Elementary row operations and some applications

1. Elementary row operations

Given an $N \times N$ matrix A, we can perform various operations that modify some of the rows of A. There are three classes of elementary row operations, which we shall denote using the following notation:

1. $R_j \leftrightarrow R_k$. This means that we interchange the jth row and kth row of A.
2. $R_j \to cR_j$, where $c \neq 0$ is a real or complex number. This means that we multiply *all* elements of the jth row by the same constant c.
3. $R_j \to R_j + cR_k$, where $c \neq 0$ is a real or complex number. This means that we add c times each element in the kth row to the corresponding element of the jth row.

To perform an elementary row operation, it suffices to multiply the matrix A from the *left* by the corresponding elementary matrix. These matrices have the following structure. For $R_j \leftrightarrow R_k$, the corresponding elementary matrix $E^{(1)}$ has non-zero matrix elements given by:

$$a_{jk} = a_{jk} = a_{nn} = 1\,, \qquad \text{for } 1 \leq n \leq N \text{ and } n \neq j,\, n \neq k\,,$$

with all other matrix elements equal to zero. That is, $E^{(1)}$ can be obtained from the $N \times N$ identity matrix by replacing the jj and kk elements (originally equal to 1) by 0 and by replacing the jk and kj elements (originally equal to 0) by 1.

For $R_j \to cR_j$, the corresponding elementary matrix $E^{(2)}$ has non-zero matrix elements given by:

$$a_{nn} = 1\,, \qquad \text{for } 1 \leq n \leq N \text{ and } n \neq j\,, \quad \text{and} \quad a_{jj} = c\,,$$

with all other matrix elements equal to zero. That is, $E^{(2)}$ can be obtained from the $N \times N$ identity matrix by replacing the jj element (originally equal to 1) by c.

For $R_j \to R_j + cR_k$, the corresponding elementary matrix $E^{(3)}$ has non-zero matrix elements given by:

$$a_{nn} = 1\,, \qquad \text{for } 1 \leq n \leq N\,, \quad \text{and} \quad a_{jk} = c\,,$$

with all other matrix elements equal to zero. That is, $E^{(3)}$ can be obtained from the $N \times N$ identity matrix by replacing the jk element (originally equal to 0) by c.

You should check that for any matrix A,

1. $E^{(1)}A$ is a matrix obtained from A by interchanging the jth and kth rows of A.
2. $E^{(2)}A$ is a matrix obtained from A by multiplying the jth rows of A by c.
3. $E^{(3)}A$ is a matrix obtained from A by adding c times the kth row of A to the jth row of A .

The following properties of the elementary matrices are noteworthy:

$$\det E^{(1)} = -1\,, \qquad \det E^{(2)} = c\,, \qquad \det E^{(3)} = 1\,. \tag{1}$$

Using the fact that $\det(AB) = (\det A)(\det B)$, it immediately follows from eq. (1) that

1. Under, $R_j \leftrightarrow R_k$, the determinant of A changes by an overall sign.
2. Under $R_j \to cR_j$, the determinant of A changes by a multiplicative factor of c.
3. Under $R_j \to R_j + cR_k$, the determinant of A is unchanged.

2. Reduced row echelon form

Given an $m \times n$ matrix A (where m is not necessarily equal to n), we can perform a series of elementary row operations by multiplication on the left by a series of elementary matrices (of the three types introduced in Section 1 above). By an appropriate set of steps, one can always reduce A into what is called *reduced row echelon form*. A matrix that is in reduced row echelon form possesses the following properties:

1. All zero rows* appear in the bottom position of the matrix. That is, no non-zero row can appear below a row of zeros.
2. Reading from left to right, the first non-zero element in a non-zero row† is a 1, which we shall call *the leading 1.*
3. For $j = 2, 3, 4, \ldots, n$, the leading 1 in row j (if it exists) appears to the right of the leading 1 in row $j-1$.
4. Any column that contains a leading 1 has all other elements in that column equal to zero.

*A zero row is defined to be a row of a matrix where all elements of the row are equal to zero. If at least one element of a row is non-zero, we call that row a non-zero row.

†In modern books on matrices and linear algebra, the first non-zero element in a non-zero row is called the *pivot*. This nomenclature will not be employed further in these notes.

To illustrate the concept, we show five matrices that are in reduced row echelon form:

$$\begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2}$$

Starting from an arbitrary $m \times n$ matrix $A = [a_{ij}]$, one can achieve the reduced row echelon form by the following series of steps known as the *Gauss-Jordan elimination* procedure. Start with column 1. If $a_{11} \neq 0$, then perform $R_1 \to R_1/a_{11}$ to get the leading 1 in the first row. If $a_{11} = 0$, then choose another row with a non-zero element of the first column (suppose this is row $k > 1$) and perform the row interchange $R_1 \leftrightarrow R_k$. Then, one can perform the row operations $R_j \to R_j - a_{j1}R_1$ for $j = 2, 3, 4, \ldots, m$ to obtain zeros in all remaining elements of the first column. Note that if all the elements of the first column are zero, then there is no leading 1 in the first column.

One can now move to the second column. If there is no leading 1 in the first column, then one repeats the previous analysis in column 2 (e.g. if $a_{21} \neq 0$, then perform $R_1 \to R_1/a_{21}$ to get the leading 1 in the second row, etc.). If there is a leading 1 in the first column, then one begins with the second row and second column. If $a_{22} \neq 0$, then perform $R_2 \to R_2/a_{22}$ to get the leading 1 in the second row. If $a_{22} = 0$, then choose another row *below* the second row (suppose this is row $\ell > 2$) and perform the row interchange $R_2 \leftrightarrow R_\ell$. Then, one can perform the row operations $R_j \to R_j - a_{j2}R_2$ for $j = 1, 3, 4, \ldots, m$ to obtain zeros in all remaining elements of the second column. Note that if $a_{j2} = 0$ for $j = 2, 3, 4, \ldots, m$, then one is left with a potentially non-zero a_{12}. But this is consistent with the reduced row echelon form, as in this case there would be no leading 1 in the second column.

One can now move to the third column. It should now be obvious how to proceed until the end. When one has carried this procedure through to the nth column, what remains after the final elementary row operations have been performed is the reduced row echelon form. Two important consequence of the procedure described above are:

1. The reduced row echelon form of a given matrix is unique.

2. If A is an $n \times n$ invertible matrix, then its reduced row echelon form is the $n \times n$ identity matrix.

To exhibit the procedure outlined above, we shall compute the reduced row echelon form of the matrix A given by:

$$A = \begin{pmatrix} 2 & 3 & -1 & -2 \\ 1 & 2 & -1 & 4 \\ 4 & 7 & -3 & 11 \end{pmatrix} \tag{3}$$

First, we interchange the first two rows, corresponding to the elementary row operation $R_1 \leftrightarrow R_2$ to obtain

$$\begin{pmatrix} 1 & 2 & -1 & 4 \\ 2 & 3 & -1 & -2 \\ 4 & 7 & -3 & 11 \end{pmatrix}.$$

Next, we perform the elementary row operations $R_2 \to R_2 - 2R_1$ and $R_3 \to R_3 - 4R_1$ to obtain

$$\begin{pmatrix} 1 & 2 & -1 & 4 \\ 0 & -1 & 1 & -10 \\ 0 & -1 & 1 & -5 \end{pmatrix} .$$

Next, we perform the elementary row operations $R_3 \to R_2 - R_2$ to obtain

$$\begin{pmatrix} 1 & 2 & -1 & 4 \\ 0 & -1 & 1 & -10 \\ 0 & 0 & 0 & 5 \end{pmatrix} .$$

Next, we perform the elementary row operations, $R_2 \to -R_2$ and $R_3 \to \frac{1}{5}R_3$, followed by $R_1 \to R_1 - 2R_2$ to obtain

$$\begin{pmatrix} 1 & 0 & 1 & -16 \\ 0 & 1 & -1 & 10 \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

Finally, we finish by performing the elementary row operations $R_1 \to R_1 + 16R_3$ and $R_2 \to R_2 - 10R_3$ to obtain

$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{4}$$

We have therefore successfully achieved the reduced row echelon form.

We end this section with a few observations about the reduced row echelon form. An examination of eq. (4) shows that the following properties are common to all matrices in reduced row echelon form. Let k be the number of non-zero rows of the reduced row echelon form (in the above example, $k = 3$). Then, the number of leading 1's is also equal to k. Moreover, if we treat the rows of A as the components of vectors, then the maximal number of linearly independent row vectors is also equal to k. Finally, we define the *basic columns* of A to be the columns corresponding to the columns of the reduced row echelon form that contain leading 1's (these are the first, second and fourth columns in the example given above). If we treat the columns of A as components of vectors, then the basic column vectors are a maximal set of linearly independent column vectors. Thus, the maximal number of linearly independent column vectors is also equal to k. This number k is called the rank of the matrix A. As an example, the computation presented at the end of Section 2 implies that the rank of the matrix A defined in eq. (3) is equal to three. Similarly, the ranks of the five matrices exhibited in in eq. (2) are easily obtained and are 2, 2, 1, 2, and 3, respectively.

In Section 4, we shall provide proofs of the statements made in the previous paragraph. Further details can be found in the references quoted at the end of these notes.

3. A method for computing the inverse of an $n \times n$ matrix

We present below the *Gauss-Jordan elimination* method for computing the inverse of a matrix. Consider an $n \times n$ invertible matrix A. Then its reduced row echelon form is the $n \times n$ identity matrix (denoted by $\mathbf{I}$), as noted in Section 2. That is, there exist a sequence of elementary matrices E_1, E_2,$E_3 \ldots E_\ell$ such that:

$$E_\ell \cdots E_3 E_2 E_1 A = \mathbf{I}\,.$$

But the inverse of A satisfies $A^{-1}A = \mathbf{I}$. Hence, we can identify

$$A^{-1} = E_\ell \cdots E_3 E_2 E_1\,. \tag{5}$$

If we apply the *identical* sequence of elementary row operations to the identity matrix, then we would find:

$$E_\ell \cdots E_3 E_2 E_1 \mathbf{I} = E_\ell \cdots E_3 E_2 E_1 = A^{-1}\,, \tag{6}$$

after using eq. (5). Thus, the following procedure can be used to compute the inverse of A. Write down the matrices A and $\mathbf{I}$ next to each other. Then perform a sequence of row operations to reduce A to reduced row echelon form. Meanwhile, perform the exact same sequence of elementary row operations on $\mathbf{I}$. If the reduced row echelon form of A is the identity matrix, then the result of applying the exact same sequence of elementary row operations on $\mathbf{I}$ will yield A^{-1} as shown in eq. (6). If the reduced row echelon form of A is not the identity matrix, then it must have at least one row of zeros, and we conclude that it is not invertible.

To exhibit the Gauss-Jordan elimination procedure for computing the inverse of a matrix, we start with the matrix B (whose inverse we wish to compute) on the left and the 3×3 identity matrix $\mathbf{I}$ on the right.

$$B = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 2 & 1 & 2 \end{pmatrix} \qquad \mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

We now perform an identical set of elementary row operations on B and $\mathbf{I}$ in order to reduce B to its reduced row echelon form. First, we take $R_2 \to R_2 - 2R_1$ and $R_3 \to R_3 - 2R_1$ to obtain:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$

Next, we interchange rows, $R_2 \leftrightarrow R_3$ to obtain

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ -2 & 0 & 1 \\ -2 & 1 & 0 \end{pmatrix}$$

Next, we take $R_3 \to R_3 - R_2$ and then multiplying the resulting third row by -1 to obtain

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad\qquad \begin{pmatrix} 1 & 0 & 0 \\ -2 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}$$

Finally, we take $R_1 \to R_1 - R_3$ to obtain

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad\qquad B^{-1} = \begin{pmatrix} 1 & 1 & -1 \\ -2 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}$$

thereby producing the desired inverse matrix.

4. The rank of a matrix and its transpose

At the end of Section 2, we defined the rank of a matrix A to be the number of non-zero rows of its reduced row echelon form. Equivalently, the rank is equal to the number of leading 1's that appear in the reduced row echelon form.

Given an $m \times n$ matrix A it is instructive to consider the elements of each row of A as the components of an n-component vector. That is, the matrix A consists of m *row vectors*. We now pose the following question: what is the maximal number of linearly independent vectors among the set of m row vectors that comprise the matrix A? Suppose that there are at most k linearly independent row vectors, which we shall denote by $\{\vec{\boldsymbol{v}}_1\,,\,\vec{\boldsymbol{v}}_2\,,\,\ldots\,,\,\vec{\boldsymbol{v}}_k\}$. This means that each of the remaining row vectors, $\{\vec{\boldsymbol{v}}_{k+1}\,,\,\vec{\boldsymbol{v}}_{k+2}\,,\,\ldots\,,\,\vec{\boldsymbol{v}}_m\}$ can be expressed as a linear combination of vectors from the set $\{\vec{\boldsymbol{v}}_1\,,\,\vec{\boldsymbol{v}}_2\,,\,\ldots\,,\,\vec{\boldsymbol{v}}_k\}$. We now demonstrate that if one performs any elementary row operation on A, then the maximal number of linearly independent row vectors of the new matrix is unchanged. By definition of linear independence, the equation

$$c_1\vec{\boldsymbol{v}}_1 + c_2\vec{\boldsymbol{v}}_1 + \cdots + c_k\vec{\boldsymbol{v}}_k = 0\,,$$

has a unique solution of $c_1 = c_2 = c_3 = \cdots = c_k = 0$. Consider the three possible classes of elementary row operations. It is obvious that $R_i \leftrightarrow R_j$ and $R_i \to cR_i$ do not modify the maximal number of linearly independent row vectors. Thus, we focus on $R_i \to R_i + cR_j$. If $\vec{\boldsymbol{v}}_i$ and $\vec{\boldsymbol{v}}_j$ are members of the linearly independent set of row vectors,[‡] then we consider the solution to the equation

$$c_1\vec{\boldsymbol{v}}_1 + c_2\vec{\boldsymbol{v}}_1 + \cdots + c_i(\vec{\boldsymbol{v}}_i + c\vec{\boldsymbol{v}}_j) + \cdots + c_j\vec{\boldsymbol{v}}_j + \cdots + c_k\vec{\boldsymbol{v}}_k = 0\,, \tag{7}$$

which can be rewritten as:

$$c_1\vec{\boldsymbol{v}}_1 + c_2\vec{\boldsymbol{v}}_1 + \cdots + c_i\vec{\boldsymbol{v}}_i + \cdots + (cc_i + c_j)\vec{\boldsymbol{v}}_j + \cdots + c_k\vec{\boldsymbol{v}}_k = 0\,, \tag{8}$$

[‡]If $\vec{\boldsymbol{v}}_i$ and/or $\vec{\boldsymbol{v}}_j$ is not a member of the linearly independent set of row vectors, the arguments above can be suitably modified to reach the same conclusion. We leave this as an exercise for the reader.

Since the vectors $\{\vec{\boldsymbol{v}}_1\,,\,\vec{\boldsymbol{v}}_2\,,\,\ldots\,,\,\vec{\boldsymbol{v}}_k\}$ are linearly independent, the unique solution to eqs. (7) and (8) is still $c_1 = c_2 = c_3 = \cdots = c_k = 0$, which implies that k row vectors of the new matrix obtained by the row operation $R_i \to R_i + cR_j$ are still linearly independent. Moreover, k is the maximal number of linearly independent row vectors, since the row vectors $\{\vec{\boldsymbol{v}}_{k+1}\,,\,\vec{\boldsymbol{v}}_{k+2}\,,\,\ldots\,,\,\vec{\boldsymbol{v}}_m\}$ are still expressible as linear combinations of the k linearly independent row vectors.

Thus, for any series of elementary row operations (which correspond to multiplication on the left by a series of elementary matrices $E_n \cdots E_3E_2E_1$), the matrix $E_n \cdots E_3E_2E_1A$ also possesses k linearly independent row vectors. In Section 2, we showed that one can perform elementary row operations to reduce A to reduced row echelon form,

$$A_E \equiv PA\,, \qquad \text{where } P \equiv E_n \cdots E_3E_2E_1\,, \tag{9}$$

by employing an appropriate series of elementary matrices. Thus, the reduced row echelon form of A must also possess k linearly independent row vectors. However, by using properties 2–4 of the reduced row echelon form, it follows that the row vectors corresponding to the non-zero rows are linearly independent. Thus, we can identify k with the number of non-zero rows of the reduced row echelon form, which by definition is equal to the rank of A. Hence, it follows that *the rank of A is equal to the maximal number of linearly independent row vectors.*

Likewise, we can consider the elements of each column of A as the components of an m-component vector. That is, the matrix A consists of n *column vectors*. Consider what happens when we row-reduce A to its reduced row-echelon form, $A_E = PA$, where P is a product of elementary matrices [cf. eq. (9)]. Using the properties of the reduced row echelon form, it follows that the column vectors of A_E that contain a leading 1, called the *basic columns*, are linearly independent. We denote this linearly independent set by $\{\vec{\boldsymbol{w}}_1\,,\,\vec{\boldsymbol{w}}_2\,,\,\ldots\,,\,\vec{\boldsymbol{w}}_k\}$, where we have noted that the number of basic column vectors is equal to the number of leading 1's, i.e. k =rank A. Moreover, it is easy to verify that any column vector without a leading 1 can be expressed as a linear combination of the basic column vectors. Since all elementary matrices are invertible, it follows that P is also invertible, and we can write $A = P^{-1}A_E$. Hence, $\{P^{-1}\vec{\boldsymbol{w}}_1\,,\,P^{-1}\vec{\boldsymbol{w}}_2\,,\,\ldots\,,\,P^{-1}\vec{\boldsymbol{w}}_k\}$ can be identified as k linearly independent *basic column vectors* of A. The linear independence of these vectors can be verified by solving the equation:

$$c_1P^{-1}\vec{\boldsymbol{w}}_1 + c_2P^{-1}\vec{\boldsymbol{w}}_1 + \cdots + c_kP^{-1}\vec{\boldsymbol{w}}_k = 0\,.$$

Multiplying this equation on the left by P yields

$$c_1\vec{\boldsymbol{w}}_1 + c_2\vec{\boldsymbol{w}}_1 + \cdots + c_k\vec{\boldsymbol{w}}_k = 0\,,$$

which possesses a unique solution, $c_1 = c_2 = \cdots c_k = 0$, by virtue of the fact that the set of vectors $\{\vec{\boldsymbol{w}}_1\,,\,\vec{\boldsymbol{w}}_2\,,\,\ldots\,,\,\vec{\boldsymbol{w}}_k\}$ is linearly independent. Moreover, the set of basic column vectors $\{P^{-1}\vec{\boldsymbol{w}}_1\,,\,P^{-1}\vec{\boldsymbol{w}}_2\,,\,\ldots\,,\,P^{-1}\vec{\boldsymbol{w}}_k\}$ contains the maximal number of linearly independent column vectors of A.[§] Thus, we have proven that *the rank of A is equal to the maximal number of linearly independent column vectors.*

The analysis presented above constitutes a proof of the Rank Theorem, which states that for any $m \times n$ matrix A, the maximal number of linearly independent row vectors is equal to the number of linearly independent column vectors. In both cases, this number is identified as the rank of A. This is a remarkable result, especially considering the fact that the total number of row vectors m is generally not equal to the total number of column vectors n. Another way to present the Rank Theorem is to introduce the transpose of the matrix A, which is denoted by A^{T}. By definition, the rows of A are the columns of A^{T} and vice versa. Clearly, the maximal number of linearly independent row vectors of A is equal to the maximal number of linearly independent column vectors of A^{T} and vice versa, by the definition of the transpose. It immediately follows from the Rank Theorem that:

$$\operatorname{rank} A = \operatorname{rank} A^{\mathsf{T}} .$$

One consequence of this result is that the rank of the $m \times n$ matrix A can be no larger than the minimum of the two numbers m and n.

References

1. Carl D. Meyer, *Matrix Analysis and Applied Linear Algebra* (SIAM, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2000). This modern textbook along with its solution manual, has been generously made available for free viewing at `http://www.matrixanalysis.com/DownloadChapters.html`.

2. Sergei Treil, *Linear Algebra Done Wrong.* This free textbook is available from `http://www.math.brown.edu/~treil/papers/LADW/LADW.pdf`.

[§]This is true since the other columns of A (if any) are of the form $P^{-1}w_{k+1}$, where w_{k+1} is one of the columns of the reduced row echelon form that does not contain a leading 1. But, we have already noted above that w_{k+1} is some linear combination of $\{\vec{\boldsymbol{w}}_1\,, \vec{\boldsymbol{w}}_2\,, \ldots\,, \vec{\boldsymbol{w}}_k\}$. It immediately follows that $P^{-1}w_{k+1}$ is the same linear combination of $\{P^{-1}\vec{\boldsymbol{w}}_1\,, P^{-1}\vec{\boldsymbol{w}}_2\,, \ldots\,, P^{-1}\vec{\boldsymbol{w}}_k\}$.