Matrices We begin by defining **addition, scalar multiplication and product** for matrices.

A matrix A is an $m \times n$ array of numbers (also called scalars with some convention) and we say $m = rownum(A)$ or number of rows of A and $n = colnum(A)$ or number of rows of A. We make a convention to write $A = A_{m\times n}$ to indicate these numbers and say that A has type (or size) $m \times n$.

In general for a given matrix A, we write A_{ij} for the entry in its ith row and jth column. Naturally, we assume $1 \leq i \leq rownum(A)$ and $1 \leq j \leq colnum(A)$. This is sometimes written as $A(i, j)$ if the expressions for i, j are complicated or if subscripts are inconvenient.

Given two matrices A, B, we define $A + B$ **only if they have the same type** and then

$$(A + B)_{ij} =_{def} A_{ij} + B_{ij}.$$

The resulting matrix has the same type as either of them.

Given a matrix A and a scalar s, we define sA by

$$(sA)_{ij} =_{def} sA_{ij}$$

and this has the same type as A.

Given matrices A, B we define the product AB **only if** $colnum(A) = rownum(B)$. If this common number is n, then we have

$$(AB)_{ij} =_{def} \sum_{k=1}^{n} (A_{ik}B_{kj}) \text{ where } 1 \leq i \leq rownum(A) \text{ and } 1 \leq j \leq colnum(B).$$

Note that $AB \neq BA$ in general. In fact it is possible that AB is defined but BA is not. Even when both are defined, they may have different sizes.

Special matrices By 0 we denote a **zero matrix,** or a matrix with all entries zero. These matrices can be of different sizes, but the same symbol is used and the size has to be deduced from the context.

By I we denote the **identity matrix** which is defined by

$$I_{ij} = 1 \text{ if } i = j \text{ and } I_{ij} = 0 \text{ if } i \neq j\ .$$

By convention I is always a square $n \times n$ matrix and we may identify the size by writing $I = I_n$.

The important property of the Identity matrix is that $AI = A$ and $IB = B$ for any matrices A, B provided the equations make sense (i.e. the I has appropriate size.)

Elementary matrices Let $p \neq q$ be chosen with $1 \leq p, q \leq n$. Let c be a number(scalar). Define a square $n \times n$ matrix $E_{pq}^n(c)$ by

- The entry in the p-th row and q-th column is c.
- The entries along the main diagonal are 1. These are respectively the $(1,1), (2,2), \cdots, (n,n)$ entries.
- All other entries are zero.

Thus, using $n = 3$ we get:

$$E_{23}^3(2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}, \ E_{13}^3(4) = \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \ E_{31}^3(5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 5 & 0 & 1 \end{pmatrix}.$$

Note: An easy way to remember these matrices is this. To understand $E_{pq}^n(c)$, start with the identity matrix I_n and perform the row operation $R_p + cR_q$ on it.

Convention. To simplify the notation, we often drop the superscript n. It is also customary to drop all notation and simply name them as $E_1, E_2, \cdots$ etc. after having identified each. This is what the book does.

Diagonal Matrices We will have need to use matrices for which all entries off the main diagonal are zero. These can be identified by just specifying the diagonal entries and we write them as $diag(c_1, c_2, \cdots, c_n)$. For example:

$$diag(1,1,3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}, diag(2,2,2) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Note that $diag(2,2,2)$ is also equal to $2I_3 = 2diag(1,1,1)$, but in general you cannot simply a diagonal matrix if the diagonal entries are different.

However, it is easy to see that the matrix $diag(c_1, c_2, \cdots, c_n)$ can be also constructed by starting with the identity I_n and multiplying its rows successively by $c_1, c_2, \cdots, c_n$.

Permutation matrices. Let $i \neq j$ be chosen with $1 \leq i, j \leq n$. We define the matrix P_{ij}^n by swapping the i-th row of I_n with its j-th row. Thus,

$$P_{23}^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \ P_{14}^4 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

As before, we drop the superscript n when convenient.

Voodoo Principle. An important principle of working with matrices is what I like to call the voodoo principle.

Row operations.

If you wish to do something to the rows of a matrix M, do it to an appropriate I_n (which has the same number of rows as the matrix M) and multiply by the result **on the left** of A.

Here are some examples. The reader should check all the details.

Let

$$M = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{pmatrix}.$$

If you wish to swap its first two rows, then multiply it by P_{12}^3. Thus you can check

$$P_{12}^3 M = \begin{pmatrix} 2 & 3 \\ 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

To perform $R_2 - 2R_1$ multiply by $E_{21}^3(-2)$.

To multiply the third row by $\frac{1}{3}$, multiply by $diag(1, 1, \frac{1}{3})$.

To make the sum of all three rows multiply by $\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$.

Column operations.

You can make column operations by the same principle. Take an identity matrix with the same number of columns as the given matrix and do the operations on it. Then multiply by the result **on right** of A.

Thus, to swap the columns of the above matrix M calculate MP_{12}^2.

Similarly, the operation $C_2 - 2C_1$ can be carried out by right multiplying by the matrix $\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$. Note that this matrix is $E_{12}^2(-2)$.

Do note how the row and column notations differ in the notation for the E-matrices.

Do consider the effect of left multiplying by $diag(1, 2, 3)$ and right multiplying by $diag(1, 2)$.

Inverses A matrix A is said to be invertible (or non singular) if:

- A is a square ($n \times n$) matrix and
- There is some matrix M such that $MA = I = I_n$. Note that by definition of matrix products, M must have type $n \times n$.

In view of our observation of the voodoo principle, we see that our usual sequence of row operations can be accomplished by left multiplying by a product of matrices which are elementary, diagonal or permutation matrices. Note that we do not allow 0 in our diagonal matrices.

Thus, if A is a square matrix of rank n, then its RREF will be I_n and we can take M to product of the row transformation matrices.

Conversely, we now show that an invertible matrix A must have rank n. If $AX = B$ is a system of linear equations and A is invertible, then we see that $MAX = MB$ or $I_n X = MB$, so that $X = B$. Thus the equation $AX = B$ is solvable for all $B \in \Re^n$.

We know that this implies that A has rank n.

Thus, we have shown that a square matrix $A = A_{n\times n}$ is invertible if and only if A has rank n.

Note. It is possible to weaken the idea of invertible by talking about left or right invertible by defining:

One sided invertibility. A matrix $H = H_{m\times n}$ is said to be left invertible if $GH = I_n$ for some matrix G (which has to be of type $n \times m$).

The matrix H is said to be right invertible if there is a matrix K such that $HK = I_m$. Note that necessarily K has to be type $n \times m$.

Let r be the rank of H. It can be shown that H is left invertible if and only if $r = n$.

Similarly, H is right invertible if and only if $r = m$.

Finding Inverses. Given a square matrix $A = A_{n\times n}$, the standard method to decide if it is invertible is this:

Start with an augmented matrix $(A|I_n)$. Perform row operations leading to RREF of A.

If you get a zero row on the A side, then A is not invertible as its rank would be less than n.

Otherwise, the RREF of A must be I_n. Thus $(A|I_n)$ transforms to $(I_n|M)$ for some matrix M. Since we know that all the row operations can be accomplished by left multiplication by some matrix, we know that the matrix used must be M, so $(I_n|M) = (MA|MI_n)$, i.e. $MA = I_n$.

At the same time, we can visualize $(A|I_n)$ as a system of equations $AX = I_n$ where X is now an $n \times n$ matrix.

Multiplying by M we see $MAX = MI_n$ or $I_n X = M$ i.e. $X = M$.

Thus, we have shown that $MA = I_n$ implies $AM = I_n$ as well. So, an invertible matrix is both left and right invertible, using the same matrix M.

Thus, it makes sense to define

Inverse of a matrix: If A is an invertible matrix, then the matrix M for which $MA = AM = I_n$ is uniqely defined and we call it the inverse of A. It is denoted by A^{-1}.

What is the proof of uniqueness? Suppose we have two different matrices M_1, M_2 satisfying equations. Think of $M_1 A M_2$. By grouping these terms differently $M_1 A M_2 = (M_1 A) M_2 = M_1 (A M_2)$, it is easy to see that the product is equal to M_1 as well M_2. Hence $M_1 = M_2$.

Special Inverses. Inverses of some special matrices can be calculated easily and should be remembered. Each is easy to check by direct calculation (or better by interpreting it as a row operation).

- All $E_{ij}^n(c)$ are invertible with inverse $E_{ij}^n(-c)$.
- A diagonal matrix $diag(c_1, c_2, \cdots, c_n)$ is invertible if and only if none of the c_i are zero. Then the inverse is $diag(c_1^{-1}, c_2^{-1}, \cdots, c_n^{-1})$.
- P_{ij}^n is its own inverse. Thus $(P_{ij}^n)^2 = I_n$. In general matrices whose powers are equal to identity are called unipotent.
- If A, B are invertible of the same type, then AB is invertible and its inverse is $B^{-1}A^{-1}$. Notice the change of order, it is crucial!

  This extends to a sequence of invertible matrices as well.

  $$(A_1 A_2 \cdots A_s)^{-1} = A_s^{-1} \cdots A_2^{-1} A_1^{-1}.$$

- For a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we set $\Delta = ad - bc$, its determinant.

  Then A is invertible if and only if $\Delta \neq 0$ and the inverse (when $\Delta \neq 0$ is given by $\frac{1}{\Delta} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.
- General square matrix A. Later, we shall define the determinants of arbitrary sizes and develop the following test for invertibility.

  Matrix $A = A_{n \times n}$ is invertible if and only if its determinant $det(A) = \Delta$ is non zero. Moreover, there is a matrix A^{adj} called the adjoint of A which can be calculated by several subdeterminants of A and the formula for the inverse will be $\frac{1}{\Delta} A^{adj}$.

  this is of great theoretical interest, but often the row reduction method is more practical. In fact, calculation of determinants is also made efficient by row transformations and hence it is not faster to calculate the adjoint from definition.
- Determinants have a product formula, $\det(AB) = \det(A)\det(B)$. This coupled with the above test for invertibility gives us the result that a product of square matrices is invertible if and only if each is!
- Challenge 1: Given that AB has an inverse M, say, what are the inverses of A and B? Make a formula in terms of A, B, M.

  Challenge 2: Assume that $A_1, \cdots A_r$ are square matrices of the same type Given $A_1 A_2 \cdots A_{r-1} A_r = I$, prove that $A_2 \cdots A_{r-1} A_r A_1 = I$

LU decomposition Suppose a matrix A is such that its REF can be obtained without swapping rows. Then the elementary matrices used to reach the REF are of a special form called "lower triangular".

We **define** a square matrix to be lower triangular if all the entries above the diagonal are zero. If, in addition, all the diagonal entries are 1, then it is said to be a unit lower triangular matrix.

An upper triangular and unit upper triangular matrix are defined similarly. All elementary matrices $E_{ij}^n(c)$ are evidently unit lower triangular if $i < j$. Of course, if $i > j$ then they are unit upper triangular.

A little thought can show the following:

- If L_1, L_2 are lower triangular matrices of the same type, then so is their product.

  Ditto for upper triangular.
- Further, if L_1, L_2 are unit lower triangular, then so is their product.

  Ditto for upper triangular.

  Thus, if A can be put into REF without swapping rows, then we can say that $L_1 A$ is in REF for some unit lower triangular matrix L_1. (The matrix L_1 is a product of unit lower triangular matrices used in our operations.) Also, it is clear that the final REF must be upper triangular. (Follows from the definition of REF, think about it). Thus we may denote $L_1 A$ as U, to denote an upper triangular matrix. **Note** that this U need not be unit upper triangular, unless our matrix was invertible to begin with.

  Thus, our row reduction algorithm proves $L_1 A = U$ or $A = LU$ where $L = L_1^{-1}$. This factorization is used in efficient solutions of systems, especially of large size. Moreover, the calculation of the matrix L does not need any additional work, it is a matter of simply recording the multipliers used in the right spots. We shall discuss this in detail later.