Chapter 2

Matrices

In this chapter, we review matrix algebra from *Linear Algebra I*, consider row and column operations on matrices, and define the rank of a matrix. Along the way prove that the "row rank" and "column rank" defined in *Linear Algebra I* are in fact equal.

2.1 Matrix algebra

Definition 2.1. A *matrix* of size $m \times n$ over a field $\mathbb{K}$, where m and n are positive integers, is an array with m rows and n columns, where each entry is an element of $\mathbb{K}$. The matrix will typically be denoted by an upper case letter, and its entries by the corresponding lower case letter. Thus, for $1 \le i \le m$ and $1 \le j \le n$, the entry in row i and column j of matrix A is denoted by a_{ij}, and referred to as the (i, j) entry of A.

Example 2.2. A column vector in $\mathbb{K}^n$ can be thought of as a $n \times 1$ matrix, while a row vector is a $1 \times n$ matrix.

Definition 2.3. We define addition and multiplication of matrices as follows.

(a) Let $A = (a_{ij})$ and $B = (b_{ij})$ be matrices of the same size $m \times n$ over $\mathbb{K}$. Then the sum $C = A + B$ is defined by adding corresponding entries of A and B; thus $C = (c_{ij})$ is given by
$$c_{ij} = a_{ij} + b_{ij}.$$

(b) Let A be an $m \times n$ matrix and B an $n \times p$ matrix over $\mathbb{K}$. Then the product $C = AB$ is the $m \times p$ matrix whose (i, j) entry is obtained by multiplying each element in the ith row of A by the corresponding element in the jth column of B and summing:
$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Remark Note that we can only add or multiply matrices if their sizes satisfy appropriate conditions. In particular, for a fixed value of n, we can add and multiply $n \times n$ matrices. Technically, the set $M_n(\mathbb{K})$ of $n \times n$ matrices over $\mathbb{K}$ together with matrix addition and multiplication is a ring (with identity). The zero matrix, which we denote by O, is the matrix with every entry zero, while the identity matrix, which we denote by I, is the matrix with entries 1 on the main diagonal and 0 everywhere else. Note that matrix multiplication is not commutative: BA is usually not equal to AB.

We already met matrix multiplication in Section 1 of the notes: recall that if $P_{B,B'}$ denotes the transition matrix between two bases of a vector space, then

$$P_{B,B'}P_{B',B''} = P_{B,B''}.$$

2.2 Row and column operations

Given an $m \times n$ matrix A over a field $\mathbb{K}$, we define certain operations on A called row and column operations.

Definition 2.4. *Elementary row operations.* There are three types:

Type 1. Add a multiple of the jth row to the ith, where $j \neq i$.

Type 2. Multiply the ith row by a non-zero scalar.

Type 3. Interchange the ith and jth rows, where $j \neq i$.

Elementary column operations. There are three types:

Type 1. Add a multiple of the jth column to the ith, where $j \neq i$.

Type 2. Multiply the ith column by a non-zero scalar.

Type 3. Interchange the ith and jth columns, where $j \neq i$.

We can describe the elementary row and column operations in a different way. For each elementary row operation on an $m \times n$ matrix A, we define a corresponding *elementary matrix* by applying the same operation to the $m \times m$ identity matrix I. Similarly for each elementary column operation we define a corresponding elementary matrix by applying the same operation to the $n \times n$ identity matrix.

We don't have to distinguish between rows and columns for our elementary matrices: each matrix can be considered either as a row or a column operation. This observation will be important later. For example, the matrix

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

corresponds to the elementary column operation of adding twice the first column to the second, or to the elementary row operation of adding twice the second row to the first. For the other types, the matrices for row operations and column operations are identical.

Lemma 2.5. *The effect of an elementary row operation on a matrix is the same as that of multiplying on the left by the corresponding elementary matrix. Similarly, the effect of an elementary column operation is the same as that of multiplying on the right by the corresponding elementary matrix.*

The proof of this lemma is somewhat tedious calculation.

Example 2.6. Let A be a 2×3 real matrix. The matrices corresponding to the elementary row operation of subtracting 4 times row 1 from row 2, and the elementary column operation of subtracting twice column 1 from column 2 are

$$\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

respectively. If A is the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix},$$

then the matrix that results from applying the above two elementary operations ought to be

$$\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & -3 & -6 \end{bmatrix}.$$

You should check that this is indeed the case.

An important observation about the elementary operations is that each of them can have its effect undone by another elementary operation of the same kind, and hence every elementary matrix is invertible, with its inverse being another elementary matrix of the same kind. For example, the effect of adding twice the first row to the second is undone by adding -2 times the first row to the second, so that

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

2.3 Rank

Recall from *Linear Algebra I* the definitions of row space, column space, row rank and column rank of a matrix.

Definition 2.7. Let A be an $m \times n$ matrix over a field $\mathbb{K}$. The *row space* of A is the vector space spanned by rows of A and the *column space* the vector space spanned by columns. The *row rank* of A is the dimension of the row space, and the *column rank* of A the dimension of the column space of A. (We regard columns or rows as vectors in $\mathbb{K}^m$ and $\mathbb{K}^n$ respectively.)

Remark 2.8. Since a maximal linearly independent set of vectors is a basis, we could alternatively define row rank as the maximum number of linearly independent rows, and the column rank analogously.

Recall also that elementary row operations preserve row rank, and elementary column operations preserve column rank. In *Linear Algebra I*, the rank of a matrix was defined as its row rank. Why? The definition privileges rows over columns, and hence seems somewhat arbitrary. In any case, why should the dimension of the row space be a significant parameter?

The next lemma goes beyond *Linear Algebra I* by showing that elementary row operations preserve column rank, not just row rank.

Lemma 2.9. *(a) Elementary column operations preserve the column space of a matrix (and hence don't change the column rank).*

(b) Elementary row operations preserve the row space of a matrix (and hence don't change the row rank).

(c) Elementary row operations don't change the column rank of a matrix.

(d) Elementary column operations don't change the row rank of a matrix.

Proof. (a) Suppose some elementary column operation is applied to matrix A to yield A'. Each column of A' is a linear combination of columns of A. Thus the column space of A' is contained in the column space of A. On the other hand, the inverse of an elementary operation is an elementary operation of the same kind taking A' to A. It follows that the column space of A is contained in the column space of A', and hence the column spaces of A and A' are equal. Since the column spaces of the matrices are equal, their column ranks are equal too.

(b) Follows by symmetry from (a).

(c) Suppose some elementary row operation is applied to matrix A to yield A'. It is not to difficult to check (some details below) that any linear dependency between the columns of A is also a linear dependency between the columns of A'. So any list of columns that is linearly dependent in A is also linearly dependent in A'. Another way of saying the same thing is that every list of columns that is linearly independent in A' is linearly independent in A. It follows that the column rank of A is greater than or equal to the column rank of A'. As before, the inverse of an elementary operation is an elementary operation of the same kind, which implies that the column ranks of A and A' are equal.

Let us check the above claim that linear dependencies are preserved by row operations. We'll just deal with Type 1 operations; the easier Type 2 and 3 operations are left as exercises. Denote the columns of A by column vectors $v_1, \ldots, v_n$. Consider some linear dependency $c_1v_1 + \cdots + c_nv_n = \mathbf{0}$ on the columns of A, where the scalars c_i not all zero. In terms of the matrix entries of A, this means, for every row k of A, that $c_1a_{k1} + \cdots + c_na_{kn} = 0$. Now apply the Type 1 operation that adds a multiple d of row j to row i. The only row of A that has changed is row i, so we just need to check that the linear dependency continues to hold for that particular row. We have

$$\begin{aligned} c_1(a_{i1} + da_{j1}) + \cdots + c_n(a_{in} + da_{jn}) &= c_1a_{i1} + \cdots + c_na_{in} \\ &\quad + d(c_1a_{j1} + \cdots + c_na_{jn}) \\ &= 0 + d \cdot 0 = 0. \end{aligned}$$

In a similar manner, you may check that if an elementary row operation of Type 2 or Type 3 is applied, then the new columns satisfy exactly the same linear relations as the old ones (that is, the same linear combinations are zero). So a linearly (in)dependent set of columns in A remains linearly (in)dependent in A' after any elementary row operation.

(d) Follows from (c) by symmetry.

□

It is important to note that elementary row operations do *not* in general preserve the column space of a matrix, only the column rank. Provide a counterexample to illustrate this fact. (An elementary row operation on a 2×2 matrix is enough for this purpose.)

By applying elementary row and column operations, we can reduce any matrix to a particularly simple form:

Theorem 2.10. *Let A be an $m \times n$ matrix over the field $\mathbb{K}$. Then it is possible to transform A by elementary row and column operations into a matrix $D = (d_{ij})$ of the* *same size as A, with the following special form: there is an $r \leq \min\{m, n\}$, such that $d_{ii} = 1$ for $1 \leq i \leq r$, and $d_{ij} = 0$ otherwise.*

The matrix D (and hence the number r), is uniquely defined: if A can be reduced to two matrices D and D', both of the above form, by different sequences of elementary operations then $D = D'$.

Definition 2.11. The number r in the above theorem is called the *rank* of A; while a matrix of the form described for D is said to be in the *canonical form for equivalence.* We can write the canonical form matrix in "block form" as

$$D = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix},$$

where I_r is an $r \times r$ identity matrix and O denotes a zero matrix of the appropriate size (that is, $r \times (n-r)$, $(m-r) \times r$, and $(m-r) \times (n-r)$ respectively for the three Os). Note that some or all of these Os may be missing: for example, if $r = m$, we just have $\begin{bmatrix} I_m & O \end{bmatrix}$.

Proof of Theorem 2.10. We first outline the proof that the reduction is possible. The proof is by induction on the size of the matrix $A = (a_{ij})$. Specifically, we assume as inductive hypothesis that any smaller matrix can be reduced as in the theorem. Let the matrix A be given. We proceed in steps as follows:

- If $A = O$ (the all-zero matrix), then the conclusion of the theorem holds, with $r = 0$; no reduction is required. So assume that $A \neq O$.

- If $a_{11} \neq 0$, then skip this step. If $a_{11} = 0$, then there is a non-zero element a_{ij} somewhere in A; by swapping the first and ith rows, and the first and jth columns, if necessary (Type 3 operations), we can bring this entry into the $(1, 1)$ position.

- Now we can assume that $a_{11} \neq 0$. Multiplying the first row by a_{11}^{-1}, (row operation Type 2), we obtain a matrix with $a_{11} = 1$.

- Now by row and column operations of Type 1, we can assume that all the other elements in the first row and column are zero. For if $a_{1j} \neq 0$, then subtracting a_{1j} times the first column from the jth gives a matrix with $a_{1j} = 0$. Repeat this until all non-zero elements have been removed.

- Now let A' be the matrix obtained by deleting the first row and column of A. Then A' is smaller than A and so, by the inductive hypothesis, we can reduce A' to canonical form by elementary row and column operations. The same sequence of operations applied to A now finishes the job.

Suppose that we reduce A to canonical form D by elementary operations, where D has r 1s on the diagonal. These elementary operations don't change the row or column rank, by Lemma 2.9. Therefore, the row ranks of A and D are equal, and their column ranks are equal. But it is not difficult to see that, if

$$D = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix},$$

then the row and column ranks of D are both equal to r. It doesn't matter which elementary operations we use to reduce to canonical form, we will always obtain the same matrix D. So the theorem is proved. □

Corollary 2.12. *For any matrix A, the row rank, the column rank, and the rank are all equal. In particular, the rank is independent of the row and column operations used to compute it.*

Example 2.13. Here is a small example. Let

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

We have $a_{11} = 1$, so we can skip the first three steps. So first we subtract 4 times the first row from the second, then subtract twice the first column from the second, and then 3 times the first column from the third. These steps yield the following sequence of matrices

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \xrightarrow{\mathrm{R}_2 - 4\mathrm{R}_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \end{bmatrix} \xrightarrow{\mathrm{C}_2 - 2\mathrm{C}_1} \begin{bmatrix} 1 & 0 & 3 \\ 0 & -3 & -6 \end{bmatrix} \xrightarrow{\mathrm{C}_3 - 3\mathrm{C}_1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & -6 \end{bmatrix}.$$

At this point we have successfully set to zero the first row and column of the matrix, except for the top left entry. From now on, we have to operate on the smaller matrix $\begin{bmatrix} -3 & -6 \end{bmatrix}$, but we continue to apply the operations to the large matrix.

Multiply the second row of the matrix by $-\frac{1}{3}$ and finally subtract twice the second column from the thord. Picking up from where we left off, this yields the sequence

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & -6 \end{bmatrix} \xrightarrow{-\frac{1}{3}\mathrm{R}_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow{\mathrm{C}_3 - 2\mathrm{C}_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = D.$$

For compactness, se are using (as in *Linear Algebra I*) shorthand such as $\mathrm{R}_2 - 4\mathrm{R}_1$ for $\mathrm{R}_2 := \mathrm{R}_2 - 4\mathrm{R}_1$ and $-\frac{1}{3}\mathrm{R}_2$ for $\mathrm{R}_2 := -\frac{1}{3}\mathrm{R}_2$. We have finished the reduction to canonical form, and we conclude that the rank of the original matrix A is equal to 2.

Theorem 2.14. *For any $m \times n$ matrix A there are invertible matrices P and Q of sizes $m \times m$ and $n \times n$ respectively, such that $D = PAQ$ is in the canonical form for equivalence. The rank of A is equal to the rank of D. Moreover, P and Q are products of elementary matrices.*

Proof. We know from Theorem 2.10 that there is a sequence of elementary row and column operations that reduces A to D. These operations correspond to certain elementary matrices. Take the matrices $R_1, R_2, \ldots, R_s$ corresponding to the row operations and multiply them together (right to left). This is the matrix $P = R_s R_{s-1} \cdots R_1$. Take the matrices $C_1, C_2, \ldots, C_t$ corresponding to the column operations and multiply them together (left to right). This is the matrix $Q = C_1 C_2 \ldots C_t$. □

Example 2.15. We illustrate the construction of P and Q in the above proof, in a continuation of our previous example. In order, here is the list of elementary matrices corresponding to the operations we applied to A. (Here, 2×2 matrices are row operations while 3×3 matrices are column operations).

$$R_1 = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix}, \quad C_1 = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$R_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1/3 \end{bmatrix}, \quad C_3 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}.$$

So the whole process can be written as a matrix equation:

$$D = R_2 R_1 A C_1 C_2 C_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1/3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} A \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

or more simply

$$D = \begin{bmatrix} 1 & 0 \\ 4/3 & -1/3 \end{bmatrix} A \begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix},$$

where, as before,

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

There is a slightly easier (for humans) method for constructing the matrices P and Q, which we examined in the lectures. Let's recall how it works in the context of computing the matrix Q. The idea is to use the same column operations we applied to A, but *starting instead with the* 3×3 *identity matrix* I_3 :

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\mathrm{C}_2 - 2\mathrm{C}_1} \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\mathrm{C}_3 - 3\mathrm{C}_1} \begin{bmatrix} 1 & -2 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\mathrm{C}_3 - 2\mathrm{C}_2} \begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} = Q.$$

Think about why this method works. It is doing essentially the same calculation, but arranging it in a more human-friendly way.

Definition 2.16. The $m \times n$ matrices A and B are said to be *equivalent* if $B = PAQ$, where P and Q are invertible matrices of sizes $m \times m$ and $n \times n$ respectively.

Remark 2.17. The relation "equivalence" defined above is an equivalence relation on the set of all $m \times n$ matrices; that is, it is reflexive, symmetric and transitive.

Corollary 2.18. *An $n \times n$ matrix is invertible if and only if it has rank n.*

Proof. Suppose that $n \times n$ matrices A and B are equivalent. Then A is invertible if and only if B is invertible. (If A is invertible and $B = PAQ$, then $Q^{-1}A^{-1}P^{-1}$ is the inverse of B, and similarly in the other direction.) We know from Theorem 2.14 that every matrix A is equivalent to some matrix D in the canonical form for equivalence. Moreover the rank of A is equal to the rank of D. Thus, we have the the following chain of implications:

$$A \text{ is invertible} \iff D \text{ is invertible} \iff D = I_n \iff A \text{ has rank } n.$$

□

Corollary 2.19. *Every invertible square matrix is a product of elementary matrices.*

Proof. If A is an invertible $n \times n$ matrix, then it has rank n and its canonical form is the identity matrix I_n. Thus there are invertible matrices P and Q, each a product of elementary matrices, such that

$$PAQ = I_n.$$

From this we deduce that

$$A = P^{-1} I_n Q^{-1} = P^{-1} Q^{-1}.$$

Since the elementary matrices are closed under taking inverses, the above is an expression for A as a product of elementary matrices. □

Corollary 2.20. *If A is an invertible $n \times n$ matrix, then A can be transformed into the identity matrix by elementary column operations alone (or by elementary row operations alone).*

Proof. We observed, when we defined elementary matrices, that they can represent either elementary column operations or elementary row operations. In the previous corollary, we saw that A can be written as a product of elementary matrices, say $A = C_1 C_2 \dots C_t$. We can transform A to the identity by multiplying on the right by $C_t^{-1}, \dots, C_2^{-1}, C_1^{-1}$ in turn. This is equivalent to applying a sequence of column operations. Equally, we can transform A to the identity by multiplying on the left by $C_1^{-1}, C_2^{-1}, \dots, C_t^{-1}$ in turn. This is equivalent to applying a sequence of row operations. □

Theorem 2.21. *Two matrices are equivalent if and only if they have the same rank.*

Proof. Suppose A and B are (not necessarily square) equivalent matrices, i.e., $B = PAQ$ for some invertible matrices P and Q. By Corollary 2.19 we can write P and Q as the product of elementary matrices. It follows that we can transform A to B by elementary row and column operations, and hence the ranks of A and B are the same. (Elementary operations preserve the rank.)

Conversely, if the ranks of A and B are the same then we can transform one to the other (e.g., via the common canonical form D) by elementary row and column operations, and hence A and B are equivalent. □

When mathematicians talk about a "canonical form" for an equivalence relation, they mean a set of objects which are representatives of the equivalence classes: that is, every object is equivalent to a unique object in the canonical form. Theorem 2.21 says that in this case there are $\min\{m, n\} + 1$ equivalence classes, and the canonical form for equivalence is a canonical form in this sense.

Remark 2.22. As with Chapter 1, the results in this chapter apply to all fields $\mathbb{K}$.