Lecture 1: Linear Equations

Shengkui Ye

September 13, 2022

1 System of linear equations

Example 1.1 *Solve*

$$\begin{aligned} x_1 - 2x_2 &= -1, \\ -x_1 + 3x_2 &= 3. \end{aligned}$$

Add the second equation to the first one, to get $x_2 = 2$. Get back to the first to get $x_1 = 3$. In this course, we will study the general case.

Definition 1.2 *Let $x_1, x_2, \cdots, x_n$ be variables. A linear equation is*

$$a_1x_1 + a_2x_2 + \cdots a_nx_n = b.$$

A system of linear equations is of the form

$$\left\{ \begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots a_{1m}x_m = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2m}x_m = b_2, \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots a_{nm}x_m = b_n. \end{array} \right. \qquad (*)$$

Fundamental questions of linear algebra are

Problem 1.3 *For the system of equations (*), how to solve it? Is there a solution? If yes, how many solutions?*

Obviously the system $(*)$ is determined by the coefficients a_{ij} and b_i $(1 \leq i \leq n, 1 \leq j \leq m)$. The answer to the above question is determined completely by these a_{ij} and b_i. For convenience, we introduce the concepts of vectors and matrices.

Definition 1.4 *A vector is an ordered tuple of real numbers:*

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix},$$

also denoted as $(u_1, u_2, \cdots, u_n)^T$*. The* n *is called the dimension of the vector* u*. We denote the set of all* n*-dimensional vectors by* $\mathbb{R}^n$*. For two* n*-dimensional vectors* u, v*, we define the sum and dot product as*

$$u + v = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix},$$

$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Using the dot product, we write the linear equation $a_1x_1 + a_2x_2 + \cdots a_nx_n = b$ as $(a_1, a_2, \cdots, a_n)^T \cdot (x_1, x_2, \cdots, x_n)^T = b$.

The system $(*)$ can be denoted as $AX = b$, where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

is called the coefficient matrix and $X = (x_1, x_2, \cdots, x_n)^T$, $b = (b_1, b_2, \cdots, b_n)^T$. The matrix $[A, b]$ is called the augmented matrix.

Example 1.5 *(Elementary row operations) In the process of solving* $AX = b$ *(or the system* $(*)$*), we can operate the following three elementary operations (to* $[A, b]$*):*

1. (Replacement) Replace one row by the sum of itself and a multiple of another row.

2 (Interchange) Interchange two rows.

3. (Scaling) Multiply all entries in a row by a nonzero constant.

Eventually, the matrix $[A, b]$ *is reduced to the form (called reduced Echelon form) satisfying:*

1. The first nonzero entry (called the leading entry) in each nonzero row is 1 *(after scalings).*

2. Each leading 1 *is the only nonzero entry in its column (after repalcements).*

3. All nonzero rows are above any rows of all zeros (after interchanges)

5. Each leading entry of a row is in a column to the right of the leading entry of the row above it (starting from the first column to the last column).

Example 1.6 *There is a standard way to reduce the matrix* $[A, b]$ *into the reduced Echelon form. The process is called Gaussian elimination (see the Textbook for an explicit explaination). The existence and uniqueness of solutions to* $AX = b$ *depend entirely on the reduced echelon form. For example,*

$$\begin{bmatrix} 1 & & \\ & 1 & \\ & & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

has solutions only when $b_3 = 0$. When $b_3 = 0$, the system has infinitely many solutions, with x_3 can be any real number. In such a case, we call x_3 the free variable.

Definition 1.7 *(linear combinations) Given vectors $v_1, v_2, \cdots, v_m \in \mathbb{R}^n$, a linear combination of these vectors is a sum*

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

for some $a_1, a_2, \cdots, a_m \in \mathbb{R}$. The set of all linear combinations is denoted by $\mathrm{Span}\{v_1, v_2, \cdots, v_m\}$.

Example 1.8 $\mathrm{Span}\{(1,1),(1,-1)\} = \mathbb{R}^2$.

Example 1.9 *The system $AX = b$ has a solution if and only if b is a linear combination of the columns of A.*

Lemma 1.10 *For any two vectors $u, v \in R^n$, $a \in \mathbb{R}$ and an $m \times n$ matrix A, we have*

$$\begin{aligned} A(u+v) &= Au + Av, \\ A(au) &= aAu. \end{aligned}$$

2 Solution sets of linear systems

The system $AX = 0$ is called homogeneous. Since $A0 = 0$, there is always a solution to the homogeneous system. A non-zero solution of $AX = 0$ is called a non-trivial solution.

Lemma 2.1 *For any two solutions X_1, X_2 to $AX = b$, the difference $X_1 - X_2$ is a solution of $AX = 0$. Fix a solution X_0 to $AX = b$. The set of all solutions to $AX = b$ is $\{X \in \mathbb{R}^n \mid AX = 0\} + X_0$.*

The following is the process of solving $AX = b$:

1. Row reduce the augmented matrix $[A, b]$ to the reduced echelon form.
2. Express each basic variable in terms of any free variables appearing in the system given by the reduced echelon form.
3. Write a typical solution X as a vector whose entries depending on the free variables, if any.
4. Decompose X into a linear combination of vectors (with numeric entries) using the free variables as parameters.

Example 2.2 *Describe all solutions of $AX = b$, where*

$$A = \begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{bmatrix}, b = \begin{bmatrix} 7 \\ -1 \\ -4 \end{bmatrix}.$$

3 Linear dependence and linear transformations

Two vectors u, v are co-line if $u = rv$ for some real number $r \in \mathbb{R}$. Three vectors u, v, w are co-plane if they lie in the same plane. The following is a general concept.

Definition 3.1 *Vectors $v_1, v_2, \cdots, v_m \in \mathbb{R}^n$ are linearly dependent if*

$$a_1 v_1 + a_2 v_2 + \cdots + a_m v_m = 0$$

for some non-zero vector $(a_1, a_2, \cdots, a_m)$. Similarly, $v_1, v_2, \cdots, v_m \in \mathbb{R}^n$ are linearly independent if $a_1 v_1 + a_2 v_2 + \cdots + a_m v_m = 0$ can only hold for $a_1 = a_2 = \cdots = 0$.

Example 3.2 *The set $\{(1,0),(0,1)\}$ is linearly independent in $\mathbb{R}^2$.*

Lemma 3.3 *Two vectors $\{u, v\}$ are linearly dependent if and only they are co-line. Three vectors $\{u, v, w\}$ are linearly dependent if and only if they are co-plane.*

Lemma 3.4 *$v_1, v_2, \cdots, v_m \in \mathbb{R}^n$ are linearly dependent if and only if one vector is a linear combination of the other vectors.*

Proof. If $a_1 v_1 + a_2 v_2 + \cdots + a_m v_m = 0$ holds for some nonzero a_i, then $v_i = -\frac{1}{a_i}(\sum_{j \neq i} a_j v_j)$, a linear combination. Conversely, if $v_i = \sum_{k \neq i} a_k v_k$, then $\sum_{k \neq i} a_k v_k - v_i = 0$. Thus $\{v_1, v_2, \cdots, v_m\}$ are linearly independent. ■

Lemma 3.5 *Let $A = [v_1, v_2, \cdots, v_m]$ be a matrix with v_i as its i-th column. $v_1, v_2, \cdots, v_m \in \mathbb{R}^n$ are linearly dependent if and only if $AX = 0$ has a non-trivial solution.*

Proof. It is obvious. ■

Corollary 3.6 *Any set $\{v_1, v_2, \cdots, v_p\} \in \mathbb{R}^n$ is linearly dependent if $p > n$.*

Recall that $\mathbb{R}^n$ is the set of all n-dimensional vectors. For two vectors $x, y \in \mathbb{R}^n$, and a real number $a \in \mathbb{R}$, we can define $x + y$ and ax.

Definition 3.7 *A linear transformation $f : \mathbb{R}^m \to \mathbb{R}^n$ is a function assigning each element $x \in \mathbb{R}^m$ an element $f(x) \in R^n$ such that*

$$f(ax + by) = af(x) + bf(y),$$

for any $a, b \in \mathbb{R}$. In other words, f assign linear combinations to linear combinations.

Example 3.8 *For an $n \times m$ matrix $A_{n \times m}$, the function $f(x) = Ax : \mathbb{R}^m \to \mathbb{R}^n$ is linear.*

Theorem 3.9 *For any linear transformation* $f : \mathbb{R}^m \to \mathbb{R}^n$*, there is a unique matrix* A *(called the standard matrix of* f*) such that* $f(x) = Ax$*. Actually,* $A = [f(e_1), f(e_2), \cdots, f(e_n)]$ *where* e_i *is the* j*-th column of the identity matrix in* $\mathbb{R}^m$*.*

Proof. Any vector $x = (x_1, x_2, \cdots, x_m)^T \in \mathbb{R}^m$ is a linear combination $x = x_1e_1 + x_2e_2 + \cdots + x_me_m$. Therefore, $f(x) = x_1f(e_1) + x_2f(e_2) + \cdots + x_mf(e_m) = [f(e_1), f(e_2), \cdots, f(e_n)]x$. ■

Example 3.10 *Let* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *be the transformation that rotates each point in* $\mathbb{R}^2$ *about the origin anti-clockwise through an angle* $\varphi \in [0, 2\pi)$*. Show that the standard matrix of* f *is*

$$\begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}.$$

Lecture 2: Matrix

Shengkui Ye

September 23, 2022

1 Matrices: sum, product and transpose

Recall that an $n \times m$ matrix $M = (a_{ij})_{1\leq i\leq n, 1\leq j\leq m}$ has a real number a_{ij} in the (i, j)-th position. For two $n \times m$ matrices $A = (a_{ij})_{n\times m}, B = (b_{ij})_{n\times m}$, we can add them together $A + B = (a_{ij} + b_{ij})_{n\times m}$. For any real number $a \in \mathbb{R}$, the scalar multiplication $aA = (aa_{ij})_{n\times m}$. We use the notation: $x_1 + x_2 + \cdots + x_m = \sum_{i=1}^{m} x_i$.

Definition 1 *For matrices $A_{n\times m}, B_{m\times k}$, the product $AB = (c_{ij})$ is an $n \times k$ matrix with (i, j)-th entry*

$$c_{ij} = \sum_{s=1}^{m} a_{is}b_{sj}.$$

Example 2 *For a matrix $A = (a_{ij})_{n\times m}$ and a vector $X = (x_1, x_2, \cdots, x_m)^T$, the product $AX = (\sum_{j=1}^{m} a_{ij}x_j)_{1\leq i\leq n}$ is an n-dimensional vector.*

Lemma 3 *For matrices $A_{n\times m}, B_{m\times k}, C_{k\times l}$, we have*
1) $(AB)C = A(BC)$;
2) $A(B_1 + B_2) = AB_1 + AB_2$, if B_1, B_2 have m rows;
3) $(A_1 + A_2)B = A_1B + A_2B$ if A_1, A_2 have m columns;
4) $a(AB) = A(aB) = (aA)B$ for any real number $a \in R$;
5) $I_nA = AI_m = A$ for identity matrice I_n, I_m (of size $n \times n, m \times m$ respectively).

Example 4 *Let $A = \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Note that $AB \neq BA$.*

Definition 5 *The transpose of an $n \times m$ matrix $A = (a_{ij})$ is the $m \times n$ matrix (b_{ij}) with $b_{ij} = a_{ji}$. Denote the transpose by A^T. In other words,*

$$A^T = \begin{bmatrix} a_{11} & a_{21} & & a_{n1} \\ a_{12} & a_{22} & & a_{n2} \\ \vdots & & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix}.$$

Example 6 *The transpose of a row vector* $(a_1, a_2, \cdots, a_m)$ *is the column vector*

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}.$$

Lemma 7 *We have* $(A^T)^T = A$, $(AB)^T = B^T A^T$, *and* $(A+B)^T = A^T + B^T$.

2 Invertible matrices

Definition 8 *A* square *matrix* $A_{n\times n}$ *is invertible if there exists a matrix* B *such that* $AB = BA = I_n$. *When* A *is invertible, we denote the inverse by* A^{-1}.

Remark 9 *The inverse is unique if it exists. Suppose* B_1, B_2 *are both inverses of* A. *Then* $B_1 = B_1 I_n = B_1(AB_2) = (B_1 A)B_2 = I_n B_2 = B_2$.

Example 10 *A matrix* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *is invertible if and only if* $ad - bc \neq 0$. *The inverse is* $\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Lemma 11 *When* A *is invertible, the system* $AX = b$ *has a unique solution* $X = A^{-1}b$.

Proof. Left mutiply both sides of $AX = b$ by A^{-1} to get that $X = A^{-1}b$. Since the inverse A^{-1} is unique, the solution $A^{-1}b$ is unique. ■

Lemma 12 *Let* A, B *be two invertible matrices of the same sizes. Then*
1) $(A^{-1})^{-1} = A$;
2) $(AB)^{-1} = B^{-1}A^{-1}$;
3) $(A^T)^{-1} = (A^{-1})^T$.

An elementary matrix is one that is obtained by performing a single elementary row operation on an identity matrix. The next example illustrates the three kinds of elementary matrices.

Example 13 *Find the inverses of the following matrices:* $\begin{bmatrix} 1 & & \\ & 1 & \\ 10 & & 1 \end{bmatrix}, \begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix}, \begin{bmatrix} 1 & & \\ & 1 & \\ & & 10 \end{bmatrix}$.
Can you replace 10 *by any nonzero real number* a?

Theorem 14 *An* $n \times n$ *matrix* A *is invertible if and only if* A *is reduced by elementary row operations to the identity* I_n. *Moreover, if* $A = E_k E_{k-1} \cdots E_1$ *then* $A^{-1} = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$.

Proof. Recall that a matrix A is reduced by elementary row operations to the reduced echelon matrix. If A is invertible, the system $AX = b$ has a unique solution. Therefore, the echelon matrix is the identity I_n. Conversely, when A is reduced to the identity matrix, the A is invertible sicne each elementary matrix is invertible. The last claim is a simply application of 2) in the previous lemma. ■

The previous theorem provides an algorithm for finding A^{-1}: reduce the matrix A in the augmented matrix $[A, I_n]$ into the identity by elementary row operations, to get $[I_n, A^{-1}]$.

Theorem 15 *For a square matrix $A_{n\times n}$, the following are equivalent:*
1) A is invertible.
2) A is reduced by elementary row operations to the identity matrix.
3) The reduced echelon form of A is the identity I_n.
4) The equation $Ax = 0$ has only the trivial solution.
5) The columns of A form a linearly independent set.
6) The linear transformation $x \mapsto Ax$ is one-to-one (injective).
7) The equation $Ax = b$ has at least one solution for each $b \in \mathbb{R}^n$.
8) The columns of A span $\mathbb{R}^n$.
9) The linear transformation $x \mapsto Ax$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$ (surjective).
10) There is an $n \times n$ matrix D such that $AD = I_n$.
11) There is an $n \times n$ matrix C such that $CA = I_n$.
12) A^T is an invertible matrix.

Proof. The previous theorem implies the equivalences 1) $\iff$ 2) $\iff$ 3). It is obvious that 3) $\iff$ 4). The equivalence 4) $\iff$ 5) is from the definition of linear independence. It is obvious that 4) $\iff$ 6) by the definition of "one-to-one". 1) $\Longrightarrow$ 7) since $x = A^{-1}b$ is a solution. It is obvious that 7) $\iff$ 8) $\iff$ 9) by the definitions of "span" and "onto". When 9) holds, the standard basis $\{e_1, e_2, ..., e_n\}$ of $\mathbb{R}^n$ has preimages. This means that there exists $x_i \in R^n$ such that $Ax_i = e_i$. Therefore, $A[x_1, ..., x_n] = I_n$. This proves 9) $\Longrightarrow$ 10). From the definition of inverse, we have 1) $\Longrightarrow$ 10), 1) $\Longrightarrow$ 11). If 11) holds, then $Ax = 0$ has $0 = CAx = x$ and thus 4) holds. Similarly, 10) implies A^T is invertible. Since $(A^T)^{-1} = (A^{-1})^T$, we have 1) $\iff$ 12). ■

Corollary 16 *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. There exists a linear transformation $g : \mathbb{R}^n \to \mathbb{R}^n$ such that such that $f \circ g = g \circ f = id_{\mathbb{R}^n}$ (the identica map) if and only if the standard matrix of f is invertible.*

Proof. It follows the uniqueness of standard matrices for $g, f \circ g$ and $g \circ f$. ■

3 Subspaces, dimensions and ranks

Definition 17 *A subspace of $\mathbb{R}^n$ is a subset H of $\mathbb{R}^n$ such that $ax + by \in H$ for any $x, y \in H$ and $a, b \in \mathbb{R}$.*

Example 18 *A line, or a plane passing* 0 *is a subspace of* $\mathbb{R}^n$*. The span of any subset* $\{v_1, v_2, \cdots, v_k\} \subset \mathbb{R}^n$ *is defined as the set of all vectors*

$$a_1v_1 + a_2v_2 + \cdots + a_kv_k,$$

for each $a_i \in \mathbb{R}$*. The span is a subspace.*

Example 19 *For a matrix* $A_{n\times m}$*, the span of columns of* A *is a vector space. The set* $\{x \in \mathbb{R}^m \mid Ax = 0\}$ *is a subspace.*

Definition 20 *A basis for a subspace* H *of* $\mathbb{R}^n$ *is a set* S *such 1)* S *is linearly independent; and 2)* S *spans* $\mathbb{R}^n$*.*

Example 21 *Let* $A_{n\times n}$ *be a square invertibe matrix. The column vectors of* A *form a basis for* $\mathbb{R}^n$*.*

Lemma 22 *Let* $S = \{s_1, s_2, \cdots, s_n\}$ *be a basis of* $\mathbb{R}^n$*. Every vector* $x \in \mathbb{R}^n$ *is a unique linear combination*

$$x_1s_1 + x_2s_2 + \cdots + x_ns_n$$

of S*. The vector* $(x_1, x_2, \cdots, x_n)$ *is called the coordinate of* x *relative to* S*.*

Proof. Since S spans $\mathbb{R}^n$, any vector x is a linear combination of S. Since S is linear independent, the linear combination is unique (i.e. suppose there are two different linear combination. take the difference to get a contradiction). ■

Definition 23 *Let* $H < \mathbb{R}^n$ *be subspace and* S *a basis of* H*. The number of elements in* S *is called the dimension* $\dim(H)$ *of* H*.*

Definition 24 *For a matrix* A*, the rank* $\text{rank}(A)$ *is the dimension of the subspace spanned by column vectors of* A*.*

Example 25 *Let*

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}.$$

Find a basis for the null space $\{x : Ax = 0\}$ *and the* $\text{rank}(A)$*.*

Lemma 26 *The rank of a matrix* A *equals to the number of leading 1s in its reduced echelon form.*

3.1 Rank theorem

Example 27 *Let*

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}.$$

Compute its rank $\text{rank}(A)$*.*

Lemma 28 *For invertible matrice $B_{m\times m}, C_{n\times n}$ and matrix $A_{m\times n}$ we have* $\operatorname{rank}(BAC) = \operatorname{rank}(A)$.

Proof. Let $A = [A_1, A_2, ..., A_n]$, where A_i is the i-th column. Then $BA = [BA_1, BA_2, ..., BA_m]$. View B as a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^n$ given by $x \to Bx$. Since B is invertible, f is bijective. This implies that

$$B\text{Span}\{A_1, A_2, ..., A_n\} = \text{Span}\{BA_1, BA_2, ..., BA_n\}.$$

If $\{x_1, x_2, \cdots, x_k\}$ is a basis of $\text{Span}\{A_1, A_2, ..., A_n\}$, then $\{Bx_1, Bx_2, \cdots, Bx_k\}$ is a basis of $\text{Span}\{BA_1, BA_2, ..., BA_n\}$. This proves that $\operatorname{rank}(BA) = \operatorname{rank}(A)$ for any $A_{m\times n}$.

For the other part, we prove that $\text{Col}(A) = \text{Col}(BC)$. For any $x \in \text{Col}(A)$, we have $x = \sum_{i=1}^{n} a_i A_i$ for numbers $a_1, a_2, ..., a_n$. Actually, $x = A \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$. But $x = AC(C^{-1} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix})$, which implies that $x \in \text{Col}(BC)$. Similarly, any $y \in \text{Col}(BC)$ has $y = ACz$ for a vector $z \in \mathbb{R}^n$. Then $y = A(Cz) \in \text{Col}(A)$. This proves that $\operatorname{rank}(AC) = \operatorname{rank}(A)$, as the dimensions are the same. ■

Corollary 29 *For matrix $A_{n\times m}$, we have* $\operatorname{rank}(A) = \operatorname{rank}(A^T)$.

Proof. Apply elementary row operations to reduce A into the reduced echelon form $C = \begin{bmatrix} I_k & * \\ 0 & 0 \end{bmatrix}$. There is an invertible matrix B (product of elementary matrices) such that $BA = C$. Lemma 28 implies that $\operatorname{rank}(A) = \operatorname{rank}(C) = k$. Since $A^T B^T = C^T$, the same lemma implies $\operatorname{rank}(A^T) = \operatorname{rank}(C^T) = \operatorname{rank}(C) = \operatorname{rank}(A)$. ■

Theorem 30 *Let $A_{n\times m}$ be a matrix. Denote by* $\text{Col}(A)$ *the vector space spanned by columns of A,* $\text{Nul}(A)$ *the subspace* $\{x \in \mathbb{R}^m \mid Ax = 0\}$. *Then*

$$\dim \text{Col}(A) + \dim \text{Nul}(A) = m.$$

Proof. Reduce A by elementary row and column operations to the reduced echelon matrix $\begin{bmatrix} I_k & * \\ 0 & 0 \end{bmatrix}$. Lemma 28 implies that $\dim \text{Col}(A)$ is k, and the dimension of $\text{Nul}(A)$ is the number of free variables for the solutions of $Ax = 0$. Therefore, we have $\dim \text{Nul}(A) = m - k$. ■

Theorem 31 *Let $A_{n\times n}$ be a square matrix. The following statements are equivalent:*

0) A is invertible.

1) The column vectors of A form a basis for $\mathbb{R}^n$.

2) $\text{Col}(A) = \mathbb{R}^n$.

3) $\dim \mathrm{Col}(A) = n$.
4) $\mathrm{rank}(A) = n$.
5) $\mathrm{Nul}(A) = 0$.

Proof. By the definitions and the previous theorem, we have 1) $\implies$ 2) $\implies$ 3) $\implies$ 4) $\implies$ 5) $\implies$ 1). The equivalence 0) $\iff$ 5) is already proved. ■

Lecture 3 : Vector spaces

Shengkui Ye

October 16, 2022

1 Vector spaces and subspaces

Let $F = \mathbb{R}$ (the set of all real numbers), $\mathbb{C}$ (the set of all complex numbers), or $\mathbb{Q}$ (the set of all rational numbers).

Definition 1 *A vector space over F is a set V, together with two operations $+$ and multiplication by F, satisfying the obvious commutativity, associativity and distribution law. Explicitly, it satisfies the 8 conditions in the textbook.*

Example 2 *The set $\mathbb{R}^n$ is a vector space; For fixed positive integers m, n, the set $M_{m\times n}(\mathbb{R})$ of all $m \times n$ matrices is a vector space. The set $V = \{f \mid f$ is a continuous function on the interval $[0, 1]\}$ is a vector space.*

Definition 3 *A subspace of a vector space V is a subset H such that $ax+by \in H$ for any $x, y \in H$ and $a, b \in \mathbb{R}$.*

Example 4 *A line, or a plane passing 0 is a subspace of $\mathbb{R}^n$. The span of any subset $\{v_1, v_2, \cdots, v_k\} \subset \mathbb{R}^n$ is a subspace.*

Example 5 *Let n be a positive integer. The set H of all diagonal $n\times n$ matrices is a subspace of $M_{n\times n}(\mathbb{R})$. The set of polynomials of degree at most n is a subspace of $\{f \mid f$ is a continuous function on the interval $[0, 1]\}$.*

Example 6 *For a matrix $A_{n\times m}$, the span of columns of A is a vector space. The set $\{x \in \mathbb{R}^m \mid Ax = 0\}$ is a subspace of $\mathbb{R}^m$.*

2 Basis and dimensions

Let V be a vector space. Simimilar to the $\mathbb{R}^n$, we can define linear combinations, linear independence, basis and dimensions for general vector spaces V, as these concepts involve only additions and scalar multiplications.

For a subset S of V, a linear combination is $a_1v_1 + a_2v_2 + \cdots + a_kv_k$ for some finitely many elements $v_1, v_2, ..., v_k \in S$ and $a_1, a_2, ..., a_k \in F$. A subset $S \subset V$ is linearly independent if any linear combination $a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0$ with $v_1, v_2, \cdots, v_k \in S$ will imply that each $a_i = 0$, $i = 1, 2, ..., k$.

Definition 7 *A basis for a subspace H of V is a set S such 1) S is linearly independent; and 2) S spans H, i.e. any vector in H is a linear combination of S.*

Example 8 *Let $A_{n\times n}$ be a square invertible matrix. The column vectors of A form a basis for $\mathbb{R}^n$.*

Lemma 9 *Let S be a subset of a vector space V and*

$$H = \{a_1v_1 + a_2v_2 + \cdots + a_kv_k \mid v_1, v_2, ..., v_k \in S, a_1, a_2, ..., a_k \in F\}$$

the subspace spanned by S.

1) Suppose that one of the vector in S, say v_k, is a linear combination of other vectors in S. Then H is the span of $S\backslash\{v_k\}$, the set of S without v_k.

2) Suppose that S is a finite set. If $H \neq 0$, then some subset of S is a basis of H.

Proof. 1) It is enough to prove that any $x \in H$ is a linear combination of elements in $S\backslash\{v_k\}$. Since $x \in H$, $x = a_1v_1 + a_2v_2 + \cdots + a_lv_l$. Suppose that $v_k = \sum_{i=1,i\neq k}^{l} b_iv_i$. Then $x = a_1v_1+a_2v_2+\cdots+\sum_{i=1,i\neq k}^{l} a_kb_iv_i = \sum_{i=1,i\neq k}^{l}(a_i+a_kb_i)v_i$, a linear combination of $S\backslash\{v_k\}$.

2) If S is linearly independent, then S is a basis by the definition. Otherwise, S is linearly dependent and one element v_k is a linear combination of $S\backslash\{v_k\}$. By (1), H is the span of $S\backslash\{v_k\}$. Continue such a process until a subset S' of S is linearly independent and H is spanned by S'. Then S' is a basis. ■

Lemma 10 *Let $S = \{s_1, s_2, \cdots, s_n\}$ be a basis of V. Every vector $x \in V$ is a unique linear combination*

$$x_1s_1 + x_2s_2 + \cdots + x_ns_n$$

of S. The vector $(x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n$ is called the coordinate of x relative to S, denoted by $[x]_S$.

Proof. Since S spans V, any vector x is a linear combination of S. Since S is linear independent, the linear combination is unique (i.e. suppose there are two different linear combinations. Take the difference to get a contradiction). ■

Definition 11 *Let V be a vector space and S a basis of V. The number of elements in S is called the dimension* $\dim(V)$ *of V. It's possible that* $\dim(V) = \infty$ *when S is infinite.*

Lemma 12 *Let V be a vector space having a basis $S = \{b_1, b_2, \cdots, b_n\}$. Then any subset S' in V containing more than n vectors must be linearly dependent.*

Proof. Suppose that $S' = \{c_1, c_2, \cdots, c_p\}$ with $p > n$. Then each c_i is a linear combination of S, where the coefficients form the coordinate $[c_i]_B \in \mathbb{R}^n$. Since

$p > n$, the set $\{[c_1]_B, [c_2]_B, \cdots, [c_p]_B\}$ is linearly dependent. Thus there exists a vector $(a_1, a_2, \cdots, a_p) \neq 0$ such that $\sum_{i=1}^{p} a_i [c_i]_B = 0$. Note that

$$\begin{aligned} c_i &= [b_1, b_2, \cdots, b_n][c_i]_B \\ [c_1, c_2, \cdots, c_p](a_1, a_2, \cdots, a_p)^T &= [b_1, b_2, \cdots, b_n][[c_1]_B, \cdots, [c_p]_B](a_1, a_2, \cdots, a_p)^T \\ &= 0 \end{aligned}$$

and thus $\sum_{i=1}^{p} a_i c_i = 0$. This proves S' is linearly independent. ∎

Corollary 13 *If a vector space V has a basis of n vectors, then any basis must consist of exactly n vectors.*

Theorem 14 *(basis extension theorem) Let V be a finite-dimensional vector space. Any linearly independent set S can be extended to be a basis of V.*

Proof. If $\text{Span}(S) = V$, then S is a basis. Otherwise, $V \supsetneq \text{Span}(S)$ and choose $0 \neq v \in V$ but $v \notin \text{Span}(S)$. Then $S \cup \{v\}$ is linearly independent (otherwise, one element is a linear combination of the previous vectors and such an element must be v). Continue such a process to get a maximal linearly independent set, which is a basis. Note that a linearly independent set has at most $\dim V$ elements by the previous lemma, and such a process must stop after at most $\dim V$ steps. ∎

Corollary 15 *(basis theorem) Let $V = \mathbb{R}^n$. Any linearly independent set consisting of n vectors is a basis of V.*

2.1 Linear Transformations

A function f from a set X to a set Y is a rule that for each (input) $x \in X$ assigns a value (output) $y = f(x) \in Y$. Here X is called the domain and Y is called the codomain of f.

Definition 16 *A linear transformation (map) $f : V_1 \to V_2$ between vector spaces V_1, V_2 is a function such that*

$$f(ax + by) = af(x) + bf(y)$$

for any $a, b \in \mathbb{R}$ and $x, y \in V_1$.

Example 17 *1) For a matrix $A_{n\times m}$, the matrix multiplication function $f : \mathbb{R}^m \to \mathbb{R}^n$,*

$$x \mapsto Ax,$$

is linear;

2) Rotations and reflections of $\mathbb{R}^2$ that fixing the origin are linear maps.

Example 18 *The kernel $\ker f = \{x \in V_1 \mid f(x) = 0\}$ and the image $\operatorname{Im} f = \{f(x) \mid x \in V_1\}$ are both vector spaces. A linear map $f : V_1 \to V_2$ is determined by its image on a spanning (or generating) set of V_1.*

Theorem 19 *(general rank theorem) Let $f : V_1 \to V_2$ be a linear map. We have*

$$\dim \ker f + \dim \operatorname{Im} f = \dim V_1.$$

Proof. Let $\{e_1, e_2, ..., e_k\}$ be a basis of $\ker f$. Extend this set to be a basis $\{e_1, e_2, ..., e_k, w_1, w_2, ..., w_l\}$ of V_1 by the basis extension theorem. It can directly checkted that $\{f(w_1), f(w_2), ..., f(w_l)\}$ is a basis of $\operatorname{Im} f$. ■

2.2 Linear maps and matrix multiplications

Let V, W be finite-dimensional vector spaces and $f : V \to W$ be a linear transformation. Fix a basis $B = \{b_1, b_2, \cdots, b_n\}$ of V and a basis $C = \{c_1, c_2, \cdots, b_m\}$ of W. Any vector $x \in V$ is a unique linear combination

$$x = x_1 b_1 + x_2 b_2 + \cdots + x_n b_n$$

of B, i.e. $x = [b_1, b_2, \cdots, b_n][x]_B$. Here

$$[x]_B = (x_1, x_2, ..., x_n)^T$$

is called the coordinate of x with respect to B. Similarly, $f(x)$ is also a linear combination

$$f(x) = y_1 c_1 + y_2 c_2 + \cdots + y_m c_m$$

of W. In other words, we have

$$f(x) = [c_1, c_2, \cdots, c_m][f(x)]_C.$$

A matrix $A = A_{f,B,C}$ is called the *representation matrix* of f with respect to bases B, C, if

$$[f(x)]_C = A[x]_B$$

for any $x \in V$.

Example 20 *When $V = W = \mathbb{R}^n$ and B is the standard basis*

$$\{(1, 0, \cdots, 0)^T, (0, 1, \cdots, 0)^T, \cdots, (0, 0, \cdots, 1)^T\},$$

the representation matrix $A_{f,B,B}$ is the standard matrix defined before. When $V = W$ and $B = C$, we simply call the representation matrix A the B-matrix of f.

Example 21 *When $V = W$ and $f = Id$, the identical map, the representation matrix $A_{Id,B,C}$ is called the transition matrix (or Change of coordinate matrix) from the basis B to the basis C. Show that $A_{Id,B,C} = A^{-1}_{Id,C,B}$.*

Lemma 22 *Let $f : V \to W$ be a linear transformation and $B = \{b_1, b_2, \cdots, b_n\}$ a basis of V, C a basis of W. The representation matrix of f with respect to B, C is*

$$A = [[f(b_1)]_C, [f(b_2)]_C, \cdots, [f(b_n)]_C].$$

Proof. It's obvious that $[[b_1]_B, [b_2]_B, \cdots, [b_n]_B]$ is the identity matrix. The claim is proved by $[f(x)]_C = A[x]_B$ for any x. ∎

Example 23 *Let* $M_2 = \{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid a, b, c, d \in \mathbb{R}\}$ *be the set of all* 2×2 *matrices. Let* $f : M_2 \to M_2$ *be given by* $f(x) = x^T$*, the transpose function. Prove that* f *is linear and find the representation matrix of* f *with respect to the basis* $\{e_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, e_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, e_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\}$.

Lemma 24 *Let* $f : V \to V$ *be a linear transformation and* B_1, B_2 *be two bases of* V*. The representation matrices* A_1, A_2 *of* f *with respect to* B_1, B_2 *are similar.*

Proof. Suppose that $B_1 = \{b_1, b_2, \cdots, b_n\}, B_2 = \{b'_1, b'_2, \cdots, b'_n\}$. According to the definition, we have

$$\begin{aligned} f(x) &= [b_1, b_2, \cdots, b_n][f(x)]_{B_1} = [b'_1, b'_2, \cdots, b'_n][f(x)]_{B_2} \\ [f(x)]_{B_1} &= A_1[x]_{B_1}, [f(x)]_{B_2} = A_2[x]_{B_2}. \end{aligned}$$

Therefore,

$$[b_1, b_2, \cdots, b_n]A_1[x]_{B_1} = [b'_1, b'_2, \cdots, b'_n]A_2[x]_{B_2}.$$

Let P be the transition matrix from B_1 to B_2, i.e. $P[x]_{B_1} = [x]_{B_2}$. Choose $x = b'_1, b'_2, \cdots, b'_n$ to get that

$$\begin{aligned} P[[b'_1]_{B_1}, [b'_2]_{B_1}, \cdots, [b'_n]_{B_1}] &= I_n, \\ [b_1, b_2, \cdots, b_n]A_1[[b'_1]_{B_1}, [b'_2]_{B_1}, \cdots, [b'_n]_{B_1}] &= [b'_1, b'_2, \cdots, b'_n]A_2. \end{aligned}$$

Note that $[b_1, b_2, \cdots, b_n][[b'_1]_{B_1}, [b'_2]_{B_1}, \cdots, [b'_n]_{B_1}] = [b'_1, b'_2, \cdots, b'_n]$. Therefore, we have

$$PA_1P^{-1} = A_2.$$

∎

Example 25 *Let* $f, g : V_1 \to V_2$ *be two linear maps. For any* $a, b \in R$*, we have a new function* $af + bg : V_1 \to V_2$ *defined by* $(af + bg)(x) = af(x) + bg(x)$ *for any* $x \in V_1$*. It can be directly checked that* $af + bg$ *is linear as well. Therefore, the set* $Hom(V_1, V_2)$ *of all linear maps is a vector space.*

Definition 26 *Two vector spaces* V_1, V_2 *are called isomorphic if there exists a bijective linear map* f *between them.*

Example 27 $M_2(\mathbb{R})$ *is isomorphic to* $\mathbb{R}^4$.

Theorem 28 *Two vector spaces* V_1, V_2 *are isomorphic if and only if* $\dim V_1 = \dim V_2$.

Proof. Choose base B_1, B_2 for V_1, V_2 respectively. If $\dim V_1 = \dim V_2$, there is a bijective $\phi : B_1 \to B_2$. Define a map $f : V_1 \to V_2$ as follows. For any $x = \sum_{b \in B_1} x_b b$, let $f(x) = \sum_{b \in B_1} x_b \phi(b)$. It's direct that f is isomorphic. The other direct is obvious. ∎

Lecture 4: Determinants

Shengkui Ye

November 7, 2022

1 Determinant: definitions

For a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the determinant is defined as $\det A = ad - bc$. Inductively, we define:

Definition 1 *For an $n \times n$ matrix A, let A_{1i} be the submatrix obtained from A by deleting the 1-th row and i-th column. The determinant*

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{n+1} a_{1n} \det A_{1n}.$$

Example 2 *Compute the determinant of* $A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$.

Similarly, we let A_{ij} be the submatrix obtained from A by deleting the i-th row and j-th column. Let $C_{ij} = (-1)^{i+j} \det A_{ij}$, called the (i,j)-cofactor.

Theorem 3 *For any $i = 1, 2, \cdots, n$, we have*

$$\begin{aligned} \det A &= a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}, \\ \det A &= a_{1i}C_{1i} + a_{2i}C_{2i} + \cdots + a_{ni}C_{ni}. \end{aligned}$$

Example 4 *Compute the determinant of* $A = \begin{bmatrix} 1 & 5 & 0 \\ 0 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$.

Example 5 *Let* $A = \begin{bmatrix} d_1 & & & \\ * & d_2 & & \\ * & & \ddots & \\ * & & * & d_n \end{bmatrix}$ *be an upper triangular matrix. Show that* $\det A = d_1 d_2 \cdots d_n$.

2 Properties

Theorem 6 *Let A be a square matrix.*

1) If two rows are exchanged to produce B, then $\det B = -\det A$.

2) If one row is multiplied by k to produce B, then $\det B = k \det A$.

3) If a multiple of one row is added to another row to produce a matrix B, then $\det A = \det B$.

Proof. Suppose that $A = (a_{ij})$.

For 1), it is obvious when the size is 2. When the size of A is larger than 2, we will prove the statement by induction. Suppose that the i, j-th ($i < j$) rows are exchanged.

Case 2.1. When i, j are both larger than 1, expand A, B along the first row to get

$$\begin{aligned} \det A &= a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}, \\ \det B &= a_{11}C'_{11} + a_{12}C'_{12} + \cdots + a_{1n}C'_{1n}. \end{aligned}$$

Here C'_{1l} is the cofactor of B. By induction, we have $C'_{1l} = -C_{1l}$ for each $l = 1, 2, \cdots, n$. Therefore, $\det A = -\det B$.

Case 2.2. When $i = 1, j = 2$. Let $\tilde{A}_{st}$ be the submatrix of A by deleting the first two rows and the s-th, t-th columns. Direct calculation shows that

$$\begin{aligned} \det A &= a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{n+1} a_{1n} \det A_{1n} \\ &= \sum_{s=1}^{n} (-1)^{1+s} a_{1s} \det A_{1s} \\ &= \sum_{s=1}^{n} (-1)^{1+s} a_{1s} \Big(\sum_{t<s} (-1)^{1+t} a_{2t} \det \tilde{A}_{st} + \sum_{t>s} (-1)^{t} a_{2t} \det \tilde{A}_{st}\Big) \\ &= \sum_{t<s} (-1)^{s+t} a_{1s} a_{2t} \det \tilde{A}_{st} + \sum_{t>s} (-1)^{1+s+t} a_{1s} a_{2t} \det \tilde{A}_{st} \\ &= -\Big(\sum_{t>s} (-1)^{s+t} a_{2t} a_{1s} \det \tilde{A}_{st} + \sum_{t<s} (-1)^{1+s+t} a_{2t} a_{1s} \det \tilde{A}_{st}\Big) \\ &= -\det B. \end{aligned}$$

Case 2.3. When $i = 1, j > 2$, we exchange the j-th and 2nd rows of B to get a matrix C. Continue to exchange the $1st, 2nd$ rows of C to get a matrix D. Exchange the j-th and 2nd rows of D to get C. By Case 2.1 and Case 2.2, we have $\det B = -\det C = \det D = -\det A$.

After exchanging rows, the 2) is obvious from the definition by expanding along the first row.

For 3), suppose that $B = (b_{ij})$ with $b_{ij} = a_{ij} + a a_{kj}$ for some i and k and any $j = 1, 2, ..., n$. Expand B along the i-th row to get

$$\det B = \sum_{j=1}^{n} b_{ij} C_{ij} = \sum_{j=1}^{n} (a_{ij} + a a_{kj}) C_{ij} = \det A + a \sum_{j=1}^{n} a_{kj} C_{ij}.$$

Note that $\sum_{j=1}^{n} a_{kj}C_{ij}$ is the derminant of the matrix C obtained from A by replacing the i-th row by the k-th row. By 1), $\det C = 0$ since exchanging i, k rows does not change C. Thus we have $\det B = \det A$. ■

Corollary 7 *1)* $\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$ *for any* $i = 1, 2, ..., n$.

2) If two rows of a matrix A *are the same, then* $\det A = 0$.

Example 8 *Let* $A = \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$. *Show that* $\det A = 15$.

Theorem 9 *A square matrix* A *is invertible if and only if* $\det A \neq 0$.

Proof. When A is invertible, A can be reduced by elementary row operations to the identity matrix and thus has non-zero $\det A$. On the other hand, when $\det A \neq 0$, the reduced echelon form of A is invertible and thus A is invertible. ■

Theorem 10 *For two square matrices* A, B, *we have* $\det AB = \det A \det B$.

Proof. Since A, B are invertible, we reduce them by elementary row operations to the identity matices. Suppose that $A = E_1E_2 \cdots E_k, B = F_1F_2 \cdots F_l$ for elementary matrices $E_i, F_j, 1 \leq i \leq k, 1 \leq j \leq l$. By Theorem 6, $\det A$ equals to $(-1)^{k_1} \det D_1$, where k_1 is the number of type 1) permutation matrices and D_1 is the product of type 2) diagonal matrices. Similarly, $\det B = (-1)^{k_2} \det D_2$ using the same notation. Since $AB = E_1E_2 \cdots E_kF_1F_2 \cdots F_l$, we have $\det AB = (-1)^{k_1+k_2} \det(D_1D_2) = \det A \det B$. ■

Corollary 11 *For a square matrix* $A_{n\times n}$, *we have* $\det A = \det A^T$. *Furthermore,*

$$\det A = a_{1i}C_{1i} + a_{2i}C_{2i} + \cdots + a_{ni}C_{ni}$$

for any $i = 1, 2, ..., n$ *(Expansion along Columns).*

Proof. If A is not invertible, then $\det A = 0$. In this case, A^T is not invertible as well and thus $\det A^T = 0$. Suppose that A is invertible and $A = E_1E_2 \cdots E_k$ for elementary matrices $E_i, 1 \leq i \leq k$. Note that $A^T = E_k^T \cdots E_2^T E_1^T$. By Theorem 6, $\det A$ and $\det A^T$ both equal to $(-1)^{k_1} \det D_1$, where k_1 is the number of type 1) permutation matrices and D_1 is the product of type 2) diagonal matrices. ■

Let $\sigma : \{1, 2, ..., n\} \to \{1, 2, ..., n\}$ be a bijection of the set consisting of n natural numbers. Usually, the bijection σ is called a permutation of the n-letters. The set S_n be the set of all bijections σ.

Corollary 12 $\det A = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) a_{\sigma(1),1}a_{\sigma(2),2} \cdots a_{\sigma(n),n}, \mathrm{sgn}(\sigma) \in \{1, -1\}$. *In particular,* $\det A$ *is a polynomial of its entries.*

Proof. Expand the determinant $\det A$ along the first column to get that $\det A = \sum_{i_1=1}^{n}(-1)^{1+i_1}a_{i_1,1}\det A_{i_1,1}$. Continue to expand $\det A_{i_1,1}$ along its first column to get that $\det A_{i_1,1} = \sum_{i_2\neq i_1}(-1)^{2+i_2}a_{i_2,2}\det(A_{i_1,1})_{i_2,1}$. Continue this process to get that

$$\det A = \sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)a_{\sigma(1),1}a_{\sigma(2),2}\cdots a_{\sigma(n),n},$$

where $(\sigma(1),\sigma(2),...,\sigma(n)) = (i_1,i_2,...,i_n)$ is a permutation of $(1,2,...,n)$ (and thus σ can be viewed as a bijection from the n-letter set $\{1,2,...,n\}$). ■

Remark 13 *Although* $\operatorname{sgn}(\sigma)\in\{1,-1\}$*, it's usually complicated to determine it explicitly. Prove that when* σ *is an interchange (i.e. there exist integers* $i\neq j\leq n$ *such that* $\sigma(i)=j,\sigma(j)=i$ *and* $\sigma(k)=k$ *for any* $k\neq i,j$*.), the sign* $sgn(\sigma)=-1$*. The proof of the previous corollary gives a practical way to calculate* $sgn(\sigma)$*.*

3 Cramer's Rule

Theorem 14 *Let* A *be an invertible* $n\times n$ *matrix. For any* $b\in\mathbb{R}^n$*, the system* $Ax=b$ *has the solution*

$$x_i = \frac{\det A_i(b)}{\det A}, i = 1,2,\cdots,n,$$

where $A_i(b)$ *is the matrix obtained from* A *by replacing the* i*-th column by* b*.*

Proof. Suppose that $I=[e_1,e_2,...,e_n]$, where the columns are the standard basis. Note that $A\times I_i(x) = [Ae_1,\cdots Ax,\cdots Ae_n] = A_i(b)$ and thus $\det AI_i(x) = \det A\cdot x_i = \det A_i(b)$, since $\det I_i(x) = x_i$. ■

Example 15 *Use cramer's rule to solve* $\begin{bmatrix}3 & -2\\ -5 & 4\end{bmatrix}x = \begin{bmatrix}6\\8\end{bmatrix}$.

By Cramer's rule, $A^{-1} = (\frac{\det A_i(e_j)}{\det A})_{1\leq i,j\leq n}$ since $AA^{-1}=I_n$.

Corollary 16 *Let* $A^* = (C_{ji})_{1\leq i,j\leq n}$*, called the adjoint matrix* A*, where* C_{ji} *is the* ji*-th cofactor. Then*

$$A^{-1} = \frac{1}{\det A}A^*.$$

Proof. It is enough to note that $\det A_i(e_j) = C_{ji}$, by expanding $A_i(e_j)$ along the j-th column. ■

Example 17 *Let* A *be an integer matrix (i.e. entries are integers) and* $\det A = 1$*. The previous corollary implies that the inverse* A^{-1} *is an integer matrix as well. Similarly, when* A *is a polynomial matrix (i.e. entries are polynomials) and* $\det A = 1$*, the inverse* A^{-1} *is a polynomial matrix.*

Corollary 18 *Let $A_{n\times n}$ be a square matrix, $n \geq 2$. If $rank(A) = n$, then $rank(A^*) = n$; if $rank(A) = n-1$, the $rank(A^*) = 1$; if $rank(A) \leq n-2$, the $rank(A^*) = 0$.*

Proof. If $rank(A) = n$, then A is invertible and thus $A^* = \det(A)A^{-1}$ is of full rank.

If $rank(A) = n-1$, then A has $n-1$ linearly independent rows. These rows form a matrix of rank $n-1$ and the submatrix thus has $n-1$ linearly independent columns. These columns form a submatrix of A with non-zero cofactor. Therefore, $A^* \neq 0$. Note that $AA^* = 0$ and the columns of A^* are the solutions of $Ax = 0$. The rank theorem implies that $Nul(A)$ has dimension 1. Therefore, $\dim rank(A^*) = 1$.

If $rank(A) \leq n-2$, then any $n-1$ rows of A are linearly dependent. Therefore, any cofactor $C_{ij} = 0$ and $A^* = 0$. ■

The following is a result relating the rank of A to the determinant of its submatrices.

Lemma 19 *Let A be a matrix. The rank of A equals to the maximal integer k such that there exists a non-zero $k \times k$ submatrix B of A with nonzero* $\det B \neq 0$.

Proof. Note that $rank(A) = \dim Col(A)$. When $k > rank(A)$, any k columns of A are linearly dependent. This means any $k \times k$ submatrix of A has linear dependent columns.

When $k = rank(A)$, choose k linearly independent columns $\{A_1, A_2, ..., A_k\}$ of A. Then $rank[A_1, ..., A_k] = k = rank[A_1, ..., A_k]^T$. There are k rows of $[A_1, ..., A_k]$, which are linearly independent. These k rows give a $k \times k$ submatrix with nonzero determinant. ■

4 Geometric meaning of determinants

Lemma 20 *If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is $|\det A|$.*

Proof. If the two rows of A are parrell, then A is not invertible and thus $\det A = 0$. We simply assume that $\det A \neq 0$ and the two rows are linearly independent. If $c = 0$, then the parallelogram has bottom $|a|$, and hight $|d|$. Thus the area is $|ad| = \det A$. Generally, when $c \neq 0$, rotate the plane anti-clockwise by degree ϕ. The corresponding linear transformation is

$$\begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} : \mathbb{R}^2 \to \mathbb{R}^2.$$

Choose an appropriate angle ϕ such that $\begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix}$ has its second component 0. Therefore,

$$\det A = \det \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}^{-1} \det \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} A$$

whose absolute value is the area of the parrellogram formed by the two rows of $\begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} A$. The proof is finished. ■

Lemma 21 *If A is a 3×3 matrix, the volume of the parallelepiped determined by the columns of A is $|\det A|$.*

Proof. The proof is similar to that of the previous lemma. Suppose that $\mathbb{R}^3$ has the ordinary coordinates x, y, z. If the third row vector of A lies on the z-coordinate, then the first two column vectors of A lie in the xoy-plane. By the previous lemma, the area of the bottom parallelogram is the absolute value of $\det A_{33} = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. Expand A along the third row to get

$$\det A = a_{33} \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

whose absolute value is the volume of the parallelepiped, whose hight is $|a_{33}|$, area of the bottom is $|\det A_{33}|$. For the general case, we rotate $\mathbb{R}^3$, such that the last column of A^T (i.e. the last row of A) lies in the z-coordinate. Since the rotation does not change volums, the proof is finished. ■

Lemma 22 *Let $S \subset \mathbb{R}^3$ be region with its volume defined. If $f : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear map with standard matrix A, then the volume of $f(S)$ is $|\det A| \mathsf{vol}(S)$. Similar result holds for $f : \mathbb{R}^2 \to \mathbb{R}^2$ on areas.*

Proof. By the definition of volumes in Calculus, the volume of S is the infimum of the sum of volums of small cubes covering S. Since f is linear, it is additive on the small cubes. Therefore, it's enough to prove the case when S is a cube. Without loss of generality, we assume that one vertex of S is the origin. Suppose that S has its three edges $(a, 0, 0)^T, (0, b, 0)^T, (0, 0, c)^T$. Then $f(S)$ is a parallelepiped, formed by the rows of $A \begin{bmatrix} a & & \\ & b & \\ & & c \end{bmatrix}$. By the geometric meaning of determinant, $f(S)$ has the volume $|\det A| abc = |\det A| \mathsf{vol}(S)$. ■

Example 23 *Let $a, b > 0$. Find the area of $\{(x, y) \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1\}$.*

Proof. Let $S = \{(x, y) \mid x^2 + y^2 \leq 1\}$. Define $f : \mathbb{R}^2 \to \mathbb{R}^2$ as

$$f(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} x/a \\ y/b \end{bmatrix} = \begin{bmatrix} \frac{1}{a} & \\ & \frac{1}{b} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Therefore, $f(\{(x, y) \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1\}) = S$, and thus $Area(S) = \frac{\pi}{\det A} = \pi ab$. ■

Lecture 5: Eigenvalues and eigenvectors

Shengkui Ye

November 22, 2022

1 Eigenvalues and eigenvectors: definitions

Definition 1 *Let A be an $n \times n$ matrix. Suppose that there exists scalar λ and nonzero vector x such that*

$$Ax = \lambda x.$$

The λ is called an eigenvalue of A and x an eigenvector of A corresponding to λ.

Example 2 *Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$ and $x = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$. Check that x is an eigenvector.*

Remark 3 *Let $\mathbb{R}x$ be the line spanned by x. If x is an eigenvector, then A (viewed as a linear transformation $\mathbb{R}^n \to \mathbb{R}^n$) maps $\mathbb{R}x$ to $\mathbb{R}x$, i.e. the line $\mathbb{R}x$ is preserved by A.*

Lemma 4 *λ is an eigenvalue of A if and only if* $\det(A - \lambda I_n) = 0$.

Proof. By the definition, $Ax = \lambda x$ if and only if $(A - \lambda I_n)x = 0$. In other words, λ is an eigenvalue of A if and only if $(A - \lambda I_n)x = 0$ has a non-zero solution, which is equivalent to $\det(A - \lambda I_n) = 0$. ■

Corollary 5 0 *is an eigenvalue of A if and only if A is not invertible.*

Example 6 *Let $A = \begin{bmatrix} a_{11} & * & & * \\ & a_{22} & & * \\ & & \ddots & \\ & & & a_{nn} \end{bmatrix}$ be an upper triangular matrix. The eigenvalues of A are diagonal entries.*

Example 7 *For a fixed eigenvalue λ, the set $V_\lambda = \{x \mid Ax = \lambda x\}$ of eigenvectors is a vector subspace, called the eigenspace of A corresponding to λ.*

Lemma 8 *If $x_1, x_2, \cdots, x_k$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_k$, then $\{x_1, x_2, \cdots, x_k\}$ are linearly independent.*

Proof. After reordering the index, we assume that $\lambda_1, \lambda_2, \cdots, \lambda_k$ are all distinct. Suppose that $\{x_1, x_2, \cdots, x_{l-1}\}$ are linearly independent, but $\{x_1, x_2, \cdots, x_l\}$ are linear dependent. For some a_i we have $a_1x_1 + a_2x_2 + \cdots + a_{l-1}x_{l-1} = x_l$. Multiplying A at both sides, we have

$$\begin{aligned} a_1Ax_1 + a_2Ax_2 + \cdots + a_{l-1}Ax_{l-1} &= Ax_l \\ a_1\lambda_1x_1 + a_2\lambda_2x_2 + \cdots + a_{l-1}\lambda_{l-1}x_{l-1} &= \lambda_lx_l. \end{aligned}$$

Therefore, $a_1(\lambda_l - \lambda_1)x_1 + \cdots + a_{l-1}(\lambda_l - \lambda_{l-1})x_{l-1} = 0$, which is a contradiction. This means that the linear independence of $\{x_1, x_2, \cdots, x_{l-1}\}$ implies the linear dependence of $\{x_1, x_2, \cdots, x_{l-1}, x_l\}$ for any l. Eventually, we have that $\{x_1, x_2, \cdots, x_k\}$ are linearly independent. ■

2 Characteristic polynomial and diagonalization

Definition 9 *For a matrix $A_{n\times n}$, the $\det(A - \lambda I_n)$ is called the characteristic polynomial of A. The roots of this polynomial are eigenvalues.*

Example 10 *Find the eigenvalues of $\begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.*

Two matrices A, B are called similar if there exists an invertible matrix P such that $PAP^{-1} = B$. Changing A into PAP^{-1} is called a similarity transformation.

Lemma 11 *Two similar matrices A, B have the same characteristic polynomials and thus the same eigenvalues.*

Proof. $\det(PAP^{-1} - \lambda I_n) = \det P(A - \lambda I_n)P^{-1} = \det P \det(A - \lambda I_n) \det P^{-1} = \det(A - \lambda I_n)$. ■

Definition 12 *Let A be an $n \times n$ matrix. If $\det(A - \lambda I_n) = (-1)^n(\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_k)^{n_k}$, then the integer n_i is called the algebraic multiplicity of the eigenvalue λ_i.*

For an eigenvalue λ, the space $V_\lambda = \{v \mid Av = \lambda v\}$ is called an eigensapce of A corresponding to λ. The dimension $\dim V_\lambda$ is called the geometric multiplicity of λ.

Example 13 *Let $A = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 1 \end{bmatrix}$. Find its eigenvalues and their algebraic and geometric multiplicities.*

A matrix A is diagonalizable if there exists invertible matrix P such that PAP^{-1} is diagonal.

Theorem 14 *An $n \times n$ matrix A is diagonalizable if and only if A has n linear independent eigenvectors.*

Proof. Suppose that there exist invertible matrix P and diagonal matrix D such that $PAP^{-1} = D$. Then $AP^{-1} = P^{-1}D$. Then the columns of P^{-1} are eigenvectors and thus linear independent.

If $\{x_1, x_2, ..., x_n\}$ are eigenvectors of A, then

$$\begin{aligned} A[x_1, x_2, ..., x_n] &= [Ax_1, Ax_2, ..., Ax_n] \\ &= [x_1, x_2, ..., x_n]\text{diag}(\lambda_1, \lambda_2, ..., \lambda_n). \end{aligned}$$

Here $\text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$ is the diagonal matrix with diagonal entries λ_i. When $\{x_1, x_2, ..., x_n\}$ are linear independent, the matrix $[x_1, x_2, ..., x_n]$ are invertible and thus $[x_1, x_2, ..., x_n]^{-1}A[x_1, x_2, ..., x_n] = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$. ■

The proof of the previous theorem shows that $A = [x_1, x_2, ..., x_n]\text{diag}(\lambda_1, \lambda_2, ..., \lambda_n)[x_1, x_2, ..., x_n]^{-1}$ for the eigenvalues λ_i and eigenvectors v_i, when A is diagonalizable.

Corollary 15 *If $A_{n\times n}$ has n distinct eigenvalues, then A is diagonalizable.*

Proof. If A has n distinct eigenvalues, then it has n linearly independent eigenvectors. ■

Example 16 *Diagonalize the following matrix, if possible*

$$A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}.$$

Theorem 17 *Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, ..., \lambda_p$.*

1. *For $1 \le k \le p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.*

2. *The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals n, and this happens if and only if*

   (i) the characteristic polynomial factors completely into linear factors and

   (ii) the geometric mutiplicity equals to the algebraic mutiplicity for each eigenvalue, i.e. the dimension of the eigenspace for each λ_k equals the algebraic multiplicity of k.

3. *If A is diagonalizable and B_k is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $B_1, B_2, \cdots, B_p$ forms an eigenvector basis for $\mathbb{R}^n$.*

Proof. (2) If A is diagonalizable, assume there exist invertible matrix P and diagonal matrix D such that $PAP^{-1} = D$. Then the characteristic polynomial of A is the same as that of D, which is a product of linear factors. The dimension of the eigenspace of D for each λ_k clearly equals the multiplicity of k. View P as an invertible linear transformation. Note that $P\{v \in \mathbb{R}^n \mid Av = \lambda_k v\}$ is the corresponding eigenspace of D.

Conversely, when the dimension of the eigenspace for each λ_k equals the multiplicity of k, we can choose k linearly independent vectors in the eigenspace V_{λ_k}. Since $\sum k = n$, we have n linearly independent eigenvectors. Therefore, A is diagonalizable.

(3) follows (2), since the dimension of the eigenspace for each λ_k equals the multiplicity of k. (1) will be proved in the next section. ∎

Example 18 *Calculate all the eigenvalues and eigenvectors of* $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. *Prove that A is not diagonalizable.*

The characteristic polynomial $p(A)$ of a matrix A with real entries has real coefficients. It does not always factor into linear factors. Sometimes, $p(A) = (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_k)^{n_k}$ has (non-real) complex roots. But the complex roots occur in conjugate pairs.

Example 19 *Let* $A = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}$. *The characteristic polynomial is* $\lambda^2 - 2(\cos\phi)\lambda + 1$. *The two roots are the conjugate pair* $\lambda_1 = \cos\phi + \sqrt{1 - \cos^2\phi}\mathrm{i}$, $\lambda_2 = \cos\phi - \sqrt{1 - \cos^2\phi}\mathrm{i}$. *When* $\cos\phi \neq 0$, *these are complex eigenvalues.*

3 Eigenvectors and linear transformations

Lemma 20 *Let $f : V \to V$ be a linear transformation and B_1, B_2 be two bases of V. The representation matrices A_1, A_2 of f with respect to B_1, B_2 are similar.*

Proof. Suppose that $B_1 = \{b_1, b_2, \cdots, b_n\}, B_2 = \{b'_1, b'_2, \cdots, b'_n\}$. According to the definition, we have

$$\begin{aligned} f(x) &= [b_1, b_2, \cdots, b_n][f(x)]_{B_1} = [b'_1, b'_2, \cdots, b'_n][f(x)]_{B_2} \\ [f(x)]_{B_1} &= A_1[x]_{B_1}, [f(x)]_{B_2} = A_2[x]_{B_2}. \end{aligned}$$

Therefore,

$$[b_1, b_2, \cdots, b_n]A_1[x]_{B_1} = [b'_1, b'_2, \cdots, b'_n]A_2[x]_{B_2}.$$

Let P be the transition matrix from B_1 to B_2, i.e. $P[x]_{B_1} = [x]_{B_2}$. Choose $x = b'_1, b'_2, \cdots, b'_n$ to get that

$$\begin{aligned} P[[b'_1]_{B_1}, [b'_2]_{B_1}, \cdots, [b'_n]_{B_1}] &= I_n, \\ [b_1, b_2, \cdots, b_n]A_1[[b'_1]_{B_1}, [b'_2]_{B_1}, \cdots, [b'_n]_{B_1}] &= [b'_1, b'_2, \cdots, b'_n]A_2. \end{aligned}$$

Note that $[b_1, b_2, \cdots, b_n][[b'_1]_{B_1}, [b'_2]_{B_1}, \cdots, [b'_n]_{B_1}] = [b'_1, b'_2, \cdots, b'_n]$. Therefore, we have

$$PA_1P^{-1} = A_2.$$

∎

Corollary 21 *Let $f : V \to V$ be a linear transformation. The eigenvalue of (a representation matrix of) f does not dependent on the choice of bases.*

The following is part (1) of Theorem 17.

Corollary 22 *Let λ be an eigenvalue of a matrix $A_{n\times n}$ and V_λ the eigenspace corresponding to λ. Then the geometric multiplicity $\dim V_\lambda \le$ the algebraic multiplicity of λ.*

Proof. Suppose that $\dim V_\lambda = p$ and choose a basis $\{v_1, v_2, \cdots, v_p\}$ of V_λ. Extend the basis to be a basis $S = \{v_1, v_2, \cdots, v_p, \cdots, v_n\}$ of $\mathbb{R}^n$. Let A' be the representation matrix of the linear transformation $A : \mathbb{R}^n \to \mathbb{R}^n$, $x \mapsto Ax$, with respect to the basis S. In other words, $A'[x]_S = [Ax]_S$ for any vector $x \in \mathbb{R}^n$. Since $Av_i = \lambda v_i$ for $i \le p$, the matrix

$$A' = \begin{bmatrix} \lambda I_p & C_1 \\ 0 & C_2 \end{bmatrix},$$

where C_1, C_2 are submatrices of appropriate sizes. Note that A and A' are similar by the previous lemma and the characteristic polynomial of A' and A are same, which is

$$\det(A' - xI_n) = (x - \lambda)^p p_1(x).$$

Here $p_1(x)$ is the characteristic polynomial of C_2. Therefore, $\dim V_\lambda = p \le$ the algebraic multiplicity of λ. ■

Lecture 6: Canonical forms and decompositions

Shengkui Ye

November 29, 2022

1 Jordan canonical forms for complex matrices

A matrix A is called nilpotent if $A^k = 0$ for some positive integer k. The Jordan block is an upper triangular matrix of the form

$$J_{d,n} = \begin{bmatrix} d & 1 & 0 & \\ & d & \ddots & 0 \\ & & \ddots & 1 \\ & & & d \end{bmatrix}_{n\times n}.$$

The direct sum (or block sum) of two matrix A, B is a block diagonal matrix $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$.

Lemma 1 *Let V be a finite-dimensional vector space and $f : V \to V$ be a linear map. There exist subspaces $V_1, V_2 < V$ such that*

1) $V = V_1 \bigoplus V_2$ (i.e. $V = V_1 + V_2$ and $V_1 \cap V_2 = 0$);

2) $f(V_1) = V_1$ and $f|_{V_1}$ is invertible.

3) $f(V_2) < V_2$, and there is an integer k such that $f^k(x) = 0$ for any $x \in V_2$ (i.e. $f|_{V_2}$ is nilpotent).

Proof. For each integer i, note that the kernels satisfy $\ker f^i \leq \ker f^{i+1}$. Since V is finite-dimensional, there is a smallest integer k such that $\ker f^k = \ker f^{k+1}$. Actually, $\ker f^k = \ker f^{k+l}$ for any integer $l \geq 0$ as the following. For any $z \in \ker f^{k+l}$, we have $0 = f^{k+l}(z) = f^{k+1+l-1}(z)$, implying $f^{l-1}(z) \in \ker f^{k+1} = \ker f^k$ and $f^{k+l-1}(z) = 0$. Repeat the argument to get $0 = f^{k+l-1}(z) = \cdots = f^k(z)$.

Note that any $x \in \ker f^k \cap \operatorname{Im} f^k$ has $x = f^k(y), f^k(x) = 0$ for some $y \in V$. This means that $f^k(f^k(y)) = 0$ and $y \in \ker f^{k+k} = \ker f^k$. Therefore, $0 = f^k(y) = x$. By the generalized rank theorem $\dim V = \dim \ker f^k + \dim \operatorname{Im} f^k$, we know that $V = \ker f + \operatorname{Im} f^k$. This finishes the proof of 1) with $V_2 = \ker f^k$ and $V_1 = \operatorname{Im} f^k$.

It's obvious that $f(V_1) \leq V_1, f(V_2) \leq V_2$. For any $x \in V_2$, we have $f^k(x) = 0$. In order to prove $f|_{V_2}$ is invertible, it is enough to prove $f|_{V_2}$ is injective since V_2

is of finite dimension. For any $z \in \operatorname{Im} f^k$ satisfying $f(z) = 0$, we have $z = f^k(y)$ for some y and $f^{k+1}(y) = 0$. This means $y \in \ker f^{k+1} = \ker f^k$ and thus $z = 0$. The injectivity of $f|_{V_2}$ is proved. ■

Lemma 2 *For a nilpotent matrix $A_{n\times n}$, the sum $I + A$ is conjugate to a direct sum of Jordan blocks with $1s$ along the diagonal.*

Proof. We prove that $V = F^n$ has a basis

$$\{a_1, Aa_1, \ldots, A^{k_1-1}a_1, a_2, Aa_2, \ldots, A^{k_2-1}a_2, \ldots, a_s, \ldots, Aa_s, \ldots, A^{k_s-1}a_s\}$$

satisfying $A^{k_i}a_i = 0$ for each i, which implies that the representation matrix of $I + A$ with respect to this basis is a direct sum of Jordan blocks with 1 along the diagonal. The proof is based on the induction of $\dim V$. When $\dim V = 1$, choose $0 \neq v \in V$. Suppose that $Av = \lambda v$. Then $A^k v = \lambda^k v = 0$ and thus $\lambda = 0$. Suppose that the case is proved for vector spaces of dimension $k < n$. Note that the subspace $AV \neq V$ (otherwise, $AV = V$ implies $A^k V = A^{k-1}V = V = 0$). By induction, the subspace AV (noting that $A(AV) \subset AV$) has a basis

$$S = \{a_1, Aa_1, \ldots, A^{k_1-1}a_1, a_2, Aa_2, \ldots, A^{k_2-1}a_2, \ldots, a_s, \ldots, Aa_s, \ldots, A^{k_s-1}a_s\}.$$

Choose $b_i \in V$ satisfying $A(b_i) = a_i$. Then A maps the set

$$S' = \{b_1, Ab_1 = a_1, \ldots, A^{k_1}b_1 = A^{k_1-1}a_1, b_2, Ab_2, \ldots, A^{k_2}b_2, b_s, \ldots, Ab_s, \ldots, A^{k_s}b_s\}$$

to the basis S. This implies that the set S' is linearly independent (Otherwise, $\sum_{j=1}^{s}(x_j b_j + \sum_{i=0}^{k_j-1} x_{ji}A^i a_j) = 0$ for some nonzero x_j, which implies $A(\sum_{j=1}^{s}(x_j b_j + \sum_{i=0}^{k_j-1} x_{ji}A^i a_j)) = \sum_{j=1}^{s}(x_j a_j + \sum_{i=0}^{k_j-1} x_{ji}A^{i+1}a_j) = 0$, a contradiction to the fact that S is a basis). Extend this set S' to be a $V's$ basis

$$S'' = \{b_1, Ab_1, \ldots, A^{k_1}b_1, b_2, Ab_2, \ldots, A^{k_2}b_2, b_s, \ldots, Ab_s, \ldots, A^{k_s}b_s, b_{s+1}, \ldots, b_{s'}\}.$$

Note that $Ab_i = 0$ for $i \geq s+1$ and $A^{k_i+1}b_i = A^{k_i}a_i = 0$ for each $i \leq s$. ■

Theorem 3 *(Jordan canonical form) Any complex matrix $A_{n\times n}$ is conjugate to a direct sum of Jordan blocks, where the diagonal entries are eigenvalues.*

Proof. Consider the linear map $A : \mathbb{C}^n \to \mathbb{C}^n$. Over the field $\mathbb{C}$ of complex numbers, we have $\det(A - \lambda I_n) = (-1)^n(\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_k)^{n_k}$, a product of distinct eigenvalues $\lambda_1, ..., \lambda_k$. View $f = A - \lambda_1 I_n$ as a linear map $\mathbb{C}^n \to \mathbb{C}^n$. Lemma 1 implies that $\mathbb{C}^n = V_1 \bigoplus V_2$, where $f|_{V_2}$ is nilpotent and $f|_{V_1}$ is invertible. Since the eigenspace $V_{\lambda_1} = \ker(A - \lambda_1 I_n) < V_2$, we see that $\dim V_2 > 0$. Lemma 2 implies that $(A - \lambda_1 I_n + I_n)|_{V_2}$ is conjugate to a direct sum of Jordan blocks with $1s$ along the diagonal. This means that $A|_{V_2}$ is conjugate to a direct sum of Jordan blocks $J_{\lambda_1, n_{1_j}}$. Consider $A|_{V_1} : V_1 \to V_1$ instead of $A : \mathbb{C}^n \to \mathbb{C}^n$ and repeat the argument. Note that there are only k eigenvalues. The proof will be finished in k steps. ■

Remark 4 *The Jordan canonical form does not hold true for real matrices. For example* $A = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$, $\phi \neq 0, \pi$, *has no real eigenvalues. The very first* V_2 *in the proof of the previous theorem would be trivial.*

Corollary 5 *(Jordan-Chevalley decomposition) Any square matrix* $A_{n\times n}$ *can be written as*

1) a sum $A = S + N$, *with* S *a diagonalizable matrix and* N *a nilpotent matrix, satisfying* $SN = NS$; *and*

2) a product $A = SU$, *with* S *a diagonalizable matrix and* $N - I_n$ *a nilpotent matrix, satisfying* $SU = US$. *Here* U *is called unipotent.*

Proof. For a Jordan block $J_{d,k}$, let $S = dI_k$, $N = J_{d,k} - S$ and $U = I_k + N$. ■

For a polynomial $p(x) = \sum_{i=0}^{n} a_i x^i$, its matrix value is $p(A) = a_n A^n + ... + a_1 A + a_0 I_n = \sum_{i=0}^{n} a_i A^i \in M_{k\times k}(F)$ for a matrix $A_{k\times k}$ with entries in a field F.

Corollary 6 *(Cayley–Hamilton Theorem) Let* $A_{n\times n}$ *be a square matrix and* $p(x) = \det(A - \lambda I_n)$ *its characteristic polynomial. We have* $p(A) = 0$.

Proof. For any invertible matrix $B_{n\times n}$, note that $(BAB^{-1})^i = BA^iB^{-1}$ and thus $p(BAB^{-1}) = Bp(A)B^{-1}$. The Jordan canonical form implies that $BAB^{-1} = D$ for some upper triangular matrix D (a direct sum of Jordan blocks) and some invertible matrix B. It is enough to prove that $p(D) = Bp(A)B^{-1} = 0$. Suppose that $p(x) = \det(A - \lambda I_n) = (-1)^n \Pi_{i=1}^{l} (\lambda - \lambda_i)^{n_i}$ for distinct roots $\lambda_1, ..., \lambda_l$. For each Jordan block J_{n_i,λ_i}, we have $J - \lambda_i I_{n_i}$ a nilpotent matrix. A direct calculation shows that $(J - \lambda_i I_{n_i})^{n_i} = 0$. In the product $p(B) = (-1)^n \Pi_{i=1}^{l} (A - \lambda_i I_n)^{n_i}$, each factor $(B - \lambda_i I_n)^{n_i}$ has the corresponding $n_i \times n_i$ block matrix zero. Therefore, $p(B) = 0$. ■

2 Real matrices

Example 7 *Any* 2×2 *matrix* $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ *is a product* $r \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$, *for* $r = \sqrt{a^2 + b^2}$ *and a suitable angle* ϕ.

Lemma 8 *Let* A *be any* 2×2 *matrix with a complex eigenvalue* $\lambda = a + b\mathbf{i}$ $(b \neq 0)$. *Then* A *is conjugate to* $r \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$ *with* $r = \sqrt{\det A}$ *and a suitable angle* ϕ.

Proof. Let $v = \mathrm{Re} + i\,\mathrm{Im}$ (viewed as a complex vector) be an eigenvector of λ, where Re is the real part and Im is the imaginary part. Denote by $\bar{v} = \mathrm{Re} - i\,\mathrm{Im}$ the complex conjugate of v. Then $Av = \lambda v$ implies that

$$A\bar{v} = \bar{\lambda}\bar{v}.$$

Since eigenvectors corresponding to different eigenvalues are linearly independent, we know that v and $\bar{v}$ are linearly independent. Since

$$[\mathrm{Re}, \mathrm{Im}] \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = [v, \bar{v}],$$

we know that $[\mathrm{Re}, \mathrm{Im}]$ are linearly independent. Note that $A\,\mathrm{Re} = a\,\mathrm{Re} - b\,\mathrm{Im}$, $A\,\mathrm{Im} = a\,\mathrm{Im} + b\,\mathrm{Re}$. Therefore,

$$\begin{aligned} A[\mathrm{Re}, \mathrm{Im}] &= [\mathrm{Re}, \mathrm{Im}] \begin{bmatrix} a & b \\ -b & a \end{bmatrix}, \\ [\mathrm{Re}, \mathrm{Im}]^{-1} A[\mathrm{Re}, \mathrm{Im}] &= \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{a^2+b^2} & 0 \\ 0 & \sqrt{a^2+b^2} \end{bmatrix} \begin{bmatrix} \frac{a}{\sqrt{a^2+b^2}} & \frac{b}{\sqrt{a^2+b^2}} \\ -\frac{b}{\sqrt{a^2+b^2}} & \frac{a}{\sqrt{a^2+b^2}} \end{bmatrix}. \end{aligned}$$

The proof is finished by taking $r = \sqrt{a^2+b^2}$ and $\phi = \arccos \frac{a}{\sqrt{a^2+b^2}}$. ∎

Theorem 9 *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *be a* 2×2 *real matrix. Then A is conjugate to*

1) a diagonal matrix; or

2) an upper triangular matrix $\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$ *or*

3) a multiple of an rotation matrix $r \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$ *with* $r = \sqrt{\det A}$ *and a suitable angle* ϕ.

Proof. Consider the characteristic polynomial $\det(A - \lambda I_2) = \lambda^2 - tr(A)\lambda + \det(A)$. When $\Delta = tr(A)^2 - 4\det(A) > 0$, there are two distinct eigenvalues and A is diagonalizable. When $\Delta = tr(A)^2 - 4\det(A) = 0$, there is only one eigenvalue λ. If $\dim V_\lambda = 2$, we know that A is diagonalizable. Otherwise, $\dim V_\lambda = 1$. Suppose that $Av = \lambda v$ for some $v \neq 0$ and $\{v, w\}$ is a basis of R^2. The representation matrix of A with respect to $\{v, w\}$ is $D = \begin{bmatrix} \lambda & x \\ 0 & \lambda \end{bmatrix}$ for some $x \neq 0$. But $D - \lambda I_2$ is nilpotent. Lemma 2 implies that D is conjugate to $\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$. When $\Delta = tr(A)^2 - 4\det(A) < 0$, there are two distinct complex eigenvalues. The previous lemma proves 3).

Corollary 10 *Let* $A_{2\times 2}$ *be a real matrix of* $\det(A) = 1$. *Then A is conjugate to either* $\begin{bmatrix} \lambda & 0 \\ 0 & \frac{1}{\lambda} \end{bmatrix}$, *or* $\pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ *or a rotation matrix* $\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$.

∎