A warmup with matrices

Ranveer[*]

"It is not knowledge, but the act of learning, not the possession of but the act of getting there, which grants the greatest enjoyment." Carl Friedrich Gauss

Most of the content in this article is a revision of what we already know. The discussion is brief but sufficient at a fundamental level. There are many 'why?', consider them as interesting exercises, it will help revising the concepts better.

1 Matrices

Let us recall our school days where we have studied the matrices, the arrangment of numbers (entries) in rows and columns. For example following is a matrix

$$Q = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

The entries of a matrix can be complex numbers but we will restrict ourselves to real numbers, in fact almost always to just integers. An another matrix

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

has the same entries as Q but the arrangement is different; the rows of the Q become corresponding columns of the above matrix. By making rows of a matrix M corresponding columns we get another matrix called as transpose of M, denoted M^T. The above matrix is Q^T. We see in Q there are three rows and two columns. In general when a matrix has m rows and n column we say the matrix has the order or the dimension $m \times n$, so Q has order 3×2. In particular when $m = n$, we say the matrix is a square matrix of order n. A matrix of order $1 \times n$ is called a row vector of order n; similarly, an $n \times 1$ matrix is a column vector of order n[1]. The set of all the column vectors of order n having real entries is denoted by $\mathbb{R}^n$. We denote the i-th row of a matrix M by $M(i, :)$, and the i-th column by $M(:, i)$. The entry that is in row $M(i, :)$ and column $M(:, j)$ is denoted

[*]Email: `ranveer@iiti.ac.in`,

[1]When we simply say a matrix M is of order n, it means M is a square matrix of order $n \times n$, and when we say a vector of order $n \times 1$, it means a column vector order n, that is, a matrix of order $n \times 1$. Although sometimes by a vector v we mean a row vector but it will be clear from the context.

by $M(i,j)$. If $i = j$, we call $M(i,j)$ a diagonal entry, else nondiagonal entry. Often we write $M = (m_{ij})$ to denote the matrix where $M(i,j) = m_{ij}$.
Recall matrix product of two matrices. The product of Q with Q^T, that is, QQ^T is the matrix

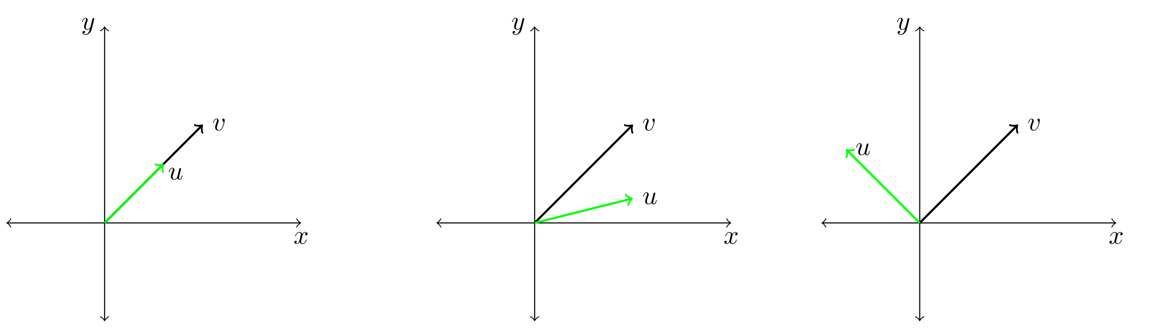
$$L = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Check whether $QQ^T = Q^TQ$?...Big no!. Matrix product is probably one of the first few counterintuitive notions in our school. Firstly, for any two matrices A, B it is not necessary that both AB and BA are defined. Second, even if they both are defined it is not necessary that $AB = BA$, that is, it is not commutative.
Now let us come back to L. Apart from being a square matrix L has an special arrangement of entries; it is transpose of itself, that is, $L = L^T$. Such a matrix is called a symmetric matrix. Symmetric matrices are very interesting and useful matrices, they are used in plethora of applications. A trivial symmetric matrix where all the nondiagonal entries are zero is called a diagonal matrix. We denote a diagonal matrix $D = (d_{ij})$ of order n by $\mathrm{diag}(d_{11}, \ldots, d_{nn})$, where d_{ii} is the i-th diagonal entry. For example $\mathrm{diag}(1,2,1)$ is the matrix

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

An typical example of a diagonal matrix is the identity matrix $I_n = \mathrm{diag}(\mathbf{1}^T)$, where $\mathbf{1}$ is all-one vector of order n.

Exercise 1.1. *Prove that for any two square matrices of order n if $AB = I_n$, then $BA = I_n$.*



(a) Dependent vectors (b) Independent vectors (c) Orthogonal vectors

Figure 1: Different possibilities for vectors u, v. If u, v are orthogonal then they are also independent.

2 Matrix multiplication as linear combinations

The multiplication of a matrix with a column vector is quite interesting. Consider Q and multiply it by column vector $\begin{bmatrix} 2 \\ 5 \end{bmatrix}$. Most likely someone will write immediately that

$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} (1 \times 2) + (0 \times 5) \\ (-1 \times 2) + (1 \times 5) \\ (0 \times 2) + (-1 \times 5) \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix}.$$

It is correct, now let us see the above product with a different way; the output vector $[2, 3, -5]^T$ can be written as

$$2 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + 5 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix},$$

that is, it is the result of adding 2 times the first column of Q with 5 times the second column. What if we multiply Q with the matrix $\begin{bmatrix} 2 & 3 \\ 5 & 2 \end{bmatrix}$? The first column of the resulting matrix will be as above but the second column will be

$$3 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

Suppose we have vectors $v_1, \ldots, v_n$ and scalars $\alpha_1, \ldots, \alpha_n$, not all zero, then the vector $\alpha_1 v_1 + \cdots + \alpha_2 v_n$ is said to be a linear combination of $v_1, \ldots, v_n$. The scalars $\alpha_1, \ldots, \alpha_n$, are the coefficients of the linear combination.

Suppose we have matrices A of order $m \times n$ and B of order $n \times p$. Its now easy to see that in AB the columns are the linear combinations of the columns of A and coefficients are given by the columns of B. We can think of product AB as linear combinations of rows rather than columns. In AB the rows are the linear combinations of the rows of B and coefficients are given by the rows of A. We will see in many occasions that visualising the matrix product as linear combinations of rows or columns is quite handy.

Exercise 2.1. *Let*

$$A = \begin{bmatrix} 2 & -1 & 3 \\ 4 & -1 & 0 \\ 0 & 3 & 1 \end{bmatrix}, \; B = \begin{bmatrix} 3 & 0 & -1 \\ 2 & 1 & -3 \\ 5 & 1 & -1 \end{bmatrix}.$$

Write down AB as (a) linear combination of rows, (b) linear combination of columns.

3 Linear independence, vector space, basis

Carrying forward the discussion on linear combination let us recall an interesting phenomenon of linear dependency of vectors. A set of vectors $v_1, \ldots, v_n$ is said to be linearly dependent set, if there exist scalars $\alpha_1, \ldots, \alpha_n$, not all zero, such that

$$\alpha_1 v_1 + \cdots + \alpha_2 v_n = \mathbf{0}. \tag{1}$$

Check that columns of Q and the output vector $[2, 3, -5]^T$ form a set of linearly dependent vectors. If the vectors $v_1, \ldots, v_n$ are not dependent, they are called independent vectors. For a given set of vectors $V = \{v_1, \ldots, v_k\}$ we can form a set S of all the vectors which are some linear combination of vectors in V. The set S is called the vector space of V. For example the vector space for a set $V = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ is the 2D space as every vector in 2D plane can written as some combination of vectors in V. But see that the set $V' = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\}$ can also generate (or span) the 2D space using linear combinations of its vectors. In this context the vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is redundant. For a vector space a maximal independent set of vectors whose linear combinations can generate all vectors in the vector space is called it basis. In the above example V' is basis for 2D while V is not a basis as it is not a maximal independent set. Note that there can be several basis for a vector space, but the number of vectors in every basis is the same. This number is called the dimension of the vector space.

Exercise 3.1. *Find a maximal set of pair-wise independent vectors from the following vectors.*

$$\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Exercise 3.2. *Find a value of x for which the following set of vectors form a basis for $\mathbb{R}^3$.*

$$V = \left\{ \begin{bmatrix} 1 \\ 0 \\ x \end{bmatrix}, \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ x \\ 1 \end{bmatrix} \right\}.$$

3.0.1 Cauchy–Schwarz inequality, orthogonal vectors

Following inequality about the dot product of two vectors is quite helpful to know. The proof is given here (Theorem 1.1).

Cauchy–Schwarz inequality

Theorem 3.3. *For any two vectors $u, v \in \mathbb{R}^n$*

$$|u^T v| \leq \sqrt{u^T u}\sqrt{v^T v},$$

where, $|.|$ denotes the absolute value. The quantity $\sqrt{u^T u}$ is known as 2-norm of u and is denoted by $||u||$. Thus Cauchy–Schwarz inequality is often stated as

$$|u^T v| \leq ||u|| ||v||.$$

Exercise 3.4. *Prove that equality in the Cauchy-Schwarz inequality holds if and only if $u = cv$, where c is some scalar.*

It is quite fruitful to know that if the dot product of two nonzero vectors u, v, that is, $u^T v$ is 0, then they are always independent (why?), such vectors are known as orthogonal vectors. Recall our 3D-coordinate space with famous x-axis, y-axis, z-axis, pick one vector along each of these axes, check that they are pairwise orthogonal. In general, if $v_1, \dots, v_n$ are pairwise orthogonal vectors, they form an independent set of vectors (why?). Corresponding to a vector space if a basis has vectors which are all pair wise orthogonal to each other, then this basis is called as orthogonal basis (its callled orthonormal basis if in addition every vector has norm 1.)

Exercise 3.5. *Find a vector $v \in \mathbb{R}^n$ which is orthogonal to the vectors*

$$\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 7 \\ 1 \end{bmatrix}.$$

Exercise 3.6. *Prove that vectors u, v are orthogonal if and only if $||u + v|| = ||u - v||$.*

Exercise 3.7. *Find a value of x for which the following set of vectors form an orthogonal basis for $\mathbb{R}^3$. Convert the orthogonal basis into a orthonormal basis.*

$$V' = \left\{ \begin{bmatrix} x \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ x \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -x \end{bmatrix} \right\}$$

4 Row space, column space, rank

Consider a general $m \times n$ matrix M with columns $v_1, \dots, v_n$, and a nonzero vector $x = [\alpha_1, \dots, \alpha_n]^T$, we can write

$$Mx = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

So $v_1, \dots, v_n, Mx$ are linearly dependent, that is, Mx is a linear combination of $v_1, \dots, v_n$. The set of all possible vectors which can be written as a linear combination of columns (rows) of M is called column (row) space of M. The dimension of column (row) space of a matrix is called its column (row) rank. Suppose M is 20×50 matrix. What could be the maximum row rank of M? Its obvious that it could be at most 20. But what about column rank; what could be the maximum column rank of M? You might tempted to say that it can be at most 50. But it is significantly wrong!. The column rank will also be at most 20. In fact even more surprise lies ahead. Suppose the row rank of M is 15, what can we say about the column rank of M? The answer is that the column rank must also be 15. For any matrix the row rank and the column rank are the same. This number is called the rank of M, denoted $\operatorname{rank} M$.

4.1 Why row rank = column rank?

Consider a $m \times n$ matrix with columns $v_1, \dots, v_n$. We can write

$$M = \begin{bmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \end{bmatrix}.$$

Suppose the column rank of M is p. That means there exist basis vectors $b_1, \dots, b_p$ and each column vector of M can be written as some linear combination of these basis vectors. Suppose $v_i = \alpha_{1i} b_1 + \dots + \alpha_{ki} b_p$, (the coefficients $\alpha_{1i}, \dots \alpha_{pi}$ can not be simultaneously zero) and let $a_i = [\alpha_{1i}, \alpha_{2i}, \dots, \alpha_{pi}]^T$, $i = 1, \dots, n$. Then we can write

$$M = \begin{bmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \end{bmatrix} = \underbrace{\begin{bmatrix} | & | & \dots & | \\ b_1 & b_2 & \dots & b_p \\ | & | & \dots & | \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} | & | & \dots & | \\ a_1 & a_2 & \dots & a_n \\ | & | & \dots & | \end{bmatrix}}_{A}.$$

Now think about the row rank of matrix M. Observe that any row of M is a linear combination of rows in A and the coefficients are given by the rows in B. (In particular, i-th row of M is a linear combination of rows of A, and the i-th row of B give the corresponding coefficients.) This implies that row rank of M is at most p which is the column rank.

Now suppose the row rank of M is q. Now visualise M in terms of its rows. Let $u_1, u_2, \dots, u_m$ be its rows. As the row rank is q, there exist basis vectors $y_1, \dots, y_q$ and each row vector of M can be written as some linear combination of these basis vectors. Suppose $u_i = \beta_{i1} y_1 + \dots + \beta_{ik} y_q$, (again, the coefficients $\beta_{i1}, \dots \beta_{iq}$ can not be simultaneously zero) and let $x_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{iq}]$, $i = 1, \dots, m$. Then we can write

$$M = \begin{bmatrix} - & u_1 & - \\ - & u_2 & - \\ \vdots & \vdots & \vdots \\ - & u_m & - \end{bmatrix} = \underbrace{\begin{bmatrix} - & x_1 & - \\ - & x_2 & - \\ \vdots & \vdots & \vdots \\ - & x_m & - \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} - & y_1 & - \\ - & y_2 & - \\ \vdots & \vdots & \vdots \\ - & y_q & - \end{bmatrix}}_{Y}$$

Now think about the column rank of matrix M. Observe that any column of M is a linear combination of columns in X and the coefficients are given by the columns in Y. (In particular, i-th column of M is a linear combination of columns of X, and the i-th column of Y gives the corresponding coefficients.) This implies that column rank of M is at most q which is the row rank.

Combining the facts that the row rank can be at most column rank, and column rank can be at most rank rank we arrive at the conclusion that both the ranks must be equal.

Theorem 4.1. *For any matrix the row rank and the column rank are the same.*

Exercise 4.2. *Find the row rank and the column rank of the following matrix and verify that both are the* *same*

$$\begin{bmatrix} 1 & -1 & 0 \\ 2 & -1 & 4 \end{bmatrix}.$$

Exercise 4.3. *Prove that if* $M = BC$*, then* $\operatorname{rank} M \leq \min(\operatorname{rank} B, \operatorname{rank} C)$.

5 Null Space

Consider the case when Mx is a zero vector, that is,

$$Mx = \mathbf{0}. \tag{2}$$

Is this possible even if M, x are nonzero matrices? Think about two real number a, b such that $ab = 0$. We know that this is possible when only at least one of a, b is zero. Can we generalize this to matrices and say that at least one of M, x must be a zero matrix for $Mx = \mathbf{0}$? No!. Multiply Q^T by all-one vector $[1, 1, 1]^T$ it gives a zero vector $[0, 0, 0]^T$. The set of all x satisfying Equation (2) is called null space of M. Vector $[1, 1, 1]^T$ is in the null space of Q^T. The minimal number of vectors whose linear combinations form the null space of M is called the nullity of M, denoted, $\operatorname{null} M$. Another fundamental fact is

$$\operatorname{rank} M + \operatorname{null} M = n,$$

look here for a proof.

Next, suppose x is a nonzero vector, under what condition $Mx = \mathbf{0}$? Notice that this is the same as the form (1). When M is a square matrix what does this imply? Here comes our old strange old mate, the determinant! It is possible only when the determinant of M, denoted $\det M$ is zero. Recall that $Mx = \mathbf{0}$ says that the columns of M are linearly dependent which implies $\det M = 0$, to figure out why? Its time to revisit our old mate, the determinant. We had already seen the applications of determinants in solving the system of linear equations, finding the area of 2D figures and volume of 3D figures. In, fact determinants can determine the areas or volumes of figures resulting after linear transformation, see Figure 2. We will see many more applications of determinants. The following quote on determinant by mathematician J. J. Sylvester tells its great importance.

For what is the theory of determinants? It is an algebra upon algebra; a calculus which enables us to combine and foretell the results of algebraical operations, in the same way as algebra itself enable us to dispense with the performance of the special operations of arithmetic. All analysis must ultimately clothe itself under this form.

6 Determinant (definition by Laplace expansion)

We know that

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc \tag{3}$$

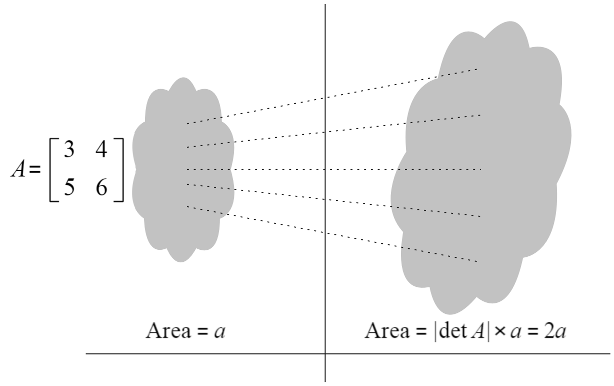


Figure 2: 2D plane: on projecting (multiplying) each point in the left shaded region with area a by matrix A and we get the shaded region on right with the area $|\det(A)|a$. (In 3D it happens for volume.)

and

$$\begin{aligned}\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} &= a \det \begin{bmatrix} e & f \\ h & i \end{bmatrix} - b \det \begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \det \begin{bmatrix} d & e \\ g & h \end{bmatrix} \\ &= aei - afh - bdi + bgf + cdh - ceg. \end{aligned} \tag{4}$$

In the above equation we expanded the determinant of the 3×3 matrix with respect to the first row. This method is called Laplace expansion which is more formally defined as follows. Let $M = (m_{ij})$ be a matrix of order n, and let $M(i|j)$ be the matrix that results from M by removing i-th row and j-th column. Then for any i the following recurrence relation holds

$$\det M = \sum_{j=1}^{n} (-1)^{i+j} m_{ij} \det M(i|j). \tag{5}$$

6.1 Elementary matrices

Recall that the following elementary row operations; these are at the heart of determinant.

1. Exchanging two rows: the determinant change its sign.
2. Multiplying a row by a nonzero constant c: the determinant get multiplied by c.
3. Adding to a row a multiple of another row: the determinant is unchanged.

If the rows (hence the columns) of M are linear dependent, by using these row operations we can make a row to be zero, which implies $\det M = 0$. Note that all the three row operations on any matrix leave the rank of the matrix unchanged.
A matrix resulting after applying one row operation to the identity matrix I is known as an elementary matrix.

Those which involve switching rows of the identity matrix are called permutation matrices. Performing any of the three row operations on a matrix A is equivalent to take the product EA, where E is the elementary matrix obtained by using the desired row operation on the identity matrix. Thus

$$\operatorname{rank} EA = \operatorname{rank} A \tag{6}$$

(similarly $\operatorname{rank} AE = \operatorname{rank} A$). Also $\det E \neq 0$ (why?). Another important fact that follows from the elementary row operations is that for any square matrix M

$$\det EM = \det E \det M. \tag{7}$$

Recall that if $\operatorname{rank} M$, that is, all the rows are independent, then there exist elementary matrices, $E_1, \ldots, E_k$ such that

$$E_k \ldots E_1 M = I. \tag{8}$$

The matrix $E_k \ldots E_1$ is known as the inverse of M, denoted M^{-1}. Thus using Equation 7 $\det M \neq 0$. For a square matrix M the inverse exists if and only $\det M \neq 0$, in this case we call M to be invertible. Since inverse of an elemantary matrix is an elementary matrix, using Equation 8 any invertible matrix can be written as the product of elemenatry matrices. Thus using 6 we see that for any invertible matrix M,

$$\operatorname{rank} MA = \operatorname{rank} A,\ \operatorname{rank} BM = \operatorname{rank} B, \tag{9}$$

where A, B are any suitable order matrices. The above discussion leads to the following important relation on determinant and determinant of the product of two square matrices A, B of the same order.

$$\det AB = \det A \det B. \tag{10}$$

Exercise 6.1. *Find the null spaces of the following $n \times n$ matrices*

$$\begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & \ldots & 1 \\ 1 & 0 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 0 \end{bmatrix}, \quad \begin{bmatrix} n & 1 & \ldots & 1 \\ 1 & n & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & n \end{bmatrix}$$

Exercise 6.2. *A minor of a $p \times q$ matrix M is the determinant of any $k \times k$ submatrix of M. Prove that $\operatorname{rank} M$ equals the size of largest nonzero minor of M.*

Exercise 6.3. *Using determinants find the rank of the following matrix.*

$$\begin{bmatrix} 1 & 1 & 2 & 3 \\ 4 & 1 & 5 & 6 \\ 7 & 1 & 8 & 9 \end{bmatrix}$$

Exercise 6.4. *If M is a nonsingular matrix, prove that $\det A^{-1} = \frac{1}{\det A}$.*

Exercise 6.5. *Let A and C be square matrices. Prove that*

$$\det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = \det A \det C.$$

Exercise 6.6. *Prove that if M is a $p \times q$ matrix, then $\det M^T M \geq 0$.*

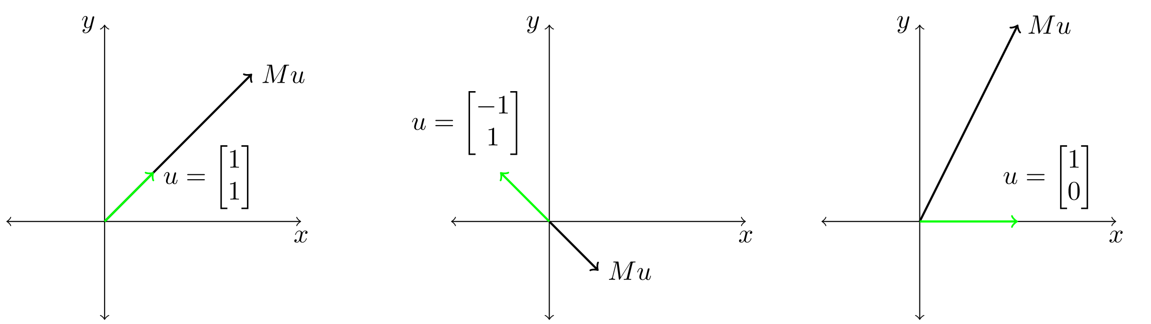
Exercise 6.7. *Prove that for a $p \times q$ matrix M, $\det M^T M > 0$ if and only if $\operatorname{rank} A = q$.*

7 Eigenvalues and eigenvectors

Till now we have seen plenty of examples for multiplying a matrix with a vector. If M is a square matrix of order n and x is a vector of order n, then Mx is some vector of order n. Usually x and Mx are oriented in different directions. But there are special vectors x such that both x and Mx have the same direction or just opposite direction. In other words, Mx is some scaling of x. These special vectors are known as eigenvectors of M, and the scaling factor is known as the eigenvalue. Formally we have

$$Mx = \lambda x. \tag{11}$$

The scalar λ is the eigenvalue, and x is a corresponding eigenvector. Notice that any nonzero multiple of x is still an eigenvector corresponding to the eigenvalue λ. Thus most often, we consider x having unit length, that is, $x^T x = 1$, also called a normal vector.



(a) u is an eigenvector, $\lambda = 3$ (b) u is an eigenvector, $\lambda = -1$ (c) u is not an eigenvector

Figure 3: Here $M = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Observe that when u is an eigenvector u and Mu are along the same line, that is, Mu is a scaled version of u.

The eigenvalues, eigenvectors play a significant role in understanding various phenomena in many branches of science. How to find out the eigenvalues and eigenvectors? In Equation (11) the only thing that we know is M; the other two are unknown. But somehow, if we know one of λ, x we can know the other. We can write

$$(M - \lambda I)x = \mathbf{0}.$$

It has the same form as Equation (2). It implies that $\det(M - \lambda I) = 0$. This gives a polynomial of λ having degree n, we call this polynomial as the characteristic polynomial. The n roots $\lambda_1, \ldots, \lambda_2$ of this polynomial are the eigenvalues of M. The set of all the eigenvalues are known as the spectrum of the matrix. We may factor the characteristic polynomial as

$$\det(M - \lambda I) = (\lambda_1 - \lambda) \dots (\lambda_n - \lambda).$$

The relationship between the eigenvalues and the determinants of submatrices is quite crucial to know. A principal submatrix of M is a submatrix formed by a set of rows and the corresponding set of columns. A principal minor of M is the determinant of a principal submatrix. We state the following theorem which tells a relationship between eigenvalues and the principal minors. A proof is given here.

Theorem 7.1. *The sum of product of the eigenvalues, of M, taken k at a time equals the sum of $k \times k$ principal minors of M.*

For $k = 1$, this implies that the sum of all the eigenvalues is equal to the trace of M (the sum of the diagonal entries). For $k = n$, this implies that the product of the eigenvalues is equal to the determinant of M.

Exercise 7.2. *Find the eigenvalues of* $\begin{bmatrix} 1 & 2 \\ k & 1 \end{bmatrix}$ *in terms of k. Also find an eigenvector corresponding to each of the eigenvalues.*

Exercise 7.3. *Find the eigenvalues and eigenvectors of the following $n \times n$ matrices*

$$\begin{bmatrix} n & 0 & \dots & 0 \\ 0 & n & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n \end{bmatrix}, \begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{bmatrix}, \begin{bmatrix} n & 1 & \dots & 1 \\ 1 & n & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & n \end{bmatrix}$$

Exercise 7.4. *Prove that, if $v_1, v_2, \dots, v_k$ are the eigenvectors of M associated with the same eigenvalue λ, then any nonzero linear combination*

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_k v_k$$

is also an eigenvector associated with λ.

Exercise 7.5. *For a square matrix M, prove that M and M^T have the same eigenvalues.*

Exercise 7.6. *For any $A_{m\times n}, B_{n\times m}$ matrices, prove that the nonzero eigenvalues of AB and BA are the same.*

8 Eigenvalues, eigenvectors of symmetric matrices

Symmetric matrices display very interesting properties with the eigenvalues and eigenvector. The eigenvalues of a symmetric matrix are real (why?). (Note that an eigenvector of a symmetric matrix can have complex entries, but we can also choose the eigenvectors to be real.)

Moreover, for distinct eigenvalues of a symmetric matrix, the orthogonality of eigenvectors comes for free. Consider a symmetric matrix M with two distinct eigenvalues α, β, and the corresponding eigenvectors x, y, respectively. That is, we have

$$Mx = \alpha x, \tag{12}$$

$$My = \beta y. \tag{13}$$

On both the side of Equation (12), take dot product with y, similarly, on both the side of Equation (13), take dot product with x, we get

$$y^T Mx = \alpha y^T x, \tag{14}$$

$$x^T My = \beta x^T y. \tag{15}$$

Taking transpose on both the sides of Equation (14), and then subtracting it from Equation (15) we get $(\alpha - \beta)(x^T y) = 0$, which implies $x^T y = 0$, that is, x, y are orthogonal.

Theorem 8.1. *For a symmetric matrix the eigenvectors associated with distinct eigenvalues are orthogonal.*

8.1 Orthogonal diagonalization

In fact for symmetric matrices we can say more than what Theorem 8.1 conveys. Even if the eigenvalues are not distinct we can choose associated eigenvectors to be orthogonal. The next theorem states this fact, a proof is given here. It is also known as the spectral theorem, and the decomposition it gives is known as eigenvalue decomposition of a symmetric matrix.

Orthogonal diagonalization (spectral theorem)

Theorem 8.2. *Let M be a symmetric matrix. Then*

$$M = \begin{bmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_n \\ | & | & \dots & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} - & v_1^T & - \\ - & v_2^T & - \\ \vdots & \vdots & \vdots \\ - & v_n^T & - \end{bmatrix}, \tag{16}$$

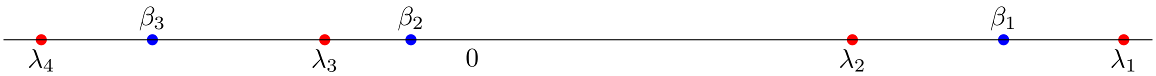
that is, $M = VDV^T$, where V is the matrix with columns $v_1, v_2, \dots, v_n$ which are the eigevectors of M corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$, and $V^T V = I$, $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$.

8.2 The interlace theorem

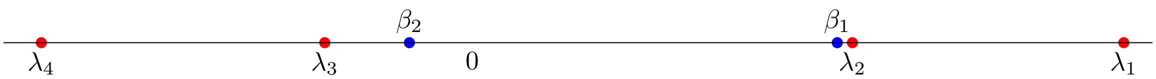
Consider the following symmetric matrix A and a matrix B which is a principal submatrix of A.

$$A = \begin{bmatrix} -2 & 4 & 2 & -4 \\ 4 & 4 & 2 & 1 \\ 2 & 2 & 0 & -3 \\ -4 & 1 & -3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & 4 & 2 \\ 4 & 4 & 2 \\ 2 & 2 & 0 \end{bmatrix}.$$

The eigenvalues of A are $\lambda_1 = 8.62, \lambda_2 = 5.03, \lambda_3 = -1.95, \lambda_4 = -5.70$, and the eigenvalues of B are $\beta_1 = 7.03, \beta_2 = -0.81, \beta_3 = -4.23$. Let us plot these eigenvalues on the number line.



Any β_i is trapped between λ_i and λ_{i+1}; we have $\lambda_{i+1} \leq \beta_i \leq \lambda_i$ for $i = 1, 2, 3$. Now let us consider another principal submatrix $\begin{bmatrix} 4 & 2 \\ 2 & 0 \end{bmatrix}$ of A. Its eigenvalues are $\beta_1 = 4.83, \beta_2 = -0.83$. Let us plot these too on the number line.



This time $\beta_1 \leq \lambda_1$ but also $\beta_1 \leq \lambda_2$. However, $\beta_1 \geq \lambda_3$. So β_1 is infact trapped between λ_1 and λ_3. The following well celebrated theorem on symmetric matrices relates the eigenvalue of a matrix with the eigenvalues of principal submatrices. A proof is given here.

8.2.1 Cauchy interlace theorem (Eigenvalue interlacing theorem)

Theorem 8.3. *Let M be a symmetric matrix of order n, and B be its principal submatrix of order m, where, $m < n$. Suppose the eigenvalues of A are $\lambda_1 \geq \cdots \geq \lambda_n$ and the eigenvalues of B are $\beta_1 \geq \cdots \geq \beta_m$. Then*

$$\lambda_i \geq \beta_i \geq \lambda_{i+n-m}, \quad \text{for } i = 1, \ldots, m.$$

In particular if $m = n - 1$, then

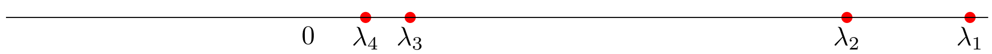
$$\lambda_1 \geq \beta_1 \geq \lambda_2 \geq \cdots \geq \beta_{n-1} \geq \lambda_n.$$

8.3 Definite, semi-definite matrices

Now consider the following symmetric matrix

$$M = \begin{bmatrix} 3 & 0 & -1 & -3 \\ 0 & 7 & 1 & -1 \\ -1 & 1 & 2 & 2 \\ -3 & -1 & 2 & 6 \end{bmatrix}.$$

Its eigenvalues are $\lambda_1 = 8.76, \lambda_2 = 7.13, \lambda_3 = 1.35, \lambda_4 = 0.76$. On the number line they are placed as follows.



All the eigenvalues of M are nonnegative. A symmetric matrix whose all the eigenvalues are nonnegative is known as positive semi-definite matrix, and when the eigenvalues are positive it is known as positive definite matrix. Check that M is positive definite. These matrices have very high importance for different applications and display beautiful properties. We will see two equivalent criteria for a matrix to be positive semidefinite or definite.

Let M be a symmetric positive semidefinite matrix of order n. Using 8.2 we can write $M = VDV^T$, where V is the matrix with columns $v_1, v_2, \ldots, v_n$ which are the eigenvectors of M corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$, and V is a orthogonal matrix. Since the eigenvalues are nonnegative we can write $D = D^{\frac{1}{2}} D^{\frac{1}{2}}$, where $D^{\frac{1}{2}}$ is a diagonal matrix with i-th diagonal entry equals to $\sqrt{\lambda_i}$. Let $\hat{V} = VD^{\frac{1}{2}}$, then we can write $M = VD^{\frac{1}{2}} D^{\frac{1}{2}} V^T = \hat{V}\hat{V}^T$. It implies that for any nonzero vector x, we have $x^T M x = x^T \hat{V}\hat{V}^T x = (\hat{V}^T x)^T (\hat{V}^T x) \geq 0$. We will see that this is a very useful information. When M is positive definite, that is, all the eigenvalues are positive, then $\hat{V}$ has rank n and so $\hat{V}^T x$ is can not be a zero vector unless x is zero vector. Since x is nonzero this implies that $x^T M x > 0$.

Now, assume that $x^T M x \geq 0$, for any symmetric matrix M and nonzero vector x. Using the fact that the smallest eigenvalue λ_n of M is equal to

$$\lambda_n = \min_{||x||=1} x^T M x = \min \frac{x^T M x}{||x||^2}.$$

(a proof is given here.) Since $x^T M x \geq 0$, we have, $\lambda_n \geq 0$ which implies that all the eigenvalues are nonnegative. If $x^T M x > 0$, it implies that all the eigenvalues are positive. Thus we state the following theorem.

Theorem 8.4. *Let M be a symmetric matrix. For any nonzero vector x,*

1. *$x^T M x \geq 0$ if and only if all the eigenvalues of M are nonnegative. (M is called positive semi-definite matrix.)*
2. *$x^T M x > 0$ if and only if all the eigenvalues of M are positive. (M is called positive definite matrix.)*

Exercise 8.5. *Prove that the rank of a symmetric matrix is equal to the number of its nonzero eigenvalues.*

Exercise 8.6. *Let A be a $n \times m$ matrix with $n \leq m$, then prove that $A^T A$ is a positive semi-definite matrix. Further show that each eigenvalue of AA^T is also an eigenvalue of $A^T A$, and the rest of the $m-n$ eigenvalues are zero.*

Exercise 8.7. *Show that* trace $A^T A = 0$ *if and only if the eigenvalues of $A^T A$ are zero.*

Exercise 8.8. *For the matrix*

$$M = \begin{bmatrix} 2 & -1 & 3 \\ -1 & 0 & 4 \\ 3 & 4 & 1 \end{bmatrix}$$

compute an orthogonal matrix V and a diagonal matrix D such that $M = VDV^T$.

Exercise 8.9. *Show that for any matrix $A_{m\times n}$, nullspaces of A and $A^T A$ are the same. Also, the column spaces of A and $A^T A$ are the same.*

Theorem 8.10. *For the matrix*

$$M = \begin{bmatrix} 0 & -1 & 2 \\ -1 & 0 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

compute the eigenvalues, and verify the Cauchy interlace theorem by computing the eigenvalues of its principal submatrices.