# Lecture Notes on Linear Algebra (MTH101)

Amit Kulshrestha
amitk@iisermohali.ac.in

November 29, 2023

# Contents

# Lecture 1

## Course logistics

August 23, 2023

### Course Hours

**Lecture.** Wednesdays and Fridays at 9:00 AM in LH-6.
**Tutorials.** Thursdays at 9:00 AM in LH-6 (Section 1), LH-5 (Section 2), LH-1 (Section 3).

**Office Hours.** Mondays from 11:00 AM to 12:00 noon and Thursdays from 11:00 AM to 1:00 PM in AB2-1F7.

### Course TAs

Team 1 (S1 : LH6)]

(i). Sumit Chandra Mishra : sumitcmishra@iisermohali.ac.in
(ii). Sandeepan Dutta : sandipandutta@iisermohali.ac.in
(iii). Deepanshi Saraf : mp19001@iisermohali.ac.in
(iv). Shruti Rastogi : mp18021@iisermohali.ac.in

Team 2 (S2 : LH5)

(i). Arup Kumar Maity : arupmaity@iisermohali.ac.in
(ii). Ramanpreet Kaur : ramanpreet@iisermohali.ac.in
(iii). Harish Kishnani : ph20021@iisermohali.ac.in
(iv). Nidhi Gupta : mp18009@iisermohali.ac.in

Team 3 (S3 : LH4)

(i). Gurleen Kaur : gurleen@iisermohali.ac.in
(ii). Chitrarekha Sahu : ph20012@iisermohali.ac.in
(iii). Aman Singh : ph22056@iisermohali.ac.in

**Grading Policy**

1. Mid semester exam : 1 x 20 = 20 points.
2. Quizzes : 3 x 5 = 15 points.
3. Assignments : 2 x 5 = 10 points.
4. Creative Assignment : 1 x 5 = 5 points.
5. End semester exam : 1 x 50 = 50 points.

**Syllabus**

- Solving systems of linear equations through row reduction
- Matrices: addition and multiplication of matrices; Using matrix multiplication to represent a system of linear equations; Row operations in terms of matrix multiplication
- Invertible matrices: Computing inverses through row reduction algorithm
- Determinants: Cramer's rule
- Permutation matrices
- Vector spaces: Subspaces; linear transformations; kernel and image of a linear transformation
- Linear dependence and independence: Span of a set, Bases of vector spaces; Dimension
- Matrix representations: vectors and linear transformations with respect to bases; Change of basis
- Rank-nullity theorem
- Direct Sums: sums of subspaces, direct sums of vector spaces
- Eigenvalues and eigenvectors, diagonalization

**Suggested readings**


- Michael Artin, *Algebra*, Prentice-Hall, Inc. 1991.
- Sheldon Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1996.

# Lecture 2

## Numbers - as we understand them

August 24, 2023

Our understanding of real numbers may not be very precise, but we assume working knowledge of real numbers. Therefore, we will proceed by our existing impression of real numbers. We will remain sensitive that laying down the foundation of real numbers has really been challenging task for generations of mathematicians. This task could be satisfactorily accomplished only during nineteenth century. However, in this course we will not bother ourselves with the fundamental questions of existence of real numbers.

The set of real numbers will be denoted by $\mathbb{R}$ and we will assume that we know how to add, subtract and multiply real numbers. We will also assume how to divide by nonzero real numbers.

# Lecture 3

# Fixing the rules

August 25, 2023

### 3.1. Vectors : let us fix their meaning

Let us assume that we understand real numbers and how to add and multiply them. We denote the set of real numbers by $\mathbb{R}$. By a *number*, we refer to a real number. To begin with, let us define a pair of numbers written as a column $v := \begin{pmatrix} a \\ b \end{pmatrix}$, to be a *column vector*. Here, $a, b \in \mathbb{R}$. The symbol $\in$ should be read as *belongs to*. The *transpose* of $v$ is just the rearrangement of $a$ and $b$ in a row. The transpose of $v$ is denoted by $v^t$ and thus, $v^t = (a \quad b)$. For convenience, $a$ and $b$ are separated by a comma and we write $v^t = (a, b)$. When a pair of numbers is written in a row, it is called a *row vector*. When we use the word 'vector', we mean a column vector.

The vector $\begin{pmatrix} a \\ b \end{pmatrix}$ is to be regarded different from $\begin{pmatrix} b \\ a \end{pmatrix}$. Thus, what we call a pair is actually an *ordered* pair. Thus, $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$, $\begin{pmatrix} \sqrt{5} \\ -2 \end{pmatrix}$ and $\begin{pmatrix} \cos(\pi/3) \\ -\sin(\pi/3) \end{pmatrix}$ are examples of vectors.

The collection of all (column) vectors is denoted by $\mathbb{R}^2$ and the collection of all row vectors is denoted by $(\mathbb{R}^2)^t$.

### 3.2. Addition and scaling of vectors

Take $v := \begin{pmatrix} a \\ b \end{pmatrix}$, $w := \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$. The *addition* of $v$ and $w$ is defined by

$$v + w := \begin{pmatrix} a + c \\ b + d \end{pmatrix}$$

and the *scaling* of $v$ by $\alpha$ is defined by $\alpha v := \begin{pmatrix} \alpha a \\ \alpha b \end{pmatrix}$.

The collection $\mathbb{R}^2$ may be thought of as a plane in the following sense: In a plane, arbitrarily fix a reference point as ″origin″ and arbitrarily fix a mutually perpendicular pair of lines as a ″coordinate frame″, or ″axes″. Now, one may label other points using elements of $\mathbb{R}^2$. In $v := \begin{pmatrix} a \\ b \end{pmatrix}$, the number $a$ is called the *$x$-coordinate* of $v$ and $b$ is called the *$y$-coordinate* of $v$.

### 3.3. Dot product of vectors

For two vectors $v := \begin{pmatrix} a \\ b \end{pmatrix}$, $w := \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{R}^2$, their *dot product* $v.w$ is a real number which is defined as follows: $v.w := ac + bd$.

In coming lectures, the motive behind such definitions will reveal itself as we begin to relate these with what we know.

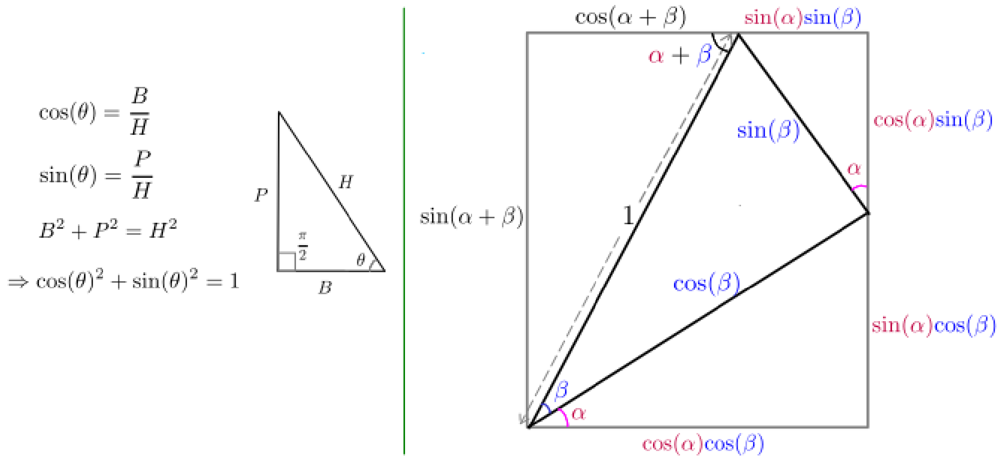
# Lecture 4

# Dot product, trigonometry and rotations

August 30, 2023

## 4.1. Back to triangles

First, refresh our trigonometry by looking at the following picture




The picture reveals that

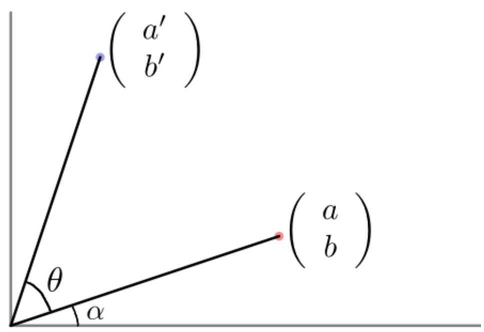
$$\sin(\alpha + \beta) = \cos(\alpha)\sin(\beta) + \sin(\alpha)\cos(\beta)$$

and

$$\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$$

This may be used to understand rotations. Let us pose the problem.

**Problem**. In the following picture, if the initial location of the red point is expressed through the vector $\begin{pmatrix} a \\ b \end{pmatrix}$, then after rotating this it by angle $\theta$ (in anti-clockwise direction) about origin, what will be the new location of this point? In other words, you are required to express $a'$ and $b'$ in terms of $a, b$ and $\theta$.




**Exercise 4.1** Using the trigonometric ideas as above, show that

$$a' = \begin{pmatrix} \cos(\theta) \\ -\sin(\theta) \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$$
$$b' = \begin{pmatrix} \sin(\theta) \\ \cos(\theta) \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$$

Therefore, the notion of dot product helps us express rotations in a plane.

### 4.2. Length of a vector

We define the *length* of a vector $v := \begin{pmatrix} a \\ b \end{pmatrix}$ by $\ell(v) := \sqrt{v.v} = \sqrt{a^2 + b^2}$. Notice that the number $v.v$ is always positive, as $v$ varies over $\mathbb{R}^2$. Thus, dot product is useful in understanding both, angles as well as lengths.

**Exercise 4.2** Let $v \in \mathbb{R}^2$, and $c \in \mathbb{R}$. Consider the following subset of $\mathbb{R}^2$

$$\mathcal{L}_{v,c} := \{w \in \mathbb{R}^2 : v.w = c\}$$

This subset is called a *line*. Why so? We refer to the notes "Thinking Mathematically : I" for an elaboration on that.

# Lecture 5

# Matrices

September 01, 2023

### 5.1. What is a matrix?

Matrices are powerful tools for a cleaner access to geometry as well as algebra. A *matrix* is an arrangement of numbers in a rectangular array. Thus, $\begin{pmatrix} 0 & 1 \\ 3 & -1 \end{pmatrix}$ is a $2 \times 2$ matrix and $\begin{pmatrix} 1 & -1 & \frac{1}{2} \\ 3 & 2 & 0 \end{pmatrix}$ is a $2 \times 3$ matrix. Vectors $\begin{pmatrix} a \\ b \end{pmatrix}$ are $2 \times 1$ matrices.

An $m \times n$ matrix $A$ consists of $m$ rows and $n$ columns. For $1 \leq i \leq m$ and $1 \leq j \leq n$, the number located at $i^{\text{th}}$ row and $j^{\text{th}}$ column is called the $ij^{\text{th}}$ entry of $A$ and is denoted by $a_{ij}$. In short, we represent a matrix as

$$A := (a_{ij})_{m \times n},$$

and called $m \times n$ the *shape* of $A$.

So, if $m = n = 2$ and $a_{ij} := \cos\left(\dfrac{\pi}{i+j}\right)$, then

$$A = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

The collection of all $m \times n$ matrices is denoted by $M_{m \times n}(\mathbb{R})$. Since vectors in $\mathbb{R}^2$ are $2 \times 1$ matrices, we have $M_{2 \times 1}(\mathbb{R}) = \mathbb{R}^2$.

**Remark 5.1** A $2 \times 2$ matrix may be thought of as a pair of vectors $v_1, v_2 \in \mathbb{R}^2$, by identifying first column of the matrix with $v_1$ and the second column with $v_2$.
Let us have a glimpse of a matrix $A := (a_{ij})_{m \times n} \in M_{m \times n}$ together with its entries.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

In the above matrix, $i^{\text{th}}$ row vector and $j^{\text{th}}$ column vector have been highlighted. This gives us an opportunity to talk about a vector beyond $\mathbb{R}^2$. A *column vector* of $m$ coordinates is an ordered $m$-tuple of real numbers. That is, we pick any $m$ real numbers (not necessarily distinct), and arrange them in a column, remembering that the ordering of these $m$ numbers *does* matter. The collection of column vectors of $m$ coordinates is denoted $\mathbb{R}^m$. *Row vectors* of $m$ coordinates are transpose of column vectors of $m$ coordinates.

### 5.2. Addition and multiplication of matrices

We begin with stating that the addition of matrices is entrywise, while the multiplication of matrices is much intricate and deeper.

**Addition**. Matrices of same shape can be added. Let $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{m\times n}$ be two matrices of shape $m \times n$. Their *addition* is again a matrix of shape $m \times n$, whose $ij^{\text{th}}$ entry is the sum of $ij^{\text{th}}$ entries of $A$ and $B$. Therefore,

$$A + B := (a_{ij} + b_{ij})_{m\times n}.$$

**Multiplication**. We first restrict our discussion to the case of square matrices, i.e. the situation when $m = n$. Let $A = (a_{ij})_{n\times n}$ and $B = (b_{ij})_{n\times n}$. The matrix multiplication of $A$ and $B$ is the matrix whose $i^{\text{th}}$ row and $j^{\text{th}}$ column has entry $v_i^t.w_j$ (dot product). Here $v_i$ is the $i^{\text{th}}$ row vector of $A$ and $w_j$ is the $j^{\text{th}}$ column vector of $B$. In long hand,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \boxed{a_{i1} \quad a_{i2} \quad \cdots \quad a_{in}} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \qquad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & \boxed{\begin{matrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{matrix}} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & & \cdots & b_{nn} \end{pmatrix}$$

and, $v_i = (a_{i1}, a_{i2}, \cdots, a_{in})$, $w_j = \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix}$. We denote the matrix thus constructed by $AB$. The matrix $AB$ is therefore an arrangement of $n^2$ dot products arranged systematically in a $n \times n$ square grid.

**Remark 5.2** Let $AB := (c_{ij})_{n\times n}$. Then, from the definition of matrix multiplication as described above, we have $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$.

**Notation**. A shorthand notation for $a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$ is $\displaystyle\sum_{k=1}^{n} a_{ik}b_{kj}$.

**Observation 5.3** If we closely observe the matrix multiplication then we conclude that the matrix multiplication may be defined even when matrices are not square matrices. The only compatibility between $A$ and $B$ that is required to define matrix multiplication $AB$ is that the number of columns of $A$ should be equal to the number of rows of $B$. In this course we shall mostly focus on $2 \times 2$ and $3 \times 3$ matrices.

In a square matrix $A := (a_{ij})_{n\times n}$, the entries $a_{ii}$ are called *diagonal entries*. Let $\alpha \in \mathbb{R}$. Consider the $n \times n$ matrix whose diagonal entries are $\alpha$, while all nondiagonal entries are $0$. Let us denote this matrix by $[\alpha]_n$ That is,

$$[\alpha]_n = \begin{pmatrix} \alpha & 0 & \cdots & 0 \\ 0 & \alpha & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \alpha \end{pmatrix}$$

Let $A := (a_{ij})_{n\times n}$. What is the matrix product $[\alpha]_n A$? Check that it is simply the matrix whose $ij^{\text{th}}$ entry is $\alpha a_{ij}$. Thus, multiplication by $[\alpha]_n$ scales entries of $A$ by the factor of $\alpha$. Thus, $[\alpha]_n$ is an example of scalar matrix. By definition, a square matrix is called *scalar matrix* if its diagonal entries are all equal and rest of the entries are $0$.

### 5.3. Scaling

Let $\alpha \in \mathbb{R}$ and $A := (a_{ij})_{m\times n} \in M_{m\times n}(\mathbb{R})$. We define the scaling of $A$ by $\alpha$ as follows:

$$\alpha.A := [\alpha]_m A$$

Observe that $[\alpha]_m A = A[\alpha]_n$. Thus, $\alpha.A$ is the matrix whose entries are the scalings of respective entries of $A$ by $\alpha$.

### 5.4. Rotation and matrix multiplication

Recall that upon rotation by an angle $\theta$, the point $\begin{pmatrix} a \\ b \end{pmatrix}$ moves to $\begin{pmatrix} a' \\ b' \end{pmatrix}$, where

$$a' = \begin{pmatrix} \cos(\theta) \\ -\sin(\theta) \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$$
$$b' = \begin{pmatrix} \sin(\theta) \\ \cos(\theta) \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$$

This can we rewritten in terms of matrix multiplication as follows.

$$\begin{pmatrix} a' \\ b' \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Thus, a rotation amounts to carrying out multiplication by a suitable matrix. The matrix

$$R_\theta := \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

is called the *rotation matrix* for angle $\theta$.

**Observation 5.4** Observe that $R_{\theta+\phi} = R_\theta R_\phi$. Thus, the expressions for $\sin(\theta+\phi)$ and $\cos(\theta+\phi)$ can simply be attributed to the obvious fact that rotation angles add up when the rotations are carried out back to back.

# Lecture 6

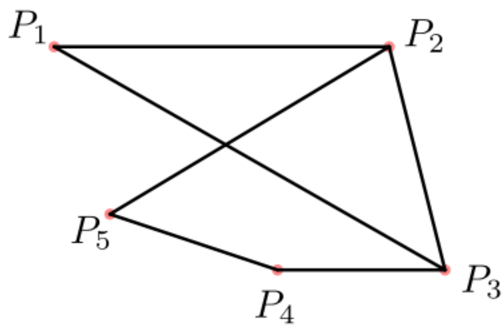
## Large and useful matrices

September 06, 2023

While the matrices may be huge - *that is to say that the number of rows and columns could be large numbers* - their occurence in practice is not clear at this point. We try to appreciate them through two examples. One concerns the networks and the other one is that of images.

### 6.1. Friends' Network and Matrices

Consider the following network of friends, where each dot represents a person, and a line connecting two dots indicates that the corresponding people are friends. Indeed, these networks are large in practice. But for the indicative purpose we are considering the following network of five people.




For each pair $i, j$ of numbers, define

$$a_{ij} := \begin{cases} 1 & \text{if } P_i \text{ and } P_j \text{ are friends} \\ 0 & \text{if } P_i \text{ and } P_j \text{ are not friends} \\ 0 & \text{if } i = j \end{cases}$$

Now, the above network can be expressed in terms of its *adjacency matrix* $A := (a_{ij})_{5\times 5}$, where $a_{ij}$ is as defined above.

**Exercise 6.1** Convince yourself that for the above network, the adjacency matrix is

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Now, let us compute the square $A^2 := AA$ for this matrix.

$$A^2 = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ \boxed{1 \;\; 1 \;\; 0 \;\; 1 \;\; 0} \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 3 & 1 & 2 & 0 \\ 1 & 1 & 3 & 0 & \boxed{2} \\ 1 & 2 & 0 & 2 & 0 \\ 1 & 0 & 2 & 0 & 2 \end{pmatrix}$$

What do the entries of $A^2$ represent, and, more importantly, why? In quest of this let us focus on $P_3$ and $P_5$. Thus, the focus of our attention is third row and fifth column of $A$ and the entry in third row and fifth column of $A^2$. Notice that this entry is

$$a_{31}a_{15} + a_{32}a_{25} + a_{33}a_{35} + a_{34}a_{55} + a_{35}a_{55}$$

Each of these five terms in this summation is either $0$ or $1$. Which of terms are $1$ is all that matters. Observe that $a_{3k}a_{k5} = 1$ if and only if $a_{3k} = 1$ and $a_{k5} = 1$. That means, $P_3$ is a friend of $P_k$, and $P_k$ is friend of $P_5$. That is to say, $k$ is a common friend of $P_3$ and $P_5$. Doesn't it suggest that the entry at $i^{\text{th}}$ row and $j^{\text{th}}$ column of $A^2$ is equal to the number of common friends of $P_i$ and $P_j$?

**Exercise 6.2** Let $A$ be the adjacency matrix of a network of $n$ friends.

(i). Let $1 \leq i \neq j \leq n$ and $A^2 = (c_{ij})_{n\times n}$. Show that $c_{ij}$ is equal to the number of common friends of $i$ and $j$.

(ii). Let $1 \leq i \leq n$ and $A^2 = (c_{ij})_{n\times n}$. What do the numbers $c_{ij}$ represent, and why?

(iii). Let $r \geq 3$. What do the entries of $A^r$ represent?

### 6.2. Grayscale images and matrices

Images can be understood in terms of matrices. An image is a rectangular array of dots (called pixels), where each pixel has a color. Let us focus on grayscale images, where each pixel is a "shade" of gray. These shades are expressed as an integer between from $0$ and $255$, where $0$ corresponds to black and $255$ corresponds to while. With that, let us look at the following $19 \times 21$ matrix, where $*$ denotes 255.

$$
A = \begin{pmatrix}
* & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * \\
* & * & * & * & * & * & * & 207 & 14 & 0 & 95 & 253 & * & * & * & * & * & * & * & * & * \\
* & * & * & * & * & * & 254 & 129 & 1 & 0 & 26 & 226 & * & * & * & * & * & * & * & * & * \\
* & * & * & * & * & * & 242 & 51 & 89 & 12 & 3 & 153 & * & * & * & * & * & * & * & * & * \\
* & * & * & * & * & * & 191 & 27 & 208 & 64 & 0 & 67 & 248 & * & * & * & * & * & * & * & * \\
* & * & * & * & * & 253 & 108 & 77 & 250 & 150 & 2 & 13 & 205 & * & * & * & * & * & * & * & * \\
* & * & * & * & * & 234 & 38 & 160 & * & 222 & 23 & 1 & 121 & 254 & * & * & * & * & * & * & * \\
* & * & * & * & * & 172 & 33 & 228 & * & 251 & 90 & 0 & 42 & 238 & * & * & * & * & * & * & * \\
* & * & * & * & 251 & 88 & 95 & 253 & * & * & 176 & 5 & 6 & 178 & * & * & * & * & * & * & * \\
* & * & * & * & 223 & 30 & 179 & * & * & * & 237 & 39 & 0 & 90 & 252 & * & * & * & * & * & * \\
* & * & * & * & 152 & 42 & 237 & * & * & * & 254 & 116 & 1 & 23 & 223 & * & * & * & * & * & * \\
* & * & * & 248 & 69 & 4 & 11 & 11 & 11 & 11 & 11 & 8 & 0 & 2 & 148 & * & * & * & * & * & * \\
* & * & * & 209 & 23 & 170 & 219 & 219 & 219 & 219 & 219 & 211 & 51 & 0 & 63 & 246 & * & * & * & * & * \\
* & * & 254 & 132 & 50 & 242 & * & * & * & * & * & 254 & 132 & 1 & 12 & 201 & * & * & * & * & * \\
* & * & 243 & 52 & 125 & 254 & * & * & * & * & * & * & 207 & 14 & 1 & 117 & 254 & * & * & * & * \\
* & * & 193 & 21 & 204 & * & * & * & * & * & * & * & 247 & 63 & 0 & 39 & 237 & * & * & * & * \\
* & 254 & 111 & 61 & 246 & * & * & * & * & * & * & * & * & 142 & 1 & 6 & 174 & * & * & * & * \\
128 & 31 & 2 & 4 & 32 & 93 & 242 & * & * & * & * & 155 & 46 & 11 & 0 & 0 & 5 & 37 & 134 & 254 & * \\
* & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * & * & *
\end{pmatrix}
$$

This matrix actually represents a grayscale image of $19 \times 21$ pixels. Each entry is actually a shade of grey at the corresponding location. Thus, Black & White images can be stored as matrices - larger the number of pixels, better the clarity of the image. The image represented by above matrix is the following.

Various transformations of images will thus correspond to certain operations on matrices.

**Think**. Let $A = (a_{ij})_{m\times n}$ be a grayscale image. Let $b_{ij}$ = integer which is closest to the average of $a_{k\ell}$ in neighbouring locations of $ij$. What are the neighbouring locations of $ij$? Well, $k\ell$ is neighbouring location of $ij$ if $|k - i| \leq 1$ and $|\ell - j| \leq 1$. The image $B = (b_{ij})_{m\times n}$ can be called the *blurring* of $A$. Think why? What do you get after blurring the identity matrix?

### 6.3. System of equations and matrices

We start with the following question.

**Question**. Can you find a vector $u \in \mathbb{R}^2$ for which

$$\begin{pmatrix} 3 & 2 \\ 1 & -1 \end{pmatrix} u = \begin{pmatrix} 21 \\ 2 \end{pmatrix} \tag{6.1}$$

Since the vector $u$ is to be determined, it is unknown to start with. We prefer to write it as a pair of two unknown real numbers.

$$v = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Thus the question concerns determining $x_1$ and $x_2$. In long hand, we rewrite (6.1) as

$$3x_1 + 2x_2 = 21 \tag{6.2}$$
$$x_1 - x_2 = 2 \tag{6.3}$$

The equations (6.2) and (6.3) together form what is called a *system of two linear equations in two variable.* Why the usage of the word *linear*? This is because each of these equations represents a line[1].

A vector $\begin{pmatrix} a \\ b \end{pmatrix}$ is to be called a *solution* of this system of equations if

$$\begin{aligned} 3a + 2b &= 21 \\ a - b &= 2 \end{aligned}$$

Thus, a solution vector $\begin{pmatrix} a \\ b \end{pmatrix}$ should lie on the intersection of the two lines represented by above equation. What if the two lines are parallel? What if the two lines are the same? These are called *inconsistent* and *degenerate* cases, respectively.

We will discuss about solving system of many linear equations in many variables later. The take home message at this point is that if $A \in M_{m\times n}(\mathbb{R})$ and $p \in \mathbb{R}^n$ then finding a $u \in \mathbb{R}^n$ such that $Au = p$ amounts to solving a system of $m$ linear equations in $n$ variables. Whether or not is the solution unique, or even existent, depends on the choice of $A$ and $p$.

We must notice that any system of linear equations boils down to a single equation $Au = p$ in terms of suitable matrices $A$ and $p$.

[1]Notice that the first line is $\mathcal{L}_{v,21}$ and the second line is $\mathcal{L}_{w,2}$, where $v = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ and $w = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Try to recall the notation $\mathcal{L}_{v,c} := \{u \in \mathbb{R}^2 : v.u = c\}$

# Lecture 7

# A bouquet of linear equations

September 14, 2023

## 7.1. Matrix notation keeps it compact

We have seen that a system of linear equations in two variables boils down to a single equation $Au = p$, for a suitable $2 \times 2$ matrix $A$, and a suitable $2 \times 1$ matrix $p$. The entries of $A$ are actually the coefficients of variables of the system of linear equations. There is nothing special about $2 \times 2$ case. Matrices are indeed capable of holding together a system of $m$ linear equations in $n$ variables.

Consider the following system of linear equations.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad \vdots \qquad \vdots \qquad &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_1 \end{aligned}$$

Rules of matrix multiplication allow us to rewrite these equations as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

This is a single matrix equation that encapsulates the above system of linear equations. Thus, if

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad u = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{and } p = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

then $Au = p$ is a compact way to express a system of linear equations. Here $A$ is a known $m \times n$ matrix and $p$ is a known $n \times 1$ matrix. The unknown here is $u$, which is to be determined. Determination of $u$ is indeed what constitutes solving the system of linear equations. How to determine such $u$? Let us pause and give some thought. How do we solve the equation $2x = 4$ in real numbers? Well, by multiplying both sides by the multiplicative inverse of $2$. So, our hope is the notion of inverse of a matrix $A$, that we wish to develop now. Our expectations from the inverse of a matrix are expressed through the following definition. We restrict ourselves to a square matrix.

### 7.2. Inverse of a matrix (if it exists)

**Definition**. Let $A \in M_{n\times n}(\mathbb{R})$ square matrix. A matrix $A' \in M_{n\times n}(\mathbb{R})$ is called an *inverse* of $A$ if $A'A = AA' = I_n$, where $I_n$ is the $n \times n$ identity matrix.

For a given $A$, why should such $A'$ exist? Simple answer is, there is no reason it should! In fact, the zero matrix (the $n \times n$ matrix whose all entries are $0$) doesn't have an inverse for the reason that *multiplication with zero matrix can never yield identity matrix*.

**Exercise 7.1** Argue that there are infinitely many $2 \times 2$ matrices which do not have inverse. What about $n \times n$ matrices for larger $n$?

A matrix is called *invertible* if it has an inverse. If $A$ is one such matrix and the matrix equation $Au = p$ is to be solved, then the solution of $Au = p$ will simply be obtained by multiplying an inverse on both sides: $A'Au = A'p$. Therefore, $u = A'p$ will solve the matrix equation, and hence the system of linear equations too.

Let us bother ourselves with the following problem. What if $A$ is an invertible matrix that has two inverse, $A'$ and $A''$? In that case, both $u = A'p$ and $u = A''p$ will be solutions of linear equations! This needs investigation.

Suppose $A'$ and $A''$ are two inverses of $A$. Then,

$$A' = A'I_n = A'(AA'') = (A'A)A'' = I_nA'' = A''$$

Aha! So, $A'$ and $A''$ will be the same. Matrix inverse, if exists, is unique. This uniqueness conveniently allows us to denote *the* inverse of an invertible matrix $A$ by $A^{-1}$.

We also note the following.

**Exercise 7.2** What definitions and properties have been used to establish above sequence of equalities?

**Remark 7.3** We note that if $Au = p$ is an equation and $B$ is an invertible square matrix whose numbers of columns is equal to the number of rows of $A$, then the system $Au = p$ has the same set of solutions as $BAu = Bp$.

Let us go back to the system of linear equations 6.2 and 6.3.

$$\begin{aligned} 3x_1 + 2x_2 &= 21 \\ x_1 - x_2 &= 2 \end{aligned}$$

This is equivalent to the matrix equation $Au = p$, where

$$A = \begin{pmatrix} 3 & 2 \\ 1 & -1 \end{pmatrix}, \quad u = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \text{and } p = \begin{pmatrix} 21 \\ 2 \end{pmatrix}$$

We verify that the matrix

$$A^{-1} = \begin{pmatrix} 1/5 & 2/5 \\ 1/5 & -3/5 \end{pmatrix}$$

is indeed the inverse of $A$ by calculating $A^{-1}A$ and $AA^{-1}$. Hence,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = u = A^{-1}p = \begin{pmatrix} 1/5 & 2/5 \\ 1/5 & -3/5 \end{pmatrix} \begin{pmatrix} 21 \\ 2 \end{pmatrix} = \begin{pmatrix} 21/5 + 4/5 \\ 21/5 - 6/5 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$$

provides us $x_1 = 5$ and $x_2 = 3$ as the solution of the system of linear equations under consideration.

Here, the matrix

$$A^{-1} = \begin{pmatrix} 1/5 & 2/5 \\ 1/5 & -3/5 \end{pmatrix}$$

has been mysteriously thrown in as the inverse of $A$. Luckily, it did work. But we should have a clear roadmap to assist us in calculating inverses. Elementary matrices play a critical role in laying down that roadmap.

### 7.3. Elementary matrices

We focus on square matrices. There are three types of elementary matrices: *swappers*, *multipliers*, and *product adders*. We begin with examples of each type in $3 \times 3$ case.

1. **Swapper matrix**.

$$S_{2,3} := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

2. **Multiplier matrix**.

$$M_2(\lambda) := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   where $\lambda$ is a nonzero element in $\mathbb{R}$.

3. **Product adder matrix**.

$$L_{1,2}(\alpha) := \begin{pmatrix} 1 & \alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   where $\alpha \in \mathbb{R}$.

The essence of these matrices lies in their interaction with other matrices. That is, if $A$ is an arbitrary $3 \times 3$ matrix, then we are curious to know how the matrix products

$$AS_{2,3} \qquad S_{2,3}A \qquad AM_2(\lambda) \qquad M_2(\lambda)A \qquad AL_{1,2}(\alpha) \qquad L_{1,2}(\alpha)A$$

look like. Let us take

$$A := \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

and carry it out.

***Multiplying elementary matrices on left***

$$S_{2,3}A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

$$M_2(\lambda)A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ \lambda a_{21} & \lambda a_{22} & \lambda a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$L_{1,2}(\alpha)A = \begin{pmatrix} 1 & \alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} + \alpha a_{21} & a_{12} + \alpha a_{22} & a_{13} + \alpha a_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Therefore, upon multiplying on left to a matrix $A$:

- Swapper matrix $S_{2,3}$ swaps rows 2 and 3 of $A$.
- Multiplier matrix $M_2(\lambda)$ multiplier row 2 of $A$ by $\lambda$.
- Product adder matrix $L_{1,2}(\alpha)$ adds scaling of row 2 of $A$ by $\alpha$ to its row 1.

Let us observe what happens when elementary matrices are multiplied on right.

***Multiplying elementary matrices on right***

$$AS_{2,3} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{13} & a_{12} \\ a_{31} & a_{23} & a_{22} \\ a_{21} & a_{33} & a_{32} \end{pmatrix}$$

$$AM_2(\lambda) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & \lambda a_{12} & a_{13} \\ a_{21} & \lambda a_{22} & a_{23} \\ a_{31} & \lambda a_{32} & a_{33} \end{pmatrix}$$

$$AL_{1,2}(\alpha) = \begin{pmatrix} a_{11} & \boxed{\begin{matrix} a_{12} \\ a_{22} \\ a_{32} \end{matrix}} & \begin{matrix} a_{13} \\ a_{23} \\ a_{33} \end{matrix} \end{pmatrix} \begin{pmatrix} 1 & \alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \begin{matrix} a_{11} \\ a_{21} \\ a_{31} \end{matrix} & \boxed{\begin{matrix} a_{12} + \alpha a_{11} \\ a_{22} + \alpha a_{21} \\ a_{32} + \alpha a_{31} \end{matrix}} & \begin{matrix} a_{13} \\ a_{23} \\ a_{33} \end{matrix} \end{pmatrix}$$

Therefore, upon multiplying on right to a matrix $A$:

- Swapper matrix $S_{2,3}$ swaps columns $2$ and $3$ of $A$.
- Multiplier matrix $M_2(\lambda)$ multiplier column $2$ of $A$ by $\lambda$.
- Product adder matrix $L_{1,2}(\alpha)$ adds scaling of column $1$ of $A$ by $\alpha$ to its column $2$.

This allows us to call multiplication by an elementary matrix to left, a *row operation*, and the multiplication by an elementary matrix to right, *column operation*.

**Observation 7.4** We make the following observation.

1. $S_{2,3}S_{2,3} = I_3$.
2. $M_2(\lambda)M_2(\lambda^{-1}) = I_3 = M_2(\lambda^{-1})M_2(\lambda)$.
3. $L_{1,2}(\alpha)L_{1,2}(-\alpha) = I_3 = L_{1,2}(-\alpha)L_{1,2}(\alpha)$.

Therefore, we conclude that these elementary matrices are invertible.

# Lecture 8

## Elementary matrices and matrix units

September 15, 2023

We have seen $3 \times 3$ elementary matrices. We now extend the notion of elementary matrices to $n \times n$ matrices.

### 8.1. $n \times n$ elementary matrices

There are three types of elementary matrices.

1. **Swapper matrices**. Let $1 \leq p \neq q \leq n$. The *swapper matrix* $S_{p,q}$ is defined by the following.
$$S_{p,q} := (a_{ij})_{n \times n}$$
where
$$a_{ij} = \begin{cases} 1 & \text{if } i = j \text{ and } i \notin \{p, q\} \\ 1 & \text{if } i = p \text{ and } j = q \\ 1 & \text{if } i = q \text{ and } j = p \\ 0 & \text{otherwise} \end{cases}$$
Observe that swapper matrix is obtained by swapping $p^{\text{th}}$ row with the $q^{\text{th}}$ row in the identity matrix $I_n$. It is not hard to check that $S_{p,q}$ does precisely what it is meant to - swap row $p$ with row $q$ when multiplied to a matrix on left, and swap column $p$ with column $q$ when multiplied to a matrix on right.

2. **Multiplier matrices**. Let $1 \leq p \leq n$ and $\lambda$ be a nonzero element of $\mathbb{R}$. The *multiplier matrix* $M_p(\lambda)$ is defined by the following.
$$M_p(\lambda) := (a_{ij})_{n \times n}$$
where
$$a_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \text{ and } i \neq p \\ \lambda & \text{if } i = j \text{ and } i = p \end{cases}$$
The multiplier matrix scales the $p^{\text{th}}$ row of a matrix by a factor of $\lambda$ when multiplied on left, and scales the $p^{\text{th}}$ column of a matrix by a factor of $\lambda$ when multiplied on right. This, again is not hard to check.

3. **Product adder matrices**. Let $1 \le p \ne q \le n$ and $\alpha$ be an element of $\mathbb{R}$. The *product adder matrix* $L_{p,q}(\alpha)$ is defined by the following.

$$L_{p,q}(\alpha) := (a_{ij})_{n\times n}$$

where

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \\ \alpha & \text{if } i = p \text{ and } j = q \\ 0 & \text{otherwise} \end{cases}$$

When multiplied on the left of a matrix $A$, the matrix $L_{p,q}$ adds the $\alpha$ times the $q^{\text{th}}$ row of $A$ to the $p^{\text{th}}$ row of $A$. Similarly, when multiplied on the right of $A$, the matrix $L_{p,q}$ adds the $\alpha$ times the $p^{\text{th}}$ column of $A$ to the $q^{\text{th}}$ column of $A$. These are routine checkings.

## 8.2. Matrix units

A matrix is called a *matrix unit* if all its entries are $0$, except one entry, and this one entry is equal to $1$. Matrix unit whose entry at $i^{th}$ row and $j^{\text{th}}$ column is equal to $1$ is denoted by $e_{ij}$. Thus, if $A = (a_{ij})_{n\times n}$ then we may rewrite it as

$$A = \sum_{1\le i,j\le n} a_{ij}e_{ij}$$

In terms of matrix units, the $n \times n$ identity matrix is written as

$$I_n = e_{11} + e_{22} + \cdots + e_{nn}$$

What is the multiplication of two matrix units? We can check that

$$e_{ij}e_{k\ell} = \begin{cases} e_{i\ell} & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}$$

This is very crucial observation. It allows us to multiply matrices without actually requiring us to focus on all matrix entries. It will be interesting the write elementary matrices in terms of matrix units.

**Exercise 8.1** Show the following.

1. $S_{p,q} = I_n + e_{pq} + e_{qp} - e_{pp} - e_{qq}$
2. $M_p(\lambda) = I_n + (\lambda - 1)e_{pp}$
3. $L_{p,q}(\alpha) = I_n + \alpha e_{pq}$

Sometimes, in calculations, this way of expressing matrices comes in handy. Let us demonstrate $M_p(\lambda)M_p(\lambda^{-1}) = I_n$ using this notation.

$$\begin{aligned}
M_p(\lambda)M_p(\lambda^{-1}) &= (I_n + (\lambda - 1)e_{pp})(I_n + (\lambda^{-1} - 1)e_{pp}) \\
&= I_n + (\lambda^{-1} - 1)e_{pp} + (\lambda - 1)e_{pp} + (\lambda - 1)(\lambda^{-1} - 1)e_{pp}e_{pp} \\
&= I_n + \left((\lambda^{-1} - 1 + \lambda - 1) + (\lambda - 1)(\lambda^{-1} - 1)\right) e_{pp} \\
&= I_n + \left(\lambda^{-1} - 1 + \lambda - 1 + \lambda\lambda^{-1} - \lambda - \lambda^{-1} + 1\right) e_{pp} \\
&= I_n
\end{aligned}$$

Along the similar lines, let us show that $L_{p,q}(\alpha)L_{p,q}(-\alpha) = I_n$

$$\begin{aligned}
L_{p,q}(\alpha)L_{p,q}(-\alpha) &= (I_n + \alpha e_{pq})(I_n - \alpha e_{pq}) \\
&= I_n - \alpha e_{pq} + \alpha e_{pq} - \alpha^2 e_{pq}e_{pq} \\
&= I_n
\end{aligned}$$

Note that if $p \neq q$, then $e_{pq}e_{pq}$ is zero matrix, and that is what we have used in the last step of above calculation.

**Exercise 8.2** Use similar ideas to show that $S_{p,q}S_{p,q} = I_n$.

This shows that every elementary matrix is invertible. Let us make a point now. We have multiplied large matrices here without using matrix entries! This is the real power of matrix unit notation.

Next, we aim at demonstrating that every invertible matrix is a product of elementary matrices. Along the way, let us introduce the notion of row echelon matrices. All this will lead us to

– solutions of systems of linear equations.

– calculation of matrix inverses.

– a systematic construction of the idea of determinants, a notion that is central to linear algebra.

### 8.3. Row echelon matrices

Let us call a matrix $A \in M_{m\times n}(\mathbb{R})$ to be a *row echelon matrix*[1] if it has the following three properties.

I. First nonzero entry in each row of $A$ is $1$. This entry is to be called the *pivot* of the row.

[1]Different books will have a variation in this definition. The various is not merely in the formulation of the definition, but in its conceptual implication too. We stick to above definition in this course.

II. The pivot of a (nonzero) row is to the right of the pivot of the preceding row. Here, by a *zero row* we mean a row whose all entries are zero. Further, if a row is zero then all the subsequent rows are also zero.

III. All entries above pivots are zero. If we read it in conjunction with II above, then we conclude that the pivot element of a row is the only nonzero element of the column it belongs to.

**Exercise 8.3** Which of the following matrices are row echelon?

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 5 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 0 & 2 & 3 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 3 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

# Lecture 9

## A road to row echelons

September 16, 2023

We begin with a question.

**Question**. Find all elementary matrices which are also row echelon matrices.

If we examine carefully, then we conclude that identity matrix is only such matrix. Our aim is to bring every matrix to a row echelon matrix after a sequence of operations. Multiplying elementary matrices *on left* of a matrix $A$ is called a *row operation* on $A$. If a matrix $A$ is not a row echelon matrix, then we can perform multiple row operations on it so that eventually we have a row echelon matrix. This is a profound idea. Let us state it in the form of a theorem.

**Theorem 9.1.** *Let $A$ be an arbitrary $m \times n$ matrix. Then there are elementary matrices $E_1, E_2, \ldots, E_r \in M_{m\times m}(\mathbb{R})$ such that $E_r E_{r-1} \ldots E_1 A$ is a row echelon matrix.*

Though we have not proved this theorem yet, let us demonstrate it through an example.

**Example**.

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix}$$

We perform successive row operations on $A$ as follows.

$$\begin{aligned}
A = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} &\xrightarrow{L_{2,1}(1)} \begin{pmatrix} 1 & -1 & 2 \\ 0 & 0 & 2 \\ 3 & 2 & 1 \end{pmatrix} \xrightarrow{L_{3,1}(-3)} \begin{pmatrix} 1 & -1 & 2 \\ 0 & 0 & 2 \\ 0 & 5 & -5 \end{pmatrix} \\
&\xrightarrow{M_3(1/5)} \begin{pmatrix} 1 & -1 & 2 \\ 0 & 0 & 2 \\ 0 & 1 & -1 \end{pmatrix} \xrightarrow{M_2(1/2)} \begin{pmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \xrightarrow{S_{2,3}} \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \\
&\xrightarrow{L_{1,2}(1)} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{L_{1,3}(-1)} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{L_{2,3}(1)} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned}$$

The following equation summarizes the entire process.

$$L_{2,3}(1)L_{1,3}(-1)L_{1,2}(1)S_{2,3}M_2\left(\frac{1}{2}\right)M_3\left(\frac{1}{5}\right)L_{3,1}(-3)L_{2,1}(1)A = I_3$$

Let us denote

$$A' := L_{2,3}(1)L_{1,3}(-1)L_{1,2}(1)S_{2,3}M_2\left(\frac{1}{2}\right)M_3\left(\frac{1}{5}\right)L_{3,1}(-3)L_{2,1}(1)$$

Then $A'A = I_3$. This makes $A'$ a good candidate to be the inverse of $A$. Remember that to ensure that $A'$ is indeed the inverse, we need to establish further that $AA' = I_3$. The following observation comes to our rescue.

**Observation 9.2** A product of elementary matrices is invertible.

*Proof.* Let $E_1, E_2, \ldots, E_r$ be elementary matrices. Consider their product

$$E := E_r E_{r-1} \ldots E_2 E_1$$

We need to show that $E$ is invertible. Since elementary matrices are invertible, $E_1^{-1}$, $E_2^{-1}$, ..., $E_r^{-1}$ makes sense. Let us look at the matrix $E' := E_1^{-1}E_2^{-1}\ldots E_{r-1}^{-1}E_r^{-1}$ obtained after multiplying these inverses (in reverse order). Then

$$\begin{aligned} E'E &= E_1^{-1}E_2^{-1}\ldots E_{r-1}^{-1}E_r^{-1}E_r E_{r-1}\ldots E_2E_1 \\ &= E_1^{-1}E_2^{-1}\ldots E_{r-1}^{-1}I_n E_{r-1}\ldots E_2E_1 \\ &= E_1^{-1}E_2^{-1}\ldots E_{r-1}^{-1}E_{r-1}\ldots E_2E_1 \end{aligned}$$

Likewise, we continue to obtain $E'E = I_n$. We now multiply in the other order.

$$\begin{aligned} EE' &= E_rE_{r-1}\ldots E_2E_1E_1^{-1}E_2^{-1}\ldots E_{r-1}^{-1}E_r^{-1} \\ &= E_rE_{r-1}\ldots E_2I_nE_2^{-1}\ldots E_{r-1}^{-1}E_r^{-1} \\ &= E_rE_{r-1}\ldots E_2E_2^{-1}\ldots E_{r-1}^{-1}E_r^{-1} \end{aligned}$$

Continuing, we have $EE' = I_n$. Therefore $E := E_rE_{r-1}\ldots E_2E_1$ is indeed invertible and its inverse is $E' = E_1^{-1}E_2^{-1}\ldots E_r^{-1}$ □

Now, let us go back to the matrix

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix}$$

Remember, we had established that for

$$A' = L_{2,3}(1)L_{1,3}(-1)L_{1,2}(1)S_{2,3}M_2\left(\frac{1}{2}\right)M_3\left(\frac{1}{5}\right)L_{3,1}(-3)L_{2,1}(1)$$

we have $A'A = I_3$

Let us carry out the following exercise.

**Exercise 9.3** The matrix $A$ is a product of elementary matrices:

$$A = L_{2,1}(-1)L_{3,1}(3)M_3(5)M_2(2)S_{2,3}L_{1,2}(-1)L_{1,3}(1)L_{2,3}(-1)$$

and hence, it is invertible by the above observation. Further,

$$A' = L_{2,3}(1)L_{1,3}(-1)L_{1,2}(1)S_{2,3}M_2\left(\frac{1}{2}\right)M_3\left(\frac{1}{5}\right)L_{3,1}(-3)L_{2,1}(1)$$

is the inverse of $A$.

# Lecture 10

## A road to row echelons : continued

September 20, 2023

Recall that we had performed row operations on the matrix

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix}$$

to obtain an identity matrix. Further, a datakeeping of the elementary matrices involved in the process of row operations enabled the calculation of the inverse of $A$.

We assert that by means of row operations every matrix can be converted into a row echelon matrix. It was stated as as Theorem 9.1. The proof of this theorem is constructive. That means, it presents an algorithm to carry out row operations on an arbitrary matrix so that at the end of the process we indeed have a row echelon matrix.

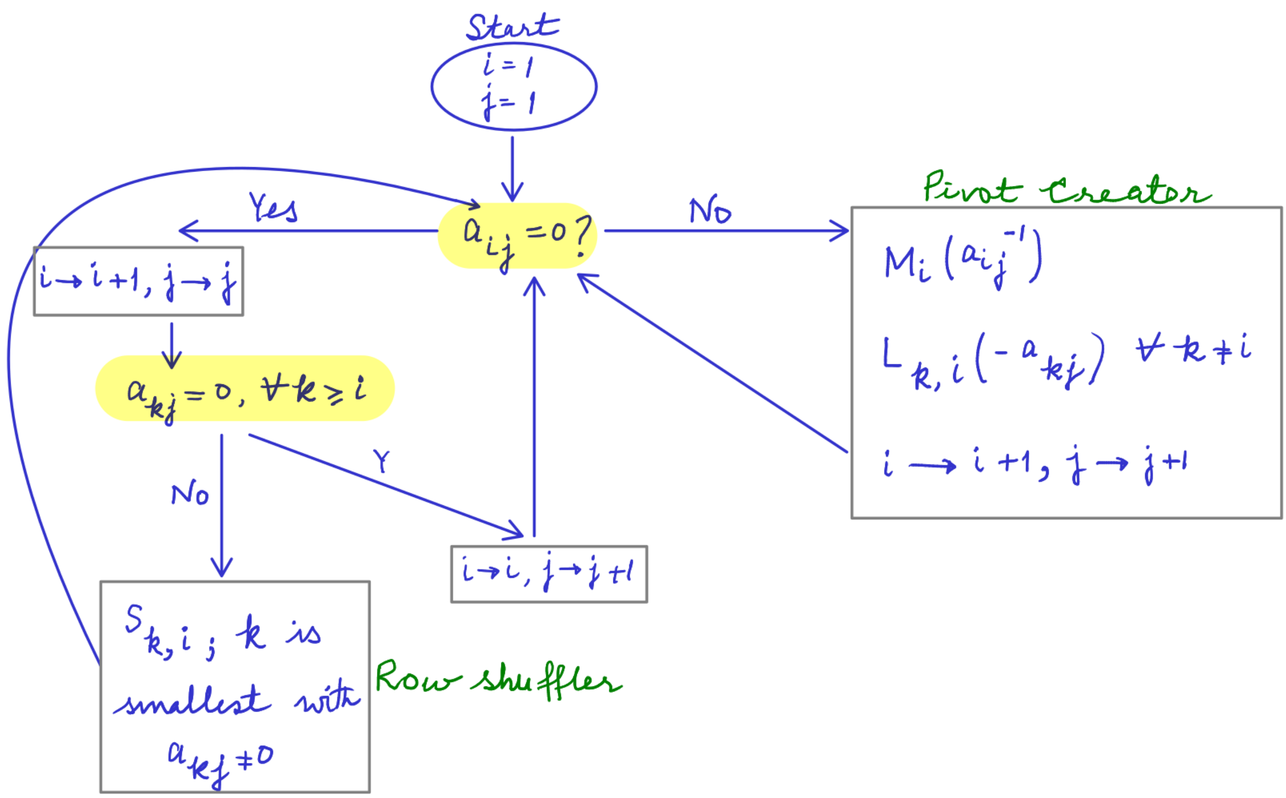


Figure 10.1: Row operations lead to a row echelon matrix

The process stops when $i$ exceeds the number of rows, or $j$ exceeds the number of columns; i.e. when traversing in search of pivots goes beyond the matrix size.

Now, we before we implement this algorithm on a matrix, let us restate the goal that we have in our mind, and make an observation.

**Goal**.

1. To simplify matrices through row operations (recall that multiplying, *on left*, by elementary matrices is called a row operation).

2. To see if a matrix is invertible.

**Observation 10.1** Let $A$ be an $m \times n$ matrix and $B$ an $m \times s$ matrix; i.e., the number of rows in $A$ and $B$ are equal. Define the matrix $(A \quad B)$ to be the $m \times (n+s)$ matrix, whose first $n$ columns are same as the $n$ columns of $A$, and last $s$ columns are same as the $s$ columns of $B$. We obsersve that if $E$ is an elementary matrix then

$$E(A \quad B) = (EA \quad EB)$$

Why? This is because elementary row operations have their *identical impact* on $A$ and $B$. Can you make it precise? Will it happen if $E$ is an arbitrary matrix?

Let us call $(A \quad B)$, the *augmentation* of $A$ by $B$. Let $I_n$ be the $n \times n$ identity matrix and $A$ be an $n \times n$ matrix. Consider the augmentation of $A$ by $I_n$. Suppose that by carrying out successive row operations on $(A \quad I_n)$ through elementary matrices $E_1, E_2, \cdots, E_r$, we attain identity matrix in first $n$ columns. That is,

$$\begin{aligned} E_rE_{r-1}\dots E_2E_1(A \quad I_n) &= (E_rE_{r-1}\dots E_2E_1A \quad E_rE_{r-1}\dots E_2E_1) \\ &= (I_n \quad E_rE_{r-1}\dots E_2E_1) \end{aligned}$$

Then,

- $E_rE_{r-1}\dots E_2E_1A = I_n$.

- The matrix $E_rE_{r-1}\dots E_2E_1$ is an invertible matrix, as it is a product of elementary matrices.

- By the exercise in the end of last lecture, $E_rE_{r-1}\dots E_2E_1$ is the inverse of $A$.

Thus, we conclude that if by carrying out successive row operations on $(A \quad I_n)$ through elementary matrices $E_1, E_2, \cdots, E_r$, we attain identity matrix in first $n$ columns, then the last $n$ columns will actually constitute the inverse of $A$. We take the matrix $A$ as in the beginning of this lecture and perform above algorithm of row operations on $(A \quad I_n)$ to obtain the inverse of $A$. The location $i, j$ is colored blue and the location $k, j$

colored red. Governed by the algorithm, we try to make $i, j$ as a pivot location and $k, j$ a location with $0$.

$$
\begin{aligned}
(A \quad I_n) &= \begin{pmatrix} \boxed{1} & -1 & 2 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{L_{2,1}(1)} \begin{pmatrix} \boxed{1} & -1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 1 & 0 \\ 3 & 2 & 1 & 0 & 0 & 1 \end{pmatrix} \\
&\xrightarrow{L_{3,1}(-3)} \begin{pmatrix} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & \boxed{0} & 2 & 1 & 1 & 0 \\ 0 & 5 & -5 & -3 & 0 & 1 \end{pmatrix} \xrightarrow{S_{2,3}} \begin{pmatrix} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & \boxed{5} & -5 & -3 & 0 & 1 \\ 0 & 0 & 2 & 1 & 1 & 0 \end{pmatrix} \\
&\xrightarrow{M_2(1/5)} \begin{pmatrix} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & \boxed{1} & -1 & -3/5 & 0 & 1/5 \\ 0 & 0 & 2 & 1 & 1 & 0 \end{pmatrix} \xrightarrow{L_{1,2}(1)} \begin{pmatrix} 1 & 0 & 1 & 2/5 & 0 & 1/5 \\ 0 & 1 & -1 & -3/5 & 0 & 1/5 \\ 0 & 0 & \boxed{2} & 1 & 1 & 0 \end{pmatrix} \\
&\xrightarrow{M_3(1/2)} \begin{pmatrix} 1 & 0 & 1 & 2/5 & 0 & 1/5 \\ 0 & 1 & -1 & -3/5 & 0 & 1/5 \\ 0 & 0 & \boxed{1} & 1/2 & 1/2 & 0 \end{pmatrix} \xrightarrow{L_{1,3}(-1)} \begin{pmatrix} 1 & 0 & 0 & -1/10 & -1/2 & 1/5 \\ 0 & 1 & -1 & -3/5 & 0 & 1/5 \\ 0 & 0 & \boxed{1} & 1/2 & 1/2 & 0 \end{pmatrix} \\
&\xrightarrow{L_{2,3}(1)} \begin{pmatrix} 1 & 0 & 0 & -1/10 & -1/2 & 1/5 \\ 0 & 1 & 0 & -1/10 & 1/2 & 1/5 \\ 0 & 0 & 1 & 1/2 & 1/2 & 0 \end{pmatrix}
\end{aligned}
$$

The preceding discussion ensures that

$$
A^{-1} = \begin{pmatrix} -1/10 & -1/2 & 1/5 \\ -1/10 & 1/2 & 1/5 \\ 1/2 & 1/2 & 0 \end{pmatrix}
$$

# Lecture 11

## Row echelon bins of $M_{m\times n}(\mathbb{R})$ : I

September 22, 2023

### 11.1. A closer look at row echelons

By now, we have attained some appreciation for elementary matrices and row echelon matrices. Let us write down all $3 \times 3$ row echelon matrices.

$$\text{No zero row} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\text{One zero row} \to \begin{pmatrix} 1 & 0 & * \\ 0 & 1 & * \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & * & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\text{Two zero rows} \to \begin{pmatrix} 1 & * & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\text{All rows zero} \to \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

So, how many $3 \times 3$ matrices are row echelon? Eight, right? Wrong! Note that matrices colored blue actually refer to infinite families of matrices, where $*$ represents arbitrary real numbers. There are eight *types* of $3$ row echelon matrices, though.

Which of these row echelon matrices are invertible? To determine that, let us make the following observation.

**Observation 11.1** If $A$ is a $3 \times 3$ row echelon matrix which is different from $I_3$ then $A$ is not an invertible matrix.

**Justification**. If $A := (a_{ij})_{3\times 3}$ is a $3 \times 3$ row echelon matrix which is different from $I_3$, then the last row of $A$ is zero. We need to show that $A$ is not invertible. We demonstrate that assuming $A$ to be invertible will give rise to a blunder, and on account of that assuming $A$ to be invertible will be wrong. That will establish that $A$ is not invertible.

So, let us assume $A$ to be invertible and look for a blunder! Under this assumption, we are bound to have an inverse of $A$. Let us call it $B := (b_{ij})_{3\times 3}$. Then,

$$AB = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

We compare the entries in the third row and third column, to obtain $0 = 1$. This is a contraction, and hence our assumption that $A$ is invertible, is wrong. Consequently, $A$ is not invertible. $\square$

Proving mathematical statements by assuming something and obtaining a contradiction, and thus concluding that our initial assumption is wrong, is very useful and frequent in mathematics. Probably the first proof that you have seen along this line of argument is that of the irrationality of $\sqrt{2}$.

Let us go back to the list of $3 \times 3$ row echelon matrices. Take two distinct row echelon matrices. Can we obtain one from the other by means of a row operation? No. This is because if we multiply a swapper matrix, multiplier matrix, or a product adder matrix to a row echelon matrix then

- it either disturbs pivot locations, and hence the resultant is not row echelon.
- if it doesn't disturb pivot locations, then it keeps the zero rows intact as well, and hence the resultant is not a *distinct* row echelon matrix.

### 11.2. Uniqueness of row echelon destinations

We now obtain the following very important theorem, which strengthens Theorem 9.1. You must understand the statement of this theorem but read the proof at your leisure. The proof is more involving and a careful compilation of many ideas that we have seen so far. In the first reading, you may skip it if you wish do. Although, giving it a try may help you consolidate many ideas that we have discussed so far.

**Theorem 11.2.** *Let $A$ be an arbitrary $m \times n$ matrix. Then there are elementary matrices $E_1, E_2, \ldots, E_r \in M_{m\times m}(\mathbb{R})$ and a **unique** row echelon $m \times n$ matrix $R$ such that $E_r E_{r-1} \ldots E_1 A = R$.*

*Proof.* Suppose $A$ can be reduced to two row echelon matrices, $R$ and $S$, by two different sequences of row operations; that is

$$R = E_r E_{r-1} \ldots E_2 E_1 A$$

and

$$S = F_s F_{s-1} \ldots F_2 F_1 A$$

where $E_i$ and $F_j$ are some elementary matrices. Then,

$$S = F_s F_{s-1} \ldots F_2 F_1 (E_r E_{r-1} \ldots E_2 E_1)^{-1} R = F_s F_{s-1} \ldots F_2 F_1 E_1^{-1} E_2^{-1} \ldots E_{r-1}^{-1} E_r^{-1} R \tag{11.1}$$

We denote the matrix $F_sF_{s-1}\dots F_2F_1E_1^{-1}E_2^{-1}\dots E_{r-1}^{-1}E_r^{-1}$ by $P$, so that $S = PR$. Note that $P$ is an invertible matrix. From this, we derive the following.

- If $R$ is the zero matrix, then so is $S$; and vice versa. Thus, if $R$ *or* $S$ is zero, then both are zero : $R = 0 = S$; and nothing remains to be proven.
- For $1 \leq j \leq n$, the $j^{\text{th}}$ column of $R$ is zero if and only if the $j^{\text{th}}$ column of $S$ is zero. This holds because $S = PR$ and $R = P^{-1}S$; hence the $j^{\text{th}}$ column of $S$ is a combination of the entries in the $j^{\text{th}}$ column of $R$, and vice versa. Consequently, the first columns of $R$ and $S$ are the same.

So, we can start with nonzero $R$ and $S$. We are required to show that $R = S$. To the contrary, we assume $R \neq S$, and look for a contradiction. Since $R \neq S$, let column number $k$ be the first column from left where mismatch occurs between $R$ and $S$. Note that $k \neq 1$, as the mismatch in the first columns would mean that one of the matrices is zero while the other one is nonzero.

We construct a matrix $R'$ from $R$ by deleting all columns after $k^{\text{th}}$ column and also deleting all columns which are not pivot columns (that is the columns which do not contain a pivot entry). We construct $S'$ from $S$ in the same manner. Then, $R'$ and $S'$ look like

$$R' = \begin{pmatrix} I_\ell & v_k \\ \mathbf{0}_{(m-\ell)\times\ell} & \mathbf{0}_{(m-\ell)\times 1} \end{pmatrix} \text{ or } \begin{pmatrix} I_\ell & v_k \\ \mathbf{0}_{1\times\ell} & 1 \\ \mathbf{0}_{(m-\ell-1)\times\ell} & \mathbf{0}_{(m-\ell-1)\times 1} \end{pmatrix}$$

and

$$S' = \begin{pmatrix} I_\ell & w_k \\ \mathbf{0}_{(m-\ell)\times\ell} & \mathbf{0}_{(m-\ell)\times 1} \end{pmatrix} \text{ or } \begin{pmatrix} I_\ell & w_k \\ \mathbf{0}_{1\times\ell} & 1 \\ \mathbf{0}_{(m-\ell-1)\times\ell} & \mathbf{0}_{(m-\ell-1)\times 1} \end{pmatrix}$$

where $\ell \geq 1$, and $\mathbf{0}$ with subscripts denotes zero matrices of appropriate sizes. Further, $v_k$ and $w_k$ are column vectors coming from the $k^{\text{th}}$ column of $R'$ and $S'$, respectively. Let $v_k = (b_1, b_2, \dots, b_\ell)^t$ and $w_k = (c_1, c_2, \dots, c_\ell)^t$.

Since, deleting corresponding columns and multiplication by elementary matrices are interchangeable processes, we indeed have $S' = PR'$. We now construct two systems of linear equations using $R'$ and $S'$, as follows.

$$R' \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_\ell \\ x_{\ell+1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \qquad S' \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_\ell \\ x_{\ell+1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \tag{11.2}$$

Since $S' = PR'$, and $P$ is an invertible matrix, the two systems have the same set of solutions (see Remark 7.3). We consider two cases.

**Case 1**. If $\lambda_1 = \lambda_2 = \cdots = \lambda_\ell = \lambda_{\ell+1} = 0$ is the only solution of these systems of equations then matrices $R'$ and $S'$ are invertible. That means, $m = \ell + 1$ and

$$R' = \begin{pmatrix} I_\ell & v_k \\ \mathbf{0}_{1\times\ell} & 1 \end{pmatrix}, \quad \text{and} \quad S' = \begin{pmatrix} I_\ell & w_k \\ \mathbf{0}_{1\times\ell} & 1 \end{pmatrix}$$

Since these are row echelon matrices, the $1$ in the last row and last column has to be a pivot entry. Which means, $v_k = 0$ and $w_k = 0$. This contradicts that the $k^{\text{th}}$ columns of $R$ and $S$ are different, making us conclude that $R = S$.

**Case 2**. If $\lambda_1, \lambda_2, \ldots, \lambda_{\ell+1}$ are solutions of equations (11.2), which are *not all zero*, that is,

$$\begin{array}{c} \lambda_1 + b_1\lambda_{\ell+1} = 0 = \lambda_1 + c_1\lambda_{\ell+1} \\ \lambda_2 + b_2\lambda_{\ell+1} = 0 = \lambda_2 + c_2\lambda_{\ell+1} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ \lambda_\ell + b_\ell\lambda_{\ell+1} = 0 = \lambda_\ell + c_\ell\lambda_{\ell+1} \end{array}$$

Then, we must have, $b_i = c_i$ for each $i$ (otherwise, each $\lambda_i = 0$). That makes $v_k = w_k$, and again we have a contradiction to the assumption that $k^{\text{th}}$ columns of $R$ and $S$ are different. Thus, in this case too we conclude that $R = S$. $\square$

In the above theorem, let us call $R$ the *row echelon destination* of $A$. The theorem asserts that an arbitrary matrix has a unique destination matrix upon implementing row operations.

Now imagine the other way round. Think of each row echelon matrix as if it is representing a bin. Arbitrary matrices are destined to fall into exactly one of the bins. Therefore, what we are imagining is partitioning of the set $M_{m\times n}(\mathbb{R})$ into bins, where each of these bins is represented uniquely by a row echelon matrix. It is a general practice in mathematics to partition a larger set into smaller disjoint subsets and study these disjoint sets one by one, so as to systematically simplify our study. It will be worth having a recollection of the notion of *relations* and *equivalence classes*.

### 11.3. Relations and equivalence classes

Consider a set $\mathcal{S}$. This set could be inspired by mathematical concepts, such as the set of all lines in a plane; or it may have its origin somewhere else, for example the set of all students in a class. A *relation* on $\mathcal{S}$ is a rule $\mathcal{R}$ through which we determine if a given *ordered* pair in the set $\mathcal{S}$ follows that rule. Consider the following examples.

**Example 11.3**

1. If the set $\mathcal{S}$ is the set of all humans and the rule $\mathcal{R}$ is *is the mother of*, then we will call this rule a relation because it allows is to determine if given an ordered pair

$(a, b)$ of two humans $a$ and $b$, the person $a$ is the mother of $b$ or not. Note that $a$ *is the mother of* $b$ is different from saying $b$ *is the mother of* $a$ - the order matters. That's why we are emphasizing on the ordered pair of elements and not just a pair of elements. If $a$ is the mother of $b$, then we write $a\mathcal{R}b$ to express it. The same set $\mathcal{S}$ may have different relations. For example on this set $\mathcal{S}$ of all humans, the rule $\mathcal{R}$ may refer to *has same birthday as*. Incidentally, in this case $a\mathcal{R}b$ holds if and only if $b\mathcal{R}a$ holds.

2. Let $\mathcal{S}$ be the set of lines in a plane. Let $\mathcal{R}$ refers to *is perpendicular to*. It is a relation on $\mathcal{S}$. Given two lines $\ell_1$ and $\ell_2$ in a plane, we can determine if $\ell_1$ is perpendicular to $\ell_2$. In this case, note that $\ell_1\mathcal{R}\ell_2$ holds if and only if $\ell_2\mathcal{R}\ell_1$ holds. That is, $\ell_1$ is perpendicular to $\ell_2$ if and only if $\ell_2$ is perpendicular to $\ell_1$.

3. Let $\mathcal{S}$ be the set of all natural numbers and $m\mathcal{R}n$ refers to $m - n$ *is divisible by* 2. It is a relation.

4. Let $\mathcal{S}$ denote a plane with a specified origin $\mathcal{O}$. Consider the relation $\mathcal{R}$, where $P_1\mathcal{R}P_2$ holds if and only if $P_1$ *can reach the current location of* $P_2$ *after a rotation about the origin*. Here, it is true that $P_1\mathcal{R}P_2$ if and only if $P_2\mathcal{R}P_1$. What about $P_1\mathcal{R}P_1$?

Note that when we pick an ordered pair $(a, b)$, the element $b$ is not required to be different from the element $a$. A relation $\mathcal{R}$ on a set $\mathcal{S}$ may have the following properties.

I. **Reflexivity**. A relation $\mathcal{R}$ is said to have *reflexivity* property, or that $\mathcal{R}$ is *reflexive*, if for every $a \in \mathcal{S}$, the $a\mathcal{R}a$ holds. That is, $a$ should be related to itself. In the above examples, the relation described in the example 1 is not reflexive - no one is the mother of itself. Is there any instance of reflexive relations in these examples?

II. **Symmetry**. A relation $\mathcal{R}$ is said to have *symmetry* property, or that $\mathcal{R}$ is *symmetric*, if for every $a, b \in \mathcal{S}$, the $a\mathcal{R}b$ holds if and only if $b\mathcal{R}a$ holds. We have seen that the example 2 above describes a symmetric relation.

III. **Transitivity**. A relation $\mathcal{R}$ is said to have *transitivity* property, or that $\mathcal{R}$ is *transitive*, if for any arbitrary triplet $a, b, c \in \mathcal{S}$, whenever it is observed that $a\mathcal{R}b$ and $b\mathcal{R}c$, then $a\mathcal{R}c$ must also be observed. In the example 2 above, if we take three lines $\ell_1, \ell_2, \ell_3$, where we observe that $\ell_1$ is perpendicular to $\ell_2$, and $\ell_2$ is perpendicular to $\ell_3$, then do we also observe that $\ell_1$ is perpendicular to $\ell_3$? No! So, the relation described in this example is not transitive.

The following table summarizes what properties hold for our examples.

| **Example** | **Reflexive** | **Symmetric** | **Transitive** |
|---:|:---:|:---:|:---:|
| Example 1 (is mother of) | $\times$ | $\times$ | $\times$ |
| Example 2 (is perpendicular to) | $\times$ | ✓ | $\times$ |
| Example 3 (2 divides $m - n$) | ✓ | ✓ | ✓ |
| Example 4 (rotation) | ✓ | ✓ | ✓ |

The example that motivated us to initiate the discussion on relations comes from matrices and the process to convert them into row echelon matrices. We shall elaborate upon that.

# Lecture 12

## Row echelon bins of $M_{m\times n}(\mathbb{R})$ : II

September 29, 2023

Let us consolidate our recent learning.

- Any matrix can be brought into row a echelon matrix by successively applying row operations.
- The row echelon matrix thus obtained is unique.
- If the initial matrix is $n \times n$ square matrix, then it is invertible if and only if the row echelon matrix obtained after row operations is the identity matrix $I_n$.

Now, we define two relations $\mathcal{R}_1$ and $\mathcal{R}_2$ on the set $\mathcal{S} := M_{m\times n}(\mathbb{R})$, as follows.

$\mathcal{R}_1$. For two matrices $A, B \in \mathcal{S}$, let us call $A$ to be a *close friend* of $B$ (written $A\mathcal{R}_1B$) if $A = EB$ for some $m \times m$ elementary matrix $E$.

$\mathcal{R}_2$. For two matrices $A, B \in \mathcal{S}$, let us call $A$ to be a *friend* of $B$ (written $A\mathcal{R}_2B$) if $A = EB$ for some $m \times m$ invertible matrix $E$.

These relations fit the following table

| **Relation** | **Reflexive** | **Symmetric** | **Transitive** |
|---|---|---|---|
| $R_1$ (is close friend of) | ✓ | ✓ | $\times$ |
| $R_2$ (is friend of) | ✓ | ✓ | ✓ |

Thus, $R_2$ is an equivalence relation but $R_1$ is not.

**Exercise 12.1** Let $A, B \in \mathcal{S}$. Then $A\mathcal{R}_2B$ if and only if $A$ and $B$ have the same row echelon destination.

The equivalence relation $\mathcal{R}_2$ allows us to partition the set $M_{m\times n}(\mathbb{R})$ into disjoint subsets. Each of these subsets is represented by a row echelon matrix. Let us call these subset *row echelon bins* of $M_{m\times n}(\mathbb{R})$. The advantages of classifying matrices into row echelon bins is that it allows us to focus on these simple looking matrices, without losing anything. In particular, solving system of linear equations becomes much faster and handy. This difference becomes more evident when the matrix size grows.

# Lecture 13

## System of linear equations : revisited

October 04, 2023

We have talked about systems of linear equations and their matrix forms in §3 of Lecture 6 and Lecture 7. We will revisit it now through some examples. Solving a linear equation is a two-step process.

Step 1. **Converting the given system into an equivalent but simpler system**. By equivalent system we mean that the set of solutions of the system, after simplification, does not change. The following constitutes the simplification process.

- Expressing the system of equations into a matrix form, and then writing it as an augmented matrix.
- Converting the augmented matrix into a row echelon form, keeping in mind that elementary matrices are invertible, and multiplication by invertible matrices leads to an equivalent system of equations (Remark 7.3).

Step 2. **Inspecting the simpler system and obtaining the solutions of the simpler system**.

We begin with the following example.

**Example 13.1** Solve the following system of linear equations.

$$\begin{aligned} x + y - z &= 4 \\ 3x + z &= 8 \end{aligned}$$

The matrix form of this system of equations is

$$\begin{pmatrix} 1 & 1 & -1 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

Let us denote

$$A := \begin{pmatrix} 1 & 1 & -1 \\ 3 & 0 & 1 \end{pmatrix}, \quad p := \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

Consider the $2 \times 4$ augmented matrix $(A \quad p)$ and apply row operations on it.

$$(A \quad p) := \begin{pmatrix} 1 & 1 & -1 & 4 \\ 3 & 0 & -1 & 8 \end{pmatrix} \xrightarrow{L_{2,1}(-3)} \begin{pmatrix} 1 & 1 & -1 & 4 \\ 0 & -3 & 4 & -4 \end{pmatrix}$$
$$\xrightarrow{M_2(-\frac{1}{3})} \begin{pmatrix} 1 & 1 & -1 & 4 \\ 0 & 1 & -\frac{4}{3} & \frac{4}{3} \end{pmatrix} \xrightarrow{L_{1,2}(-1)} \begin{pmatrix} 1 & 0 & \frac{1}{3} & \frac{8}{3} \\ 0 & 1 & -\frac{4}{3} & \frac{4}{3} \end{pmatrix}$$

The final matrix is indeed in the row echelon form and we interpret it as the following system of equations, which ought to be equivalent to the original system of equations.

$$\begin{aligned} x + \frac{z}{3} &= \frac{8}{3} \\ y - \frac{4z}{3} &= \frac{4}{3} \end{aligned}$$

It seems to have many solutions. Let us examine. If $z = 2$, then $x = 2$ and $y = 4$. That gives a solution. If $z = 5$, then $x = 1$ and $y = 8$ gives yet another solution. In fact, for any choice of $z$, say $z = t$, where $t$ is an arbitrary real number, $x = \frac{8-t}{3}$ and $y = \frac{4(1+t)}{3}$ provides a solution of the given system of equations. In other words,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{8-t}{3} \\ \frac{4(1+t)}{3} \\ t \end{pmatrix}$$

is the entire set of solutions, as $t$ varies over real numbers. Thus, we see that the given system has infinitely many solutions which are *parameterized* by $t$.

We now take another example.

**Example 13.2** Solve the following system of linear equations.

$$\begin{aligned} x + 2y - z &= 4 \\ 3x + y + z &= 7 \\ x - 3y + 3z &= 1 \end{aligned}$$

The matrix form of this system is

$$\begin{pmatrix} 1 & 2 & -1 \\ 3 & 1 & 1 \\ 1 & -3 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \\ 1 \end{pmatrix}$$

We apply row operations on the corresponding $3 \times 4$ augmented matrix

$$\begin{pmatrix} 1 & 2 & -1 & 4 \\ 3 & 1 & 1 & 7 \\ 1 & -3 & 3 & 1 \end{pmatrix} \xrightarrow{L_{2,1}(-3)} \begin{pmatrix} 1 & 2 & -1 & 4 \\ 0 & -5 & 4 & -5 \\ 1 & -3 & 3 & 1 \end{pmatrix}$$
$$\xrightarrow{L_{3,1}(-1)} \begin{pmatrix} 1 & 2 & -1 & 4 \\ 0 & -5 & 4 & -5 \\ 0 & -5 & 4 & -3 \end{pmatrix} \xrightarrow{M_2\left(-\frac{1}{5}\right)} \begin{pmatrix} 1 & 2 & -1 & 4 \\ 0 & 1 & -\frac{4}{5} & 1 \\ 0 & -5 & 4 & -3 \end{pmatrix}$$
$$\xrightarrow{L_{1,2}(-2)} \begin{pmatrix} 1 & 0 & \frac{3}{5} & 2 \\ 0 & 1 & -\frac{4}{5} & 1 \\ 0 & -5 & 4 & -3 \end{pmatrix} \xrightarrow{L_{3,2}(5)} \begin{pmatrix} 1 & 0 & \frac{3}{5} & 2 \\ 0 & 1 & -\frac{4}{5} & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$
$$\xrightarrow{M_3\left(\frac{1}{2}\right)} \begin{pmatrix} 1 & 0 & \frac{3}{5} & 2 \\ 0 & 1 & -\frac{4}{5} & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{L_{1,3}(-2)} \begin{pmatrix} 1 & 0 & \frac{3}{5} & 0 \\ 0 & 1 & -\frac{4}{5} & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$
$$\xrightarrow{L_{2,3}(-1)} \begin{pmatrix} 1 & 0 & \frac{3}{5} & 0 \\ 0 & 1 & -\frac{4}{5} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Again, the final matrix is a row echelon matrix. We rewrite is in terms of a system of linear equations as follows.

$$\begin{aligned} x + 0y + \frac{3z}{5} &= 0 \\ 0x + y - \frac{4z}{5} &= 0 \\ 0x + 0y + 0z &= 1 \end{aligned}$$

The last equation in this system is contradictory, as it leads to $0 = 1$. This contraception suggests that the given system of equation has no solution, or that the system is *inconsistent*.

We examine such examples closely and arrive at the following.

**Theorem 13.3.** *A system of linear equations is represented by the matrix equation $Au = p$. Then exactly one of the following possibility occurs.*

1. *The system has no solution.*
2. *The system has exactly one solution.*
3. *The system has infinitely many solutions.*

Proof of this theorem will be presented in these notes shortly.

# Lecture 14

## $A \rightsquigarrow d(A)$ : The three rules

October 06, 2023

We start by assigning ourselves a task. The motive of doing so will become apparent at a later stage. All matrices in this lecture are $n \times n$ square matrices. Let us fix some notation before we proceed.

$$e_i := i^{\text{th}} \text{ row of the identity matrix } I_n$$

Thus, $e_i = (0, 0, \ldots, 1, \ldots, 0, 0)$, where the sole $1$ in the $n$-tuple $e_i$ is at the $i^{\text{th}}$ location. We then write the identity matrix as follows.

$$I_n = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Along these lines, the swapper matrix $S_{p,q}$, where $p < q$, is written as

$$S_{p,q} = \begin{pmatrix} e_1 \\ \vdots \\ e_q \\ \vdots \\ e_p \\ \vdots \\ e_n \end{pmatrix}$$

We introduce one more notation, for $p \neq q$

$$R_{p,q} := \begin{pmatrix} e_1 \\ \vdots \\ e_q \\ \vdots \\ e_q \\ \vdots \\ e_n \end{pmatrix}$$

to denote the matrix whose all rows except $p^{\text{th}}$ row are same as the rows of the identity matrix, and the $p^{\text{th}}$ row is equal to $e_q$. Note that $R_{p,q} \neq R_{q,p}$.

### 14.1. The partial row sum

Let $1 \le i \le n$. Let $A$ and $B$ be $n \times n$ matrices, whose all rows *except possibly* the $i^{\text{th}}$ row, are equal. Let us call such pairs of matrices $i^{\text{th}}$ *row sum compatible matrices*. For such matrices, we define the operation $+_i$ as follows: $A +_i B$ is the matrix whose $i^{\text{th}}$ row is the sum of $i^{\text{th}}$ rows of $A$ and $B$, and rest of the rows are the same as the rows of $A$ (or $B$, remember that these rows are the same for $A$ and $B$). The operation $+_i$ is called the *partial row sum* with respect to $i^{\text{th}}$ row. Following are some illustrations of this operation.

1. $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} +_1 \begin{pmatrix} 0 & \alpha \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$
2. More generally, $L_{p,q}(\alpha) = I_n +_p M_p(\alpha) R_{p,q}$.
3. $I_n +_p I_n = M_p(2) I_n$.
4. $(I_n +_p R_{p,q}) +_q (S_{p,q} +_p R_{q,p}) = \begin{pmatrix} e_1 \\ \vdots \\ e_p + e_q \\ \vdots \\ e_p + e_q \\ \vdots \\ e_n \end{pmatrix}$.
5. $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} +_2 \begin{pmatrix} -1 & -2 & -3 \\ -4 & -5 & -6 \\ -7 & -8 & -9 \end{pmatrix}$ is not a valid operation. This is because there are rows, *apart from the second row*, of these matrices which do not match. In other words, $A$ and $B$ are not second row sum compatible matrices.
6. $A +_i B = B +_i A$, whenever $A$ and $B$ are $i^{\text{th}}$ row sum compatible matrices.

### 14.2. The three rules

**The task**. To each $n \times n$ square matrix $A$, we wish to associate a number $d(A)$ in a consistent manner that is determined by the following rules.

(I). $d(I_n) = 1$.

(II). $d$ is *linear in each row*. This means the following

II(a). **(Row additivity)**

$$d(A +_i B) = d(A) + d(B)$$

for each row $i$, and for every pair $A, B$ of $i^{\text{th}}$ row sum compatible matrices.

II(b). **(Scaling)**

$$d(M_i(\lambda)A) = \lambda d(A)$$

for each row $i$ and for every $n \times n$ matrix $A$.

(III). If two rows of $A$ are equal, then $d(A) = 0$.

The first and the foremost question - *do these rules actually allow us to consistently define $d(A)$ for every $A \in M_{n\times n}(\mathbb{R})$*? And, if there is indeed a consistent way of following these rules, then we have our second question ready - *is it possible that following these rules in two different sequences, we are lead to two different number for some matrix $A$*?

Thus, we are raising a doubt on the existence as well as the uniqueness of the assignment $A \rightsquigarrow d(A)$. What follows now, ***assumes*** the existence of an assignment $A \rightsquigarrow d(A)$ that follows the three rules. **We *fix one such assignment* and denote it as** $A \rightsquigarrow D(A)$.

While we will venture into the questions of existence and uniqueness later, let us try to follow these rules for the assignment $A \rightsquigarrow D(A)$, that we have fixed.

**Example 14.1** Consider the matrix $A := \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$. Then

$$\begin{aligned} D(A) &= D\left(\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} +_1 \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\right) \\ &= D\left(M_1(2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} +_1 \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\right) \\ &= 2D\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) + D\left(\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\right) \\ &= 2D(I_2) + D\left(\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\right) \\ &= (2 \times 1) + 0 = 2 \end{aligned}$$

**Example 14.2** For the matrix $R_{p,q}$, we have $D(R_{p,q}) = 0$. This follows straightforwardly from Rule III, as the rows $p$ and $q$ of $R_{p,q}$ are identical.

In the next three examples, we apply the three rules for elementary matrices.

**Example 14.3** For $E = M_p(\lambda)$, a multiplier matrix, we have $D(E) = \lambda$. This comes directly from II(b) by putting $A = I_n$.

**Example 14.4** For a product adder matrix $E = L_{p,q}(\alpha)$, we recall from §14.1 that

$$L_{p,q}(\alpha) = I_n +_p M_p(\alpha)R_{p,q}$$

Therefore,

$$\begin{aligned} D(L_{p,q}(\alpha)) &\overset{\text{R II(a)}}{=} D(I_n) + D(M_p(\alpha)R_{p,q}) \\ &\overset{\text{R II(b)}}{=} D(I_n) + \alpha D(R_{p,q}) \\ &\overset{\text{R I, III}}{=} 1 + (\alpha \times 0) = 1 \end{aligned}$$

In the last step, we have used that the matrix $R_{p,q}$ has two identical rows.

**Example 14.5** We now consider the swapper matrix $S_{p,q}$. Recall that

$$(I_n +_p R_{p,q}) +_q (S_{p,q} +_p R_{q,p}) = \begin{pmatrix} e_1 \\ \vdots \\ e_p + e_q \\ \vdots \\ e_p + e_q \\ \vdots \\ e_n \end{pmatrix}$$

Since the matrix on the right hand side has repeated rows,

$$\begin{aligned} 0 &= D((I_n +_p R_{p,q}) +_q (S_{p,q} +_p R_{q,p})) \\ &= D(I_n +_p R_{p,q}) + D(S_{p,q} +_p R_{q,p}) \\ &= D(I_n) + D(R_{p,q}) + D(S_{p,q}) + D(R_{q,p}) \\ &= 1 + 0 + D(S_{p,q}) + 0 \end{aligned}$$

And it follows that $D(S_{p,q}) = -1$. These were indeed crucial examples. How do row operations affect the association $A \rightsquigarrow D(A)$? The following proposition provides an answer.

**Proposition 14.6.** *For an arbitrary $n \times n$ matrix $A$, we have the following*

*(i).* $D(M_p(\lambda)A) = \lambda D(A)$. *This is just Rule II(b).*

*(ii).* $D(S_{p,q}A) = -D(A)$.

*(iii).* $D(L_{p,q}(\alpha)A) = D(A)$.

*Therefore, $D(EA) = D(E)D(A)$, for every $n \times n$ elementary matrix $E$.*

*Proof.* We prove this proposition essentially by imitating the idea used while working out above examples. The following observations would lead to the proof: *Whenever $B$ and $C$ are $i^{\text{th}}$ row sum compatible matrices for some $i$, and $A$ is any $n \times n$ matrix, then $BA$ and $CA$ are also $i^{\text{th}}$ row sum compatible matrices. Further, $(B +_i C)A = BA +_i CA$.*

We then consider the identities

$$L_{p,q}(\alpha) = I_n +_p M_p(\alpha) R_{p,q}$$

and

$$(I_n +_p R_{p,q}) +_q (S_{p,q} +_p R_{q,p}) = \text{ a matrix with repeated rows}$$

Multiplying these identities by $A$ on right, and making use of the above observation we obtain

$$L_{p,q}(\alpha)A = A +_p M_p(\alpha) R_{p,q} A,$$

and

$$(A +_p R_{p,q}A) +_q (S_{p,q}A +_p R_{q,p}A) = \text{ a matrix with repeated rows}$$

We now systematically apply Rules I, II, II to complete the proof. □

# Lecture 15

## $A \rightsquigarrow d(A)$ : The three rules continued

October 11, 2023

We assume that there is indeed an association $A \rightsquigarrow D(A)$ that satisfies the three rules. Consider the matrix $A := S_{2,3}M_2(\lambda)L_{1,2}(\alpha) \in M_{3\times 3}(\mathbb{R})$. What would be $D(A)$? We applying Proposition 14.6 to obtain

$$D(A) = D(S_{2,3}M_2(\lambda)L_{1,2}(\alpha)) = -D(M_2(\lambda)L_{1,2}(\alpha)) = -\lambda D(L_{1,2}(\alpha)) = -\lambda$$

We extend this line of thought to derive the following from Proposition 14.6.

**Proposition 15.1.** *Let $A := E_n E_{n-1} \cdots E_2 E_1$, where $E_i$ are elementary matrices. Then,*

$$D(A) = (-1)^r \prod_{i=1}^{s} (\lambda_i)$$

*where $r$ denotes the number of swapper matrices among $E_i$, and $M_{i_1}(\lambda_{i_1}), \cdots, M_{i_s}(\lambda_{i_s})$ are all multiplier matrices among $E_i$.*

This proposition takes care of invertible matrices. What about those which are not invertible? Let us start with a matrix $A$ that contains a row which is entirely zero.

**Proposition 15.2.** *Let $A$ be a square $n \times n$ matrix whose some row is zero. Then, $D(A) = 0$.*

*Proof.* If $A$ is the zero matrix then $D(A) = 0$ by Rule III. So, we may assume that $A$ is not the zero matrix. Let $p^{\text{th}}$ row of $A$ is nonzero, and $q^{\text{th}}$ row is zero. Then, $A = L_{q,p}(-1)L_{q,p}(1)A$, and therefore

$$\begin{aligned} D(A) &= D(L_{q,p}(-1)L_{q,p}(1)A) \\ &= D(L_{q,p}(1)A) = 0 \end{aligned}$$

Here, the last equality holds because $p^{\text{th}}$ row of $L_{q,p}(1)A$ is same as its $q^{\text{th}}$ row. $\square$

As a consequence, the following remarkable theorem holds.

**Theorem 15.3.** *A square matrix is invertible if and only if $D(A) \neq 0$.*

*Proof.* Let $A$ be a square matrix. We write $R = E_n E_{n-1} \cdots E_2 E_1 A$, where $E_i$ are elementary matrices and $R$ is a row echelon matrix. By repeatedly applying Proposition 14.6, $D(R) = D(E_n)D(E_{n-1}) \cdots D(E_2)D(E_1)D(A)$. Since each $D(E_i) \neq 0$, we have that $D(A) \neq 0$ if and only if $D(R) \neq 0$.

Now, $D(R) \neq 0$ if and only if $R = I_n$. This is because if $R \neq I_n$, then $R$ will be a non-identity row echelon matrix whose last row is zero, and hence $D(R) = 0$. The case $R = I_n$ is precisely the case when $A$ is invertible. $\square$

# Lecture 16

## $A \rightsquigarrow d(A)$ : further properties

October 12, 2023

We continue with the assumption that there is indeed an assignment $A \rightsquigarrow d(A)$ that obeys the three rules. We denote one such assignment by $A \rightsquigarrow D(A)$.

### 16.1. Multiplicativity of $D$ and its invariance under transpose

**Theorem 16.1.** *Let $A$ and $B$ two square matrices. Then $D(AB) = D(A)D(B)$.*

*Proof.* We consider the two cases.

**Case 1.** *$A$ is invertible.* In this case $A = E_k E_{k-1} \dots E_2 E_1$, where $E_i$ are elementary matrices. By successively applying Proposition 14.6, we have

$$D(A) = D(E_k E_{k-1} \dots E_2 E_1) = D(E_k) D(E_{k-1}) \dots D(E_2) D(E_1)$$

and

$$D(AB) = D(E_k E_{k-1} \dots E_2 E_1 B) = D(E_k) D(E_{k-1}) \dots D(E_2) D(E_1) D(B)$$

Therefore $D(AB) = D(A)D(B)$

**Case 2.** *$A$ is not invertible.* By Theorem 15.3, $D(A) = 0$. Thus, $D(A)D(B) = 0$. We now show that $D(AB) = 0$. Since $A$ is not invertible, we have elementary matrices $E_1, E_2, \dots, E_k$ and a row echelon matrix $R$ whose last row is zero, such that $R = E_k E_{k-1} \dots E_2 E_1 A$. Thus, $RB = E_k E_{k-1} \dots E_2 E_1 AB$. Notice that the last row of $RB$ is zero. By Proposition **??**,

$$0 = D(RB) = D(E_k E_{k-1} \dots E_2 E_1 AB)$$

Thus,

$$\begin{aligned} 0 &= D(E_k E_{k-1} \dots E_2 E_1 AB) \\ &= D(E_k) \dots D(E_1) D(AB) \end{aligned}$$

Since each $D(E_i) \neq 0$, we conclude that $D(AB) = 0$. □

This is a remarkable theorem, and it is reworded as "$D$ is multiplicative". The following is an easy exercise.

**Exercise 16.2** If $A$ and $B$ are two $n \times n$ matrices, then $(AB)^t = B^t A^t$. Here $A^t$ is the transpose of $A$; $B^t$ is the transpose of $B$; and, $(AB)^t$ is the transpose of $AB$.

**Exercise 16.3** Let $E$ be an elementary matrix. Then, $D(E^t) = D(E)$.

The above exercise can be done by observing that $S_{p,q}^t = S_{p,q}$, $M_p(\lambda)^t = M_p(\lambda)$, and $L_{p,q}(\alpha)^t = L_{q,p}(\alpha)$.

**Exercise 16.4** Let $R$ be a row echelon $n \times n$ matrix. Then, $D(R^t) = D(R)$.

Here is a solution. If $R = I_n$, then $D(R^t) = D(R) = 1$ is evident. Otherwise, the last row of $R$ is zero, and so is the last column of $R^t$. It is easy to show, along the lines of the justification of Observation 11.1, that $R$ and $R^t$ are not invertible. Now, by Theorem 15.3, $D(R^t) = D(R) = 0$.

Combining the three exercises, we obtain the following.

**Proposition 16.5.** *Let $A$ be a square matrix. Then $D(A^t) = D(A)$.*

*Proof.* We write $R = E_k E_{k-1} \dots E_2 E_1 A$, where $E_k$ are suitable elementary matrices and $R$ is a row echelon matrix. Then, by Exercise 16.2, $R^t = A^t E_1^t \cdots E_k^t$. Now, $D(R^t) = D(R)$ (Exercise 16.4) gives,

$$D(A^t E_1^t \cdots E_k^t) = D(E_k E_{k-1} \dots E_2 E_1 A)$$

We now use Theorem 16.1 and Exercise 16.3, and the fact the $D(E) \neq 0$, if $E$ is an elementary matrix, to obtain $D(A^t) = D(A)$. $\square$

### 16.2. $D$ for $3 \times 3$ matrices

We start with the following $3 \times 3$ matrix.

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{pmatrix}$$

We note that first row and first column will not be involved in converting $A$ into a row echelon matrix. Thus, converting $A$ into a row echelon matrix would constitute similar sequence of elementary matrices as it would take to convert the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ into a row echelon matrix. Therefore, $D(A) = D\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = ad - bc.$

Now, we modify the matrix a bit.

$$A = \begin{pmatrix} 1 & 0 & 0 \\ r & a & b \\ s & c & d \end{pmatrix}$$

We note that

$$L_{3,1}(-s)L_{2,1}(-r)\begin{pmatrix} 1 & 0 & 0 \\ r & a & b \\ s & c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{pmatrix}$$

We use Proposition 15.1, and the fact that $D(L_{2,1}(-r)) = 1 = D(L_{3,1}(-3))$, to obtain
$D(A) = D\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{pmatrix}\right) = ad - bc.$

Consider the following matrices.

$$AS_{1,2} = \begin{pmatrix} 0 & 1 & 0 \\ a & r & b \\ c & s & d \end{pmatrix}, \quad AS_{1,2}S_{2,3} = \begin{pmatrix} 0 & 0 & 1 \\ a & b & r \\ c & d & s \end{pmatrix}$$

Thanks to Theorem 16.1 that we have

$$D\left(\begin{pmatrix} 0 & 1 & 0 \\ a & r & b \\ c & s & d \end{pmatrix}\right) = D(AS_{1,2}) = D(A)D(S_{1,2}) = -D(A) = -(ad - bc)$$

and

$$D\left(\begin{pmatrix} 0 & 0 & 1 \\ a & b & r \\ c & d & s \end{pmatrix}\right) = D(AS_{1,2}S_{1,3}) = D(A)D(S_{1,2})D(S_{1,3}) = D(A) = ad - bc$$

We compile it carefully and use Rules I, II, III to obtain the following.

**Theorem 16.6.** *Let $A \rightsquigarrow D(A)$ be an assignment that follows Rules I, II, III. Then for an arbitrary $3 \times 3$ matrix*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

*we are bound to have*

$$D(A) = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

*Proof.* Let us denote

$$A_1 := \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad A_2 := \begin{pmatrix} 0 & a_{12} & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad A_3 := \begin{pmatrix} 0 & 0 & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

and write $A$ as the following partial row sum of matrices with respect to the first row.

$$A := A_1 +_1 A_2 +_1 A_3,$$

Therefore, by Rule II(a), $D(A) = D(A_1) + D(A_2) + D(A_3)$. Let us calculate $D(A_1), D(A_2)$ and $D(A_3)$.

First we look at $D(A_1)$. If $a_{11} = 0$, then the first row of $A_1$ is zero and by Proposition 15.2, $D(A_1) = 0 = a_{11}(a_{22}a_{33} - a_{23}a_{32})$. So, we assume that $a_{11} \neq 0$, and write

$$A_1 = M_1(a_{11}) \begin{pmatrix} 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

By Rule II(b), $D(A_1) = a_{11} D \left( \begin{pmatrix} 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \right) = a_{11}(a_{22}a_{33} - a_{23}a_{32})$

Thus, in both cases ($a_{11} = 0$ and $a_{11} \neq 0$), we have $D(A_1) = a_{11}(a_{22}a_{33} - a_{23}a_{32})$.

Calculations for $D(A_2)$ and $D(A_3)$ are similar and lead to

$$D(A) = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

□

We see that in case of $2 \times 2$ and $3 \times 3$ matrices, the expression for $D(A)$ depends *only on* the entries of $A$. This expression is the *unique outcome* of Rules I, II, III, and it indeed follows these three rules. For matrices of larger size similar conclusions can be made using induction. We summarize all our discussion into the following

**Theorem 16.7.** *For every square matrix $A$, there is a unique number $D(A)$ so that the assignment $A \rightsquigarrow D(A)$ follows the following three rules.*

*(I).* $D(I_n) = 1$.

*(II).* $D$ *is* linear in each row. *This means the following*

*II(a).* ***(Row additivity)***

$$D(A +_i B) = D(A) + D(B)$$

*for each row $i$, and for every pair $A, B$ of $i^{\text{th}}$ row sum compatible matrices.*

*II(b).* ***(Scaling)***

$$D(M_i(\lambda)A) = \lambda D(A)$$

*for each row $i$ and for every $n \times n$ matrix $A$.*

*(III).* *If two rows of $A$ are equal, then $D(A) = 0$.*

The number $D(A)$ is called the *determinant* of $A$.

# Lecture 17

## Cramer's rule and functions

October 25, 2023

Before we start talking about functions, let us utilize the multiplicativity of determinant Theorem 16.1 to a good use.

### 17.1. Cramer's Rule

Cramer's rule allows us to solve a system of linear equations that has a unique solution.

**Theorem 17.1** (Cramer's Rule)**.** *Consider the following system of $n$ linear equations in $n$ variables.*

$$\begin{aligned}
a_{11}x + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1\\
a_{21}x + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2\\
\vdots \qquad \vdots \qquad \vdots \qquad &\vdots\\
a_{31}x + a_{32}x_2 + \cdots + a_{nn}x_n &= b_n
\end{aligned}$$

*Let*

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

*be the matrix of coefficients of the equations in the system, and $B_1$ be the matrix obtained by replacing the first column of $A$ by the column vector*

$$B := \begin{pmatrix} b_1\\ b_2\\ \vdots\\ b_n \end{pmatrix}$$

*That is,*

$$B_1 := \begin{pmatrix} b_1 & a_{12} & \cdots & a_{1n}\\ b_2 & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \vdots & \vdots\\ b_n & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

*Similarly, let and $B_j$ be the matrix obtained by replacing the $j^{\text{th}}$ column of $A$ by the column vector $B$. If $A$ is invertible then the given system of equations has a unique solution, which is $x_j = D(B_j)D(A)^{-1}$. Here, $D(B_j)$ denotes the determinant of $B_j$ and $D(A)$ denotes the determinant of $A$.*

*Proof.* We illustrate the proof for $n = 3$, and $j = 1$. In this case,

$$A := \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad B_1 := \begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix}$$

Let us define

$$X := \begin{pmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{pmatrix}$$

Observe that $AX = B_1$. Now, $D(B_1) = D(AX) = D(A)D(X) = D(A)x_1$, where the second equality is attributed to the multiplicativity of determinant (Theorem 16.1), and the last equality is simply the calculation of the determinant of $X$. Since $A$ is an invertible matrix, $D(A) \neq 0$ (Theorem 15.3). Therefore, $x_1 = D(B_1)D(A)^{-1}$. Other cases can be handled similarly. $\square$

While Cramer's rule is interesting and it is not difficult to implement when matrices are of smaller size, solving a system of linear equations using row reduction is much efficient. As $n$ grows, calculating determinants become very tedious task. That is what affects the efficient of Cramer's rule for large matrices.

Determinant is an example of functions. Determinant is a process, or a rule, that converts a matrix into a real number. We now study functions in detail.

### 17.2. Functions

Let $X$ and $Y$ be two sets. A *function* $f$ from $X$ to $Y$ is a rule that assigns to each element $x \in X$, an element $y \in Y$ (depending on the choice of $x$). The element $y \in Y$ which is assigned to $x$ by the function $f$ is denoted by $f(x)$; i.e. $y = f(x)$, and is called the *image* of $x$. A function $f$ from $X$ to $Y$ is denoted by $f : X \to Y$. The set $X$ can be thought of as the set of inputs and the set $Y$ can be thought of as that of outputs.

In a function $f : X \to Y$, the set $X$ is called the *domain* of $f$, and the set $Y$ is called the *codomain* of $f$. For the determinant function $D$ on $n \times n$ matrices, the domain is the set $M_{n\times n}(\mathbb{R})$ and its codomain is the set $\mathbb{R}$ of real numbers.

The *image* of a function $f : X \to Y$ is the following subset of $Y$.

$$\text{image}(f) := \{f(x) : x \in X\}$$

A function $f : X \to Y$ is called *surjective* if $Y = f(X)$. The determinant function $D : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ is an example of surjective function. This is because every element in $\mathbb{R}$ occurs as the determinant of some $n \times n$ matrix.

A function $f : X \to Y$ is called *injective* if every element in $X$ is completely identifiable through its image. That is, whenever $a, b \in X$ are such that $f(a) = f(b)$, we necessarily have $a = b$. Again, we take the determinant function $D : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$. This time we conclude that determinant is not an injective function as two distinct matrices may have the same determinant.

**Example 17.2** Let us have a look at examples of interesting functions that we have already have encountered in this course.

1. Fix two integers $p, q$ so that $1 \leq p, q \leq n$. Consider $L_{p,q} : \mathbb{R} \to M_{n\times n}(\mathbb{R})$ that assigns to each $\alpha$ in $\mathbb{R}$, the product adder matrix $L_{p,q}(\alpha)$. The function $L_{p,q}$ thus defined is injective, but not surjective.
2. Consider $\mathrm{re} : M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ that assigns to each matrix $A$, the row echelon matrix obtained by performing row operations on $A$. Thus, in terms of this function such a row echelon matrix is to be denoted by $\mathrm{re}(A)$. This function is neither surjective nor injective.
3. Consider $\mathrm{double} : M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ that assigns to each matrix $A$, the matrix $A + A$. This function is both injective as surjective.
4. Consider $\mathrm{square} : M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ that assigns to each matrix $A$, the matrix $A^2$. This function is not injective. The matrices $I_n$ and $-I_n$ have the same image under this function. The function is not surjective either. This is because the matrices whose determinant is negative do not lie in $\mathrm{image}(\mathrm{square})$.

# Lecture 18

## Binary operations

October 26, 2023

We continue our discussion on functions. Some functions of our interest are to be called binary operations. For that, we first define direct product of two sets.

Let $X$ and $Y$ be two sets. The *direct product* of $X$ and $Y$ is the set that consists of pairs of elements, where the first one is an element of $X$ and the second one is an element of $Y$. This set of pairs is denoted by $X \times Y$. Thus,

$$X \times Y := \{(x, y) : x \in X \text{ and } y \in Y\}$$

Sometimes, the matter of our interest demands $Y = X$. One such example that we would encounter often is $Y = X = M_{n\times n}(\mathbb{R})$. The direct product $M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R})$ consists of ordered pairs of $(A, B)$ of $n \times n$ matrices $A$ and $B$.

Here are two example of functions whose domain is $M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R})$.

**Example 18.1**

1. **(Matrix addition)** $+ : M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ defined by
$$+((A, B)) := A + B.$$
2. **(Matrix multiplication)** $\bullet : M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ defined by
$$\bullet((A, B)) := AB.$$

These are examples of binary operations.

### 18.1. Binary operations and abelian groups

By definition, a *binary operation* $*$ on a set $X$ is a function $* : X \times X \to X$. Thus, a binary operation takes an ordered pair from a set $X$ and outputs an element of the same set. In a binary operation $* : X \times X \to X$, the image of an ordered pair $(a, b)$ is denoted by $a * b$. That is, $a * b := *((a, b))$. This notation provides simplicity in expressing some of the properties of binary operations.

**Definition**. A binary operation $* : G \times G \to G$ on a set $G$ is called an *abelian group* if it has all of the following properties.

I. **Associativity**.
$$*((*((a, b)), c)) = *((a, *((b, c))))$$
for every triplet $a, b, c$ in $G$. The notation $a * b := *((a, b))$ allows us to express associativity condition as $(a * b) * c = a * (b * c)$.

II. **Existence of a neutral element**. There exists an element $n \in G$ such that for every $a \in G$,
$$\begin{aligned} a * n &= a, \\ \text{and,} \quad n * a &= a \end{aligned}$$

III. **Existence of inverse**. For every $a \in G$, there exists $a' \in G$ such that
$$\begin{aligned} a * a' &= n, \\ \text{and,} \quad a' * a &= n \end{aligned}$$

IV. **Commutativity**. For every $a, b \in G$ we have $a * b = b * a$.

A binary operation $* : G \times G \to G$ is required to have all these properties in order to be called an abelian group. However, it may have only some of these properties, and it is called a

- **semigroup**, if (I) holds.
- **monoid**, if (I) and (II) hold.
- **group**, if (I), (II) and (III) hold.

Let us have a look at some examples.

**Example 18.2**

1. **Addition of matrices**.
$$+ : M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$$
defined by
$$+((A, B)) := A + B$$
is an abelian group. Here, neutral element is the zero matrix.

2. **Multiplication of matrices**.

$$\bullet : M_{n\times n}(\mathbb{R}) \times M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$$

defined by

$$\bullet((A, B)) := AB$$

is a monoid but not a group. Here, $n \times n$ identity matrix $I_n$ is the neutral element. But there are matrices, for example the zero matrix, which do not have inverse.

3. **Addition of natural numbers**. Recall that positive integers are called *natural numbers*. The set $\mathbb{N} := \{1, 2, 3, \cdots\}$ consisting of natural numbers carries the binary operation of addition. Natural numbers with this binary operation form a semigroup, but not a monoid.

4. **Multiplication of natural numbers**. On the set $\mathbb{N} := \{1, 2, 3, \cdots\}$ of natural numbers, equipped with the the binary operation of multiplication, forms a monioid. The natural number $1$ is the neutral element. However there are elements, for example the natural number $2$, which do not have an inverse.

# Lecture 19

## More abelian groups

October 27, 2023

We continue with examples of binary operations $*: G \times G \to G$ which are abelian groups.

**Example 19.1**

1. **Integers with addition**. The set of integers is $\mathbb{Z} := \{\cdots, -2, -1, 0, 1, 2, \cdots\}$ On the set $\mathbb{Z}$, the binary operation of addition of integers forms an abelian group. However, if we consider the binary operation of multiplication of integers, then it is a monoid but not a group.

2. **Positive reals with multiplication**. Consider the set of positive reals:
$$\mathbb{R}^+ := \{x \in \mathbb{R} : x > 0\}$$
Under the binary operation $a * b := ab$. It is an abelian group. Here, $1 \in \mathbb{R}$ is the neutral element and for $\alpha \in \mathbb{R}$, the real number $1/\alpha$ is its inverse.

3. **Negative reals with strange multiplication**. Consider the set of negative reals:
$$\mathbb{R}^- := \{x \in \mathbb{R} : x < 0\}$$
Under the binary operation $a * b := -ab$. It is an abelian group. What is the neutral element here? What about inverses?

4. **Addition in clock time**. Consider the set of labels on a clock:
$$G := \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$
On this, we define the a binary operation $*$ to be the clock hour addition. Thus, $10 * 3 = 1$, $11 * 9 = 8$. The clock time addition is a binary operation. It indeed forms an abelian group. Neutral element is $12$. Inverse is given by "hours remaining to hit next 12". So, the inverse of $8$ is $4$, the inverse of $6$ is $6$, and the inverse of $12$ is... 12.

5. For $\alpha \in \mathbb{R} \setminus \{0\}$, define $A_\alpha$ to be the following $2 \times 2$ matrix:
$$A_\alpha := \begin{pmatrix} \alpha & \alpha \\ \alpha & \alpha \end{pmatrix}$$

Consider the following subset of $M_{2\times 2}(\mathbb{R})$.

$$G := \{A_\alpha : \alpha \in \mathbb{R} \setminus \{0\}\}$$

On $G$, we define

$$* : G \times G \to G$$

to be the matrix multiplication. We note that $A_\alpha * A_\beta = A_\alpha A_\beta = A_{2\alpha\beta}$. For the binary operation $*$, the neutral element is the matrix $A_{1/2}$, and for $A_\alpha \in G$, the inverse if $A_{1/4\alpha}$. Voila! Here is an abelian group.

6. **Rotations in** $\mathbb{R}^2$. Recall that for $\theta \in \mathbb{R}$, the $2 \times 2$ rotation matrix $R_\theta$ is defined by

$$R_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Consider the set of rotation matrices $\mathrm{Rot}(\mathbb{R}^2)$.

$$\mathrm{Rot}(\mathbb{R}^2) := \{R_\theta : \theta \in \mathbb{R}\}$$

Under matrix multiplication, the set $\mathrm{Rot}(\mathbb{R}^2)$ is an abelian group.

7. **Polynomials over** $\mathbb{R}$. Let $n \neq 0$ be an integer. A *polynomial* of degree $n$ in one variable $x$ over $\mathbb{R}$ is an expression of the following form

$$p(x) := a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

where each $a_i \in \mathbb{R}$. Thus, upon taking $n = 0$, a degree $0$ polynomial is nothing but a real number $a_0$. The real number $a_i$ is called the *coefficient* of $x_i$ in $p(x)$. Two polynomials can be added "termwise". That is, if $p_1(x)$ and $p_2(x)$ are two polynomials then $p_1(x) + p_2(x)$ is the polynomial for which the coefficient of each $x^i$ is equal to the coefficients of $x^i$ in $p_1(x)$ and $p_2(x)$. Define

$$\mathbb{R}[x] := \{p(x) : p(x) \text{ is a polynomial of some degree over } \mathbb{R}\}$$

Under the addition of polynomials, $\mathbb{R}[x]$ is an abelian group. The zero polynomial is the neutral element for the binary operation of polynomial addition.

# Lecture 20

## Vector spaces

November 01, 2023

On some abelian groups it is possible to put an extra structure of multiplication by a scalar which is compatible with the binary operation of the abelian group. Before we define vector spaces, let us look at a motivating example.

**Example 20.1** Consider $\mathbb{R}^3 := M_{3\times 1}(\mathbb{R})$, and put the binary operation of addition on it. So,

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} * \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} := \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{pmatrix}$$

This is an abelian group. We define the scaling of elements of this abelian group by real numbers, as follows.

$$\bullet : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$$

where for each $\alpha \in \mathbb{R}$ and $v := \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, the operation $\bullet$ is defined by

$$\bullet(\alpha, v) := \begin{pmatrix} \alpha a_1 \\ \alpha a_2 \\ \alpha a_3 \end{pmatrix}$$

We denote $\bullet(\alpha, v)$ by $\alpha.v$, and observe that for every $\alpha, \beta \in \mathbb{R}$ and $v, w \in \mathbb{R}^3$, the following holds.

(i). $1.v = v$.

(ii). $\alpha.(\beta.v) = (\alpha\beta).v$, where $\alpha\beta$ is the multiplication of real numbers.

(iii). $\alpha.(v * w) = \alpha.v * \alpha.w$.

(iv). $(\alpha + \beta).v = \alpha.v * \beta.v$, where $\alpha + \beta$ is the addition of real numbers.

The abelian group $(\mathbb{R}^3, *)$ together with $\bullet$ (the rule of scaling), is to be called a vector space as it satisfies the following definition.

**Definition**. A *vector space* $(V, *, \bullet)$ over real number consists of the following.

I. An abelian group $(V, *)$.

II. A rule of scaling elements of $V$ by real numbers, which is compatible with $*$. By this, we mean a function $\bullet : \mathbb{R} \times V \to V$ satisfying the following four conditions for every $\alpha, \beta \in \mathbb{R}$ and $v, w \in \mathbb{V}$. Here we denote $\alpha.v := \bullet(\alpha, v)$.

   (i). $1.v = v$.

   (ii). $\alpha.(\beta.v) = (\alpha\beta).v$, where $\alpha\beta$ is the multiplication of real numbers.

   (iii). $\alpha.(v * w) = \alpha.v * \alpha.w$.

   (iv). $(\alpha + \beta).v = \alpha.v * \beta.v$, where $\alpha + \beta$ is the addition of real numbers.

Let $n \in V$ denote the neutral element of $(V, *)$. While we have not explicitly insisted in the four conditions that $0.v = n$ for every $v \in V$, it is actually their consequence. Let us see how.

$$0.v = (0 + 0).v \overset{\text{(iv)}}{=} 0.v * 0.v$$

From this, the conclusion that $0.v = n$ holds because if $w \in V$ is such that $w * w = w$, then $w = n$. This can be obtained through the following.

$$w * w = w \Rightarrow (w * w) * w' = w * w' \Rightarrow w * (w * w') = w * w' \Rightarrow w * n = n \Rightarrow w = n$$

Here $w'$ is the inverse of $w$ for the binary operation $*$.

Now, let us have a look at more examples of vector spaces.

**Example 20.2** All of the following are examples of vector spaces.

1. (i). $V = M_{m \times n}(\mathbb{R})$ : the set of $m \times n$ matrices over $\mathbb{R}$.

   (ii). $* = +$ : the addition of matrices

   (iii). The scaling rule $\bullet : \mathbb{R} \times M_{m \times n}(\mathbb{R}) \to M_{m \times n}(\mathbb{R})$ is given by

$$*(\alpha, (a_{ij})_{m \times n}) := (\alpha a_{ij})_{m \times n}$$

2. (i). $V = \mathbb{R}[x]$, the set of all polynomials in one variable of all degrees, with coefficients in $\mathbb{R}$.

   (ii). $* = +$ : the addition of polynomials.

   (iii). The scaling rule $\bullet : \mathbb{R} \times \mathbb{R}[x] \to \mathbb{R}[x]$ given by

$$*(\alpha, p(x)) := \alpha\text{ı}(x)$$

   where $\alpha p(x)$ is the polynomials obtained by multiplying each coefficient of $p(x)$ by $\alpha \in \mathbb{R}$.

3. Consider a $n \times n$ square matrix $A$, and the system of linear equations given by $AX = 0$. Here $0$ stands for the $n \times 1$ zero matrix.

   We observe that if $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^n$ are solutions of $AX = 0$, then $v + w$ is also a solution of $AX = 0$. Further, for every $\alpha \in \mathbb{R}$, the element $\alpha v \in \mathbb{R}^n$ is also a solution of $AX = 0$.

   We denote the set of solutions of $AX = 0$ by $\mathrm{null}(A)$. Clearly, $\mathrm{null}(A)$ is a subset of $\mathbb{R}^n$. In fact, $\mathrm{null}(A)$ is an abelian group under the binary operation of vector addition. Further, the scaling $\bullet$ of vectors in $\mathbb{R}^n$ by real numbers retains an element of $\mathrm{null}(A)$ within $\mathrm{null}(A)$.

   In fact, $(\mathrm{null}(A), +, \bullet)$ is a vector space over $\mathbb{R}$. We refer to this vector space by the *null space* of the matrix $A$. We note that if $A$ is invertible then the null space consists only of the neutral element, which is the zero column vector.

4. Let $X$ be any set. Let $FX$ denote the set of all functions whose domain is $X$ and codomain is $\mathbb{R}$. That is
$$FX = \{f : X \to \mathbb{R}\}$$
   We put a binary operation $* : FX \times FX \to FX$ by defining $f * g$ to be the function whose value at $x \in X$ is $f(x) + g(x)$. This makes $(FX, *)$ an abelian group. In this abelian group the zero function (the function that assigns $0$ to every element of $X$) is the neutral element.

   For the scaling rule $\bullet : \mathbb{R} \times FX \to FX$ let is define $\bullet(\alpha, f)$ to be the function whose value at $x \in X$ is the product of real numbers $\alpha$ and $f(x)$. With this set up, $(FX, *, \bullet)$ is a vector space.

# Lecture 21

## Vector subspaces

November 03, 2023

Before we define what subspaces of vector space are, let us see one more example of vector spaces.

**Example 21.1** Consider the following.

(i). $V = \mathbb{R}^+$, the set of positive reals.

(ii). $* = .$ : the multiplication of real number

(iii). The scaling rule $\bullet : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^+$ defined by $\bullet(\alpha, x) := x^\alpha$.

It turns out that $(\mathbb{R}^+, *)$ is an abelian group and $(\mathbb{R}^+, *, \bullet)$ is a vector space over reals.

### 21.1. Subspaces

Let us consider the following subsets of $\mathbb{R}^2$.

1. $X \cup Y := \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 : a = 0 \text{ or } b = 0 \right\}$. We call it the *union of axes*.
2. $\mathbb{Z}^2 := \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 : a \in \mathbb{Z} \text{ and } b \in \mathbb{Z} \right\}$. We call it *integer lattice* in $\mathbb{R}^2$.
3. $\mathcal{L}_{(1,-1)} := \left\{ \begin{pmatrix} a \\ a \end{pmatrix} \in \mathbb{R}^2 : a \in \mathbb{R} \right\}$. It is simply the line $y = x$.
4. $C := \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 : a^2 + b^2 = 1 \right\}$. We will call it a *circle*.

To appreciate why these subsets are under consideration, let us first look at the following definition.

**Definition**. Let $(V, *, \bullet)$ be a vector space over reals. A subset $S \subseteq V$ is said to be *closed under* $*$ if every $v * w \in S$, whenever $v \in S$ and $w \in S$. A subset $S \subseteq V$ is said to be *closed under scaling* if for every $\alpha \in \mathbb{R}$ and $v \in S$, we have $\alpha \bullet v \in S$.

Among the above subsets of $\mathbb{R}^2$, where the ambient vector space is $(\mathbb{R}^2, +, \cdot)$, only $\mathbb{Z}^2$ and $\mathcal{L}_{(1,-1)}$ are closed under $*$; and only $X \cup Y$ and $\mathcal{L}_{(1,-1)}$ are closed under scaling. The circle $C$ is neither closed under $*$, nor under scaling.

**Definition**. Let $(V, *, \bullet)$ be a vector space over reals. A subset $S \subseteq V$ is called a *vector subspace* of $(V, *, \bullet)$ if

(i). $n \in S$, where $n$ is the neutral element of the abelian group $(V, *)$.

(ii). $S$ is closed under $*$.

(iii). $S$ is closed under scaling.

Among the above subsets of $\mathbb{R}^2$, where the ambient vector space is $(\mathbb{R}^2, +, \cdot)$, only $\mathcal{L}_{(1,-1)}$ is a vector subspace of $(\mathbb{R}^2, +, \cdot)$. We notice that

$$\mathcal{L}_{(1,-1)} = \text{null}(A)$$

where

$$A = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$$

Recall that $\text{null}(A)$ is the null space of $A$, as defined earlier.

**Remark 21.2** Let $(V, *, \bullet)$ be a vector space over reals and $S \subseteq V$ be a subspace. Then the binary operation $*$ and the scaling rule $\bullet$ may be restricted to the set $S$. Through these restricted operations, $(S, *, \bullet)$ is a vector space.

**Exercise 21.3** For the vector space $(\mathbb{R}^2, +, \cdot)$, the only subspaces are the following subsets.

(i). $S = V$.

(ii). The subset consisting only of the zero vector.

(iii). $\mathcal{L}_{(\alpha,\beta)} := \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 : \alpha a + \beta b = 0 \right\}$

Note that $V$ is the null space of the zero matrix, the zero space is the null space of any invertible matrix, and $\mathcal{L}_{(\alpha,\beta)}$ is the null space of the matrix

$$\begin{pmatrix} \alpha & \beta \\ 0 & 0 \end{pmatrix}$$

Therefore, each subspace of $(\mathbb{R}^2, +, \cdot)$ occurs as a null space for a suitable matrix. In fact, this holds for $(\mathbb{R}^n, +, \cdot)$ for every $n$.

**Exercise 21.4** Consider the subset $S := \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2 : b = [a] \right\}$. Here, $[x]$ denotes the largest integer that is smaller than or equal to $x$. Then, $S$ is not a subspace of the vector space $(\mathbb{R}^2, +, \cdot)$. In fact, $S$ is not closed under addition.

## 21.2. Space of sequences

A *sequence* $(a_n)$ in $\mathbb{R}$ is an arrangement of some real numbers (not necessarily distinct), so that for every natural number is assigned a real number. Thus, in the sequence

$$(a_n) := a_1, a_2, a_3, \cdots, a_r, \cdots$$

the natural number $r$ is assigned the real number $a_r$. We say that $a_r$ is the $r^{\text{th}}$ member of this sequence. Let $\operatorname{seq}(\mathbb{R})$ denote the set of all sequences in $\mathbb{R}$.

We put the following binary operation on $\operatorname{seq}(\mathbb{R})$: For two sequences $(a_n), (b_n) \in \operatorname{seq}(\mathbb{R})$, we define $(a_n) * (b_n)$ to be the sequence whose $r^{\text{th}}$ member is $a_r + b_r$, for each $r \in \mathbb{N}$. Then $(\operatorname{seq}(\mathbb{R}), *)$ is an abelian group. The neutral element is the sequence whose each member is $0$.

For $\alpha \in \mathbb{R}$ and $(a_n) \in \operatorname{seq}(\mathbb{R})$, let us define the scaling rule by $\bullet(\alpha, (a_n)) := (\alpha a_n)$. Then $(\operatorname{seq}(\mathbb{R}), *, \bullet)$ is a vector space. Consider the following subsets of $\operatorname{seq}(\mathbb{R})$.

(i). $\operatorname{seq}(\mathbb{Q})$ : the subset of $\operatorname{seq}(\mathbb{R})$ consisting of sequences $(a_n)$, where each $a_1, a_2, a_3, \cdots$ is a rational number. This subset contains neutral element of $\operatorname{seq}(\mathbb{Q})$. It is also closed under addition. However, it is not a subspace of $(\operatorname{seq}(\mathbb{R}), *, \bullet)$ because it is not closed under scaling.

(ii). $\operatorname{ffib}(\mathbb{R})$ : the subset of sequences $(a_n) := a_1, a_2, \cdots, a_r, \cdots$ in which $a_{r+2} = a_{r+1} + a_r$ for every $r \geq 1$. Let us call such a sequence *a friend of Fibonacci*. The well known Fibonacci sequence $0, 1, 1, 2, 3, 5, 8, \cdots$ is a friend of Fibonacci. It is a good exercise to show that $\operatorname{ffib}(\mathbb{R})$ is a subspace of $(\operatorname{seq}(\mathbb{R}), *, \bullet)$.

# Lecture 22

## Linear Combinations

November 08, 2023

We fix a vector space $(V, +, \bullet)$. Stay alert that $+$ is a notation for the binary operation on the vector space $(V, +)$ in context. It does not make much sense in demanding that the binary operation is 'addition'. The reason we wish to denote the binary operation by a '$+$' is that in most of the examples we would encounter, the binary operation is going to be obtain via 'entry-wise', or 'component-wise' or coefficient-wise' addition of real numbers. Another alert is that for $\alpha \in \mathbb{R}$ and $v \in V$, the element $\alpha \bullet v$ obtained after scaling $v$ by $\alpha$ is going to be denoted by $\alpha v$. Further, the neutral element $n \in V$ is going to be denoted by $\mathbf{0}$.

Let $v_1, v_2, \cdots, v_r \in V$. A *linear combination* of $v_1, v_2, \cdots, v_r$ is an element of $V$, which is of the form $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r$, for some $\alpha_1, \alpha_2, \cdots, \alpha_r \in \mathbb{R}$. Therefore, the neutral element $\mathbf{0}$ is a linear combination of *any* collection $v_1, v_2, \cdots, v_r$ of elements of $V$.

If $S$ is a subset of $V$, then a *linear combination in $S$* is an element of $V$, which is of the form $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r$, for some $\alpha_1, \alpha_2, \cdots, \alpha_r \in \mathbb{R}$ and for some $v_1, v_2, \cdots, v_r \in V$.

**Observation 22.1** Subspaces of $(V, +, \bullet)$ are closed under linear combination. That is, if $S \subseteq V$ is a subspace of $(V, +, \bullet)$, then every linear combination in $S$ lies within $S$.

What if $S$ is not a subspace and only a subset of $V$? In that case, some linear combinations in $S$ must lie outside $S$. The collection of all linear combinations in $S$ is called the *span* of $S$ and is denoted by $\operatorname{span}(S)$. Therefore,

$$\operatorname{span}(S) := \{\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r : r \in \mathbb{N}, \alpha_i \in \mathbb{R}, v_i \in S\}$$

The following observation is immediate.

**Observation 22.2** Let $(V, +, \bullet)$ be a vector space over $\mathbb{R}$ and $S$ be a subset of $V$. Then $\operatorname{span}(S)$ is a subspace of $(V, +, \bullet)$.

Let us have some examples.

**Example 22.3** Consider the vector space $(\mathbb{R}^2, +, \cdot)$. Let

$$S_1 := \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

and

$$S_2 := \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

Then, a typical element in $\mathrm{span}(S_1)$ is

$$\alpha_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

where $\alpha_1, \alpha_2 \in \mathbb{R}$ are arbitrary. In other words, an arbitrary element of $\mathbb{R}^2$ lies in $\mathrm{span}(S_1)$. Thus $\mathrm{span}(S_1) = \mathbb{R}^2$.

We now work out $\mathrm{span}(S_2)$, where a typical element is

$$\alpha_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} \alpha_1 + \alpha_2 \\ \alpha_1 - \alpha_2 \end{pmatrix}$$

For an arbitrary element

$$\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$$

the choice $\alpha_1 = \frac{1}{2}(a+b)$ and $\alpha_2 = \frac{1}{2}(a-b)$ indeed ensures that

$$\alpha_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

Thus, again we have $\mathrm{span}(S_2) = \mathbb{R}^2$.

The following is an interesting exercise.

**Exercise 22.4** Consider the vector space $(\mathbb{R}^2, +, \cdot)$. Then, for a subset

$$S := \left\{ \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} \right\} \subseteq \mathbb{R}^2$$

we have $\mathrm{span}(S) = \mathbb{R}^2$ if and only if the matrix

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

is invertible.

**Example 22.5** Consider the vector space $(M_{2\times 2}(\mathbb{R}), +, \cdot)$ and the subset

$$S := \{e_{11}, e_{12}, e_{21}, e_{22}\} \subseteq M_{2\times 2}(\mathbb{R})$$

consisting of matrix units. Recall that $e_{ij} \in M_{2\times 2}(\mathbb{R})$ is the matrix whose $(i,j)^{\text{th}}$ entry is $1$ and other entries are $0$. Then, $\text{span}(S) = M_{2\times 2}(\mathbb{R})$.

There are many subsets $S$ of $M_{2\times 2}(\mathbb{R})$ for which $\text{span}(S) = M_{2\times 2}(\mathbb{R})$. The subset of all invertible $2 \times 2$ matrices is one such. To justify this, you may like to establish the following interesting statement.

**Exercise 22.6** Every matrix $A \in M_{n\times n}(\mathbb{R})$ can be written as the sum of two invertible matrices $B$ and $C$.

In fact, $B$ can be taken as an invertible matrix in which all entries below diagonal are $0$, and $C$ can be taken as an invertible matrix in which all entries above diagonal are $0$. Here is a situation in $2 \times 2$ case.

$$\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \frac{1}{2} & 0 \\ 2 & 1 \end{pmatrix}$$

Now, pick this as a hint and try to finish the above exercise.

**Exercise 22.7** Show that there are four invertible matrices $A_1, A_2, A_3, A_4 \in M_{2\times 2}(\mathbb{R})$ such that

$$\text{span}(\{A_1, A_2, A_3, A_4\}) = M_{2\times 2}(\mathbb{R})$$

Here the vector space in question is $(M_{2\times 2}(\mathbb{R}), +, \cdot)$.

# Lecture 23

## Basis of a vector space

November 10, 2023

We fix a vector space $(V, +, \cdot)$. Let $S_1$ and $S_2$ be two subsets of $V$. By *sum* of $S_1$ and $S_2$ we mean the following subset of $V$.

$$\{v_1 + v_2 : v_1 \in S_1 \text{ and } v_2 \in S_2\}$$

Here $+$ in the expression $v_1 + v_2$ refers to the binary operation on the vector space under consideration. We denote the sum of two subsets $S_1$ and $S_2$ by $S_1 + S_2$. If there are many subsets $S_1, S_2, \cdots, S_r$ of $V$ then their sum is defined by

$$S_1 + S_2 + \cdots + S_r := \{v_1 + v_2 + \cdots + v_r : v_i \in S_i; 1 \leq i \leq r\}$$

**Example 23.1** Consider the following subsets of $(\mathbb{R}^2, +, \cdot)$.

1. $S_1 := \left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix} : \alpha \in \mathbb{R} \right\} \subseteq \mathbb{R}^2$
2. $S_2 := \left\{ \begin{pmatrix} 0 \\ \beta \end{pmatrix} : \beta \in \mathbb{R} \right\} \subseteq \mathbb{R}^2$
3. $S_3 := \left\{ \begin{pmatrix} \lambda \\ \lambda \end{pmatrix} : \lambda \in \mathbb{R} \right\} \subseteq \mathbb{R}^2$

Then, we have

$$S_1 + S_2 = S_2 + S_3 = S_1 + S_3 = S_1 + S_2 + S_3 = \mathbb{R}^2$$

Note that the operation of 'sum of subsets' *does not* have cancellation property. That is, if $S_1 + S_2 = S_1 + S_3$, then $S_2 = S_3$ *does not* follow.

Recall that subspaces are those subsets which contain $\mathbf{0}$ (that is, the neutral element of $(V, +)$), are closed under $+$, and are closed under scaling rule. We note that if $S_1$ and $S_2$ are subspaces of a vector space $(V, +, \cdot)$, then their sum $S_1 + S_2$ is also a subspace of $(V, +, \cdot)$.

### 23.1. Linear relations

We fix a vector space $(V, +, \cdot)$. Let $\mathbf{0}$ denote the neutral element of $(V, +)$. We begin with the following exercise.

**Exercise 23.2** Let $v \in V$. Then $0.v = \mathbf{0}$.

This exercise is an immediate consequence of the property $(\alpha + \beta).v = \alpha.v + \alpha.v$. Since

$$0.v + 0.v = (0 + 0).v = 0.v,$$

composing with the inverse of $0.v$ on both sides indeed shows that $0.v = \mathbf{0}$.

Our aim is to understand linear relations. This is an important aspect of learning vector spaces. Let $(V, +, \cdot)$ be a vector space and $S$ be a subset of $V$. A *linear relation* in $S$ is a linear combination

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r$$

that is equal to $\mathbf{0}$, where each $v_i \in S$. A trivial example of linear relation is when we choose $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$, and thus obtain

$$0.v_1 + 0.v_2 + \cdots + 0.v_r = \mathbf{0}$$

Are there situations when we have a nontrivial example of a linear relation? By a nontrivial example, we mean a linear relation where *not all* coefficients $\alpha_i$ are $0$. That means, at least one coefficients is nonzero. Let us look at the following.

**Example 23.3** Consider the vector space $(\mathbb{R}^2, +, \cdot)$. Then,

$$1.\begin{pmatrix} 1 \\ 1 \end{pmatrix} + (-1).\begin{pmatrix} 0 \\ 1 \end{pmatrix} + (-1)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

is a nontrivial example of a linear relation.

**Definition**. A subset $S \subseteq V$ is called a *linearly dependent* set if it exhibits a nontrivial linear relation. In other words, linearly dependent subsets $S$ are those subsets of $V$ where it is possible to find $v_1, v_2, \cdots, v_r \in S$, and nonzero real numbers $\alpha_1, \alpha_2, \cdots, \alpha_r$, such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r = \mathbf{0}$$

You should be able to observe that any subset that contains $\mathbf{0}$ is a linearly dependent set.

Why such a terminology? Suppose, $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \cdots + \alpha_r v_r = \mathbf{0}$, where $\alpha_1 \neq 0$. Then, using vector space axioms, we obtain

$$v_1 = (-\alpha_1^{-1}\alpha_2)v_2 + (-\alpha_1^{-1}\alpha_3)v_3 + \cdots + (-\alpha_1^{-1}\alpha_r)v_r$$

We read it as: "$v_1$ depends on $v_2, v_3, \cdots, v_3$". Therefore, in a linearly dependent set you are destined to identify elements which depend on other elements of the set.

A subset $S \subseteq V$ is called *linearly independent* if it is not linearly dependent. Therefore, linearly independent sets are those subsets of $V$ which have no nontrivial linear relation. If we are given some elements $v_1, v_2, \cdots, v_r$ of a linearly independent set, and are also given a linear relation $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r = \mathbf{0}$, then we must conclude that each $\alpha_i = 0$.

**Definition**. A subset $\mathcal{B}$ of $V$ is called a *basis* of $(V, +, \cdot)$ if

(i). $\operatorname{span}(\mathcal{B}) = V$.

(ii). $\mathcal{B}$ is a linearly independent set

Basis is not unique. A vector space has many bases[1]. Say, for instance, if $\{v_1, v_2, \cdots, v_r\}$ is a basis of $(V, +, \cdot)$, then for every nonzero $\alpha \in \mathbb{R}$, the subset $\{\alpha v_1, \alpha v_2, \cdots, \alpha v_r\}$ is also a basis of $(V, +, \cdot)$.

[1]Plural of basis is spelled as bases

# Lecture 24

# Dimension

November 15, 2023

Again, we fix a vector space $(V, +.\cdot)$. A subset $S \subseteq V$ is to be called a *spanning set* of $(V, +, \cdot)$ if $\mathrm{span}(S) = V$. Thus, a spanning set $S$ has the property that every element of $V$ is a linear combination of elements of $S$. This definition allows us to make the following statement.

*A linearly independent spanning set is called a basis.*

**Example 24.1** We shall now look at examples of basis for some vector spaces.

1. **Vector space of polynomials of small degree**. Let $\mathbb{R}[x]_3$ denote the set of polynomials whose degree is at most 3 and coefficients are in $\mathbb{R}$. Stay alert that elements of $\mathbb{R}$ are to be treated as polynomials of degree 0. Thus $\mathbb{R} \subseteq \mathbb{R}[x]$. Then, $\mathbb{R}[x]_3$ is a subspace of the vector space $(\mathbb{R}[x], +, \cdot)$, where $+$ refers to the addition of polynomials and the scaling $\cdot$ of polynomials is coefficient-wise. Neutral element of this vector space is the zero polynomial $\mathbf{0}$, that is, the polynomial whose all coefficients are 0. Since $\mathbb{R}[x]_3$ is a subspace, $(\mathbb{R}[x]_3, +, \cdot)$ is itself a vector space. We shall exhibit some bases in this vector space.

   Take $\mathcal{B}_1 := \{1, x, x^2, x^3\}$. Since every element in $\mathbb{R}[x]_3$ is a linear combination in $\mathcal{B}$, the subset $\mathbb{B}_1$ is a spanning set of $(\mathbb{R}[x]_3, +, \cdot)$. That it is linearly independent follows because, if a polynomial $\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 = \mathbf{0}$, then it is a must that each $\alpha_i = 0$. Thus, $\mathcal{B}_1$ is a basis.

   We tweak $\mathcal{B}_1$ a bit and see that $\mathcal{B}_2 := \{1, 2x, 5x^2, 3x^3\}$ is also a basis of $(\mathbb{R}[x]_3, +, \cdot)$. We now exhibit another basis of this vector space.

   Consider $\mathcal{B}_3 := \{1, (1+x), (1+x)^2, (1+x)^3\}$. A typical element in $\mathrm{span}(\mathcal{B}_3)$ is

   $$\alpha_0 + \alpha_1(1+x) + \alpha_2(1+x)^2 + \alpha_3(1+x)^3$$

   Is is true that every polynomial $\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \in \mathbb{R}[x]_3$ is of this form, for suitable $\alpha_0, \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$? To examine this, we rewrite a typical element

   $$\alpha_0 + \alpha_1(1+x) + \alpha_2(1+x)^2 + \alpha_3(1+x)^3$$

of $\mathrm{span}(\mathcal{B}_3)$ as

$$(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3) + (\alpha_1 + 2\alpha_2 + 3\alpha_3)x + (\alpha_2 + 3\alpha_3)x^2 + \alpha_3 x^3$$

and check if for arbitrary $\beta_0, \beta_1, \beta_2, \beta_3 \in \mathbb{R}$, the system of equations

$$\begin{aligned}
\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 &= \beta_0 \\
\alpha_1 + 2\alpha_2 + 3\alpha_3 &= \beta_1 \\
\alpha_2 + 3\alpha_3 &= \beta_2 \\
\alpha_3 &= \beta_3
\end{aligned}$$

has a solution? Here $\alpha_i$ are the unknowns to be determined. Solving it is not difficult. In fact,

$$\begin{aligned}
&\alpha_3 = \beta_3 \\
&\alpha_2 = \beta_2 - 3\beta_3 \\
&\alpha_1 = \beta_1 - 2(\beta_2 - 3\beta_3) - 3\beta_3 = \beta_1 - 2\beta_2 + 3\beta_3 \\
&\alpha_0 = \beta_0 - (\beta_1 - 2\beta_2 + 3\beta_3) - (\beta_2 - 3\beta_3) - \beta_3 = \beta_0 - \beta_1 + \beta_2 - \beta_3
\end{aligned}$$

is a solution of this system of equations[1]. Thus, an arbitrary element of $\mathbb{R}[x]_3$ is indeed there in $\mathrm{span}(\mathbb{R}[x]_3)$. Since this solution of system of linear equations is unique, it follows that if $\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 = \mathbf{0}$, then each $\alpha_i = 0$. This establishes the linear independence of $\mathcal{B}_3$. Hence, $\mathcal{B}_3$ is a basis of $(\mathbb{R}[x], +, \cdot)$.

Observe that we could have rewritten the above solution of linear equations in the following matrix form.

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

The invertibility of the matrix of coefficients would have ensured a unique solution of this system. The existence of a solution would then imply that $\mathcal{B}_3$ is a spanning set and the uniqueness would imply that $\mathcal{B}_3$ is a linearly independent set.

2. **The vector space of $2 \times 2$ matrices**. Here, the vector space in question is

$$(M_{2\times 2}(\mathbb{R}), +, \cdot)$$

We have seen that $\mathcal{B} := \{e_{11}, e_{12}, e_{21}, e_{22}\}$ consisting of matrix units is a spanning set. This set is also linearly independent. Linear independence here is merely a restatement that all entries of the zero matrix are $0$. Thus, matrix units form a basis of the vector spaces of $2 \times 2$ matrices. In fact, this holds for the vector space of $m \times n$ matrices too.

[1] The solution is, in fact, unique.

**Proposition 24.2.** *Let $(V, +, \cdot)$ be a vector space. Suppose it has a finite basis*

$$\mathcal{B} := \{v_1, v_2, \cdots, v_n\}$$

*Then for each $v \in V$, there are* unique *scalars $\alpha_1, \alpha_2, \cdots, \alpha_n \in \mathbb{R}$ such that*

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

*Proof.* Since $\mathcal{B}$ is a spanning set, the existence of scalars $\alpha_1, \alpha_2, \cdots, \alpha_n \in \mathbb{R}$ such that $v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$ is certain. Now, if $\beta_1, \beta_2, \cdots, \beta_n \in \mathbb{R}$ are (*possibly other* scalars) such that $v = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n$, then

$$(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n) + (-1)(\beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n)v + (-1)v = \mathbf{0}$$

We use vector space axioms to arrive at

$$(\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \cdots + (\alpha_n - \beta_n)v_n = \mathbf{0}$$

Now, we invoke linear independence of $\mathcal{B}$ to conclude that each $\alpha_i - \beta_i = 0$. This concludes the uniqueness of scalars for a given $v$. □

**Definition**. In the set up of Proposition 24.2, for $v \in V$ we denote

$$v_{\mathcal{B}} := \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{R}^n$$

and call $v_{\mathcal{B}}$ the *coordinate vector* of $v$ with respect to $\mathcal{B}$. Entries of the coordinate vector $v_{\mathcal{B}}$ are called *coordinates* of $v$ with respect to $\mathcal{B}$.

The following theorem is very important in understanding of basis.

**Theorem 24.3.** *In a vector space $(V, +, \cdot)$, let $S \subseteq V$ and $L \subseteq V$ be finite subsets such that $S$ is a spanning set and $L$ is a linearly independent set. Then $|L| \leq |S|$. Here, $|S|$ denotes the number of elements in $S$ and $|L|$ denotes the number of elements in $L$.*

We omit the proof here, that depends on our understanding of solutions of a system of linear equations. The following is am immediate corollary.

**Corollary 24.4.** *Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two finite sets, which are basis of a vector space $(V, +, \cdot)$. Then $|\mathcal{B}_1| = |\mathcal{B}_2|$.*

This shows that, while a vector space does not have a unique basis, the number of elements in a basis, if finite, is a well defined number. We refer to this number as the *dimension* of the vector space $(V, +, \cdot)$. It is denoted by $\dim_{\mathbb{R}}(V)$, or simply by $\dim(V)$.

**Exercise 24.5**

1. Show that the dimension of the vector space $(M_{mn}(\mathbb{R}), +, \cdot)$ is $mn$.
2. Show that the dimension of the vector space $(\mathbb{R}[x]_3, +, \cdot)$ is $4$.
3. Show that the dimension of the vector space $(\mathbb{R}^3, +, \cdot)$ is $3$.
4. Show that the dimension of the vector space $(\mathbb{R}^n, +, \cdot)$ is $n$.
5. Show that the vector space $(\mathbb{R}[x], +, \cdot)$ does not have a spanning set which has finitely many elements. In a such situation, we say that the vector space has *infinite dimension*.

# Lecture 25

## Linear Transformations

November 17, 2023

*Linear transformations preserve vector space structure across two vector spaces.*

We start with an example. Consider the vector space $(\mathbb{R}[x], +, \cdot)$ - the vector space of polynomials over reals. Recall that $\mathcal{B} := \{1, x, x^2, x^3, \cdots\}$ is a basis of this vector space, and hence its dimension is infinite. On this vector space consider the function *differentiation with respect to* $x$.

$$\frac{d}{dx} : \mathbb{R}[x] \to \mathbb{R}[x]$$

that maps a polynomial $p(x)$ to its derivative $\dfrac{dp(x)}{dx}$. Thus,

$$\frac{d}{dx}(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_1 + 2a_2x + 3a_3x^2 + \cdots + na_nx^{n-1}$$

We observe the following holds for every $p_1(x), p_2(x), p(x) \in \mathbb{R}$ and evert $\alpha \in \mathbb{R}$.

(i). $\dfrac{d(p_1(x) + p_2(x))}{dx} = \dfrac{dp_1(x)}{dx} + \dfrac{dp_2(x)}{dx}$. We refer to it by saying that *differentiation preserves addition*.

(ii). $\dfrac{d(\alpha p(x))}{dx} = \alpha\dfrac{dp(x)}{dx}$. We refer to it by saying that *differentiation preserves scalar multiplication*.

This makes the function $\dfrac{d}{dx} : \mathbb{R}[x] \to \mathbb{R}[x]$, a linear transformation. Here is the formal definition.

**Definition**. Let $(V, +_1, \cdot)$ and $(W, +_2, \cdot)$ be two vector spaces. A *linear transformation* between them is a function

$$T : V \to W$$

such that

(i). $T(v_1 +_1 v_2) = T(v_1) +_2 T(v_2)$ for every $v_1, v_2 \in V$.

(ii). $T(\alpha v) = \alpha T(v)$ for every $\alpha \in \mathbb{R}$ and every $v \in V$.

Observe that if $T$ is a linear transformation between $(V, +_1, \cdot)$ and $(W, +_2, \cdot)$, then $T(\mathbf{0}_V) = \mathbf{0}_W$. Here $\mathbf{0}_V$ is the neutral element of $(V, +_1)$ and $\mathbf{0}_W$ is the neutral element of $(W, +_2)$. To see this, we write

$$T(\mathbf{0}_V) + T(\mathbf{0}_V) = T(\mathbf{0}_V + \mathbf{0}_V) = T(\mathbf{0}_V)$$

and compose both sides with $(-1)T(\mathbf{0}_V)$ to obtain

$$T(\mathbf{0}_V) = T(\mathbf{0}_V) + (-1)T(\mathbf{0}_V) = (1 + (-1))T(\mathbf{0}_V) = 0T(\mathbf{0}_V) = \mathbf{0}_W$$

Before we see their examples, let us ponder why should such functions be called *linear transformations*. For $v_1, v_2 \in V$, define the set

$$L_{v_1,v_2} := \{(1-\alpha)v_2 + \alpha v_1 : \alpha \in \mathbb{R}\}$$

to be the *line in V passing through* $v_1$ and $v_2$. What does a linear transformation $T : V \to W$ do to this line? We check the effect of $T$ on a typical element on the line $L_{v_1,v_2}$:

$$T((1-\alpha)v_2 + \alpha v_1) = (1-\alpha)T(v_2) + \alpha T(v_1)$$

This element in $W$ lies on the line in $W$ passing through $T(v_1)$ and $T(v_2)$. This we conclude that, under a linear transformation, the image of a line in $V$ is a line in $W$. In other words, linear transformations map lines to lines. This is the reason behind their nomenclature.

**Example 25.1** Consider the vector space $(M_{2\times 2}(\mathbb{R}), +, \cdot)$ of $2 \times 2$ matrices. Then the determinant map

$$D : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}$$

is *not* a linear transformation.

**Example 25.2** Consider the vector space $(\mathbb{R}^n, +, \cdot)$. Fix a matrix $A \in M_{n\times n}(\mathbb{R})$, and define a function

$$T_A : \mathbb{R}^n \to \mathbb{R}^n$$

by $T_A(v) = Av$. Here $Av$ is multiplication of $n \times n$ matrix with a $n \times 1$ matrix. This is a linear transformation - thanks to the properties of matrix multiplication.

Let us consider linear transformations $T : V \to V$ from a vector space $(V, +, \cdot)$ to itself. Then we define *image* of $T$ to be the subset

$$\mathrm{image}(T) := \{T(v) : v \in V\}$$

of $V$. We define *kernel* of $T$ to be the subset

$$\mathrm{kernel}(T) := \{v \in V : T(v) = \mathbf{0}\}$$

of $V$

**Exercise 25.3** The subsets $\mathrm{image}(T)$ and $\mathrm{kernel}(T)$ are subspaces of $(V, +, \cdot)$.

**Theorem 25.4** (Rank-nullity theorem). *Let $(V, +, \cdot)$ be a vector space and $T : V \to V$ be a linear transformation. Then,*

$$\dim(\mathrm{image}(T)) + \dim(\mathrm{kernel}(T)) = \dim(V)$$

Observe that for the Example 1, the kernel $\mathrm{image}(T_A)$ is the null space as defined in Example 20.2(3).

**Exercise 25.5** By rank-nullity theorem, we can show that if a linear transformation on a finite dimensional vector space is surjective, then it is injective as well.

# Lecture 26

# Eigenvectors

November 22, 2023

*Eigenvectors for a linear transformation are the lines passing through the neutral element which do not experience a change due the transformation. The change experienced by elements on this line is merely scaling.*

### 26.1. Matrix of a linear transformation

Throughout, $(V, +, \cdot)$ will denote a vector space. We keep the Example in mind which highlights that for an $n \times n$ matrix $A$, the function $T_A : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T_A(v) := Av$ is a linear transformation. What we wish to state is that every linear transformation $T : V \to V$ on an arbitrary vector space is expressible through a matrix in this manner, provided we fix a basis of $V$.

**Definition**. Let $(V, +, \cdots)$ be a finite dimensional vector space. Let us fix a basis $B := \{v_1, v_2, \cdots, v_n\}$ of this vector space. Then, for a given linear transformation $T : V \to V$, we define the *matrix of $T$ with respect to $\mathcal{B}$* to be matrix whose $j^{\text{th}}$ column is the coordinate vector $(T(v_j))_{\mathcal{B}}$. Recall the definition of coordinate vector from the discussion following Proposition 24.2. We denote this matrix by $A_{\mathcal{B}}$. Therefore,

$$A_{\mathcal{B}} = \text{ the matrix whose } j^{\text{th}} \text{ column is } \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix},$$

where $T(v_j) = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$. In other words, if we know that the matrix of a linear transformation $T : V \to V$ with respect to a basis $\mathcal{B}$ is $A_{\mathcal{B}} := (a_{ij})$, then

$$T(v_j) = a_{1j}v_1 + a_{2j}v_2 + \cdots + a_{ij}v_i + \cdots + a_{nj}v_n$$

This allows us to reconstruct a given linear transformation from the information about $A_{\mathcal{B}}$ and $\mathcal{B}$. In that sense, Example 1 captures all linear transformation.

**Example 26.1** Consider the vector space $(\mathbb{R}[x]_3, +, \cdot)$, its basis $\mathcal{B} := \{1, x, x^2, x^3\}$ and the linear transformation $\dfrac{d}{dx}$, which is the differentiation with respect to $x$. Then,

$$\frac{d}{dx}(1) = 0.1 + 0.x + 0.x^2 + 0.x^3$$
$$\frac{d}{dx}(x) = 1.1 + 0.x + 0.x^2 + 0.x^3$$
$$\frac{d}{dx}(x^2) = 0.1 + 2.x + 0.x^2 + 0.x^3$$
$$\frac{d}{dx}(x^3) = 0.1 + 0.x + 3.x^2 + 0.x^3$$

Hence, the matrix of $\frac{d}{dx}$ respect to the basis $\mathcal{B}$ is

$$A_{\mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We must note that matrix of a linear transformation depends on the choice of basis of the underlying vector space.

**Exercise 26.2** Work out the above example, for the basis $\{1, 1 + x, (1 + x)^2, (1 + x)^3\}$.

### 26.2. Eigenvectors and eigenvalues

Let $T : V \to V$ be a linear transformation on a vector space $(V, +, \cdot)$. An *eigenvector* for $T$ is a nonzero vector $v \in V$ that transforms, under $T$, to a scalar multiple of itself; i.e., for some $\alpha \in \mathbb{R}$, we have that $T(v) = \alpha v$.

If $v$ is an eigenvector for $T$, then the scalar $\alpha$ such that $T(v) = \alpha v$ is called the *eigenvalue of $v$ for $T$*. The set of *eigenvalues of $T$* is the collection of these eigenvalues, as $v$ varies over all eigenvectors.

The idea of studying eigenvectors is to understand the linear transformations along ”simple looking” directions, with the hope that this will help in understanding the linear transformation in its entirety. However, there are linear transformation which do not have any eigenvector. Thus, this approach to understand a linear transformation completely through eigenvectors has its limitations.

**Exercise 26.3** There are linear transformations $T : V \to V$, where $V$ has a basis $\mathcal{B}$ in which every element is an eigenvector for $T$ Such linear transformations are called *diagonalizable*. Show that if $T$ and $\mathcal{B}$ are as described above, then the matrix $A_{\mathcal{B}}$ of $T$ will be a diagonal matrix.

**Example 26.4** Let us consider the linear transformation $\frac{d}{dx} : \mathbb{R}[x]_3 \to \mathbb{R}[x]_3$ of Example 26.1 and try to find eigenvectors. If $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \in \mathbb{R}[x]_3$ is an eigenvector, then we are required to have a scalar $\alpha \in \mathbb{R}$ such that

$$\frac{d}{dx}(p(x)) = a_1 + 2a_2x + 3a_3x^2 = \alpha a_0 + \alpha a_1 x + \alpha a_2 x^2 + \alpha a_3 x^3$$

This suggests that if $\alpha \neq 0$, then $a_3 = 0, a_2 = 0, a_1 = 0$, and $a_0 = 0$; i.e. $p(x)$ is the zero polynomial. Since a zero polynomial is prohibited to be an eigenvector, the only possible eigenvalue is the scalar $0$. The corresponding eigenvector is obtained by solving for $p(x)$ in the following.

$$\frac{d}{dx}(p(x)) = 0 \in \mathbb{R}[x]_3$$

This suggests that nonzero constants in $\mathbb{R}[x]_3$ are the only eigenvectors for the transformation $\frac{d}{dx}$. Recall that the polynomials for which the coefficients of $x, x^2, x^3$ are all zero, are called *constants*.

The following example will take us through the ideas required to determine eigenvalues and eigenvectors of a linear transformation.

**Example 26.5** The linear transformation of this example is the 'rotation by $\theta$' function in two dimensions. Hence, the vector space in question is $(\mathbb{R}^2, +, \cdot)$, and the linear transformation $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$R_\theta\left(\begin{pmatrix} a \\ b \end{pmatrix}\right) = \begin{pmatrix} a\cos\theta - b\sin\theta \\ a\sin\theta + b\cos\theta \end{pmatrix}$$

So, we are interested if there is some $\alpha \in \mathbb{R}$ such that

$$\begin{aligned} a\cos\theta - b\sin\theta &= \alpha a \\ a\sin\theta + b\cos\theta &= \alpha b \end{aligned}$$

for some

$$\begin{pmatrix} a \\ b \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \mathbb{R}^2$$

If $\sin\theta = 0$, that is to say that if $\cos\theta = 1$, then $\alpha = \cos\theta \in \{-1, 1\}$ will be an eigenvalue of $\mathbb{R}_\theta$ and every nonzero element of $\mathbb{R}^2$ will be an eigenvector. This is because in that case the linear transformation will be scaling by $\cos\theta$ for *all* vectors.

So, we focus on that case when $\sin\theta \neq 0$. Assuming there is some such $\alpha$, we rewrite this system of equations in the following form.

$$\begin{aligned} a(\cos\theta - \alpha) - b\sin\theta &= 0 \\ a\sin\theta + b(\cos\theta - \alpha) &= 0 \end{aligned}$$

Here $a$ and $b$ are unknowns. To solve for these unknowns, we substitute

$$b = \frac{a(\cos\theta - \alpha)}{\sin\theta}$$

in the second equation to get

$$a\sin\theta + \frac{a(\cos\theta - \alpha)^2}{\sin\theta} = 0,$$

which boils down to

$$a(\alpha^2 - 2\cos\theta\alpha + 1) = 0$$

We notice that if $a = 0$, then $b = 0$. But, the hunt for an eigenvector demands at least one of these to be nonzero. Thus, we are forced to have $a \neq 0$. That will imply

$$\alpha^2 - 2\cos\theta\alpha + 1 = 0$$

Existence of one such $\alpha \in R$ is not possible, unless $\cos^2\theta = 1$. But we have excluded that case.

From this discussion we conclude that if $\sin\theta \neq 0$, then the rotation $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ does not have any eigenvector.

We approach this example in an alternative manner so as to allow us to generalize the process of looking for eigenvectors.

First, we notice that if we choose the basis $\mathcal{B} := \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ then the matrix of $R_\theta$ with respect to $\mathcal{B}$ will be

$$A_{\mathcal{B}} := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

and an eigenvector $v \in \mathbb{R}^2$ will be a solution of the equation

$$A_{\mathcal{B}} v = \alpha v$$

for some $\alpha \in \mathbb{R}$. This is equivalent to

$$(A_{\mathcal{B}} - \alpha In \times n)v = 0$$

where $\alpha I_{n\times n}$ is the $n \times n$ matrix whose all diagonal entries are $\alpha$ and all nondiagonal entries are $0$. If the matrix $A_{\mathcal{B}} - \alpha In \times n$ were invertible, the only solution would be $v = 0$. Since eigenvectors are nonzero, the existence of an eigenvector would demand the non-invertibility of the matrix $A_{\mathcal{B}} - \alpha In \times n$. This is same as forcing $D(A_{\mathcal{B}} - \alpha In \times n) = 0$. Here $D$ specifies the determinant. Equating this determinant to $0$, we obtain eigenvalues. If $\alpha_0$ is en eigenvalue then the corresponding eigenvectors are precisely the elements in the null space of the matrix $A_{\mathcal{B}} - \alpha_0 I_{n\times n}$.

The above discussion has all the features that lead to the following theorem for arbitrary situation.

**Theorem 26.6.** *Let $(V, +, \cdot)$ be a vector space and $T : V \to V$ be a linear transformation. Let $\mathcal{B} := \{v_1, v_2, \cdots, v_n\}$ be a basis of $V$ and $A_{\mathcal{B}}$ be the matrix of $T$ with respect to $\mathcal{B}$. Then, eigenvalues of $T$ are precisely the solutions of the polynomial equation $D(A_{\mathcal{B}} - xI_{n\times n}) = 0$. Further, if $\alpha_0 \in \mathbb{R}$ is one such solution then the corresponding eigenvectors are precisely the elements $\beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n \in V$, where*

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} \in \mathbb{R}^n$$

*varies over the null space of the matrix $A_{\mathcal{B}} - \alpha_0 I_{n\times n}$.*

We end with the following.

**Definition**. The polynomial $p_T(x) := D(A_{\mathcal{B}} - I_{n\times n}) \in \mathbb{R}[x]$ in the set up of above theorem is called the *characteristic polynomial* of the linear transformation $T : V \to V$. This polynomial has plenty of hidden information about the linear transformation. We leave further explorations of $T$ using the characteristic polynomial for the next time we meet linear algebra.

# Thanks

The following readers are acknowledged for careful reading and pointing out errors.

- Aniruddha Mishra ⟨ms23023@iisermohali.ac.in⟩ – corrected a typo in Remark 7.3.
- Aditya Rao ⟨ms23195@iisermohali.ac.in⟩ – corrected a typo in §3
- Omansh Malhotra ⟨ms23122@iisermohali.ac.in⟩ – corrected a typo in §1 of Lecture 21.
- Karthik Nair ⟨ms23253@iisermohali.ac.in⟩ – corrected a typo in the title of Lecture 17.
- Ankan Mandal ⟨ms23243@iisermohali.ac.in⟩ – corrected a typo in a remark in Lecture 21.
- Parvathy Shine ⟨ms23022@iisermohali.ac.in⟩ – corrected a typo in an exercise in Lecture 21.
- Aishmeen Kaur ⟨ms23097@iisermohali.ac.in⟩ – corrected two typos in Lecture 23, a typo in Lecture 26, and another typo in Lecture 22.
- Satyabrata Patra ⟨ms23131@iisermohali.ac.in⟩ – corrected a typo in Lecture 25.
- Arani Chatterjee ⟨ms23098@iisermohali.ac.in⟩ – corrected two typos in Lecture 16 and another typo in Lecture 19.
- Ankush Chakraborty ⟨ms23254@iisermohali.ac.in⟩ – corrected a typo in Lecture 25.
- Tushti Govind ⟨ms23144@iisermohali.ac.in⟩ – corrected a typo in Lecture 24.
- Akshat Bhardwaj ⟨ms23001@iisermohali.ac.in⟩ – corrected two typos in Lecture 26, and another typo in the spelling of typo!
- Md Anas Rahman ⟨ms23246@iisermohali.ac.in⟩ – corrected a typo in Lecture 23.
- Tanmayii Tushar Kolhe ⟨ms23034@iisermohali.ac.in⟩ – corrected a typo in Lecture 26.
- Adeshbir Singh Benipal ⟨ms23129@iisermohali.ac.in ⟩ – corrected a typo in Lecture 26.