> The note of UIUC course MATH 416 Honor Abstruct Linear Algebra in Fall 2023, by Zory Zhang. In case of any broken math renderring in Github preview, please open this markdown file using your own markdown editor. PDF version (maybe not up-to-date) can be found here.

Textbook

- Linear Algebra via Exterior Products (2020)
- Linear Algebra Done Right (2023)
- Linear Algebra Done Wrong (2021)
- Linear Algebra (Stephen H. Friedberg, Arnold J. Insel etc.) (2021) [main]

Reminder

1. Carefully look at "dependent" or "independent".

Notation / Convention

1. ϵ_n is the standard basis of F^n.
2. a_{ij} stands for entries of A; A_{ij} stands for minor; $\hat{a}_{ij}$ stands for cofactor of entry a_{ij}.
3. Basis $\beta = (s_1, \dots s_n)$ for V, $\alpha = (t_1, \dots, t_m)$ for W.
4. Unless specified, the vector space is finite-dimensional.

Ch1 Vector Spaces

1.2 Vector Space

Def. (Vector space V on field F)

A non-empty set with vector addition and scalar multiplication, with the following axioms:

1. Additive commutativity;
2. Additive and scalar multiplicative associativity;

3. Additive identity and scalar multiplicative identity;
4. Additive inverse;
5. Vector and scalar additive distributivity.

Rmk. This definition gives rise to a few special vector space, e.g. $\mathbb{R}^n$ and $\mathcal{P}^n$, which will compose others by standard procedure introduced later.

Thm. (1.1 Cancellation law for vector addition) By playing inverse (rule 4).

Cor. a) $\exists ! \underline{0}$; b) $\exists ! \underline{-x}$; c) $0 \cdot \underline{x} = \underline{0}$; d) $(-\lambda) \cdot \underline{x} = -(\lambda \underline{x}) = \lambda \cdot \underline{-x}$; e) $\lambda \cdot \underline{0} = \underline{0}$.

1.3 Subspaces

Def. (Subspace W of vector space V)

A non-empty subset of V, such that:

1. $\underline{0} \in W$;
2. Closed under vector addition and scalar multiplication.

Thm. (1.4) Subspace is closed under arbitrary intersection.

1.4 Linear combination

Def. (Span)

For a set $S \subset V$, $span(S) := \bigcap_{S \subset \text{subspace } W \subset V} W$.

Rmk. If $S_1 \subset S_2$, then $span(S_1) \subset span(S_2)$.

Prop. $span(S)$ is the set of linear combination of elements in S.

1.5 Linear independence

Def. (Linear dependent)

n distinct s_i, there exists $\lambda_1 \ldots \lambda_n$ that are not all zero, such that $\sum \lambda_i s_i = 0$.

Thm. S are linear independent set of vectors, $v \in V \setminus S$, then $S \cup \{v\}$ are linear dep. iff $v \in span(S)$.

1.6 Bases and dimention

Def. (Basis)

Minimal (defined in the subset inclusion sence, not in size sence) spanning set.

Cor. $span(S) = V$, then it's basis iff it's linear indep.

Thm. (**Replacement thm**) V has a basis $s_1, \dots, s_n$ of size n, let $\{x_1, \dots, x_i\}$ of size i be linear indep. and $i \leq n$, then $\{x_1, \dots, x_i, s_{i+1}, \dots, s_n\}$ (some of si is replaced by xi) is a basis.

Cor. card(linear indep) <= card(basis) <= card(spanning set)

Cor. Basis has the same cardinality.

Cor. If $|S| = \dim V$, then TFAE: a) spanning; b) linear indep; c) basis.

Thm. (1.11) $W \subset V$, $\dim W \leq \dim V$, then $\dim W = \dim V$ iff $W = V$.

Cor. $\dim V < \infty$, $W \subset V$, then W posseses a compliment.

Def. (Quotient space)

Given subspace W, define $x \sim y$ if $x - y \in W$,
$[x] := \{y : x \sim y\} =: \{x + w | w \in W\} =: x + W$, and $\{[x]\} := V_{/W}$ is a vector space called quotient space, by the intuitive definition of addition and scalar multiplication: $[v] = [\sum \lambda_i s_i] := \sum \lambda_i [s_i]$ and $\lambda[x] := [\lambda x]$, e.g. $-[x] = [-x]$.

Prop. $\dim(V_{/W}) = \dim V - \dim W$.

Thm. Given subspace W, there's a bijection between $\{H : subspace\ H, W \subset H\}$ and $\{\bar{H} \in V_{/W} : subspace\ H\}$, where the $\bar{H} := H_{/W} = \{[x] \in V_{/W} : x \in H\}$.

Rmk. This together with the usage of flags give another proof for Cor 1.11.

Def. (Direct sum)

$W_1 \oplus W_2$ if $W_1 + W_2 = V$ and $W_1 \cap W_2 = \emptyset$.

Cor. $\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2)$, by showing $\dim \bar{V} = \dim \bar{W}_1 + \dim \bar{W}_2$.

1.7 Maximal linear independent subset

Def. (Chain / nest / tower)

A collection of elements that are totally ordered.

Thm. (**Hausdorff maximal principle / the axiom of choice**) Every partially ordered set has a maximal linearly / totally ordered subset. It's the same as the next thm.

Thm. (**Zorn's lemma**) For a partially ordered set $(X, \leq)$, for any $C \subset X$ be totally ordered. Suppose $\exists x_c \in X, s.t., \forall x \in C, x \leq x_c$ (every chain has a top), then $\exists x_m, s.t., \forall y \in X, x_m \leq y \to x_m = y$ (maximum exists).

Def. (Maximal linear independent set)

Again, maximal with respect to set inclusion.

Lemma. A set is a maximal linear independent set iff it's a basis.

Thm. For any linearly independent subset S of a vector space V, there's a basis that contains S.

Proof. Construct X to be the collection of independent sets containing S. For any chain C in X, we need to find a top of it in X. This can be done by taking union of sets in C, which means it's a top and therefore containing S. Also, it's independent, since for any u_i for $i = 1 \ldots n$, we can find a set in C such that it contains all these vectors, therefore they're linearly independent.

Cor. Every vector space has basis.

Thm. Subspace $W \subset V$, then $\exists W', s.t. V = W \oplus W'$.

Ch2. Linear Transformations and Matrices

2.1 Rank-nullity

Def. (Linear map)

$T: V \to W$ with $T(av + bw) = aT(v) + bT(w)$.

Rmk.

1. $T(0) = 0$.
2. $Ker(T) \subset V, Ran(T) \subset W$ are subspaces, called **null space / kernel** and **range / image**, and their dimention is called **nullity** and **rank**.
3. (2.4) T is 1-1 iff $Ker(T) = \{0\}$.

Thm. (Dimension thm)

For linear $T: V \to W$, and V is finite-dimensional, then $nullity(T) + rank(T) = \dim(V)$.

Thm. T is isomorphism iff $\exists T^{-1}, s.t., T \circ T^{-1} = id_V. T^{-1} \circ T = id_W, T^{-1}\ linear.$

Thm. (2.19) $T: V \to W$, $\dim V = n < \infty$, then V is isomorphic to W iff $\dim W = n$.

Cor. T is isomorphism, subspace $V' \subset V$, then $T|_{V'}: V' \to T(V')$ is still isomorphism.

Thm. $T: V \to W$ induces isomorphism $\bar{T}: V_{/KerT} \to R(T)$ by letting $\bar{T} := [x] \mapsto T(x)$.

Cor. $\dim V < \infty$, $\dim KerT + \dim R(T) = \dim V$.

Cor. If $V = R(T) + Ker(T)$, then it's direct sum.

Ex. If $T \circ T = T$, then the above is true, and further more, $T = \pi_{R(T)}$.

Ex. Consider subspace $W' \subset W$, then $T^{-1}(W') \subset V$ is a subspace, and another induced linear quotient map $\bar{T}: V_{/T^{-1}(W')} \to W_{/W'}$ can be given by $\bar{T}: [x] \to [T(x)]$. When T is onto, it's bijective.

2.2 Matrix and map

Lemma. For linear map $T : F^n \to F$, there's a unique tuple (a_i), such that $T(x) = \sum_{i=1}^{n} a_i x_i$. Constructively, $a_i = T(e_i)$.

Thm.

For linear map $T : F^n \to F^m$, there's a unique $m \times n$ matrix $A = (a_{ji})$ such that $T(x) = (T_1(x), T_2(x), \ldots)$ and $T_j(x) = \sum_{j=1}^{n} a_{ji} x_i$. We use L_A to refer to T. Further more, $T(e_i) = (a_{1i}, a_{2i}, \ldots)$ is the i-th column of A.

Rmk. We define matrix as a compact representation of a linear transformation between euclidean spaces. Matrix A is defined to be $[L_A]_{\epsilon_n}^{\epsilon_m}$.

Thm. (2.20)

$T : V \to W$, V and W respectively possess ordered bases $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)$, then $T(\beta_i) = \sum_{j=1}^{m} a_{ji} \alpha_j$. Further more, given β, α, there's an isomorphism between T and $[T]_\beta^\alpha = (a_{ji})$. This can be done since $\phi_\beta : V \to F^n, \phi_\alpha : W \to F^n$, we have $L_A \phi_\beta = \phi_\alpha T$.

Ex. Given a complete flag $\mathcal{F}$: $\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n = V$ in V so that $\dim(V_i/V_{i-1}) = 1, \forall i$. We say T is **upper triangular** w.r.t. $\mathcal{F}$ if $T(V_i) \subset V_i, \forall i$. In this case, let β be any ordered basis that can generate the flag, then matrix $[T]_\beta$ will also be **upper triangular** in matrix sense. At the same time, the induced quotient map $\bar{T}_i : V_i/V_{i-1} \to V_i/V_{i-1}, \bar{T}_i : [x] \mapsto [T(x)]$ is given by multiplication by a unique $\lambda_i \in F$. In this case, T is invertible iff $\forall i, \lambda_i \neq 0$, or $a_{ii} \neq 0$ for $[T]_\beta$. If invertible, T^{-1} and $[T^{-1}]_\beta$ also upper triangular w.r.t. $\mathcal{F}$.

Thm. (2.11)

$[S \circ T]_\beta^{\beta''} = [S]_{\beta'}^{\beta''} [T]_\beta^{\beta'}$.

Def. (Nilpotent)

For a non-zero matrix A, it's called nilpotent if $\exists n \in \mathbb{N}, s.t., A^n = 0$.

> Prop. Multiplicative property of A: non-communative, no cancellation, and there exist nilpotent matrix.

> Prop. If $T: V \to W, \dim V = \dim W = n$, T.F.A.E:
>
> 1. T is an isomorphism;
> 2. $\exists \beta$ as a basis of V, s.t. $T(\beta)$ is a basis of W.
> 3. $\forall \beta$ as a basis of V, $T(\beta)$ is a basis of W.

Proof. (1->3) We know $card(T(\beta)) \leq n$, and since T is onto, $T(\beta)$ spans W.

Thm. (2.22 Change of basis)

Say $\dim V = n$,

1. $A = [Id_V]_\beta^\alpha \in M_{n\times n}$ is invertible;
2. Fix β/α, then any invertible A is $[Id_V]_\beta^\alpha$ for some unique α/β.

Proof.

1. The inverse is $[Id_V]_\alpha^\beta$;
2. Say fix $\beta = (s_1, \dots, s_n)$, for invertible $A = [A_1, \dots, A_n]$, find unique $\alpha = (t_1, \dots, t_n)$. Let $\phi_\beta, \phi_\alpha : F^n \to V$ be the translation isomorphism. Since $\{A_j\}$ is a basis of F^n, $\phi_\beta(\{A_j\})$ is a basis of F^n. Let $t_i = \phi_\beta(A_i) = \phi_\beta(L_A(e_i)) = \sum_j a_{ji}s_j$. Then we can write $Id(t_i) = \sum_j a_{ji}s_j$, which means $[Id]_\alpha^\beta = A$, and so $[Id]_\beta^\alpha = A^{-1}$.
3. So if we apply the above construction with A^{-1} in place of A to construct $\alpha' = \{\phi_\beta(A_i^{-1})\}$, then $[Id]_\beta^{\alpha'} = A$. This gives the existence of α in original statement.
4. If $[Id]_\beta^\alpha = [Id]_\beta^\gamma$, then $[Id]_\alpha^\beta = [Id]_\gamma^\beta$, and then $\alpha = \gamma$, which shows the uniqueness.

Rmk. $V \xrightarrow{T} W, \dim V = n, \dim W = m$, then $B := [T]_{\beta'}^{\alpha'} = [Id_W]_\alpha^{\alpha'}[T]_\beta^\alpha[Id_V]_{\beta'}^\beta =: QAP$. This inspires an equivalence relation on $M_{m\times n}$, i.e., $A \sim B$ iff $B = [L_A]_{\beta'}^{\alpha'}$ for some ordered bases β' of V and α' for W.

> Prop. If $rk(A) = r$, then there're invertible matrices P, Q s.t.
>
> $$PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \in M_{m\times n}$$

Rmk. That means there's only $\min(n, m) + 1$ many equivalence classes. Also, that means $rkA \leq \min(n, m)$. This result will be justified later.

Proof.

1. By replacement thm, we can pick α so that $R(L_A) = span(t_1, \ldots, t_r)$. For $i \leq r$, since $t_i \in R(L_A)$, we can find s_i s.t. $L_A(s_i) = t_i$.
2. Claim $\{s_i\}_{i=1}^r$ are independent. Since $\bar{L}_A : F^n/ker(T) \to R(L_A)$ is isomorphism and $R(L_A) = span(t_1, \ldots, t_r)$, we have $\{[s_i]\}$ forming a basis in $F^n/ker(T)$.
3. Now let $W = span(s_1, \ldots, s_r)$. Claim $W \cap ker(L_A) = 0$, otherwise contradicts with the independence.
4. By rank-nullity, $F^n = W \oplus ker(L_A)$. Merge them into one basis β we seek.

2.3 Duality

Def. (Dual space)

$V^* := \mathcal{L}(V, F)$.

Rmk.

1. $(F^n)^* \cong F^n$.
2. Further more, when $\dim V = n$ and a basis is given, $V^* := \mathcal{L}(V, F) \cong \mathcal{L}(F^n, F) =: (F^n)^* \cong F^n$.
3. Which means although the "all linear functionals" looks scary, the cardinality doesn't increase.

Def. (Dual basis)

$s_i^* : V \to F$ defined by $s_i^*(s_j) = \mathbb{1}(i = j)$. Then $\beta^* := \{s_i^*\}$ is a basis of V^*.

E.g. $e_i^*(e_j) = \mathbb{1}(i = j) =: \delta_{ij}$. Then e_i^* is the functional that essentially picks the i-th coordinate.

Def. (Dual map)

$T^*: W^* \to V^*, T^*: \phi \mapsto \phi \circ T.$

Rmk. $V(= span(\beta)) \overset{T}{\to} W(= span(\alpha)) \overset{\phi \in W^*}{\to} F.$

Thm. (Transpose)

If $A = [T]_\beta^\alpha$, then $A^T = [T^*]_{\alpha^*}^{\beta^*}$.

Proof.

1. It suffices to show $T^*(t_i^*) = \sum_i (A^T)_{ji} s_j^* = \sum_i a_{ij} s_j^*$.
2. $T^*(t_i^*)(s_k) = t_i^* \circ T(s_k) = t_i^*(\sum_j a_{jk} t_j) = a_{ik}$.
3. $\sum_i a_{ij} s_j^*(s_k) = a_{ik}$.

> Lemma. $rkT = rkL_A$.

Proof. Since $R(T) = R(T \circ \varphi_\beta) = R(\varphi_\alpha \circ L_A) = R(L_A)$.

Thm.

$rkA = rkA^T$.

Proof.

1. It suffices to show that $rkT = rkT^*$, where $T: V \to W$.
2. $kerT^* = \{\varphi \in W^*: \varphi \circ T = 0\}$. It's usually denoted as $W^\perp$. Write $\varphi = \sum_i a_i t_i^*$.
3. $\varphi \circ T = 0$ iff $\varphi(R(T)) = 0$, pick basis so that $R(T) = span(t_1 \dots t_r)$, then iff $a_1 \dots a_r = 0$ iff $kerT^* = span(t_{r+1} \dots t_m)$. Then $rkT^* = m - r = rkT$.

Thm. (Double dual)

> There's a canonical isomorphism between V and V^{**} that doesn't depends on choice of bases, given by $hat: V \to \mathcal{L}(V^*, F)$ and $hat: x \mapsto \hat{x}$, where $\hat{x}: V^* \to F$, $\hat{x}: \varphi \mapsto \varphi(x)$.

Proof.

1. $\hat{x}$ is linear;
2. hat is linear;
3. hat is bijective. The case of infinite dimension is #NotCovered . Otherwise, $\dim V^{**} = \dim V$, we need only 1-1 or onto. We show 1-1 here. Whenever $\hat{x} = 0$, i.e. $\forall \varphi \in \mathcal{L}(V, F), \varphi(x) = 0$. Suppose $\exists x_0 \neq 0$ follows the above condition. When it's non-zero, one thing we can tell by replacement thm is that we can pick a basis in V as $\beta := (x_0, \ldots)$, then we got $x_0^*(x_0) = 1 \neq 0, x_0^* \in \mathcal{L}(V, F)$, contradicts.

> Cor. If $\dim V < \infty$, then for any basis γ of V*, there's a basis β of V, s.t. $\beta^{}=\gamma$.*

Proof. It's nice to be able to regard linear transformation as elements of vector space. For $\gamma := (\varphi_1, \ldots)$, we can generate $\gamma^* := (\varphi_1^*, \ldots)$, and find unique $\beta := (x_1, \ldots), s.t. \hat{x}_i = \varphi_i^*$. It's what suggested by the notation since $\varphi_i(x_j) =: \hat{x}_j(\varphi_i) = \varphi_j^*(\varphi_i) := \delta_{ji} = \delta_{ij}$.

Ch3. Elementary Matrix Operations and Systems of Linear Equations

3.1 Elementary Matrix Operations

> **Def. (Elementary operations on row)**
>
> $A \in M_{m\times n}$,
>
> 1. Interchanging two rows;
> 2. Multiplying each element in a row by a non-zero number;
> 3. Adding a scalar λ multiple of j-th row to i-row ($E = I_m + \lambda e_{ij}$, A'=EA).

Elementary matrix is a matrix obtained by performing an elementary operation on I_n.

Thm. (3.1)

Performing an elementary operation on a matrix is equivalent to multiplying the matrix by an elementary matrix.

Thm. (3.2)

Inverse and transpose of elementary matrix are still elementary of the same type.

Thm. (3.4)

If P, Q are invertible, then $rk(PAQ) = rk(A)$.

Proof. We can express PAQ into $[Id]_\epsilon^\alpha [L_A]_\epsilon^{\epsilon'} [Id]_\beta^{\epsilon'}$, where α, β can be given by thm2.22. Then $PAQ = [L_A]_\beta^\alpha$, and then $rk(PAQ) = rk(A)$.

Thm. (3.6, Rank-preserving)

There're invertible matrix P, Q that are product of elementary matrices, s.t. $PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$, where $r = rkA$. Thus we have rank-preserving matrix operations.

Proof. Constructive induction. Transform A into $\begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix}$, then $R(A) = span(e_1) \oplus R(B)$, and doing row and column transformations on B won't affect the first row and column, thus induction works.

> Cor. Every invertible matrix in $M_{n\times n}$ is a product of elementary matrices.

Proof. $PAQ = I_n$, then $A = Q^{-1} I_n P^{-1}$ is a product of elementary matrices.

> Cor. $rkA = rkA^T$. Thus rows and columns generate subspaces of the same dimension.

Proof. Since transposing $PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ gives $Q^T A^T P^T = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$.

> Cor. (3.7) $rk(AB) \leq rk(A), rk(B)$.

Proof. The first is trivial, and second is because
$rk(AB) = rk(B^T A^T) \leq rk(B^T) = rk(B)$.

3.2 Systems of Linear Equations

Def. (System of linear equations)

$Ax = b$, where x is the variable vector.

Rmk. Based on results from Ch2, we know the system of linear equations has a solution if $b \in R(L_A)$. If so, i.e., $\exists s, s.t.\ As = b$, then the solution space as a preimage of b is $s + KerL_A$.

> Thm. (3.11) If the solution set is nonempty, the system is called **consistent**. It's consistent iff $rkA = rk(A|b)$.

> Thm. (3.13) Given an invertible matrix P, then $Ax = b$ iff $PAx = Pb$.

Def. (Row Reduced Echelon Form, RREF)

1. Any nonzero row precedes zero row;
2. The first nonzero entry in each row is 1;
3. The first nonzero entry in each row is the only nonzero entry in its column and to the right of the first 1 in the preceding row.

Thm. (3.16)

$\forall A \in M_{n\times n}$ with rank r:

1. $\exists P, s.t.\ PA$ is RREF;

2. Say $B = PA$ with $rkB = r$. For each $i = 1 \dots r$, there's a column $b_{j_i} = e_i$. We claim that $\{a_{j_i}\}$ is a basis of column space of A;
3. If both P_1A, P_2A are RREF, then $P_1A = P_2A$.

Proof.

1. Gaussian elimination produce a product of elementary matrices;
2. $\sum_i c_i a_{j_i} = 0 \implies \sum_i c_i M a_{j_i} = 0 \implies \sum_i c_i e_i = 0$, thus $c_i = 0$.
3. Since B has only r nonzero rows, every column of $B = (Pa_1, Pa_2, \dots)$ has the form $b_k = Pa_k = (d_1, \dots d_r, 0, \dots)$, then $a_k = P^{-1}b_k = P^{-1}(\sum_{i=1}^r d_i e_i) = P^{-1}(\sum_{i=1}^r d_i b_{j_i}) = \sum_{i=1}^r d_i a_{j_i}$. Since d_i are uniquely dependent on A, B is unique.

Ch4 Determinant

Def. (Bilinear, alternating)

1. $B : V \times W \to F$ s.t. $\forall w \in W, v \mapsto B(v, w)$ and $\forall v \in V, w \mapsto B(v, w)$ are linear. E.g. Dot product.
2. When B is bilinear with $V = W$, if $\forall v, B(v, v) = 0$, then it's **alternating**.

> Lemma. If B is bilinear alternating, then $\forall v_1, v_2 \in V, B(v_1, v_2) = -B(v_2, v_1)$. The converse is true if $\frac{1}{2} \in F$.

Def. (Multilinear, alternating)

1. Multilinear if $M : V_1 \times \dots \times V_n \to F$ is linear on each entry.
2. When B is multilinear with $V_1 = V_2 \dots$, it's **alternating** if $(\exists i, v_i = v_{i+1}) \implies \delta(v_1 \dots v_n) = 0$.

Ex. Say H is the vector space of all function of the form $V_1 \dots \times V_n \to F$, then S the set of multilinear func is a subspace of H with dimension $\prod_i dimV_i$.

> Lemma. (1) -> (2) -> (3), and (3) -> (1) if $\frac{1}{2} \in F$:

> 1. $(\exists i, v_i = v_{i+1}) \implies \delta(v_1 \dots v_n) = 0$;
> 2. $(\exists i, j, v_i = v_j) \implies \delta(v_1 \dots v_n) = 0$;
> 3. $\forall i, j, \delta(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -\delta(v_1, \dots, v_j, \dots, v_i, \dots, v_n)$

Def. (Determinant)

Multilinear, alternating, and $\det I_n = 1$.

> Lemma. A is invertible only if $\det A \neq 0$.

> Lemma. $\det AE = -\det A, \lambda \det A, \det A$ respectively if E is of type 1, 2, and 3.

> Cor. $d(E) = -1, \lambda, 1$ respectively if E is of type 1, 2, and 3, by letting $A = I$.

> Cor. $\det AE = \det A \det E$.

> Cor. $\det AB = \det A \det B$. Therefore $\det AB = \det BA$.

Proof. If B is not invertible, then $rk(AB) \leq rk(B) < n$, AB is not invertible either, then both sides are 0; If B is invertible, according to cor of thm3.6 (Linear algebra > ^377d64), it is a product of elementary matrices.

> Cor. A is invertible iif $\det A \neq 0$. In that case, $\det A^{-1} = (\det A)^{-1}$.

> Cor. (**Uniqueness**) There's at most one determinant for each n. Since the values on elementary matrices are already fixed.

> Lemma. $\det A = \det A^T$. Since $\det E = \det E^T$.

Thm. (Cumer's rule)

For $Ax = b$, consider $M_k := (A_1, \dots, A_{k-1}, b, A_{k+1}, A_n)$, then $x_k = \frac{\det M_k}{\det A}$.

Def. (Minor matrix, cofactor of entry)

$A \in M_{n\times n}$, then a **minor** matrix is $A_{ij} \in M_{n-1,n-1}$ that removes i-th row and j-th column of A, and the **cofactor of entry** a_{ij} is $\hat{a}_{ij} := (-1)^{i+j} \det A_{ij}$.

Thm. (Existence)

For fixed j, $\det A = \sum_{i=1}^{n}(-1)^{i+j} a_{ji} \det A_{ji} = \sum_{\sigma}(-1)^{sgn\sigma} \prod a_{\sigma(i),i}$.

Proof. Expend the first row first, and then the column:

$$\det A = \det \begin{pmatrix} \sum_i a_{1i} e_i^T \\ r_2 \dots \\ r_n \end{pmatrix} = \sum_i a_{1i} \det \begin{pmatrix} e_i^T \\ r_2 \\ \dots \\ r_n \end{pmatrix},$$

$$\begin{aligned} \det \begin{pmatrix} e_i^T \\ r_2 \\ \dots \\ r_n \end{pmatrix} &= \det\left(c_1, \dots, e_1 + \textstyle\sum_{j=2}^{n} a_{ji} e_j, \dots, c_n\right) \\ &= \det\left(c_1, \dots, e_1, \dots, c_n\right) \\ &= (-1)^{i-1} \det\left(e_1, c_1, \dots, c_n\right), \\ &\text{where } c_{10}, \dots c_{n0} = 0, \end{aligned}$$

$$\text{thus } (e_1, c_1, \dots, c_n) \text{ is multilinear alternating, thus is determinant}$$

$$\det A = \sum_{i=1}^{n} (-1)^{i+1} a_{1i} \det(A_{1i})$$

> Cor. For upper triangular matrix A, $\det A = \prod a_{ii}$.

> Notation. $\det T := \det[T]_\alpha^\alpha$ for any choice of α, since $\det[T]_\alpha^\alpha = \det([I]_\alpha^\beta [T]_\beta^\beta [I]_\beta^\alpha) = \det[T]_\beta^\beta$.

Thm. (4.1)

$\det \begin{pmatrix} B & C \\ 0 & D \end{pmatrix} = \det B \det D$.

> Cor. Subspace $W \subset V, \beta = (s_1, \dots, s_r, \dots, s_n)$ is a basis of V while $\alpha = (s_1, \dots, s_r)$ is a basis of W and $\bar{\alpha} = ([s_{r+1}], \dots, [s_n])$ is a basis of V/W. Let $[T]_\beta^\beta = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$. Then $B = [T]_\alpha^\alpha, D = [\bar{T}]_{\bar{\alpha}}^{\bar{\alpha}}$, and thus $\det T = \det(T|_W) \det \bar{T}$.

Ch5 Eigenvalue

5.1 Polynomials

Def. (Irreducible over F)

Polynomial $p(t)$ is irreducible over F if $p(t) = q(t)s(t) \to$ "either $q(t)$ or $s(t)$ is constant". E.g. the polynomial $x^2 - 2$ is irreducible over the integers but not over the reals.

> Prop. $\forall p(t), \exists q(t), s(t), s.t.\ q(t)s(t) = p(t)$ and $q(t), s(t)$ are irreducible. If all $p(t), q(t), s(t)$ have leading 1, then $q(t), s(t)$ are unique up to reordering.

Def. (Split)

A polynomial over F is split if there is a factorization $p(t) = a(t - \lambda_1)\dots(t - \lambda)$ for some $a \in F \setminus \{0\}, \lambda_i \in F$. We do not require that λ_i are all distinct.

(Fundamental Theorem of Algebra, FTOA)

1. If $F = \mathbb{C}$, then $p(t)$ is irreducible iff $p(t)$ is linear;
2. If $F = \mathbb{R}$, then $p(t)$ is irreducible iff $p(t)$ is linear or a degree 2 polynomial with no real solution;

> Lemma. (**Polynomial division**) For polynomials $p, s, \exists! q, r$, s.t., $p = q \cdot s + r, (r = 0 \wedge deg(r) < deg(s))$.

> Cor. $p(\lambda) = 0 \iff p = (t - \lambda) \cdot q$.

> Cor. $p(\lambda) = 0, p \neq 0$, then $\exists! k \geq 1, s.t., p = (t - \lambda)^k \cdot q, q(\lambda) \neq 0$. k is called the **multiplicity** of λ.

> Prop. (5.1) If p split, $p = q \cdot r$, then q, r are split.

> Cor. If p split, say $p = a_n \prod (t - \lambda_i)^{k_i}$, then $\sum k_i = n$.

5.2 Eigenvalue

Def. (Eigenline)

An eigenline of a linear transformation is a subspace charaterized by $span(\{v\})$, s.t., $\exists \lambda, Tv = \lambda v$. Such $v \in span(\{v\})$ is called **eigenvecter**, and λ is called **eigenvalue**.

Rmk. If we regard matrix/transformation W as a space movement in Euclidean space, we need to apply it on certain vector to examine its feature. E.g. When A is the adjacency matrix, $(A\vec{v})_i = \frac{1}{deg_i}\sum_{j\in N(i)} v_j$. When $L = I - D^{-1}A$, the Laplacian matrix, $(L\vec{v})_i = \frac{1}{deg_i}\sum_{j\in N(i)}(v_i - v_j)$. What if we try to apply it multiple times?

$$\vec{v} = \sum_i \alpha_i \vec{u}_i$$
$$W^k\vec{v} = \sum_i \alpha_i W^k \vec{u}_i = \sum_i \alpha_i \lambda_i^k \vec{u}_i$$

We find out that the largest eigenvalue corresponding eigenvector will eventually dominate as k getting larger and larger. That's why we would like to conclude:

- first principle eigenvalue(largest) indicates the movement speed;
- first principle eigenvector indicates the movement direction.

Def. (Charateristic polynomial)

$p_A(t) := \det(A - tI_n)$, a polynomial of t.

Rmk.

1. $p_A(t)$ can be written as $(-1)^n t^n + (-1)^{n-1} trA \cdot t^{n-1} + \sum_{i=1}^{n-2} a_i t^i + \det A \cdot t^0$.
2. If $B = PAP^{-1}$, then $p_B(t) = \det(PAP^{-1} - \lambda I_n) = \det(P(A - \lambda I_n)P^{-1}) = p_A(t)$.
3. (Characteristic polynimial of linear map) $p_T(t) := \det([T]_\beta^\beta - tI_n)$ for any choice of β.
4. $$\begin{aligned}\exists v, Tv = \lambda v \iff \exists v, (T-\lambda I)v = 0 \iff Ker(T-\lambda I) \neq 0 \\ \iff \det(A - \lambda I_n) = 0 \iff p_A(t) \text{ has root } \lambda.\end{aligned}$$

   Therefore, if $p_A(t)$ split, $p_A(t) = \prod(\lambda_i - t)$. Then $\det A = \prod \lambda_i$.
5. T is invertible iff 0 is not an eigenvalue of T;

6. Invertible T, then λ is an eigenvalue of T iff λ^{-1} is an eigenvalue of T^{-1}.
7. $p_A(t) = p_{A^T}(t)$. This is because $\det(A - tI) = \det(A - tI)^T = \det(A^T - tI)$.

Def. (Generalized eigenvector)

v is a **generalized eigenvector** of $\lambda \in F$ if $(T - \lambda I)^k v = 0$ for some $k \geq 1$.

Rmk.

1. If $k = 1, v \neq 0$, then v is an eigenvector;
2. If $v \neq 0$, then λ is an eigenvalue, since $\exists k' \in [0, k-1], w := (T - \lambda I)^{k'} v \neq 0, (T - \lambda I)w = 0$.

Def. (Eigenspace)

Subspace $E_\lambda := \{v \in V : (T - \lambda I)v = 0\}$ is the **eigenspace** for λ; Subspace $V_\lambda := \{v \in V : \exists k \geq 1, s.t., (T - \lambda I)^k v = 0\}$ is the **generalized eigenspace** for λ. Note that $E_\lambda \subset V_\lambda, T(V_\lambda) \subset V_\lambda$.

Rmk.

1. For an uppertriangular matrix, the numbers on the diagonal are the eigenvalues, by showing that $Ker(A - a_{ii}I) \neq 0$ since it's no longer invertible but this is linear operator.
2. $T(V_\lambda) \subset V_\lambda$: $\forall v, s.t. (T - \lambda I)^n v = 0$, then $(T - \lambda I)^n (Tv) = T((T - \lambda I)^n v) = 0$. The commutivity comes from the nature of polynomial of transformation.
3. As said before, A and A^t have the same charateristic polynomial, where multiplicity is known to be the same. Now furthermore, $\dim E_\lambda = \dim E'_\lambda$. Thus if A is diagonalizable, so is A^t.

> Lemma. $\dim V < \infty, \exists! W$ with decomposition $V = V_0 \oplus W$, s.t. $T|_W : W \to W$ is an isomorphism, where V_0 is the generalized eigenspace for 0. In fact, $W = R(T^n), V_0 = KerT^n$.

Proof.

1. (Existence) Consider the flag $V \supset R(T) \supset \cdots \supset R(T^n) \supset \{0\}$. If all n+1 containing relations are strict, it contradicts with $dimV = n$. Thus $\exists k$ s.t. $R(T^{k-1}) = R(T^k)$. Since $T : R(T^{k-1}) \to R(T^k)$ is also onto, it is an isomorphism. Thus we can take $W := R(T^n)$, where $R(T^n) = \ldots R(T^k)$ due to the same dimension. It's a decomposition because $R(T^n) \cap Ker(T^n) = 0$, thus $R(T^n) \oplus Ker(T^n)$. According to rank-nullity thm, their direct sum have the same dim as V. Thus $KerT^n \oplus W = V$. By definition, $KerT^n \subset V_0$. $\forall v \in V, v = v_0 + w$ where $v_0 \in KerT^n, w \in W$ based on the decomposition. Now suppose $T^n v = 0$, then $0 = T^n v = T^n v_0 + T^n w = T^n w$, then $w = 0$, thus $T^n v = 0 \to v \in V_0$, that is $KerT^n \subset V_0$, thus $KerT^n = V_0$.
2. (Uniqueness) Suppose W is not unique, i.e. exists $W' \neq W, V = V_0 \oplus W', T(W') \subset W'$. Then $\forall v, v = v_0 + w'$, then $T^n v = T^n w'$, yet LHS in W and RHS in W', thus $W \subset W'$.

> Cor. $\dim V < \infty$. For any eigenvalue λ, $\exists!$ decomposition $V = V_\lambda \oplus W$, s.t. $(T - \lambda I)|_W : W \to W$ is an isomorphism and $T(W) \subset W$, where V_λ is the generalized eigenspace for eigenvalue λ. In fact, $W = R((T - \lambda I)^n), V_\lambda = Ker(T - \lambda I)^n$.

Proof. Let $S := T - \lambda I$, apply the lemma, we get $W = R(S^n)$, $S|_W : W \to W$ is isomorphism, $V = V_\lambda \oplus W, V_\lambda = Ker((T - \lambda I)^n)$. What we need to show is $T(W) \subset W$, which is trivial.

Thm. (5.2.1)

$\dim V = n < \infty, T : V \to V$, T.F.A.E:

1. $p_T(t) = (-1)^n \prod_{i=1}^r (t - \lambda_i)^{n_i}$ split;
2. $V = \bigoplus_{i=1}^r V_{\lambda_i}$;
3. T is **triangularizable**, i.e. $\exists$ ordered basis β, $[T]_\beta^\beta$ is upper triangular.

Proof.

1. (1=>2) If $p_T(t)$ split, pick a root λ_1. Since $V = V_{\lambda_1} \oplus W$ and each is T-invariant, they are $(T - \lambda_1 I)$-invariant. According to thm4.1 (Linear algebra > ^061166), $p_T = p_{T|_{V_{\lambda_1}}} p_{T|_W}$, and both are split according to prop5.1 (Linear algebra > ^c619e8).

2. $p_{T|_W}$ has no root λ_1, otherwise $\det(T|_W - \lambda_1 I) = 0$, yet it is invertible. OTAH, $p_{T|_{V_{\lambda_1}}}$ has only root / eigenvalue λ_1. Suppose not, $\exists \mu \in F, v \in V_{\lambda_1} \setminus \{0\}$, s.t. $Tv = \mu v$, then $(T - \lambda_1 I)v = Tv - \lambda_1 v = (\mu - \lambda_1)v$, yet by definition, $(T - \lambda_1 I)^n v = 0$. Then $(\mu - \lambda_1)^n v = 0$, which means $\mu = \lambda_1$.
3. Now we can apply induction to have $W = \bigoplus_{i=2}^{r} W_{\lambda_i}$. Since by definition, $W_{\lambda_i} = V_{\lambda_i} \cap W$, all we need to show is $W \supset V_{\lambda_i}$. $\forall x \in V_{\lambda_i}$, basic-decomp $v = x + y$, where $x \in V_{\lambda_1}, y \in W$. Now $0 = (T - \lambda_i)^n v = (T - \lambda_i)^n x + (T - \lambda_i)^n y$. Since $(T - \lambda_i)^n$ is isomorphism on V_{λ_1}, this force $(T - \lambda_i)^n y \in V_{\lambda_1} \cap W = \{0\}$, thus $x = 0$.
4. (2=>3) $[T]_\beta^\beta$ is upper triangular iff $\forall i, T(s_i) \in span\{s_1 \dots s_i\}$. Since each is T-invariant, we only need to show each $T|_{V_\lambda}$ is triangularizable. WLOG, $V = V_\lambda$. Pick any s_1 so that $s_1 \neq 0, TS_1 = \lambda s_1$. Then consider quotient $\bar{V} := V/span(s_1), \bar{T} : \bar{V} \to \bar{V}$. Note that $(\bar{V})_\lambda = \bar{V}$. Apply induction on $\dim V$ to get $\forall j, \bar{T}[s_j] = \lambda[s_j] + \sum_{i=2}^{j} a_{ij}[s_i] = [\sum_{i \le j} a_{ij} s_i]$. Then $Ts_j = \sum_{i \le j} a_{ij} s_i + a_{1j} s_1$.
5. (3=>1) Note that $p_T(t) = \prod (a_{ii} - t)$, thus split.

Cor. If $p_T(t)$ split, $\dim V_{\lambda_i} = multi_p(\lambda_i)$, i.e. $p_T(t) = \prod (\lambda_i - t)^{\dim V_{\lambda_i}}$.

#TODO

Thm. (5.2.2)

$\dim V = n < \infty, T : V \to V$, T.F.A.E:

1. $p_T(t) = (-1)^n \prod_i^r (t - \lambda_i)^{n_i}$ split, and $\forall \lambda \in F, \dim(E_\lambda) = multi(\lambda)$;
2. $V = \bigoplus_{i=1}^{r} E_{\lambda_i}$;
3. T is **diagonalizable**, i.e. $\exists$ ordered basis β, $[T]_\beta^\beta$ is diagonal.

Proof. #TODO

> Lemma. $\exists$ polynomial $q(t)$ over F s.t. $q(T) = 0$.

Proof. $q(T) \in \mathcal{L}(V, V)$. #TODO

> Def. (**T-cyclic subspace of V generated by x**) #TODO

Thm. (Cayley-Hamilton thm)

$Ann(T) := \{q(t) \in F(t) : q(T) = 0\}$. Then $P_T(t) \in Ann(T)$.

E.g. If $V = V_\lambda$, $P_T(t) = \det(T - tI)^n = (t-\lambda)^n \in Ann(T)$.

Proof.

1. Method 1: prove it on $F = \mathbb{C}$ first. #TODO
2. Method 2: more elementary. #TODO

> Cor. T is nilpotent iff $p_T(t) = (-1)^n t^n$.

Proof. (=>) Again, consider the flag $V \supset R(T) \supset \cdots \supset R(T^n) \supset \{0\}$ like we did in lemma (Linear algebra > ^81ced0). If it is nilpotent, $R(T^n) = 0$, which means $V_0 = KerT^n = V$, then T is triangularizable, and the only eigenvalue is 0. Then when it is triangularized, the diagonal will be zero, thus $p_T(t) = (0-t)^n$.

Ex. $\dim V < \infty, T(W) \subset W$, suppose that $v_1 \dots v_k$ are eigenvectors with distinct eigenvalues, then $\sum v_i \in W \implies \forall i, v_i \in W$.

Cor. Diagonalizable/triangularizable T, $T(W) \subset W$, then $T|_W$ is also diagonalizable/triangularizable.

Ch6 Inner Product

6.1 Inner product

Motivation. Dot product is a bilinear form, with the properties $B(x,x) \geq 0, B(x,x) = 0 \iff x = 0$. The consequences are:

1. $\mathbb{R}^n \to (\mathbb{R}^n)^*, x \mapsto (y \mapsto B(x,y))$ is an isomorphism;
2. Length $||x|| := \sqrt{B(x,x)}$;
3. $v \in V, c \in F, ||cv||^2 = |c|^2 \cdot ||v||^2$.

Background. (**Conjugate**) The following properties hold:

1. $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$;
2. $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$;
3. $\overline{(\overline{z})} = z$;

4. $\sqrt{z \cdot \bar{z}} = |z|$;

Def. (Inner Product)

A function $< \cdot, \cdot >: V \times V \to F$ that has the following properties:

1. $<,>$ is linear in the first variable;
2. $< x, y > = \overline{< y, x >}$;
3. $< x, x > \in \mathbb{R}^{\geq 0}$ and $< x, x > = 0 \iff x = 0$.

Rmk.

1. (Antilinear) $< x, \lambda y_1 + y_2 > = \overline{\lambda} < x, y_1 > + < x, y_2 >$. Thus when $\lambda \in \mathbb{R}$, the computation is the same as bilinear.
2. Usually when there is a term that is going to be unpacked, try to put it in the first variable.

E.g. When $V = \mathbb{C}, < x, y > := \sum x_i \overline{y_i}$.

Def. (Matrix Adjoint)

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

Rmk. (**Notational simplification**) For $x, y \in F^n, < x, y > = y^* x$.

E.g. (**Frobenius inner product**) $< A, B >$ can be defined as $tr(AB^*)$.

> Lemma. (**Pythagorean theorem**) $< u, v > = 0 \implies ||u + v|||^2 = ||u||^2 + ||v||^2$.

> Prop. (6.1.1) $F = \mathbb{R}, \mathbb{C}, M \in M_{n \times n}(\mathbb{R}), M = M^t$, then $\forall$ eigenvalue $\lambda \in \mathbb{R}$.

Proof. $Mv = \lambda v$, then $(Mv)^* = v^* M^* = v^* M$ equals to $(\lambda v)^* = \bar{\lambda} v^*$. Then on one hand, $v^* M v = (v^* M) v = (\bar{\lambda} v^*) v = \bar{\lambda} v^* v$, OTAH, $v^* M v = v^*(Mv) = v^*(\lambda v) = \lambda v^* v$, therefore $\lambda = \bar{\lambda}$.

Thm. (Cauchy-Schwarz)

$| < u, v > | \leq ||u|| \cdot ||v||$.

Proof. The case of $v=0$ is trivial. Otherwise, for any c, we have $0 \leq < x-cy, x-cy >=< x,x > -\bar{c} < x,y > -c < y,x > +c\bar{c} < y,y >$. Now let $c = \frac{<x,y>}{<y,y>}$, then we have $0 \leq < x,y > - \frac{|<x,y>|^2}{<y,y>}$.

> Cor. (**Triangle inequality**) $||u+v|| \leq ||u|| + ||v||$.

Proof. $||u+v||^2 =< u+v, u+v >= ||u||^2 + ||v||^2 + 2Re(< u,v >)$, which is less than $||u||^2 + ||v||^2 + 2||u|| \cdot ||v| = (||u|| + ||v||)^2$.

E.g. Given basis $\beta = (s_1, \dots, s_n)$, then for $x = \sum a_i s_i, y = \sum b_i s_i$, $< x,y >$ can defined as $\sum a_i \overline{b_i}$.

Thm. (Essentially one inner product)

For any inner product $<,>$, $\exists$ basis $\beta = (s_1, \dots, s_n)$ s.t. $< x = \sum a_i s_i, y = \sum b_i s_i >= \sum a_i \overline{b_i}$.

Proof. #TODO

6.2 Orthogonal

Def. (Orthogonal and orthonormal)

1. The set $\{s_1, \dots, s_m\}$ is **orthogonal** / **perpendicular** if all $e_i \neq 0$ and $\forall i \neq j, < s_i, s_j >= 0$;
2. The set $\{s_1, \dots, s_m\}$ is **orthonormal** if it is orthogonal and $||s_i|| = 1$.

> Lemma. $v \neq 0, c = \frac{<u,v>}{||v^2||}, w = u - cv$, then $< v,w >= 0$.

Thm. (6.3)

For orthogonal set $s_1, \dots, s_k$, 1) if $w \in span(s_1, \dots, s_k)$, then $w = \sum_{i=1}^{k} \frac{<w,s_i>}{||s_i||^2} s_i$; The coefficients are called **Fourier coefficients** when the set is orthonormal. 2) $s_1, \dots, s_k$ are independent.

Proof. Say $w = \sum a_j s_j$, then $< w, s_i >= a_i ||s_i||^2$.

> Cor. (6.2.1) $T : V \to W$ with orthonormal ordered basis $\beta = (s_1, \dots, s_n)$ for V and $\alpha = (t_1, \dots, t_n)$. Let $A = [T]_\beta^\alpha$, then $A_{ij} =< T(s_j), t_i >_W$.

Thm. (Gram-Schmidt process)

Given linearly independent $\{s_i\}$. Suppose $v_1 = s_1, v_i = s_i - \sum_{j=1}^{i-1} \frac{<s_i, v_j>}{||v_j||^2} v_j$. Then $\forall i, span(s_1, \dots, s_i) = span(v_1, \dots, v_i)$, i.e. forming the same flags, and each $\{v_1, \dots, v_i\}$ is orthogonal.

Proof.

1. Proof by induction. Since $span(s_1, \dots, s_i) \supset span(v_1, \dots, v_i)$, by thm 6.3, we only need to show 1) all $v_i \neq 0$; 2) the independence by orthogonality.
2. If $v_i = 0$, independence of $\{s_i\}$ fails.
3. Assume it's true on $i - 1$. $\forall k < i, < v_i, v_k >=< s_i, v_k > - \sum_{j<i} \frac{1}{||v_j||^2} < s_i, v_j >< v_j, v_k >= 0$.

> Cor. For $W \subset V, \exists$ orthonormal basis $(e_1, \dots, e_n)$, s.t. $W = span(e_1, \dots, e_m)$. Furthur more, W can has $< \sum a_i e_i, \sum b_i e_i >_W= \sum a_i \overline{b_i}$.

Def. (Orthogonal complement)

$S^\perp := \{x : \forall y \in S, < x, y >= 0\}$.

> Prop. (6.6) $V = W \oplus W^\perp$.

Proof. Inspired by the above corollary, WTS $W^\perp = span(e_{m+1} \dots e_n)$. To show the latter contains the former, consider any element in the former. According to thm6.3, it is not in the span of basis of W.

> Cor. $\dim W < \infty, W \subset V, \forall x \notin W, \exists y \in W^\perp, < x, y >\neq 0$.

Proof. For $x = w + w', w' \neq 0$, we can let $y = w'$, then $< x, y >= ||y||^2 \neq 0$.

> Cor. $\dim V < \infty, (W^\perp)^\perp = W$.

Proof. Last corollary implies that for $x \in V, \forall y \in W^{\perp}, < x, y >= 0 \implies x \in W$. Then it follows.

> Cor. $R(T^*)^{\perp} = N(T)$. When $\dim V < \infty$, $R(T^*) = N(T)^{\perp}$.

> Cor. (Least distance function) For $W \subset V$, let $v \in V$, then $\exists! w \in W, w' \in W^{\perp}, v = w + w'$ and $||v - w|| \leq ||v - u||$ for all u, and are equal only when $w = u$.

Proof. $||v - w||^2 = ||w'||^2$, OTAH, $||v - u||^2 = ||w - u + w'||^2 = ||w - u||^2 + ||w'||^2$.

6.3 Adjoint

> Lemma. For $(V, <, >)$, $\dim V < \infty$, let $\phi \in V^*$, then $\exists! w \in V, s.t. \phi =< \cdot, w >$.

Proof.

1. Existence: pick orthonormal basis, $\phi(v = \sum a_i e_i) = \sum a_i \phi(e_i)$. Now $< v = \sum a_i e_i, w = \sum b_i e_i >= \sum a_i \overline{b_i}$. Then $b_i = \overline{\phi(e_i)}$ gives a valid w.
2. Uniqueness: usual way.
3. Thus $L : \phi \mapsto w$ is a function. Furthermore, it's linear, onto and 1-1.

> **Thm. (Adjoint)**
>
> For $T : V \to W, \exists! T^* : W \to V, s.t. \forall v \in V, w \in W, < Tv, w >_W =< v, T^* w >_V$.

Proof.

1. For $w \in W$, define $\phi_w := v \mapsto < Tv, w >, \phi_w \in V^*$. IOW, $< Tv, w >= \phi_w(v)$.
2. For any ϕ_w, by lemma, $\exists! y \in V, s.t. \phi_w =< \cdot, y >$. IOW, $< v, y >= \phi_w(v)$.
3. Then $T^* := w \mapsto y = (w \mapsto \phi_w) \circ (\phi_w \mapsto y) = (w \mapsto \phi_w) \circ L$ satisfies the desired equality.
4. It can be shown that it is linear, and further, an isomorphism.
5. T^* is also unique since y is unique in step 2.

> Cor. $< w, Tv >=< T^* w, v >$, by taking conjugate.

> Cor. $[L_A^*]_\alpha^\beta = A^*$.

Proof. By cor6.2.1 (Linear algebra > ^bc4c8f), $a_{ij} =< T(s_j), t_i >= \overline{< T^*(t_i), s_j >} = \overline{b_{ji}}$.

> Prop.
>
> 1. $(T+U)^* = T^* + U^*$;
> 2. $(\lambda T)^* = \bar{\lambda} T^*$;
> 3. $I^* = I$;
> 4. If T, U composible, $(TU)^* = U^* T^*$.

Rmk. (Relationship between dual and adjoint)

$$\begin{aligned}
&T: V \to W, T^*: W^* \to V^*, T^*: (W, <,>_W) \to (V, <,>_V) \\
&\phi \in W^*, T^*(\phi) := \phi \circ T \\
&\phi_w :=< \cdot, w >_W \in W^* \\
&(\text{dual}) T^*(\phi_w) = \phi_w \circ T =< T\cdot, w >_W =< \cdot, T^* w >_V = \psi_{T^* w} \in V^* (\text{adjoint})
\end{aligned}$$

Deeper result

Ex. Orthonormal basis β, let Q be the matrix whose columns are β, then $Q^* = Q^{-1}$.

Ex. $\begin{aligned} T: V \to V, \dim V < \infty &\implies N(T^*T) = N(T) \implies rk(T^*T) = rk(T) \\ &\implies rk(T) = rk(T^*) \implies rk(TT^*) = rk(T) \end{aligned}$.

6.4 Normal and self-adjoint

Def. (Self-adjoint, normal)

- **Self-adjoint / Hermitian**: if $T^* = T$;
- **Normal**: if $T^*T = TT^*$.

Ex. $V = W \oplus W^{\perp}$, then $proj_W : V \to W$ is self-adjoint.

> Thm. (6.15) T is normal operator, then
>
> 1. $\forall v, ||Tv|| = ||T^*v||$;
> 2. $Tx = \lambda x \implies T^*x = \overline{\lambda} x$;

> Prop. Normal T is self-adjoint iff all eigenvalues are real.

Lemma. (Schus's lemma)

$T : V \to V, p_T(t)$ split, then $\exists$ orthonormal basis β s.t. $[T]_\beta$ is upper triangular, i.e. T-invariant flag.

Proof. Since it splits, there is a basis producing T-invariant flag, i.e. $T(s_i) = \sum_{j=1}^{i} a_i s_i$. Gram-Schmidt process is a precedure that automatically keeps this flag.

Thm. (6.4.1)

TFAE when $F = \mathbb{C}$:

1. T is normal;
2. $V = \bigoplus_{i=1}^{r} E_{\lambda_i}$ and $\forall v \in E_{\lambda_i}, w \in E_{\lambda_j}, i \neq j \to < v, w >= 0$, i.e. $E_{\lambda_i}^{\perp} = \bigoplus_{j\neq i} E_{\lambda_j}$;
3. V has an orthonormal basis β of eigenvectors of T, i.e. $[T]_\beta$ is diagonal.

Proof.

1. (1=>3) By Schus's lemma (Linear algebra > ^a80b9a), there is $\beta = (s_1 \dots s_n)$, s.t. $[T]_\beta = (a_{ij})$ being upper triangular. Then $[T^*]_\beta$ is lower triangular. According to thm6.15 (Linear algebra > ^34affa), $||Ts_1||^2 = ||a_{11}s_1||^2 = |a_{11}|$ should be equal to $||T^*s_1||^2 = ||\sum_{j=1}^{n} \overline{a_{1j}} s_1||^2 = \sum_{j=1}^{n} |\overline{a_{1j}}|$. Thus $\forall j > 1, a_{1j} = 0$. Apply induction.
2. (3=>1) multiplication of two diagonal matrices commutes
3. (3=>2) Since $[T]_\beta$ is diagonal, by thm5.2.2 (Linear algebra > ^2f8c8e), we have $V = \bigoplus_{i=1}^{r} E_{\lambda_i}$. For any $E_\lambda = span\{s_i\}, E_\sigma = span\{t_j\}$, orthogonal to each other.
4. (2=>3) pick orthonormal basis in each eigenspace, thus are eigenvectors, and then combine them.

Thm. (6.4.2)

TFAE when $F = \mathbb{R}$:

1. T is self-adjoint;
2. $V = \bigoplus_{i=1}^{r} E_{\lambda_i}$ and $\forall v \in E_{\lambda_i}, w \in E_{\lambda_j}, i \neq j \to < v, w >= 0$, i.e. $E_{\lambda_i}^{\perp} = \bigoplus_{j\neq i} E_{\lambda_j}$;
3. V has an orthonormal basis β of eigenvectors of T, i.e. $[T]_\beta$ is diagonal.

Proof. Only need to fix (1=>3). WTS p_T splits over field $\mathbb{R}$. Choose orthonormal basis β to have $A := [T]_\beta$ s.t. $A = A^t$. Consider the vecter space $(\mathbb{C}^n, <,>)$ over field $\mathbb{C}$, where $L_A : \mathbb{C}^n \to \mathbb{C}^n, L_A^* = L_{\overline{A^t}} = L_A$ is normal. By last thm, $[L_A^*]_\alpha = [L_A]_\alpha$ is diagonal, which means all entries are real. Thus p_{L_A} splits over field $\mathbb{R}$, which is the same charateristic poly as A and T. Now the original proof can be applied.

Rmk. (**Spectral thm**) An important implication is the spectral decomposition $T = \sum \lambda_i T_i, s.t. I = \sum T_i, T_i \circ T_j = \delta_{ij} T_i$.

> Cor. $F = \mathbb{C}$, then T is normal iff $T^* = p(T)$ for some poly p.

Proof. (=>) Consider the projection mappings $\pi_i : v \to v_i \in E_{\lambda_i}$. Due to the direct sum decomposition, $T = \sum \lambda_i \pi_i$. #TODO

> Ex. $T : V \to V, W \subset V. T(W) \subset W$. Then $T^*(W^\perp) \subset W^\perp$. If further that $T^*(W) \subset W$, then $(T|_W)^* = (T^*)|_W$. If once further that T is normal, then $T|_W$ is normal.

> Lemma. (6.3.1) $S : V \to V$ self-adjoint ($F = \mathbb{R}$) or normal ($F = \mathbb{C}$), then $\forall v, < Sv, v >= 0$ implies $S = 0$.

Proof. There is an orthonormal basis of eigenvectors. For any eigenvector $v, 0 =< Sv, v >= \lambda ||v||^2$, thus $\lambda = 0$. Since S is diagonalizable, it has to be 0.

> Ex. In fact, when $F = \mathbb{C}$, even if S is jsut a general operator, if $\forall v, < Sv, v >= 0$, then $S = 0$. This can be proved by expending the case of $v = x + y, v = x + iy$ respectively.

> Ex. If T is normal, $\dim V < \infty$, then $Ker(T) = Ker(T^*), R(T) = R(T^*)$.

> Cor. $F = \mathbb{C}. \forall v \in V, < Tv, v >\in \mathbb{R}$ iff T is self-adjoint.

Proof. #TODO also proved in note

> **Thm.**
>
> $M \in M_{n\times n}(F)$ self-adjoint ($F = \mathbb{R}$) or normal ($F = \mathbb{C}$), let $v_1 \dots v_n$ be the orthonormal eigenvector basis given thm6.4.2 (Linear algebra > ^fe98ba), then $M = \sum_i \lambda_i v_i v_i^*$.

Proof. It suffices to show that $\forall j, Mv_j := \lambda_j v_j = (\sum_i \lambda_i v_i v_i^*) v_j$, where $RHS = \sum_i \lambda_i v_i < v_j, v_i >= \lambda_j v_j$.

> Cor. M self-adjoint ($F = \mathbb{R}$) or normal ($F = \mathbb{C}$),
>
> 1. $M^k := \sum_i \lambda_i^k v_i v_i^*$;
> 2. $M^{-1} = \sum_i \frac{1}{\lambda_i} v_i v_i^*$;
> 3. $tr(M) = \sum_i \lambda_i$;
> 4. $tr(M^* M) = \sum_i |\lambda_i|^2$.

Proof.

1. λ^k, v are eigenvalue, eigenvector for M^k.
2. $$\left(\sum_i \frac{1}{\lambda_i} v_i v_i^*\right)\left(M = \sum_j \lambda_j v_j v_j^*\right) = \sum_{i,j} \frac{\lambda_j}{\lambda_i} v_i (v_i^* v_j) v_j^* = \sum_i v_i v_i^* = I$$
3. This can be shown by taking orthonormal standard basis $\epsilon = (e_1 \dots e_n)$ and orthonormal eigenvector basis $\beta = (s_1 \dots s_n) \subset F^n$, and then by cor6.2.1 (Linear algebra > ^bc4c8f),
$$\begin{aligned} tr(M) = tr([L_M]_\epsilon) &= \sum_i < Me_i, e_i >= \sum_i e_i^* M e_i \\ &= \sum_i e_i^* \left(\sum \lambda_j v_j v_j^*\right) e_i = \sum_{i,j} \lambda_j (e_i^* v_j)(v_j^* e_i) \\ &= \sum_{i,j} \lambda_j < v_j, e_i >< e_i, v_j > \\ &= \sum_j \lambda_j < (\sum_i < v_j, e_i > e_i), v_j > \\ &= \sum_j \lambda_j < v_j, v_j >= \sum_j \lambda_j \end{aligned}$$
4. Similarly, $e_i = \sum_j < e_i, v_j > v_j, Me_i = \sum_j \lambda_j < e_i, v_j > v_j$,
$$\begin{aligned} tr(M^* M) = tr([L_{M^* M}]_\epsilon) &= \sum_i < M^* M e_i, e_i >= \sum_i < Me_i, Me_i > \\ &= \sum_i \sum_j \lambda_j \overline{\lambda_j} < e_i, v_j >< v_j, e_i > \\ &= \sum_j |\lambda_j|^2 \end{aligned}$$

6.5 Isometry

Def. (Isometry)

$T : V \to V$ is an **isometry** if $\forall v, ||Tv|| = ||v||$. It's also called **orthogonal** when $F = \mathbb{R}$ and **unitary** when $F = \mathbb{C}$.

Thm. TFAE:

1. T is an isometry;
2. T is invertible and $T^* = T^{-1}$;
3. $\forall u, v, < Tu, Tv >=< u, v >$;
4. $\forall$ orthonormal $\beta, T(\beta)$ is still orthonormal;
5. $\exists$ orthonormal $\beta, T(\beta)$ is still orthonormal;
6. T^* is isometry.

Proof.

1. (1=>2) $< v, v >=< Tv, Tv >=< v, T^*Tv >$, then $< v, (I - T^*T)v >= 0$. Let $S := I - T^*T$, which is self-adjoint. By lemma6.3.1 (Linear algebra > ^98666c), $S = 0$. The other way is the same.
2. (2=>3) Trivial.
3. (3=>4) Trivial.
4. (4=>5) G-S process.
5. (5=>2) Say $\beta = (s_1 \dots s_n)$.
6. (3=>1) #TODO
7. (1=>6) #TODO
8. (6=>1) Apply (1=>6) one more time.

> Cor. $A \in M_{n\times n}(F)$, A is self-adjoint/normal iff $\exists$ isometry U s.t. UAU^{-1} is diagonal.

Proof. Self-adjoint/normal iff orthonormal diagonalizable iff isometry diagonalizable, since $UAU^{-1} = [I]_\epsilon^\beta [L_A]_\epsilon^\epsilon [I]_\beta^\epsilon = [L_A]_\beta$ and $[I]_\beta^\epsilon$ is isometry by definition.

> Prop.

> 1. If T is isometry, T is normal (given form2);
> 2. If $F = \mathbb{C}$, then T is isometry iff $\exists$ orthonormal basis β s.t. $[T]_\beta$ diagonal and $\forall i, |\lambda_i| = 1$.

Def. (Matrix Isometry)

$U \in M_{n\times n}(F)$ is an isometry if columns orthonormal.

> Lemma. $U \in M_{n\times n}(F)$ is isometry iff $L_U : F^n \to F^n$ in $(F^n, <, >_{F^n})$ is isometry.

Proof. #TODO

Ex. If M is isometry and upper triangular, then M is diagonal.

Proof. $s_i := Me_i = \sum_{j=1}^{i} m_{ji}e_i$, then $1 =< s_i, s_i >= \sum_{j=1}^{i} |m_{ji}|^2$; $\forall i < k, 0 =< s_i, s_k >= \sum_{j=1}^{i} m_{ji}\overline{a_{jk}}$. Induction.

Ex. (**QR-decomposition** for solving linear sys) Every invertiable A can be decomposed as the product of a isometry matrix Q and an upper triangular matrix R.

Def. (Orthogonal projection)

$T : V \to V, R(T)^\perp = N(T), N(T)^\perp = R(T)$.

6.6 Positivity

Def. (Positive semidefinite operator)

- For $F = \mathbb{C}, \forall v, < Tv, v >\geq 0$;
- For $F = \mathbb{R}, T$ is self-adjoint.

Thm.

TFAE:

1. T is positive semidefinite;
2. $T = T^*$ and $\forall$ eigenvalue $\lambda \in \mathbb{R}^{\geq 0}$;

3. $\exists$ positive semidefinite R, $R^2 = T$;
4. $\exists R = R^*, R^2 = T$;
5. $\exists R, RR^* = T$.

Proof. #TODO

Rmk. Positive semidefinite $\implies$ self-adjoint. In other word, when $F = \mathbb{C}$, it requires more than when $F = \mathbb{R}$.

Ex. T^*T, TT^* are positive semidefinite and $rk(TT^*) = rk(T^*T)$.

6.7 SVD and Pseudo-inverse

Thm. (Singular Value thm)

$T : V \to W$, then $\exists$ orthonormal basis $(v_1 \dots v_n)$ of V and $(u_1 \dots u_m)$ of W, s.t. $T(v_i) = \sigma_i u_i \mathbb{1}(1 \le i \le r)$ and $\sigma_i \searrow, \sigma_i \in \mathbb{R}^{>0}$ for $1 \le i \le r$. Conversely, when such condition satisfied, v_i is an eigenvector of T^*T with eigenvalue $\lambda_i \mathbb{1}(1 \le i \le r)$. In other word, the singular value set $\{\sigma_i\}_{i=1}^r$ uniquely determined by T.

Proof.

1. Existence: T^*T are positive semidefinite, then $\forall$ eigenvalues $\lambda \ge 0$. Put them into order to get $T^*Tv_i = \lambda_i v_i \mathbb{1}(1 \le i \le r)$. Now define $\sigma_i := \sqrt{\lambda_i}$ and $u_i := \frac{1}{\sigma_i} T(v_i)$, so that $< u_i, u_j >_W = \frac{\sigma_i^2}{\sigma_i \sigma_j} < v_i, v_j >= \delta_{ij}$. Extend it by gram-schmitd and normalization to get orthonormal basis of W. For $i > r, T^*Tv_i = 0$, then $< Tv_i, Tv_i >=< v_i, T^*Tv_i >=< v_i, 0 >= 0$ confirming the equation $\sigma_i \searrow, \sigma_i \in \mathbb{R}^{>0}$.
2. Uniqueness: just verify the statement about eigenvector via inner product.

> Cor. (Singular Value Decomposition, **SVD**) $A \in M_{m\times n}(F), \exists$ isometry $U \in M_{m\times m}, V \in M_{n\times n}$ s.t. $A = U\Sigma V^*$ where $diag(\Sigma) = (\sigma_1, \sigma_2, \dots, \sigma_r, 0, \dots, 0)$ with $\sigma_i \searrow, \sigma_i \in \mathbb{R}^{>0}$.

Proof. Get singular values and orthonormal bases from theorem applied on L_A. U and V are isometry by def. Now $AVe_i = Av_i = \sigma_i u_i \mathbb{1}(i \le r)$ and $U\Sigma e_i = U\mathbb{1}(i \le r)\sigma_i e_i = \mathbb{1}(i \le r)\sigma_i u_i$, therefore $AV = U\Sigma$, i.e. $A = U\Sigma V^{-1} = U\Sigma V^*$.

Rmk. This decomposition can be regarded as eigenvalue decomposition with extra demand: two change of basis operators should preserve volume. The benefit is that Σ will give a characterization of the size of A, while the downside is that now we have weaker but still nice form of Σ (compared to Linear algebra > ^4f3137).

Def. (Psuedo-inverse)

$T^\dagger : W \to V, T^\dagger|_{R(T)} := (T|_{(KerT)^\perp})^{-1}, T^\dagger|_{R(T)^\perp} := 0.$

Motiv. Consider $V = KerT \oplus KerT^\perp$ and $W = R(T)^\perp \oplus R(T)$. T maps one subspace in LHS to RHS correspondingly, but only $R(T)$ can be inverted back. Since actually $T(KerT) = 0$, we can define its psuedo-inverse on $R(T)^\perp$ be simply a zero map, thus our definition of psuedo-inverse.

Rmk. $T^\dagger T : V \to V$ is $proj^V_{KerT^\perp}$, $TT^\dagger : W \to W$ is $proj^W_{R(T)^\perp}$. When T is 1-1, $T^\dagger T = I_V$. When T is onto, $TT^\dagger = I_W$. Therefore, psuedo-inverse is close to inverse as much as possible.

Ch7 Jordan Normal/Canonical Form

Rmk. This charpter assumes all charateristic polys split.

Def. (Jordan form)

$$J_{r,\lambda} := \begin{bmatrix} \lambda & 1 & \\ & \ddots & \ddots \\ & & \lambda \end{bmatrix}$$

Thm. (Jordan normal form)

$A \in M_{n\times n}(F), p_A(t)$ split, then

1. $A \sim \begin{bmatrix} J_{r_1,\lambda_1} & & \\ & \ddots & \\ & & J_{r_t,\lambda_t} \end{bmatrix};$

2. If $A \sim \begin{bmatrix} J_{s_1,\sigma_1} & & \\ & \ddots & \\ & & J_{s_k,\sigma_t} \end{bmatrix}$, then $k = t$, and up to renumbering, $\forall i, (s_i, \sigma_i) = (r_i, \lambda_i)$.

Proof. #TODO

> Cor. $A \sim B$ iff $\exists$ Jordan form J s.t. $A \sim J \sim B$.

> Cor. $A \sim A^t$ when $p_A(t)$ split.

Rmk. This is in fact always true, but that's a much deeper result.

Proof. Find $J = PAP^{-1}$. Then $J^T = (P^{-1})^T A^T P^T = (P^T)^{-1} A^T P^T$. It suffices to show $J \sim J^T$. Only need to look at each nilpotent $S = J_{\lambda_i, r_i} - \lambda_i I_{r_i} = \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & \ddots & \ddots \end{bmatrix}$.

Note that $Se_i = e_{i-1}$ and $S^T e_i = e_{i+1}$. Therefore $S \sim S^T$.

Ex. $T : V \to V, \dim V < \infty, T^2 = Id$, then T is diagonalizable.

Rmk. An example is $V = \{M_{n \times n}\}$ and T is taking transpose/adjoint. Then E_1 are the symmatric matrices, and E_{-1} are the skew-symmetric matrices.

Proof. (Trick of invariant cyclic subspace)

1. First we show $p_T(t)$ split by looking at the invariant cyclic subspace $U = span\{v, Tv, T^2 v \dots\}$ to decompose the space/polynomial. Note that $\dim U \le 2$. In the case of $\dim U = 2$, $p_{T|_U} = \det \begin{bmatrix} -t & 1 \\ 1 & -t \end{bmatrix} = t^2 - 1$ split. Let $\bar{T} : V/W \to V/W, p_T(t) = p_{T|_U}(t) p_{\bar{T}}(t)$, then since $\dim(V/W) < \dim V$ and $(\bar{T})^2[x] = [T^2 x] = [x] \implies (\bar{T})^2 = Id$, apply induction to show that $p_T(t)$ split.
2. The roots of $p_T(t)$ are contained in $\{-1, 1\}$. Simply since $Tv = \lambda v \implies v = T^2 v = \lambda^2 v$.
3. Since split, take the JNF of T, $J = (J_{r_1, \pm 1} \dots)$. Being diagonalizable means $\forall i, r_i = 1$. Since $T^2 = Id$, $J^2_{r_i, \pm 1} = I$. OTAH, write $J_{r_i, \pm 1} = S + N$, where $S = \pm I$ and N is nilpotent, then contradiction $I = J^2_{r_i, \pm 1} = S^2 + N^2 + 2SN \ne I$ holds whenever $r_i > 1$.

4. For the last step, alternatively, note that $V = V_{-1} \oplus V_1$. Show every generalized eigenvector is also an eigenvector.

Summary up to now

1. Equivalence relations
    1. $PAQ \sim A$: only $n+1$ equivalence classes by rank (Linear algebra > ^4f3137).
    2. $PAP^{-1} \sim A$: conjugate (which is the default one)
    3. $PA \sim A$, where P is invertible: RREF
    4. $UAV^* \sim A$: SVD

Later

https://courses.cs.washington.edu/courses/cse521/16sp/521-lecture-8.pdf

Rayleigh Quotient: Variational Characterization of Eigenvalues

$$\text{symmetric real } M_{n\times n},\ \text{eigenvalue } \lambda_1 \le \lambda_2 \ldots \le \lambda_n$$
$$\text{Rayleigh quotient } R_M(\boldsymbol{x}) = \frac{\boldsymbol{x}^T M \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}$$
$$\lambda_k = \min_{\forall V, \dim V = k} \max_{\boldsymbol{x} \in V - \{\mathbf{0}\}} R_M(\boldsymbol{x})$$

proof: https://blog.csdn.net/a358463121/article/details/100166818

证明V里面一定存在向量使得Rayleigh quotient时，只需要取$\lambda_1, \lambda_2, \ldots, \lambda k$对应的$v1, v_2, \ldots, v_k$组成的空间$V$即可。 https://zhuanlan.zhihu.com/p/80817719