Linear algebra

Course MATH 416 Honor@UIUC

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

1. ϵ_n is the standard basis of F^n.

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 $\lambda_1 \dots \lambda_n$ 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)

$T : V \to W.$

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 isomorphic 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 T is isomorphism iff $\dim W = n$.

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

Thm. $T : V \to W$ induces linear $\bar{T} : V_{/KerT} \to R(T)$ 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 $\beta = (s_1, \ldots, s_n)$, for invertible $A = [A_1, \ldots, A_n]$, find unique $\alpha = (t_1, \ldots, t_n)$. 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.
>
> $$PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \in M_{m\times n}$$
>
> That means there's only $\min(n, m)$ 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 $R(L_A) = span(t_1, \dots, t_r)$. 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 $R(L_A) = span(t_1, \dots, t_r)$, we have $\{[s_i]\}$ forming a basis in $F^n/ker(T)$.
3. Now let $W = span(s_1, \dots, s_r)$. 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. $V(= span(\beta)) \xrightarrow{T} W(= span(\alpha)) \xrightarrow{\phi \in W^*} F$.

Thm. (Transpose)

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

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 $rkT = rkT^*$.
2. $kerT^* = \{\varphi \in W^* : \varphi \circ T = 0\}$. Write $\varphi = \sum_i a_i t_i^*$.
3. $\varphi \circ T = 0$ iff $\varphi(R(T)) = 0$, pick basis so that $R(T) = span(t_1 \ldots t_r)$, then iff $a_1 \ldots a_r = 0$ iff $kerT^* = span(t_{r+1} \ldots t_m)$. Then $rkT^* = rkT$.

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

Def. (System of linear equations)

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

Prop. Given an invertible matrix P, then $Ax = b$ iff $PAx = Pb$.

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).

Prop. Inverse and transpose of elementary matrix are still elementary.

Lemma. 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.

There're invertible matrix P, Q that are product of elementary matrices, s.t. $PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$.

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 them.

Cor. $rkA = rkA^T$.

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. $rk(AB) \le rk(A), rk(B)$.

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

Eigenvalue

Eigenvalue, eigenvector and movement

For a matrix $A_{n\times n}$, consider all $(\vec{u}, \lambda)$ pair such that: $A\vec{u} = \lambda\vec{u}$ We call them **eigenvalues** and **eigenvectors** of matrix A. There're totally n pairs of $(\vec{u}_i, \lambda_i)$ for diagonalizable linear transformation, and the eigenvectors form a basis(some λ_i might be the same).

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. 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

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, $(L\vec{v})_i = \frac{1}{deg_i} \sum_{j \in N(i)} (v_i - v_j)$

How to find them?

When the transformation A is normal operator, which means orthogonal diagonalizable, then:

$$A = P \Lambda P^{-1}$$

where Λ stretchs (eigenvalues), P rotates (orthonormal eigenvectors). Further more, when A is symmetric real matrix(e.g. adjacency and Laplacian matrix), then it is hermitian/self-adjoint, which means all eigenvalues are real.

THM

Symmetric real matrix M $M := \sum_i \lambda_i v_i v_i^T$, #TODO (upd) dim V=K We may use the same eigenvectors in M^k, such that $M^k := \sum_i \lambda_i^k v_i v_i^T$ claim: $M^{-1} := \sum_i \frac{1}{\lambda_i} v_i v_i^T$, $M^{-1}M = I$ proof: substitute

thm2: $tr(M) = \sum_i \lambda_i$ https://courses.cs.washington.edu/courses/cse521/16sp/521-lecture-8.pdf

Variational Characterization of Eigenvalues

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

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

Hilbert space

conjucture symmetric

Reference

Basic knowledge in Spectral Theory.