{R}_{\geq 0} \) such that \( \|v\| = 0 \) iff \( v = 0 \), \( \|\alpha v\| = |\alpha| \|v\| \) and \( \|u + v\| \leq \|u\| + \|v\| \). A **normed space** is a pair \( (V,\|\cdot\|) \).

**Lemma 141.** Let \( \|\cdot\| \) be a norm on \( V \). Then the function \( d(u,v) = \|u - v\| \) is a metric.

**Exercise 142.** The map \( \|\cdot\| \rightarrow d \) is a bijection between norms on \( V \) and metrics on \( V \) which are (1) translation-invariant \( d(u,v) = d(u + w,v + w) \) and (2) 1-homogenous: \( d(\alpha u,\alpha v) = |\alpha| d(u,v) \).

The restriction of a norm to a subspace is a norm.
3.2.1. Finite-dimensional examples.

Example 143. Standard norms on \( \mathbb{R}^n \) and \( \mathbb{C}^n \):

1. The supremum norm \( \|v\|_\infty = \max \{ |v_i| \}_{i=1}^n \), parametrizing uniform convergence.
2. \( \|v\|_1 = \sum_{i=1}^n |v_i| \).
3. The Euclidean norm \( \|v\|_2 = \left( \sum_{i=1}^n |v_i|^2 \right)^{1/2} \), connected to the inner product \( \langle u, v \rangle = \sum_{i=1}^n u_i v_i \) (prove \( \Delta \) inequality from this by squaring norm of sum).
4. For \( 1 < p < \infty \), \( \|v\|_p = \left( \sum_{i=1}^n |v_i|^p \right)^{1/p} \).

Proof. These functions are clearly homogeneous, and clearly are non-zero if \( v \neq 0 \); the only non-trivial part is the triangle inequality (“Minkowsky’s inequality”). This is easy for \( p = 1, \infty \), well-known for \( p = 2 \). Other cases left to supplement.

Exercise 144. Show that \( \lim_{p \to \infty} \|v\|_p = \|v\|_\infty \).

We have a geometric interpretation. The unit ball of a norm is the set \( B = B(0, 1) = \{ v \in V | \|v\| \leq 1 \} \). This determines the norm ( \( \frac{1}{\|v\|} \) is the largest \( \alpha \) such that \( \alpha v \in B \)). Now applying a linear map to \( B \) gives a the ball of a new norm.

Exercise 145. Draw the unit balls of

Proposition 146 (Pullback). Let \( T : U \to V \) be an injectivel linear map. Let \( \|\cdot\|_V \) be a norm on \( V \). Then \( \|u\|_U \overset{\text{def}}{=} \|Tu\|_V \) defines a norm on \( U \).

Proof. Easy check.

3.2.2. Infinite-dimensional examples. Now the norm comes first, the space second.

Example 147. For a set \( X \) let \( \ell^\infty(X) = \{ f \in F^X | \sup \{ |f(x)| : x \in X \} < \infty \} \), \( \|f\|_\infty = \sup_{x \in X} |f(x)| \).

Proof. The map \( \|\cdot\|_\infty : F^X \to [0, \infty] \) satisfies the axioms of a norm, suitably extended to include the value \( \infty \). That the set of vectors of finite norm is a subspace follows from the scaling and triangle inequalities.

Remark 148. A vector space with basis \( B \) can be embedded into \( \ell^\infty(B) \) (we’ve basically seen this).

Example 149. \( \ell^p(\mathbb{N}) = \{ a \in F^\mathbb{N} : \sum_{i=1}^\infty |a_i|^p < \infty \} \) with the obvious norm.

In the continuous case we a construction from earlier in the course:

Definition 150. \( L^p(\mathbb{R}) = \{ f : \mathbb{R} \to F \text{ measurable} | \int_{\mathbb{R}} |f(x)|^p \, dx < \infty \} \) with the natural norm.

Remark 151. The quotient is essential: for actualy functions, can have \( \int |f(x)|^p \, dx = 0 \) without \( f = 0 \) exactly. In particular, elements of \( L^p(\mathbb{R}) \) don’t have specific values.

Fact 152. In each equivalence class in \( L^p(\mathbb{R}) \) there is at most one continuous representative.

So part of PDE is about whether an \( L^p \) solution can be promoted to a continuous function. We give an example theorem:

Theorem 153 (Elliptic regularity). Let \( \Omega \subset \mathbb{R}^2 \) be a domain, and let \( f \in L^2(\Omega) \) satisfy \( \Delta f = \lambda f \) distributionally: for \( g \in C^\infty_c(\Omega) \), \( \int_\Omega \Delta f g = \lambda \int f g \). Then there is a smooth function \( \tilde{f} \) such that \( \Delta \tilde{f} = \lambda \tilde{f} \) pointwise and such that \( f = \tilde{f} \) almost everywhere.
3.2.3. **Converges in the norm.** While there are many norms on $\mathbb{R}^n$, it turns out that there is only one notion of convergence.

**Lemma 154.** Every norm on $\mathbb{R}^n$ is a continuous function.

**Proof.** Let $M = \max_i \| e_i \|$. Then
\[
\| x \| = \left\| \sum_{i=1}^n x_i e_i \right\| \leq \sum_{i=1}^n |x_i| \| e_i \| \leq M \| x \|_1.
\]
In particular,
\[
\| x \| - \| y \| \leq \| x - y \| \leq M \| x - y \|_1.
\]

**Definition 155.** Call two norms equivalent if there are $0 < m \leq M$ such that $m \| x \| \leq \| x \|' \leq M \| x \|$ holds for all $x \in V$.

**Exercise 156.** This is an equivalence relation. The norms are equivalent iff the same sequences of vectors satisfy $\lim_{n \to \infty} x_n = 0$.

**Theorem 157.** All norms on $\mathbb{R}^n$ (and $\mathbb{C}^n$) are equivalent.

**Proof.** It is enough to show that they are all equivalent to $\| \cdot \|_1$. Accordingly let $\| \cdot \|$ be any other norm. Then the Lemma shows that there is $M$ such that
\[
\| x \| \leq M \| x \|_1.
\]
Next, the “sphere” $\{ x \mid \| x \|_1 = 1 \}$ is closed and bounded, hence compact. Accordingly let $m = \min \{ \| x \| \mid \| x \|_1 = 1 \}$. Then $m > 0$ since $\| 0 \|_1 = 0 \neq 1$. Finally, for any $x \neq 0$ we have
\[
\frac{\| x \|}{\| x \|_1} = \frac{x}{\| x \|_1} \geq m
\]
since $\frac{\| x \|}{\| x \|_1} = 1$. It follows that
\[
m \| x \|_1 \leq q \| x \| \leq M \| x \|_1.
\]

3.3. **Norms on matrices**

**Definition 158.** Let $U, V$ be normed spaces. A map $T : U \to V$ is called bounded if there is $M \geq 0$ such that $\| Tu \|_V \leq M \| u \|_U$ for all $u \in U$. The smallest such $M$ is called the (operator) norm of $T$.

**Remark 159.** Motivation: Let $U$ be the space of initial data for an evolution equation (say wave, or heat). Let $V$ be the space of possible states at time $t$. Let $T$ be “time evolution”. Then a key part of PDE is finding norms in which $T$ is bounded as a map from $U$ to $V$. This shows that solution exist, and that they are unique.

**Example 160.** The identity map has norm 1. Now consider the matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ acting on $\mathbb{R}^2$. 

30
(1) As a map from $\ell^1 \to \ell^1$ we have
\[
\| A \begin{pmatrix} x \\ y \end{pmatrix} \|_1 = |x + y| + |y| \leq 2 \left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\|_1,
\]
with equality if $x = 0$. Thus $\|A\|_1 = 2$.

(2) Next,
\[
\left\| A \begin{pmatrix} x \\ y \end{pmatrix} \right\|_2^2 = |x + y|^2 + |y|^2 \leq \frac{3 + \sqrt{5}}{2} |x^2 + y^2|.
\]

(3) Finally,
\[
\left\| A \begin{pmatrix} x \\ y \end{pmatrix} \right\|_\infty = \max \{ |x + y|, |y| \} \leq 2 \max \{ |x|, |y| \}
\]
with equality if $x = y$, thus $\|A\|_\infty = 2$.

**Example 161.** Consider $D_\chi : C_c^\infty (\mathbb{R}) \to C_c^\infty (\mathbb{R})$. This is not bounded in any norm (consider $f(x) = e^{2\pi i kx}$).

**Lemma 162.** Every map of finite-dimensional spaces is bounded.

**Proof.** Identify $U$ with $\mathbb{R}^n$. Then the $\|\cdot\|_U$ is equivalent with $\|\cdot\|_1$, so there is $A$ such that $\|u\|_1 \leq A \|u\|_U$. Now the map $u \mapsto \|Tu\|_V$ is 1-homogenous and satisfies the triangle inequality, so by the proof of Lemma 154 there is $B$ so that $\|Tu\|_V \leq B \|u\|_1 \leq (AB) \|u\|_U$. □

**Lemma 163.** Let $S, T$ be bounded and composable. Then $ST$ is bounded and $\|ST\| \leq \|S\| \|T\|$.

**Proof.** For any $u \in U$, $\|STu\|_W \leq \|S\| \|Tu\|_V \leq \|S\| \|T\| \|u\|_U$. □

**Proposition 164.** The operator norm is a norm on $\text{Hom}_b(U, V)$, the space of bounded maps $U \to V$.

**Proof.** For any $S, T \in \text{Hom}_b(U, V)$, $|\alpha| \|T\| + \|S\|$ is a bound for $\alpha T + S$. Since the zero map is bounded it follows that $\text{Hom}_b(U, V) \subset \text{Hom}(U, V)$ is a subspace, and setting $\alpha = 1$ gives the triangle inequality. If $T \neq 0$ then there is $u$ such that $Tu \neq 0$ at which point
\[
\|T\| \geq \frac{\|Tu\|}{\|u\|} > 0.
\]
Finally, $\| (\alpha T) u \| = |\alpha| \|Tu\| \leq |\alpha| \|T\| \|u\|$ so $\|\alpha T\| \leq |\alpha| \|T\|$. But then
\[
\|T\| = \left\| \frac{1}{|\alpha|} \alpha T \right\| \leq \frac{1}{|\alpha|} \|\alpha T\|
\]
gives the reverse inequality. □

### 3.4. Example: eigenvalues and the power method (Lecture, 17/)

Let $A$ be diagonalable. Want eigenvalues of $A$. Raising $A$ to large powers selects the eigenvalue with largest component.

- Algorithm: multiply by $A$ and renormalize.
- Advantage: if $A$ sparse only need to multiply by $A$.
- Rate of convergence related to spectral gap.
3.5. Sequences and series of vectors and matrices

3.5.1. Completeness (Lecture 19/3/2014).

DEFINITION 165. A metric space is complete if

EXAMPLE 166. \( \mathbb{R}, \mathbb{R}^n \) in any norm. \( \text{Hom}(\mathbb{R}^n, \mathbb{R}^m) \) (because isom to \( \mathbb{R}^{mn} \)).

FACT 167. Any metric space has a completion. [note associated universal property and hence uniqueness]

THEOREM 168. Let \((U, \| \cdot \|_U), (V, \| \cdot \|_V)\) be normed spaces with \( V \) complete. Then \( \text{Hom}_b(U, V) \) is complete with respect to the operator norm.

PROOF. Let \( \{T_n\}_{n=1}^\infty \) be a Cauchy sequences of linear maps. For fixed \( u \in U \), the sequence \( \{T_nu\} \) is Cauchy: \( \| (T_nu - T_mu) \|_V \leq \| T_n - T_m \|_U \| u \|. \) It is therefore convergent – call the limit \( T_u \). This is linear since \( \alpha T_nu + T_mu' \) converges to \( \alpha Tu + Tu' \) while \( T_n (\alpha u + u') \) converges to \( T (\alpha u + u') \).

Since \( \| T_n \| - \| T_m \| \leq \| T_n - T_m \| \), the norms themselves are a Cauchy sequences of real numbers, in particular a convergent sequence. Now for fixed \( u \), we have \( \| Tu \|_V = \lim_{n \to \infty} \| T_n u \|_V \). We have the pointwise bound \( \| T_n u \| \leq \| T_n \| \| u \|_U \). Passing to the limit we find

\[
\| Tu \|_V \leq \left( \lim_{n \to \infty} \| T_n \| \right) \| u \|_U
\]

so \( T \) is bounded. Finally, given \( \varepsilon \) let \( N \) be such that if \( m, n \geq N \) then \( \| T_n - T_m \| \leq \varepsilon \). Then for any \( u \in U \),

\[
\| T_n u - T_m u \| \leq \| T_n - T_m \| \| u \|_U \leq \varepsilon \| u \|_U.
\]

Letting \( m \to \infty \) and using the continuity of the norm, we get that if \( n \geq N \) then

\[
\| T_n u - Tu \| \leq \varepsilon \| u \|_U.
\]

Since \( u \) was arbitrary this shows that \( \| T_n - T \| = \varepsilon \) for \( n \geq N \) and we are done. \( \square \)

EXAMPLE 169. Let \( K \) be a compact space. Then \( C(K) \), the space of continuous functions on \( K \), is complete w.r.t \( \| \cdot \|_\infty \).

PROOF. Continuous functions on a compact space are bounded. Let \( \{f_n\}_{n=1}^\infty \subset C(K) \) be a Cauchy sequence. Then for fixed \( x \in X \), \( \{f_n(x)\}_{n=1}^\infty \subset \mathbb{C} \) is a Cauchy sequence, hence convergent to some \( f(x) \in \mathbb{C} \). To see the convergence is in the norm, give \( \varepsilon > 0 \) let \( N \) be such that \( \| f_n - f_m \|_\infty \leq \varepsilon \) for \( n, m \geq N \). Then for any \( x \),

\[
|f_n(x) - f_m(x)| \leq \varepsilon.
\]

Letting \( m \to \infty \) we find for all \( n \leq N \) that \( |f_n(x) - f(x)| \leq \varepsilon \), that is

\[
\| f_n - f \|_\infty \leq \varepsilon.
\]

Finally, we need to show that \( f \) is continuous. Given \( x \in X \) and \( \varepsilon > 0 \) let \( N \) be as above and let \( n \geq N \). For any \( x \), the continuity of \( f_n \) gives a neighbourhood of \( x \) where \( |f_n(x) - f_n(y)| \leq \varepsilon \). Then

\[
|f(x) - f(y)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| \leq 3\varepsilon
\]

in that neighbourhood, so \( f \) is continuous at \( x \). \( \square \)
3.5.2. Series of vectors and matrices (Lecture 21/3/2014). Fix a complete normed space $V$.

**Definition 170.** Say the series $\sum_{n=1}^{\infty} v_n$ converges *absolutely* if $\sum_{n=1}^{\infty} \|v_n\|_V < \infty$.

**Proposition 171.** If $\sum_{n=1}^{\infty} v_n$ converges absolutely it converges, and $\|\sum_{n=1}^{\infty} v_n\|_V \leq \sum_{n=1}^{\infty} \|v_n\|_V$.

**Proof.** Standard. \qed

**Theorem 172 (M-test).** Let $X$ be a (topological) space, $f_n : X \to V$ continuous. Suppose that we have $M_n$ such that $\|f_n(x)\|_V \leq M_n$ holds for all $x \in X$. Suppose that $M = \sum_{n=1}^{\infty} M_n < \infty$. Then $\sum_n f_n$ converges uniformly to a continuous function $F : X \to V$.

**Remark 173.** This can be interpreted as $C_b(X, V)$ (continuous functions $X \to V$ with $\|f(x)\|_V$ bounded) being complete with respect to the norm $\|f\|_\infty = \sup \{ \|f(x)\|_V : x \in X \}$.

We will apply this to power series of matrices.

**Example 174.** Let $\|\cdot\|$ be some operator norm on $M_n(\mathbb{R})$, and let $A \in M_n(\mathbb{R})$. For $0 < T < \frac{1}{\|A\|}$ (any $T > 0$ if $A = 0$) and $z \in \mathbb{C}$ with $|z| \leq T$ consider the series

$$\sum_{n=0}^{\infty} z^n A^n.$$ 

We have $\|A^n\| \leq \|A\|^n$ (operator norm!) so that $\|z^n A^n\| \leq (T \|A\|)^n$. Since $\sum_{n=0}^{\infty} (T \|A\|)^n$ converges, we see that our series converges and the sum is continuous in $z$ (and in $A$). Taking the union we get convergence in $|z| < \frac{1}{\|A\|}$. The limit is $(\text{Id} - zA)^{-1}$ (incidentally showing this is invertible).

**Remark 175.** In fact, the radius of convergence is $\frac{1}{\rho(A)}$.


**Lemma 176 (Limit arithmetic).** Let $U, V, W$ be normed spaces. Let $u_f(x) : X \to U$, $\alpha_f(x) : X \to F$, $T_f(x) : X \to \text{Hom}_F(U, V)$, $S_f(x) : X \to \text{Hom}_F(V, W)$. Then, in each case supposing the limits on the right exist, the limits on the left exist and equality holds:

1. $\lim_{x \to x_0} (\alpha_f(x) u_f(x) + \alpha_2(x) u_2(x)) = (\lim_{x \to x_0} \alpha_1(x)) (\lim_{x \to x_0} u_1(x)) + (\lim_{x \to x_0} \alpha_2(x)) (\lim_{x \to x_0} u_2(x))$.
2. $\lim_{x \to x_0} T_f(x) u_f(x) = (\lim_{x \to x_0} T_f(x)) (\lim_{x \to x_0} u_f(x))$.
3. $\lim_{x \to x_0} S_f(x) T_f(x) = (\lim_{x \to x_0} S_f(x)) (\lim_{x \to x_0} T_f(x))$.

**Proof.** Same as in $R$, replacing $|\cdot|$ with $\|\cdot\|_V$. \qed

We can also differentiate vector-valued functions (see Math 320 for details)

**Definition 177.** Let $X \subset \mathbb{R}^n$ be open. Say that $f : X \to V$ is *strongly differentiable* at $x_0$ if there is a bounded linear map $L : \mathbb{R}^n \to V$ such that

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - Lh\|_V}{\|h\|_{\mathbb{R}^n}} = 0.$$ 

In that case we write $Df(x_0)$ for $L$.

It is clear that differentiability at $x_0$ implies continuity at $x_0$. 33
Lemma 178 (Derivatives). Let $U, V, W$ be normed spaces. Let $u_j(x): X \to U$, $T(x): X \to \text{Hom}_b(U, V)$, $S(x): X \to \text{Hom}_b(V, W)$ be differentiable at $x_0$. Then the derivatives on the left exist and take the following values:

1. $D(u_1 + u_2)(x_0) = Du_1(x_0) + Du_2(x_0)$.
2. $D(Tu)(x_0)(h) = (DT(x_0)(h) \cdot u(x_0)) + T(x_0) \cdot Du(x_0)(h)$.
3. $D(ST)(x_0)(h) = (DS(x_0)(h) \cdot T(x_0)) + (S(x_0) \cdot DT(x_0)(h))$.

Proof. Same as in $R$, replacing $|\cdot|$ with $\|\cdot\|_V$. □
The Holomorphic Calculus

4.1. The exponential series (24/3/2014)

We prove in the last lecture:

**Theorem 179.** \( f_n : X \to V \) cts, \( \| f_n(x) \|_V \leq M_n \). Then if \( \sum_n M_n < \infty \), \( \sum_n f_n \) converges uniformly to a cts function \( X \to V \).

We apply this to power series:

**Corollary 180.** Let \( \sum_n a_n z^n \) be a power series with radius of convergence \( R \). Then \( \sum_n a_n A^n \) converges absolutely if \( \| A \| < R \), uniformly in \( \{ \| A \| \leq R - \varepsilon \} \).

**Proof.** Let \( X = V = \text{End}_b(V) \), \( f_n(A) = a_n A^n \), so that \( \| f_n(A) \| \leq |a_n| \| A \|^n \). For \( T < R \) we have \( \sum_n |a_n| T^n < \infty \) and hence uniform convergence in \( \{ \| A \| \leq T \} \). \( \square \)

We therefore fix a normed space \( V \), and plug matrices \( A \in \text{End}_b(V) \) into power series.

**Example 181.** \( \exp(A) = \sum k A^k / k! \) converges everywhere.

**Lemma 182.** \( \exp(tA) \exp(sA) = \exp((t+s)A) \).

**Proof.** The series converge absolutely, so the product converges in any order. We thus have

\[
\exp(tA) \exp(sA) = \left( \sum_{k=0}^{\infty} \frac{(tA)^k}{k!} \right) \left( \sum_{l=0}^{\infty} \frac{(sA)^l}{l!} \right) = \sum_{k,l} \frac{t^k s^l A^{k+l}}{k! l!} \\
= \sum_{m=0}^{\infty} \sum_{k+l=m} \frac{t^k s^l A^{k+l}}{k! l!} = \sum_{m=0}^{\infty} A^m \sum_{k+l=m} \frac{m!}{k! l!} t^k s^l \\
= \sum_{m=0}^{\infty} A^m \frac{(t+s)^m}{m!} = \exp((t+s)A)
\]

\( \square \)

**Corollary 183.** \( \frac{d}{dt} \exp(tA) = A \exp(tA) = \exp(tA) A \).

**Proof.** At \( t = 0 \) we have \( \frac{\exp(hA) - \text{Id}}{h} = A + \sum_{k=1}^{\infty} \frac{h^k}{(k+1)!} A^{k+1} \) and

\[
\left\| \sum_{k=1}^{\infty} \frac{h^k}{(k+1)!} A^{k+1} \right\| \leq \sum_{k=1}^{\infty} \frac{|h|^k}{(k+1)!} \| A \|^k \leq \exp(|h| \| A \|) - 1 - \| A \| |h| \to 0 \quad (h \to 0).
\]

In general we have

\[
\frac{\exp((t+h)A) - \exp(tA)}{h} = \exp(tA) \frac{\exp(hA) - \text{Id}}{h} \to \exp(tA) A
\]

That \( A \exp(tA) = \exp(tA) A \). \( \square \)
4.1.1. Application: differential equations with constant coefficients. Consider the system of differential equations
\[ \begin{cases} \frac{d}{dt} \mathbf{v}(t) = A \mathbf{v}(t) \\ \mathbf{v}(0) = \mathbf{v}_0 \end{cases} \]
where \( A \) is a bounded map.

**Proposition 184.** The system has the unique solution \( \mathbf{v}(t) = \exp(At) \mathbf{v}_0 \).

**Proof.** We saw \( \frac{d}{dt} \exp(At) \mathbf{v}_0 = A (\exp(At) \mathbf{v}_0) \). Conversely, suppose \( \mathbf{v}(t) \) is any solution. Then
\[ \frac{d}{dt} \left( e^{-At} \mathbf{v}(t) \right) = \left( e^{-At} (-A) \right) (\mathbf{v}(t)) + \left( e^{-At} \right) (A \mathbf{v}(t)) \]
\[ = e^{-At} (-A + A) \mathbf{v}(t) = 0. \]
It remains to prove: □

**Lemma 185.** Let \( f: [0,1] \to V \) be strongly differentiable. If \( f'(t) = 0 \) for all \( t \) then \( f \) is constant.

**Proof.** Suppose \( f(t_0) \neq f(0) \). Let \( \varphi \in V' \) be a bounded linear functional such that \( \varphi (f(t_0) - f(0)) \neq 0 \). Then \( \varphi \circ f: [0,1] \to \mathbb{R} \) is differentiable and its derivative is 0:
\[ \lim_{h \to 0} \frac{\varphi (f(t+h)) - \varphi (f(t))}{h} = \lim_{h \to 0} \varphi \left( \frac{f(t+h) - f(t)}{h} \right) = \varphi \left( \lim_{h \to 0} \frac{f(t+h) - f(t)}{h} \right) = \varphi (f'(t)). \]
But \( (\varphi \circ f)(t_0) - (\varphi \circ f)(0) = \varphi (f(t_0) - f(0)) \neq 0 \), a contradiction. □

**Remark 186.** If \( V \) is finite-dimensional, every linear functional is bounded. If \( V \) is infinite-dimensional the existence of \( \varphi \) is a serious fact.

Now consider a linear ODE with constant coefficients:
\[ \begin{cases} \frac{d^n}{dt^n} u(t) = \sum_{k=0}^{n-1} a_k u^{(k)}(t) \\ u^{(k)}(0) = w_k \quad 0 \leq k \leq n-1. \end{cases} \]
We solve this system via the auxiliary vector
\[ \mathcal{v}(t) = \left( u(t), u'(t), \ldots, u^{(n-1)}(t) \right). \]
We then have
\[ \frac{d\mathcal{v}(t)}{dt} = A \mathcal{v} \]
where \( A \) is the companion matrix
\[ A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} \end{pmatrix}. \]
(companion to the polynomial \( x^n - \sum_{k=0}^{n-1} a_k x^k \)). It follows that
\[ \mathcal{v}(t) = e^{At} \mathcal{w}. \]

**Idea:** bring \( A \) to Jordan form so easier to take exponential.
4.1.2. Diagonal matrices. HW: \( f(\text{diag}(a_1, \ldots, a_n)) = \text{diag}(f(a_1), \ldots, f(a_n)) \).

4.2. 26/3/2014

**Definition 187.** Let \( f(z) = \sum_{n} a_n z^n \). Define \( f(A) = \sum_{n=0}^{\infty} a_n A^n \).

**Lemma 188.** \( Sf(A)S^{-1} = f(SAS^{-1}) \).

**Proposition 189.** \( f \circ g)(A) = f(g(A)) \) if it all works.

**Theorem 190.** \( \det(\exp(A)) = \exp(\text{Tr}(A)) \).

4.3. Invertibility and the resolvent (31/3/2014)

Say we have a matrix \( A \) we’d like to invert. Idea: write \( A = D + E \) where we know to invert \( D \).

Then \( A = D(I + D^{-1}E) \), so if \( \|D^{-1}E\| < 1 \) we have

\[
(I + D^{-1}E)^{-1} = \sum_{n=0}^{\infty} (-D^{-1}E)^n
\]

and

\[
A^{-1} = \sum_{n=0}^{\infty} (-D^{-1}E)^n D^{-1}
\]

(in particular, \( A \) is invertible).

4.3.1. Application: Gauss-Seidel and Jacobi iteration.

4.3.2. Application: the resolvent. Let \( V \) be a complete normed space. Let \( T \) be an operator on \( V \). Define the resolvent set of \( T \) to be the set of \( z \in \mathbb{C} \) for which \( T - z \text{Id} \) has a bounded inverse. Define the spectrum \( \sigma(T) \) to be the complement of the resolvent set. This contains the actual eigenvalues (\( \lambda \) such that \( \text{Ker}(T - \lambda) \) is non-trivial) but also \( \lambda \) where \( T - \lambda \) is not surjective, and \( \lambda \) where an inverse to \( T - \lambda \) exists but is unbounded).

**Theorem 191.** The resolvent set is open, and the function ("resolvent function") \( \rho(T) \to \text{End}_b(V) \) given by \( z \mapsto R(z) = (z \text{Id} - T)^{-1} \) is holomorphic.

**Proof.** Suppose \( z_0 - T \) has a bounded inverse. We need to invert \( z - T \) for \( z \) close to \( z_0 \). Indeed, if \( |z - z_0| < \frac{1}{\|z_0 - T\|} \) then

\[
-\sum_{n=0}^{\infty} (T - z_0)^{n+1} (z - z_0)^n
\]

converges and furnishes the requisite inverse. It is evidently holomorphic in \( z \) in the indicated ball. \( \square \)

**Example 192.** Let \( \Omega \subset \mathbb{R}^2 \) be a bounded domain with nice boundary, \( \Delta = \frac{d^2}{dx^2} + \frac{d^2}{dy^2} \) the Laplace operator (say defined on \( f \in C^\infty(\Omega) \) vanishing on the boundary). Then \( \Delta \) is unbounded, but its resolvent is nice. For example, \( R(i\varepsilon) \) only has eigenvalues. It follows that the spectrum of \( \Delta \) consists of eigenvalues, that is for \( \lambda \in \sigma(\Delta) \) there is \( f \in L^2(\Omega) \) with \( \Delta f = \lambda f \) (and \( f \in C^\infty \) by elliptic regularity).
CHAPTER 5

Vignettes

Sketches of applications of linear algebra to group theory.
Key Idea: linearization – use linear tools to study non-linear objects.

5.1. The exponential map and structure theory for GL\(_n\)(\(\mathbb{R}\)) (2/4/2014)

Our goal is to understand the (topologically) closed subgroups of \(G = \text{GL}_n(\mathbb{R})\).
Idea: to a subgroup \(H\) assign the logarithms of the elements of \(H\). If \(H\) was commutative this would be a subspace.

**Definition 193.** \(\text{Lie}(H) = \{X \in M_n(\mathbb{R}) \mid \forall t : \exp(tX) \in H\}\).

**Remark 194.** Clearly this is invariant under scaling. In fact, enough to take small \(t\), and even just a sequence of \(t\) tending to zero (since \(\{t \mid \exp(tX) \in H\}\) is a closed subgroup of \(\mathbb{R}\)).

**Theorem 195.** \(\text{Lie}(H)\) is a subspace of \(M_n(\mathbb{R})\), closed under \([X,Y]\).

**Proof.** For \(t \in \mathbb{R}\) and \(m \in \mathbb{Z}_{\geq 1}\), \(\left(\exp\left(\frac{tX}{m}\right) \exp\left(\frac{tY}{m}\right)\right)^m = \left(\text{Id} + \frac{tX + tY}{m} + O\left(\frac{1}{m^2}\right)\right)^m \xrightarrow{m \to \infty} \exp(tX + tY)\).
Thus If \(X,Y \in \text{Lie}(H)\) then also \(X + Y \in \text{Lie}(H)\). \(\square\)

**Theorem 196.** Bijection between closed connected subgroups of \(G\) and subalgebras of the Lie algebra.

Classify subgroups of \(G\) containing \(A\) by action on Lie algebra and finding eigenspaces.

5.2. Representation Theory of Groups

**Example 197 (Representations).**
1. Structure of \(\text{GL}_n(\mathbb{R})\): let \(A\) act on \(M_n(\mathbb{R})\).
2. \(M\) manifold, \(G\) acting on \(M\), thus acting on \(H_k(M)\) and \(H^k(M)\).
3. Angular momentum: \(O(3)\) acting by rotation on \(L^2(\mathbb{R}^3)\).
Bibliography