MATH 227

LINEAR ALGEBRA II

NOLAN ZUREK

Topic 1 - Determinants

Lecture 1

Carrol determinant formula:
$\det A \times \boxed{A} = \text{bottom-left-removed} \times \text{top-right-removed} - \text{bottom-right-removed} \times \text{top-left-removed}$

- Here, $\boxed{A}$ means that the first and last columns and rows have been removed

$M_{n\times n}(\mathbb{F})$ is a *ring* and *group under addition*

$\mathrm{GL}_n(\mathbb{F})$ is the **group of invertible matrices**, and is a *group under matrix multiplication*

Here, the **field** $\mathbb{F}$ is defined as a *group under addition* ($\mathbb{F}^+$) and $\mathbb{F}^\times = \mathbb{F} \setminus \{0\}$ is a group under multiplication

> **Theorem**
>
> $\det : \mathrm{GL}_n(\mathbb{F}) \to \mathbb{F}^\times$ is a *group homomorphism*

Cofactor expansion is a strategy for computing the determinant

- Move along a row (or column), and for each element, add to the sum $\text{that element} \times \text{det of matrix found by removing that row and col} \times (-1)^{i+j}$
- Formally, $\det A = \sum_{i=1}^{n} A_{ij} C_{ij}$, where C_{ij} is the **cofactor matrix** of A, which is found by removing row i and col j from matrix A.
- Cofactor expansion is inefficient to compute, but conceptually and algebraically useful because it defined the determinant *recursively*. Thus, proofs about the determinant can inherit the structure of the cofactor definition; this is particularly suited to proofs by induction.

Lecture 2

The determinant $\det$ is the *unique* function $M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ that is

1. Linear in each row, i.e. for any row, $\det \begin{bmatrix} c\vec{a} + \vec{b} \\ \vec{r_1} \\ \vdots \end{bmatrix} = c \cdot \det \begin{bmatrix} \vec{a} \\ \vec{r_1} \\ \vdots \end{bmatrix} + \det \begin{bmatrix} \vec{b} \\ \vec{r_1} \\ \vdots \end{bmatrix}$
2. Equals 0 when two rows are equal
3. Maps $I_n \mapsto 1$

Each of these properties can be shown/proven inductively using cofactor expansion. Uniqueness is harder to prove

Lecture 3

For $A \in M_{n\times n}(R)$, $\det A \in R$, where R is a (not necessarily commutative) *ring*. This follows from the ring axioms and the cofactor expansion definition

Adjugate formula: For invertible (i.e. $\det A \neq 0$) $A \in M_{n\times n}(\mathbb{F})$, $A^{-1} = \dfrac{1}{\det A}\mathrm{Adj}(A)$, where $\mathrm{Adj}(A)_{ij}$ is the (j,i)th cofactor matrix C_{ji}, i.e. $\mathrm{Adj}(A) = C^\intercal$

- This motivates the useful and well-known identity $A^{-1} = \dfrac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
  We can use this rule to find a particular entry of $A{-}1$ without calculating all of it: $A^{-1} = \dfrac{1}{\det A}\mathrm{Adj}(A)$ implies that $A^{-1}_{ij} = \dfrac{1}{\det A}\mathrm{Adj}_{ij} = \dfrac{1}{\det A}C_{ji}$

Lecture 4

Proof: $(A \cdot \mathrm{Adj}(A))_{ij} = \displaystyle\sum_{k=1}^{n} A_{ik}C_{jk}$; cofactor expand along jth row of the matrix you get by replacing the jth row of A with its ith row

- $i = j \to$ matrix is just $A \to \det A$
- $i \neq j \to$ matrix has two identical rows $\to \det A = 0$

So, $A \cdot \mathrm{Adj}(A) = \det A \cdot I_n$

For commutative ring R, $A \in M_{n\times n}(R)$ has a *multiplicative inverse* in $M_{n\times n}(R) \iff \det A$ has inverse in $R \to$ then adjugate formula holds

- R is a field $\to$ any member (including $\det A$) is invertible by defn
- $R = \mathbb{Z} \to$ invertible elements are ± 1
- $R = \mathbb{Z}_n \to$ invertible elements are any $\pm m$ that doesn't share factors with n

Proof: ($\to$) If $A^{-1} \in M_{n\times n}(R)$, then $1 = \det AA^{-1} = \det A \det A^{-1}$, so $\det A$ is invertible. ($\leftarrow$) If $\det A$ is invertible, proof follows from adjugate formula

Lecture 5

We can check if a matrix A is invertible in field $\mathbb{F}$ by checking if $\det A$ is invertible in $\mathbb{F}$. E.g. $\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}$ is *not* invertible in $\mathbb{Z}_{12}$ since $-2 = \det\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix} \mid 12$, but it *is* invertible in $\mathbb{Z}_9$ since $-2 \nmid 9$.

- For $\mathbb{Z}_n$, m is invertible in $\mathbb{Z}_n$ if and only if m does not share any factors with n.
- Polynomials can only be invertible if ring/field/etc over which they are defined have the special 0 property; only constant terms can be invertible

Cramer's Rule: For $A\vec{x} = \vec{b}$ (where A is invertible), $x_i = \dfrac{\det \tilde{A}}{\det A}$, where $\tilde{A}$ is the matrix formed by replacing the ith column of A with $\vec{b}$

- Proof: $\vec{x} = A^{-1}\vec{b} = \dfrac{1}{\det A}\mathrm{Adj}(A)\vec{b} \implies x_i = \dfrac{1}{\det A}\sum_{j=1}^{n}\mathrm{Adj}(A)_{ij}b_j = \dfrac{1}{\det A}\sum_{j=1}^{n}C_{ji}b_j$
  Expanding the cofactor formula for $\det\left(A \in M_{3\times 3}(\mathbb{F})\right)$, we get 6 terms, which are all the permutations of having one of each row and column in the indices

Lecture 6

Permutation: *invertible* function $\sigma : \{1, 2, \dots n\} \to \{1, 2, \dots n\}$, i.e. maps from self to self uniquely

Symmetric Group (denoted S_n or $\mathrm{Sym}(n)$): set of all permutations (functions) on set of size n

- Group operation is *function composition* $\sigma_1 \circ \sigma_2 = \sigma_1(\sigma_2(i))$; this always leads to another permutation, since this is equivalent to permuting the set twice in succession
- Alternate representation: $\begin{bmatrix}a_1 & \dots & a_n\\ b_1 & \dots & b_n\end{bmatrix}$ represents $\sigma : a_1 \mapsto b_1, \dots a_n \mapsto b_n$

Symmetric groups are *not abelian (i.e. not commutative)*

Inverses of symmetric groups can be found by flipping the rows, since the "inverse" is just "undoing" the permutation by "moving everything back to where it was"

S_n trivially has $n!$ elements

Permutation formula for determinant: $\det A = \sum_{\sigma \in S_n} [\pm 1] \cdot A_{1,\sigma(1)}A_{2,\sigma(2)}, \dots A_{n,\sigma(n)}$, where $A_{i,j}$ is the (i, j)th entry of $A \in M_{n\times n}(\mathbb{F})$

- This is derived from the cofactor formula
- The "sign" term comes from $\det P(\sigma)$; these must be $\in \{\pm 1\}$ since they are invertible

Permutation matrix: For standard basis vectors $\vec{e_1} \dots \vec{e_n}$, $P(\sigma) = \begin{bmatrix}\vec{e_{\sigma(1)}} & \vec{e_{\sigma(1)}} & \dots & \vec{e_{\sigma(n)}}\end{bmatrix}$

- I.e. $P(\sigma)$ for $\sigma = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}$ is $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

  Multiplication acts like composition: $P(\sigma)P(\sigma') = P(\sigma \circ \sigma')$. This is because, formally, permutation matrices are a *subgroup* of $\mathrm{GL}(\mathbb{Z})$ and $P : S_n \to P(\sigma_n)$ is a *group homomorphism* .

Leibniz formula for determinant $\det A = \sum_{\sigma \in S_n} \det P(\sigma) \cdot A_{1,\sigma(1)} A_{2,\sigma(2)}, \ldots A_{n,\sigma(n)}$

- $\det P(\sigma)$ is the **sign** of σ and is the number of row swaps for $P(\sigma) \to I_n$
- Aside: infinite-loop recursive determinant definition??

Lecture 7

The **transposition** operator τ_{ij} represents the $i \leftrightarrow j$ row swap; the composition $\tau_{i_1 j_1} \circ \tau_{i_2 j_2} \circ \cdots \circ \tau_{i_n j_n}$ represents successive swaps.

Transpositions are the building blocks of S_n

Determinant test: $\vec{v_1} \ldots \vec{v_n} \in \mathbb{F}^n$ form a basis $\iff \det [\vec{v_1} \quad \ldots \quad \vec{v_n}] \neq 0$ ($\iff \vec{v_1} \ldots \vec{v_n}$ are linearly independent) ($\iff \operatorname{span}(\vec{v_1} \ldots \vec{v_n}) = V$)

- This instance of the "make your life easy theorem" is great for checking if members of a vector space form a basis
- *Coordinate* vectors are used to calculate the determinant

Lecture 8

Geometric properties of the determinant: measure of the "parallelogram" defined by sides $\vec{v_1} \ldots \vec{v_n}$ in $\mathbb{R}^n$ is $\det \begin{bmatrix} \vec{v_1} \\ \vdots \\ \vec{v_n} \end{bmatrix}$ (where the vectors $\vec{v_1} \ldots \vec{v_n}$ form the *rows*)

The *linear transformation* $T : \mathbb{R}^n \to \mathbb{R}^n$ increases area by a factor of $\det T$

Topic 2 - Equivalence Classes and Quotient Spaces

Lecture 9

$\sim$ is an **equivalence relation** if it satisfies

1. **Reflexivity**: $a \sim a$
2. **Symmetry**: $a \sim b \implies b \sim a$
3. **Transitivity**: $a \sim b \wedge b \sim c \implies a \sim c$

E.g. $m \sim n \iff 7 \mid (m - n)$ is an equivalence relation

Each $\sim$ splits the domain over which it is defined into **equivalence classes** $\overline{a}$. These sets **partition** (i.e. fully cover in aggregate) the domain D, and are defined as the groups of elements that are all equivalent to each other

- Clearly, each element has only one such class
- For $m \sim n \iff 7 \mid (m - n)$, the equivalence classes are $\{\overline{1} \dots \overline{7}\}$

Inversely, every possible partition has an associated equivalence relation, defined as $a \sim b$ if a and b are in the same partition.

- Aside: what structure is implied here?

Lecture 10

For set X, $X/\sim$ is the partition of X by $\sim$, i.e. X is partitioned into the equivalence classes defined by $\sim$ for elements in X

- $\overline{x}$ is the equivalence class that contains $x \in X$
- E.g. $\mathbb{Z}/m \sim n \iff m \equiv n \mod 3$ is $\{\overline{0}, \overline{1}, \overline{2}\}$; note that $\overline{0} = \overline{3} = \overline{6} \dots$
- E.g. in $\mathbb{R}^3$, $(x, y, z) \sim (x', y', z') \iff z = z'$, $\mathbb{R}/\sim$ has equivalence classes $\overline{(0, 0, z)} = \{(x, y, z) : \forall x, y \in \mathbb{R}\}$ for any $z \in \mathbb{R}$. Note that the basis for $\mathbb{R}/\sim$ is $(0, 0, 1)$. Essentially, we only care about z, so x and y are free to be any value, since they will be equivalent. Thus, the equivalence classes depends only on z, thus the set of equivalence classes is isomorphic to $\mathbb{R}$.

We have a *vector space structure* for equivalence classes over vector spaces; the equivalence classes inherit the structure of the domain.

- Vectors are equivalence classes
- Addition and scaling are closed (and well-defined)

- This structure can also be a field, ring, etc. if the domain has that structure

Lecture 11

> **Quotient Theorem 1**
>
> Let V be a vector space over $\mathbb{F}$, and $S \subseteq V$. Define $\vec{u} \sim \vec{v} \iff \vec{u} - \vec{v} \in S$. Then
>
> 1. $\sim$ is an equivalence equivalence relation if and only if S is a group under vector addition, i.e. is closed under $+$, $\vec{0} \in S$, and $\vec{u} \in S \iff -\vec{u} \in S$
> 2. Define $\overline{\vec{u}} + \overline{\vec{v}} = \overline{\vec{u} + \vec{v}}$ and $k\overline{\vec{u}} = \overline{k\vec{u}}$. Then $V/\sim$ is a vector space if and only if S is a *subspace* of V
>
> Informally, $V/\sim$ "plays nice" iff $\sim$ is an equivalence relation, meaning S is a subgraoup, etc.

Equivalence classes are also known as **coset**s

Quotient of V by S: V/S, defined by $V/\sim$ where for $a, b \in V$, $a \sim b \iff a - b \in S$

- Equivalence classes are thusly defined as $\overline{\vec{v}} = \vec{v} + S$, i.e. for any $\vec{s} \in S$
- V/S is the set of *equivalence classes* induced on V by $\sim$ (S is like a parameter)
- E.g. $\mathbb{Z}_n$ is defined as the quotient $\mathbb{Z}/\{0, 1, \ldots, n-1\}$.

Lecture 12

For set X, equivalence relation $\sim$, and function $f : X \to X$, f is **well-defined** on $X/\sim$ iff $x \sim y \implies f(x) = f(y)$, i.e. two equivalence elements of X must be mapped to the same value by f

> **Quotient Theorem 2**
>
> Let V be a finite-dimensional vector space and S be a subspace. Then V/S has dimension $\dim V - \dim S$.
>
> A basis can be found for V/S by finding a basis for S, then adding necessary vectors to make it a basis for V. The vectors that need to be added are the basis for V/S

Aside: This gives insight into what a quotient space actually is. It is the "complimentary" subspace you get when "forcing the structure of S onto V". This is why vectors that are different in S are the equivalence classes (V/S) in V, by the definition of the equivalence relation. V/S is the space that you can use to "build" V from just S; each equivalence class in V/S "contains" the structure of S. The basis construction is the best way to give insight into what V/S.

- Basically, the "essence" of S is removed from V to form V/S.
- The $\mathbb{Z}_n$ example is another good motivator of understanding

Aside: a semantically better notation for V/S might be $V - S$, since this makes the nature of the relationship between V, S, and V/S more clear

Lecture 13

E.g. if S is the xy plane in $\mathbb{R}^3$, then the equivalence classes for $\mathbb{R}^3/S$ are the planes parallel to the xy plane

Lecture 14

Quotient Theorem 3

We have linear map $T : V \to V/S$ defined by $T(\vec{v}) = \overline{\vec{v}}$.

We know this is linear since the equivalence relation for S is well-defined, i.e.
$T(a\vec{u} + \vec{v}) = \overline{a\vec{u} + \vec{v}} = \cdots = a\overline{\vec{u}} + \overline{\vec{v}} = aT(\vec{u}) + T(\vec{v})$

Aside: I think the number of equivalence classes of V/S are $\#(V/S) = \#V/\#S$, perhaps this is why the notation is is V/S?

E.g. Let S be all $A \in M_{n\times n}(\mathbb{F})$ where $A \sim A^\intercal$, i.e. for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $b = c$. Clearly a basis for S is $\{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\}$, since the transpose "flips" the matrix along the tl-br diagonal, so any element there can be the same. To get a full basis for $M_{n\times n}(\mathbb{F})$, we can add another matrix $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ (there are other options). So, the basis for $M_{n\times n}(\mathbb{F})/S$ is $\{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\}$, and so the equivalence classes are $\overline{\begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix}}$ for all $a \in \mathbb{F}$. We also have from the basis sizes $\dim M_{n\times n}(\mathbb{F})/S = \dim M_{n\times n}(\mathbb{F}) - \dim S$

Lecture 15

We can think of $\overline{\vec{u}} + \overline{\vec{v}} = \overline{\vec{u} + \vec{v}}$ and $k\overline{\vec{u}} = \overline{k\vec{u}}$ in two ways

1. A linear transformation T defined by $T : V \to V/S$ and $T : \vec{v} \mapsto \overline{\vec{v}}$

2. $(\vec{u}+S)+(\vec{v}+S)=(\vec{u}+\vec{v})+(S+S)=\vec{u}+\vec{v}+S$ (since S is closed under addition)

Sum vector space: $V \oplus W=\{(\vec{v}, \vec{w})\}$, i.e. combining the vectors element-wise

$S=\vec{0} \oplus W$ is a subspace of $V \oplus W$, so $V \oplus W /(\vec{0} \oplus W) \cong V$ has equivalence classes $\overline{(\vec{v}, \vec{w})}=\overline{(\vec{v}, 0)}$

So V / S could be called "direct difference" $V \ominus S$, since $(V \oplus W) \ominus(\vec{0} \oplus W) \cong V$

T has image V / S (onto) and kernel S

Quotient Kernel Theorem

Let $T: V \rightarrow W$ be linear. Then $\operatorname{Image}(T) \cong V / \operatorname{ker} T$, implying $\operatorname{dim} \operatorname{Image}(T)=\operatorname{dim} V-\operatorname{dim} \operatorname{ker} T$

- If $\operatorname{dim} V=\operatorname{dim} W$, $\operatorname{dim} \operatorname{ker} T=0$ and $\operatorname{Image}(T)$ has dimension v, so we have $v=v-0$.
- E.g. if T maps down a dimension, then $\operatorname{ker} T$ must be 1 and $\operatorname{Image}(T)=v-1$, etc

Quotient Theorem 1 for Groups

Let G be a group, and $S \subseteq G$. Define relation $g \sim g^{\prime}$ as $g^{-1} g^{\prime} \in S$.

1. $\sim$ is an equivalence relation if and only if S is a subgroup
2. Let S be a subgroup. Define $\bar{g} \bar{h}=\overline{g h}$. This is well-defined and defined a group structure $G / \sim$ if and only if S satisfies $g s g^{-1} \in S$ for all $g \in G$, $s \in S \rightarrow$ sub subgroups are **normal subgroups**

Topic 3 - Eigenstuff

Lecture 17

Let $A \in M_{n\times n}(\mathbb{F})$. A number λ is an **eigenvalue** of A with **eigenvector** $\vec{v}$ if $A\vec{v} = \lambda\vec{v}$; the set $\{\lambda, \vec{v}\}$ is an **eigenpair**.

- Essentially, eigenvector is an input to the transformation defined by A where the transformation simply *scales* the vector $\vec{v}$, namely by λ
- $\{0, \vec{0}\}$ is always trivially an eigenpair

Eigenstuff is widely applicable:

- *Markov processes*: the eigenvalues of a transition matrix representing a *Markov process* are the (probabilistic) eventual state(s) of the system
  - Aside: for Markov processes, all eigenvalues are 1 (because the process doesn't "fizzle out"), and the determinant encodes how fast the eventual state is reached
- *Quantum mechanics*: energy levels of the hydrogen atom are eigenvalues of a $\infty \times \infty$ matrix
- *Dynamical systems*: eigenpairs encode the steady state of the system

We have $A\vec{v} = \lambda\vec{v} \iff (A - \lambda I)\vec{v} = \vec{0}$. So, $\text{Null}(A - \lambda I)$ is the set (and subspace) of eigenvectors corresponding to eigenvalue λ

- If $A - \lambda I$ is invertible, then $\vec{v} = \vec{0}$ is the only solution for that given λ (since otherwise, $\vec{v}$ would be a linear combination of the rows of $A - \lambda I$ summing to $\vec{0}$, implying non-invertibility). This is trivially the case for any λ, so this doesn't tell us anything, so we search for nonzero eigenvectors $\vec{v} \neq \vec{0}$.
- So, if we have solutions with $\vec{v} \neq \vec{0}$ eigenvectors, $A - \lambda I$ can't be invertible, so $\det(A - \lambda I) = 0$
- Aside: this is connected to the characteristic polynomial by the relation $(A - \lambda I)\vec{v} = \vec{0}$ between eigenvalues and eigenvectors. $\det A - \lambda I$ finds the values of λ that make this equation 0 (eigenvalues), whereas $\text{Null}(A - \lambda I)$ finds the corresponding values of $\vec{v}$ that make this equation $\vec{0}$ (eigenvectors).

Thus, to find all eigenvalues λ of A, we solve the polynomial $\det(A - \lambda I) = 0$ to find all *eigenvalues* λ, then find the null space $\text{Null}(A - \lambda I)$ to find all *eigenvectors* $\vec{v}$.

- Aside: this means that the basis $\text{Null}(A - \lambda I)$ is also a basis for the *eigenspace*, so they have the same dimension, i.e. $\dim \text{Null}(A - \lambda I) = \dim E_\lambda$

Lecture 18

The set of *eigenvectors* of A corresponding to *eigenvalue* λ (denoted E_λ) is a *subspace*, namely the **eigenspace** of λ. We know that $E_\lambda = \mathrm{Null}(A - \lambda I)$.

- Proof: We know $\vec{0}$ is trivially an eigenvector for any λ; closure under addition and scalar multiplication follow from the definition $A\vec{v} = \lambda\vec{v}$

The **characteristic polynomial** $C_A(\lambda)$ of A is $\det(A - \lambda I) = 0$; its roots are the eigenvalues of A

We can **diagonalize** A if we can find a *diagonal matrix* D and *invertible matrix* P such that $P^{-1}AP = D \iff PDP^{-1} = A$.

- We have $P^{-1}AP = D \iff AP = PD$, so for *diagonal* $D = \begin{bmatrix} d_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d_n \end{bmatrix}$ and $P = [\vec{v}_1 \quad \dots \quad \vec{v}_n]$, we get $A\vec{v}_i = d_i\vec{v}_i$. So, the diagonal entries of D are the *eigenvalues* of A, and the column vector entries of P are the *corresponding eigenvectors* of A

Lecture 19

The **algebraic multiplicity** a_λ of *eigenvalue* λ is its *multiplicity* as a root of the *characteristic polynomial* $\det(A - \lambda I) = 0$

The **geometric multiplicity** g_λ of *eigenvalue* λ is defined as $\dim E_{\lambda_i} = \dim \mathrm{Null}(A - \lambda I)$, i.e. the number of vectors that form a basis for the *eigenspace* E_{λ_i}

For any eigenvalue of any matrix, it is true that $a_{\lambda_i} \geq b_{\lambda_i}$.

Diagonalizability Criterion: If some *characteristic polynomial* for matrix A factors completely into linear factors, then A is diagonalizable if and only if the *geometric multiplicity* and the *algebraic multiplicity* are equal

- Eigenvectors for different eigenvalues are linearly independent, since if this weren't the case, we've have $A\vec{v} = \lambda_1\vec{v}$ and $A\vec{w} = \lambda_2\vec{w}$ where $\vec{w} = c\vec{v}$, so $Ac\vec{v} = \lambda_2 c\vec{v} \implies A\vec{v} = \lambda_2\vec{v}$ by dividing by c, contradicting $\lambda_1 \neq \lambda_2$
- If A is diagonalizable, i.e. $A = PDP^{-1}$, then A and D have the same determinant (follows from the structure of $A = PDP^{-1}$), so $\det A - \lambda I = \det D - \lambda I = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\dots(\lambda_n - \lambda)$, where each $(\lambda_i - \lambda)$ happens g_{λ_i} times. This represents the algebraic multiplicity exactly (since it's the characteristic polynomial), so the two are equal

We always have $\sum_\lambda a_\lambda = n$ (i.e. $\det A - \lambda I$ has degree n, so an $n \times n$ matrix has n eigenvalues), but this isn't always the case for g_{λ_i}.

- E.g. $\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ has $\lambda = \{3\}$, so $a_\lambda = 2$. $E_\lambda = \mathrm{Null}\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\dim E_\lambda = 2$. However, $\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ also has $\lambda = \{3\}$, so $a_\lambda = 2$, but $E_\lambda = \mathrm{Null}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, $\dim E_\lambda = 1$.

Corollary: if all the roots of a characteristic polynomial are distinct, the matrix is diagonalizable, implying almost all matrices are diagonalizable

Lecture 20

Some matrices may not be *diagonalizable* over their own *field*. In particular, there exist matrices in $\mathbb{R}$ that can only be diagonalized in $\mathbb{C}$, e.g. $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ has *characteristic polynomial* $C_A(\lambda) = \lambda^2 + 1$, which has solutions $\{i, -i\} \not\subset \mathbb{R}$, meaning A is diagonalizable over $\mathbb{C}$, but not $\mathbb{R}$.

We note that for matrix A in $\mathbb{R}$ with *eigenpair* $\{\lambda, \vec{v}\}$, we have $A\vec{v} = \lambda\vec{v} \iff \overline{A\vec{v}} = \overline{\lambda\vec{v}}$ $\iff \overline{A}\overline{\vec{v}} = \overline{\lambda}\overline{\vec{v}} \iff A\overline{\vec{v}} = \overline{\lambda}\overline{\vec{v}}$ (since A is real), meaning $\{\overline{\lambda}, \overline{\vec{v}}\}$ is also an eigenpair of A.

- Aside: this is related to the conjugate root theorem, since $C_A(\lambda)$ is a polynomial in $\mathbb{R}[x]$

The characteristic polynomial can be used for the following sanity checks:

1. C_A is of degree n (for $A_{n\times n}$)
2. The leading coefficient $a_n = (-1)^n$, since $C_A = (\lambda - \lambda_i)^n$ for various i
3. The constant term a_0 is the product of A's eigenvalues (since the constant term must be divisible by all the factors, which are all the eigenvalues)
4. $a_{n-1} = (-1)^{n-1}\mathrm{Trace}(A) = (-1)^{n-1}\sum_{i=0}^{n}\lambda_i$ (follows from expanding polynomial multiplication)

We have $\boxed{\det A = \prod_{i=1}^{n}\lambda_i}$ since $\det A = C_A(0) = a_0$, which is the product of eigenvalues as shown in 3)

Diagonalizability is important because, for diagonalizable A, we have $A^n = PD^nP^{-1}$ where $D = \begin{bmatrix} \lambda_1^n & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^n \end{bmatrix}$, since D is diagonal.

- Proof: $A^n = (PDP^{-1})\dots(PDP^{-1}) = PD(PP^{-1})D\dots DP^{-1} = PDP^{-1}$
- So, since for $A \to A^n$, $D \to D^n$ and P, P^{-1} stay the same, by their definitions with respect to *eigenvalues* and *eigenvectors*, we see that $A\vec{v} = \lambda\vec{v} \implies A^n\vec{v} = \lambda^n\vec{v}$

We can also define $\sqrt{A}$ as $P\sqrt{D}P^{-1}$, where $D = \begin{bmatrix} \sqrt{d_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sqrt{d_n} \end{bmatrix}$, i.e. A^q is defined for $q \in \mathbb{Q}$

Lecture 21

Since A^n is defined for $n \in \mathbb{N}$, we can define the **exponential function** e^A or $\exp A$ as $\sum_{n=1}^{\infty} \frac{A^n}{n!}$

- We have $e^A = \sum_{n=1}^{\infty} \frac{PD^nP^{-1}}{n!} = P\left(\sum_{n=0}^{\infty} \frac{D^n}{n!}\right)P^{-1}$, where $\sum_{n=0}^{\infty} \frac{D^n}{n!}$ has (i,i)th entry $\frac{(D_{ii})^n}{n!}$
- Aside: this seems to imply that any function that can be approximated with a Taylor series (i.e. any n-times differentiable real-valued function) can be defined for matrices, since the operations that form power series ($+$ and $\cdot$) are defined over vector spaces

Aside: Recall that $\frac{d}{dx}$ is a *linear operator* (linear map on a *space of differentiable functions* like $\mathbb{R}[x]$) that can be described by matrix A. So, we can define *another* linear operator $e^A = e^{\frac{d}{dx}}$ sends "vector" $f(x)$ to $\sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!}$. This is the taylor expansion of $f(x+1)$, so the transformation T_A described by matrix A maps $f(x) \mapsto f(x+1)$; it is the *translation operator*

- Aside: obviously this works for polynomials because they form a vector space, but can this work over spaces that aren't vector spaces?
- Aside: is the entire space of functions 1) a thing that makes sense 2) a vector space?
- Aside: what class of things does this "translation operator" belong to? What other operators exist in this class?
- Aside: the way we can define a function/operator using a series (like e^A), compose it with another function/operator (like $\frac{d}{dx}$), then get back a taylor series we can interpret as a *different* operator seems like a powerful pattern of derivation

If A is *diagonalizable* and all of A's eigenvalues satisfy $|\lambda| < 1$, then $A^n \to \mathcal{O}$ for $n \to \infty$

Let F_n define the nth Fibonacci number, and define $\vec{v_n} = \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix}$. Then $\vec{v_{n+1}} = \begin{bmatrix} F_{n+2} \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} F_{n+1} + F_n \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \vec{v_n}$, so it follows that $\vec{v_n} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ since $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ are the initial conditions of the Fibonacci series. $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ has eigenvalues $\frac{1 \pm \sqrt{5}}{2} = \{\phi, 1-\phi\}$. Since we can find the closed form expression of F_n (component of) $\vec{v_n} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, where we find $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n$ by diagonalizing it; $F_n = \frac{1}{\sqrt{5}}(1+\sqrt{5})^n - \frac{1}{\sqrt{5}}(1-\sqrt{5})^n$

- The eigenvalue ϕ relates to the fact that the ratio of successive Fibonacci terms approaches ϕ.
- Aside: the $\pm$ in $\frac{1 \pm \sqrt{5}}{2}$ seems to have an analog to complex conjugation trick for diagonalization, which is given as true. When we have two fields (here, $\mathbb{Q} \subset \mathbb{R}$), there are symmetries we can exploit
- Aside: encoding recurrence relations as vector-matrix multiplications (a case of the broader encoding functions as matrix multiplications) → eigenvalue (steady state) and possibly

determinant have useful interpretations

- Aside: eigenstuff somehow encodes self-relation limits, i.e. $L = \dfrac{1}{1+L}$ for ϕ.

We found that since a matrix $A\vec{v}$ represents a transformation (i.e. a function/operator) applied to $\vec{v}$ by vector multiplication, $A^n\vec{v}$ corresponds to it being repeated (*composed*) n times

- Aside: composition $\circ$ and multiplication $\cdot$ seem to be equivalent/isomorphic quite often in linear algebra. Why? And between what other structures is this isomorphism defined?

Aside: since every function can be described perfectly (in the limit case) by a Taylor polynomial (within the radius of convergence) and a series can be represented as a series, and thus an ∞-dimensional vector, can every function be represented as a vector? What do linear transformations between these vectors correspond to? presumably non-linear functions? can every operator be described this way?

Topic 4 - Bases

> based!

Lecture 22

Change of variables is a common technique in all of math, where a *new variable* that is more convenient to use is defined as a function of *existing variables* in an expression. When the variables we change define a coordinate system (i.e. x, y, etc.), this is a *coordinate change*; the coordinate change in linear algebra is a **change of basis**, since a basis is (like) a coordinate system.

Recall that $[T(\vec{v})]_{\beta'} = [T]_{\beta' \leftarrow \beta}[\vec{v}]_\beta$ (**change of basis for transformation**) and $[S \circ T]_{\beta'' \leftarrow \beta} = [S]_{\beta'' - \beta}[T]_{\beta' - \beta}$ (**change-of-basis composition**) for bases β of V, β' of W, β'' of U, and $T : V \to W, S : W \to U$

$P_{\beta' \leftarrow \beta} = [\mathrm{Id}_V(\vec{v})]_{\beta' \leftarrow \beta}$ is the **change of basis matrix** from β to β'. Here, $\mathrm{Id} : \vec{v} \mapsto \vec{v}$ is the *identity transformation* for vector space V $\mathrm{Id} : \vec{v} \mapsto \vec{v}$, where $T_{\mathrm{Id}} = I$

- This implies that $P_{\beta' \leftarrow \beta} = [\vec{e_1}_{\beta'}, \ldots, \vec{e_n}_{\beta'}]$ for *standard basis vectors* $\vec{e_1} \ldots \vec{e_n}$
- If $\beta = \beta'$, $P_{\beta' \leftarrow \beta}$ will be the *identity matrix*; for $\beta \neq \beta'$, it won't be

The *change of basis matrix* can change the basis of a *vector* $\vec{v}$ (here, from β to β'): $[\vec{v}]_{\beta'} = [\mathrm{Id}(\vec{v})]_{\beta'} = [\mathrm{Id}]_{\beta' \leftarrow \beta}[\vec{v}]_\beta = P_{\beta' \leftarrow \beta}[\vec{v}]_\beta$. So, $\boxed{[\vec{v}]_{\beta'} = P_{\beta' \leftarrow \beta}[\vec{v}]_\beta}$

Change of basis matrices P *reverse under inversion*, i.e. $(P_{\beta' \leftarrow \beta})^{-1} = P_{\beta \leftarrow \beta'}$

- This follows from $I = [\mathrm{Id}]_{\beta' \leftarrow \beta} = [\mathrm{Id} \circ \mathrm{Id}]_{\beta' \leftarrow \beta} = [\mathrm{Id}]_{\beta \leftarrow \beta'}[\mathrm{Id}]_{\beta' \leftarrow \beta} = P_{\beta \leftarrow \beta'}P_{\beta' \leftarrow \beta}$

Let $T : V \to W$, and β_V, β'_V be bases of V, and β_W, β'_W be bases of W. Then $[T]_{\beta'_W \leftarrow \beta'_V} = [\mathrm{Id}_W \circ T \circ \mathrm{Id}_V]_{\beta'_W \leftarrow \beta'_V} = [\mathrm{Id}_W]_{\beta'_W \leftarrow \beta'_V}[T]_{\beta_W \leftarrow \beta_V}[\mathrm{Id}_V]_{\beta'_V \leftarrow \beta_V} = P_{\beta'_W \leftarrow \beta'_V}[T]_{\beta_W \leftarrow \beta_V}P_{\beta'_V \leftarrow \beta_V}$

- I.e. change of basis "chains" through composition
- When $V = W$, i.e. $T : V \to V$, then we get $[T]_{\beta \leftarrow \beta'} = (P_{\beta' \leftarrow \beta})^{-1}[T]_{\beta \leftarrow \beta}P_{\beta \leftarrow \beta'}$

Note: so far, we've just been treating matrices as immutable things with one representation. But, *bases* add an "extra dimension" to this, where matrices are represented differently in different bases. So far, we've essentially been assuming matrices are expressed in the *standard basis* and have omitted the notation.

Note: the concept of bases (and converting between different bases) for vectors is similar to the concept of expressing natural numbers in a base other than 10. In each, the underlying system describing how to calculate the "pure" value changes, but the value itself doesn't.

Lecture 23

E.g. take $V = \mathbb{F}[x], D = \frac{d}{dx}, \beta = (x, 1), \beta' = (1, 2x)$. Then, we have $[D]_{\beta\leftarrow\beta} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ (i.e. the definition of D), and $[D]_{\beta'\leftarrow\beta'} = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$ (since the order of basis vectors (and thus cols) is swapped, and we have $2x$ instead of x). We find by inspection that $P_{\beta\leftarrow\beta'} = \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$, so $P_{\beta'\leftarrow\beta} = (P_{\beta\leftarrow\beta'})^{-1} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & 0 \end{bmatrix}$, we have $\begin{bmatrix} 0 & 1 \\ \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$ as desired

$n \times n$ matrices A and B are **similar** iff there exists *invertible matrix* P such that $B = P^{-1}AP$

- *Similarity* is an *equivalence relation*, i.e. the set of all $n \times n$ matrices can be *partitioned* into sets of *mutually similar matrices* called **similarity classes**
- Aside: *diagonalizing* can be seen as finding a "nice" representative of a matrix in the same *similarity class*

$[T]_{\beta'\leftarrow\beta'} = (P_{\beta'\leftarrow\beta})^{-1}[T]_{\beta\leftarrow\beta}P_{\beta\leftarrow\beta'}$, shows that for bases β and β', $[T]_{\beta'\leftarrow\beta'}$ and $[T]_{\beta\leftarrow\beta}$ are *similar*. Conversely, given matrix B similar to $[T]_{\beta\leftarrow\beta}$, we can find basis β' such that $[T]_{\beta'\leftarrow\beta'} = B$

Lecture 24

If we have $D = P^{-1}AP$ for diagonal A and $D = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}$, we can choose P to be a *permutation matrix* (since A and D are both diagonal). Then, A will be $D = \begin{bmatrix} \lambda_{\pi(1)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_{\pi(n)} \end{bmatrix}$, where $\pi(n)$ is the *permutation* described by P.

- So, if a diagonal matrix is in a similarity class, then any *permutation* of its diagonal entries is also in that similarity class

Every matrix, even non-diagonalizable ones, is similar to an *upper-triangular* matrix, i.e. every matrix where the entry below the diagonal is 0.

Every non-diagonalizable 2×2 matrix is in the form $\begin{bmatrix} a & b \\ 0 & a \end{bmatrix}$ for $b \neq 0$ (since if the diagonal entries (eigenvalues) were different, they'd be distinct $\rightarrow$ diagonalizable). With $P = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{b} \end{bmatrix}$, this is similar to $\begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}$.

Topic 5 - Jordanstuff and Generalized Eigenstuff

Lecture 24

The **Jordan block** of λ, denoted $J_n(\lambda)$, is the $n \times n$ matrix equivalent of $\begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix}$, where all the diagonal entries are λ and the entries directly above them are 1 (note that this matrix is *upper triangular*)

We defined the **direct sum/block sum** $A \oplus B$ as the "composite" matrix $\begin{bmatrix} [A] & \mathcal{O}_n \\ \mathcal{O}_n & [B] \end{bmatrix}$, where $\mathcal{O}_n$ is the $n \times n$ *zero matrix* and $[A], [B]$ are the entries of A and B.

- E.g. $[1] \oplus [1] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$, $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \oplus \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & e & f \\ 0 & 0 & g & h \end{bmatrix}$

Lecture 25

For $n \times n$ matrices A, A' and $m \times m$ matrices B, B', the *direct sum* is "distributive":

- $(A \oplus B) + (A' \oplus B') = (A + A') \oplus (B + B')$
- $(A \oplus B) \cdot (A' \oplus B') = (AA') \oplus (BB')$
- Aside: what property is this? distributivity? structure looks kinda like De Morgan's laws. What algebra is at work here?

The *Jordan block* $J_m(\lambda)$ has characteristic polynomial $(\lambda - x)^m$, so it has one eigenvalue (namely, $x = \lambda$) with *algebraic multiplicity* m and *geometric multiplicity* 1

- The geometric multiplicity follows from the fact that $J_m(\ell) - \ell I$ is the $n \times n$ version of $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, whose null space clearly has one dimension (corresponding to the first column $\vec{0}$)

Lecture 26

Jordan Canonical Form Theorem

If A is a square matrix, then its *similarity class* has a matrix of the form $(J_{m_{1,1}}(\lambda_1) \oplus J_{m_{1,2}}(\lambda_1) \oplus \cdots \oplus J_{m_{1,r}}(\lambda_1)) \oplus (J_{m_{2,1}}(\lambda_2) \oplus \cdots \oplus J_{m_{2,r}}(\lambda_2)) \oplus \cdots \oplus$ $(J_{m_{k,1}}(\lambda_j) \oplus \cdots \oplus J_{m_{j,r}}(\lambda_k))$ where $\lambda_1 \ldots \lambda_k$ are the *unique* eigenvalues of A; this is the **Jordan canonical form** of A.

- λ_i has *algebraic multiplicity* $m_{i,1} + \cdots + m_{i,r_i}$ and *geometric multiplicity* r_i
- $J_m(\lambda)$ is the $m \times m$ *Jordan block* for λ

- Informally, every square matrix is *similar* to a direct sum of Jordan blocks, particularly of those built from the matrix's eigenvalues
- JCF replaces the diagonal matrix when A is *not diagonalizable*
- Essentially, the JCF is any form where the matrix is created from Jordan blocks of its eigenvalues. If an eigenvalue is repeated, any partition of these into different Jordan blocks is allowed (see assignment 8 q2), but the "canonical" one is the one with eigenvalues in increasing order.
- If we require that $m_{i,1} \geq m_{i,2} \geq \ldots$, then the JCF is *unique* (i.e. it is unique up to the ordering of eigenvalues).

A matrix A is *diagonalizable* iff all $m_{i,j} = 1$, i.e. if all the Jordan blocks are as small as possible, and JCF of A is a (the) diagonal matrix.

Minimal Polynomial

The **minimal polynomial** $m_A(x)$ of matrix A is the unique* polynomial with the following properties

1. m_A is *monic*, i.e. its leading coefficient is 1
2. $M_A(A) = \mathcal{O}$
3. m_A is has the smallest degree possible while satisfying both 1) and 2)

- If A is diagonal with entries λ_1, λ_2, etc, then $m_A(x) = (x - \lambda_1)(x - \lambda_2) \ldots$, where multiplicity is ignored (since the extra multiplicity of a factor doesn't change whether m_A has a zero there).
  - So, the degree of the minimal polynomial is that of the *characteristic polynomial* minus the number of multiplicity repeats. Aside: does this imply some sort of quotient structure?
- If A is *Jordan block* $J_m(\lambda)$, then $m_A(x) = (x - \lambda)^m$
- $m_{A \oplus B}(x) = m_A(x) m_B(x)$, i.e. the minimal polynomial of a *direct sum* of matrices is the product of the minimal polynomials of the matrices.

"Minimal Polynomial Theorem"

If A is an $n \times n$ matrix, we have

1. The *minimal polynomial* $m_A(x)$ of A exists an is *unique*
2. For any $p(x)$ where $p(A) = 0$, then $m_A(x)$ divides $p(x)$. So, $p(x) = m_A(x)q(x)$
3. $A^{d-1}, A^{d-2}, \ldots, A, I$ are *linearly independent* in $M_{n\times n}(\mathbb{F})$ where $m_A(x)$ has *degree* d. However, $A^d, A^{d-1}, \ldots$ are *not* linearly independent (found by plugging A into m_A)
4. Any matrix *similar* to A has the *same minimal polynomial*

Lecture 27

The proof for the *minimal polynomial theorem* is found here. The proofs for each point can be sketched as follows

1. Somewhat trivial
2. We must have $p(x) = m_A(x)q(x) + r(x)$ (number theory!), plugging in A turns $p(x)$ and $m_A(x)$ to 0, deriving the point
3. Proof straightforward from definition of linear dependence (also shows uniqueness for 1.)
4. Plug in $B = P^{-1}AP$ into the polynomial form of $m_A(x)$, P^{-1} and Ps "telescope" to cancel

Cayley-Hamilton Theorem

For $n \times n$ matrix A with *characteristic polynomial* $C_A(\lambda)$, then $C_A(A) = \mathcal{O}$

- Proof sketch: Because plugging a Jordan block $J_m(\lambda)$ into the *characteristic polynomial* C_B of a matrix B in JCF with λ as an eigenvalue is 0 (proof here, follows from the structure of a Jordan block), $C_B(B)$ is the direct sum of 0-matrices, i.e. is $\mathcal{O}$. Since any matrix A with JCF B has the same characteristic polynomial (because A and B must be similar), $C_A(A) = \mathcal{O}$ for any matrix A.

Lecture 28

Since addition and scaling "distribute" over direct summation $\oplus$, "taking a polynomial" does as well, i.e. $p(A \oplus B) = p(A) \oplus p(B)$ for *polynomial* p.

For polynomials $p(x)$, $q(x)$ and matrix A, $p(A)q(A) = q(A)p(A)$, i.e. multiplication of matrix-polynomial compositions is *commutative*

- This is significant because $\lozenge[AB \neq BA]$ for matrices in general
- The proof follows from inheriting commutativity from the "summation" expression of a polynomial, i.e. that switching the order of the sums adds the same polynomial terms in a different order

The **generalized eigenspace** GE_λ of matrix A and *arbitrary (existentially quantified)* number $\lambda \in \mathbb{F}$ is defined as the space of $\vec{v}$ where $(A - \lambda I)^\ell \vec{v} = \vec{0}$ for some $\ell \in \mathbb{N}$.

- So, $GE_\lambda(A) = \bigcup_{\ell=0}^{\infty} \text{Null}((A - \lambda I)^\ell)$
- $GE_\lambda(A)$ is a *subspace* of $\mathbb{F}^n$ that trivially contains the "regular" eigenspace $E_\lambda(A)$, which corresponds to $\ell = 1$, i.e. $E_\lambda(A) \subseteq GE_\lambda(A)$

Lecture 29

The proof that GE_λ is a subspace is sketched as follows:

1. Nonempty: clearly $\vec{0} \in GE_\lambda$
2. Closure under $+$: whichever $\vec{u}, \vec{v}$ has the lowest corresponding ℓ is in the generalized eigenspace of the other; closure is inherited because GE_λ is a null space
3. Closure under $\cdot$: inherited directly from the null space definition as well

For any matrix A over *field* $\mathbb{F}$ with characteristic polynomial $(\lambda_1 - x)^{m_1}(\lambda_2 - x)^{m_2} \dots (\lambda_k - x)^{m_k}$, $GE_{\lambda_1} \oplus GE_{\lambda_2} \oplus \dots \oplus GE_{\lambda_k} = \mathbb{F}^n$

- So, any $\vec{v} \in \mathbb{F}^n$ can be written as a sum of vectors from generalized eigenspaces, i.e. $\forall \vec{v} \in \mathbb{F}^n, \vec{v} = \sum_{i=1}^{n} \vec{v}_i \text{ for } \vec{v}_i \in GE_{\lambda_i}$
- Also, if β_i is a basis of GE_{λ_i}, then $\beta = \bigcup_{i=1}^{n} \beta_i$ will be a basis for $\mathbb{F}^n$
- Finally, if P is a *change of basis matrix* from β to the standard matrix of a transformation T, then $P^{-1}AP$ will be a *direct sum* of matrices, one for each GE_{λ_i}

Lecture 30

If a matrix N consists of a diagonal set of 1s (after the main diagonal of the matrix, like $J_n(\lambda) - \lambda I$), then N^2 is a diagonal set of 1s on the next diagonal, N^3 on the one after that, etc. until $N^n = \mathcal{O}$.

- The *basis* for the null space of some N^k are the first k vectors of the standard basis, i.e. the ones corresponding to 0-columns in N^k
- Note that matrices of the form $J_m(\lambda) - \lambda I$ are of this form
- So, $E_\lambda = \text{Null}(N) \subset \text{Null}(N^2) \subset \dots \subset \text{Null}(N^n) = GE_\lambda$

Finding the JCF of a matrix ("dots" method)

For a matrix A, we find the *characteristic polynomial* $C_A(\lambda)$ and factor it to find eigenvalues $\lambda_1 \dots \lambda_k$ (factoring may need to include $\mathbb{C}$). For each eigenvalue λ

- Compute $\dim(A - \lambda I), \dim((A - \lambda I)^2), \ldots, \dim((A - \lambda I)^{m_a})$ where m_a is the algebraic multiplicity of λ
- Draw m_a dots in a (horizontal) row
- For each dot (say n), write $\dim((A - \lambda I)^{n+1}) - \dim((A - \lambda I)^n)$ dots in a horizontal row under it (vertically). So, the first dot will have $\dim((A - \lambda I)^2) - \dim((A - \lambda I)^1)$ dots under it

The vertical groupings of dots correspond to the Jordan blocks that correspond to a given λ. So, the JCF of A is the *direct sum* of all of the Jordan blocks for all of the eigenvalues.

Since the JCF J of A is similar to A, we have $A = P^{-1}JP$. The columns of P are formed from the vectors that need to be added to the basis of $\mathrm{Null}((A - \lambda I)^n)$ to form $\mathrm{Null}((A - \lambda I)^{n+1})$

- More formally, the columns of P are the bases of the corresponding *quotient space* $\mathrm{Null}((A - \lambda I)^{n+1})/\mathrm{Null}((A - \lambda I)^n)$ for eigenvalue λ and $n < m_{a_\lambda}$
- So if this difference is a, then the basis for that "direct difference" (quotient) has a vectors (and thus a columns in P), and the corresponding Jordan block contributing to the JCF is $a \times a$

If $\vec{v_1}, \vec{v_2}, \ldots$ are columns in P that correspond to the same eigenvalue, then we must have $\vec{v_n} = N\vec{v_{n+1}}$ where $N = J_n(\lambda) - \lambda I$

- For $T = T_{J_n(\pi)}$ is the linear transformation described by $T(\vec{v}) = J_n(\pi)\vec{v}$. Let basis $\beta = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$. We can show that $[T]_{\beta \leftarrow \beta}$ is $J_n(\pi)$

Topic 6 - Inner Products and Inner Proudct Spaces

Lecture 31

Inner Product

For *vector space* V over $\mathbb{R}$, the pairing $\langle \vec{u}, \vec{v} \rangle$ is an **inner product** if it has the following properties

1. **Symmetry**: $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$
2. **Bilinearity**: $\langle \vec{u}, a\vec{v} + b\vec{w} \rangle = a\langle \vec{u}, \vec{v} \rangle + b\langle \vec{u}, \vec{w} \rangle$
    - This can be generalized (using symmetry) to apply to both arguments
3. **Positivity**: $\langle \vec{v}, \vec{v} \rangle > 0$ for all nonzero $\vec{v}$

Inner Product Space

A vector space V with *inner product* $\langle \vec{u}, \vec{v} \rangle$ form an **inner product space**

For $\mathbb{R}^n$, $\langle \vec{u}, \vec{v} \rangle = \sum_{i=1}^{n} u_i v_i$ is a common *inner product* known as the **dot product**, denoted $\vec{u} \cdot \vec{v}$

- E.g. for $V = \mathbb{R}^2$, the *dot product* is $\langle \vec{u}, \vec{v} \rangle = u_1 v_1 + u_2 v_2$.

For vector $\vec{v}$, we define its **length** (or **norm**) as $\sqrt{\langle \vec{v}, \vec{v} \rangle}$, denoted $\|\vec{v}\|$

- Recall the *norm* of a vector
- E.g. For the above example, we find by the *pythagorean theorem* that $\langle \vec{v}, \vec{v} \rangle = v_1 v_1 + v_2 v_2 = {v_1}^2 + {v_2}^2$ is the squared *length* (or *euclidean distance*) of $\vec{v}$

We can derive the following from the inner product definition

- *Linearity of the first argument*: $\langle a\vec{u} + b\vec{v}, \vec{w} \rangle = a\langle \vec{u}, \vec{w} \rangle + b\langle \vec{v}, \vec{w} \rangle$ from *symmetry and bilinearity*
- $\langle \vec{0}, \vec{v} \rangle = \langle \vec{v}, \vec{0} \rangle = 0$ for all $\vec{v}$

We have the following examples and non-examples of inner products for other domains

- $V = M_{n \times n}$ has an inner product $\langle A, B \rangle = \mathrm{Tr}(AB^\top)$, where Tr is the *trace* of the matrix
- $V = \mathbb{R}[x]_2$ has an inner product $\langle p(x), q(x) \rangle = \int_0^1 p(x)q(x)\, dx$
- $V = \mathbb{C}$ has inner product $\langle \vec{u}, \vec{v} \rangle = u_1 \overline{v_1} + u_2 \overline{v_2}$, where $\overline{z}$ is the *complex conjugate*

- $\langle \vec{u}, \vec{v} \rangle = u_1 v_1 + u_2 v_2$ violates positivity over $\mathbb{C}$, since $\left\langle \begin{bmatrix} i \\ 0 \end{bmatrix}, \begin{bmatrix} i \\ 0 \end{bmatrix} \right\rangle = -1 < 0$
- Non-example: $V = \mathbb{R}^4$ with "inner product" $\langle \vec{u}, \vec{v} \rangle = u_1 v_1 + u_2 v_2 + u_3 v_3 - c^2 u_4 v_4$ violates positivity. This is used in Einstein's theory of relativity; c is the speed of light

The Lecture notes have an example for determining $\|\vec{u} + \vec{v}\|$ given the inner product definition, $\|\vec{u}\|$ and $\|\vec{v}\|$. Generally, this is done by expanding (using bilinearity) and possibly using symmetry to match.

Lecture 32

Cauchy-Schwarz Inequality

Let V and $\langle \vec{u}, \vec{v} \rangle$ be an inner product space. The **Cauchy-Schwarz inequality** states that $-\|\vec{u}\|\|\vec{v}\| \geq \langle \vec{u}, \vec{v} \rangle \geq \|\vec{u}\|\|\vec{v}\|$

- Proof: by positivity and linearity, $0 \leq \langle \vec{u} - x\vec{v}, \vec{u} - x\vec{v} \rangle = \|\vec{u}\|^2 - 2x\langle \vec{u}, \vec{v} \rangle + x^2\|\vec{v}\|^2$. Choosing $x = \dfrac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{v}\|^2}$ simplifies to $\langle \vec{u}, \vec{v} \rangle^2 \leq \|\vec{u}\|\|\vec{v}\|$, which alternatively states Cauchy-Schwarz.

The **angle** α between vectors $\vec{v}, \vec{u}$ in *inner space* V is the *unique* $0 \leq \alpha \leq \pi$ such that $\cos \alpha = \dfrac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\|\|\vec{v}\|}$

- By Cauchy-Schwarz, $\dfrac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\|\|\vec{v}\|} \in [-1, 1]$, guaranteeing an α exists for any $\vec{u}, \vec{v}$
- α is not defined for $\vec{u}, \vec{v} = 0$

Vectors $\vec{u}, \vec{v}$ are **parallel** if $\langle \vec{u}, \vec{v} \rangle = \|\vec{u}\|\|\vec{v}\|$ (i.e. $\alpha = 0$) and **orthogonal** if $\langle \vec{u}, \vec{v} \rangle = 0$ (i.e. $\alpha = \dfrac{\pi}{2}$)

Aside: for the dot product in $\mathbb{R}^2$ (inner space), in polar coordinates, we get $\vec{u} = \begin{bmatrix} \|\vec{u}\| \cos\theta \\ \|\vec{u}\| \sin\theta \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} \|\vec{v}\| \cos\varphi \\ \|\vec{v}\| \sin\varphi \end{bmatrix}$, so $\vec{u} \cdot \vec{v} = \cdots = \|\vec{u}\|\|\vec{v}\| \cos(\varphi - \theta)$. This is where this definition of the dot product comes from, and implies that the dot product measures how close the angles of vectors are.

- More succinctly, we have $\|\vec{u} \cdot \vec{v}\| = \|\vec{u}\|\|\vec{v}\| \cos\theta$
- Aside: just like we define an angle using the *dot product*, so too can we define other "types of angle" with other inner products.

A *basis* $\beta = (\vec{b}_1, \vec{b}_1, \ldots, \vec{b}_n)$ is an **orthogonal basis** if $\langle \vec{b}_i, \vec{b}_j \rangle = 0$ for all $i \neq j$

- E.g. the *standard basis* is an orthogonal basis

  We define **coordinate vector** of vector $\vec{v}$ in basis β as $[\vec{v}]_\beta = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ where $\sum_{i=1}^{n} c_i \vec{b}_i = \vec{v}$

  In an orthogonal basis, we have $c_i = \dfrac{\langle \vec{v}_i, \vec{b}_i \rangle}{\langle \vec{b}_i, \vec{b}_i \rangle}$
- Proof: follows from the definition of an orthogonal basis

Lecture 33

"Orthogonal Decomposition Theorem"

Let V be an *inner product space* with dimension n; let $\vec{v}_1, \dots, \vec{v}_n$ be orthogonal, so any pairwise $\langle \vec{u}_i, \vec{v}_i \rangle = 0$. Then $(\vec{v}_1, \dots, \vec{v}_n)$ is an *orthogonal basis* if and only if all $\vec{v}_i \neq \vec{0}$

- Proof sketch: we wish to show that $\vec{v}_1, \dots, \vec{v}_n$ are linearly independent (make-your-life-easy theorem does the rest). Let $c_1\vec{v}_1 + \cdots + c_n\vec{v}_n = \vec{0}$. Taking the inner product of both sides of the $=$ yields 0 by (inductive) orthogonality. Any $\langle \vec{v}_i, \vec{v}_i \rangle$ satisfy $\langle \vec{v}_i, \vec{v}_i \rangle > 0$ due to positivity, so we can divide both sides by it to yield $c_i = 0$. We know that if some $\vec{v}_i = \vec{0}$, then $\vec{v}_1, \dots, \vec{v}_n$ would not be linearly independent to begin with.

Cauchy-Schwarz refinement

If $\langle \vec{u}, \vec{v} \rangle = \|\vec{u}\|\|\vec{v}\|$ and $\vec{v} \neq \vec{0}$, then $\vec{u} = a\vec{v}$ for some scalar $a \geq 0$, i.e. $\vec{u}$ and $\vec{v}$ are *linearly dependent*.

- Proof: $\langle \vec{u}, \vec{v} \rangle = \|\vec{u}\|\|\vec{v}\| \implies \langle \vec{u} - x\vec{v}, \vec{u} - x\vec{v} \rangle = 0 \implies \vec{u} - x\vec{v} = \vec{0} \implies \vec{u} = x\vec{v}$

We can use Cauchy-Schwarz to define *optimization problems* by reinterpreting the *equation* whose parameter we want to optimize as an (instance of an) inner product, using Cauchy-Schwarz to define bounds, then using *Cauchy-Schwarz refinement* to find an expression based on the bound(s).

- E.g. which polynomial $q(x)$ satisfying $= \int_0^1 (3x-1)q(x)\,dx = \frac{1}{2}$ has the smallest $\int_0^1 q(x)^2\,dx$ (i.e. smallest "size")? We define the inner product $\langle p(x), q(x) \rangle = \int_0^1 p(x)q(x)\,dx$, and rephrase the first equation as $\langle 3x-1, q(x) \rangle = \frac{1}{2}$; by Cauchy-Schwarz, we find the bound $\frac{1}{2} \leq \|3x-1\|\|q\|$. Evaluating, we find $\|3x-1\| = 1$, so $\frac{1}{2} \leq \|q\|$. By Cauchy-Schwarz refinement, we find that

$q(x) = a(3x-1)$ for some a. Finally, we now know $\frac{1}{2} = \langle 3x-1, a(3x-1)\rangle = a \times \|3x-1\|^2 = a$. So, $q(x) = \frac{1}{2}(3x-1)$.

- Find the smallest $\vec{v} \in \mathbb{R}^3$ where$\langle (1,2,3), \vec{v}\rangle = -1$ where we use the dot product. From the dot product definition, we get $-1 = \langle (1,2,3), \vec{v}\rangle = \|(1,2,3)\| \|\vec{v}\| \cos(\theta)$. To minimize $\vec{v}$, we pick the smallest possible $\cos(\theta)$, i.e. $\theta = \pi$. We now know $\vec{v} = a(1,2,3)$. So, $-1 = \langle (1,2,3), (1,2,3)\rangle = a(1^2+2^2+3^2) = 14a$, so $\vec{v} = -\frac{1}{14}(1,2,3)$.

Lecture 33

The **Gram-Schmidt process** is used to construct an *orthogonal basis* from an basis (or even any set of vectors) in an inner product space.

Gram-Schmidt Process

For basis $\vec{v_1}, \ldots, \vec{v_n}$ in an inner product space, we can construct a basis $\vec{u_1}, \ldots, \vec{u_n}$ for $\mathrm{Span}(\vec{v_1}, \ldots, \vec{v_n})$ that is orthogonal in that inner product space. We do this with the following process:

- $\vec{u_1} = \vec{v_1}$,
- $\vec{u_2} = \vec{v_2} - \frac{\langle \vec{v_2}, \vec{u_1}\rangle}{\|\vec{u_1}\|^2}\vec{u_1}$,
- $\vec{u_3} = \vec{v_3} - \frac{\langle \vec{v_3}, \vec{u_1}\rangle}{\|\vec{u}\|^2}\vec{u_1} - \frac{\langle \vec{v_3}, \vec{u_2}\rangle}{\|\vec{u_2}\|^2}\vec{u_2}$
- Etc.

In general, we have $\vec{u_k} = \vec{v_k} - \sum_{\ell=1}^{k-1} \frac{\langle \vec{v_k}, \vec{u_\ell}\rangle}{\|\vec{u_\ell}\|^2}\vec{u_\ell}$

- Proof sketch: First, note that $\langle \vec{u_1}, \vec{u_2}\rangle = \left\langle \vec{u_1}, v_2 - \frac{\langle \vec{v_2}, \vec{u_1}\rangle}{\|\vec{u_1}\|^2 \vec{u_1}} \right\rangle = \ldots \text{bilinearity expansion} = 0$. Perform induction on $k \to \ldots n$ to show that $\langle \vec{u_1}, \vec{u_k}\rangle = 0$. Knowing this, we can perform induction on j to show that $\langle \vec{u_j}, \vec{u_k}\rangle = 0$, showing that all the $\vec{u}$ terms are orthogonal.
- Proof: [Lecture 34 Slides]

Lecture 35

- E.g. In the inner product space of polynomials with $\langle p(x), q(x)\rangle = \int_0^1 p(x)q(x)\,dx$, an orthogonal basis for $(1, x, x^2)$ (which interestingly is *not* orthogonal here) is $(1, x - \frac{1}{2}, x^2 - x + \frac{1}{6})$

If we perform the Gram-Schmidt process on *linearly dependent* $\vec{v_1}, \ldots, \vec{v_n}$, then at least one $\vec{u}$ will be $\vec{0}$.

Application: the **least-squares solution** for system of equations $A\vec{x} = \vec{b}$ is the $\hat{\vec{x}}$ that minimizes the *error*, namely $\|A\hat{\vec{x}} - \vec{b}\|$. If this system is consistent, then $A\vec{x}$ is in the *column space* of A, and a "full" solution exists. When it's not consistent (i.e. in almost every practical application), $A\hat{\vec{x}} = \vec{b}$ must be orthogonal to the column space, why???, i.e. to all the columns of A. So, the least squares solution is any solution to $A^\top(A\hat{\vec{x}} - \vec{b}) = \vec{0} \implies A^\top A\hat{\vec{x}} = A^\top\vec{b}$.