Some of the notations below are enhanced versions of the book notations. We also include material taught in the course which may be slightly different from the book.

1. **Conventions.** A linear equation in variables $x_1, \cdots, x_n$ will be assumed to be **prearranged** as $a_1x_1 + \cdots + a_nx_n = b$ where the coefficients $a_1, a_2, \cdots, a_n, b$ are assumed to be numbers (or scalers) in a field F. If convenient, we may also assume parameters involved in these coefficients, either by an agreement or by assuming the field F to having been enlarged to accept them.

   Thus $2x + 3y = 5$ as well as $tx + (1 - t^2)y = t^2 + t + 5$ may both be accepted as linear equations in x, y. The first may also be considered linear in x, y, t, but not the second one!

2. **Augmented matrix** To each linear equation in $x_1, \cdots, x_n$, we associate a row vector with $n + 1$ entries

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_n & b \end{pmatrix}.$$

   Given a system of several equations, we stack their corresponding rows into a matrix called the augmented matrix of the system of linear equations. Sometimes, the last column of the augmented matrix is separated by a vertical line to indicate the equality signs, but the book avoids this marker. We may sometimes add it for convenience.

3. **Pivots and related concepts.** We shall make the following formal notations for precision.

   - Given a row which has at least one non zero entry, its pivot position is the column number of the first non zero entry in it. For a row full of zeros, the pivot position is defined to be ∞.
     Thus, the pivot position for $\begin{pmatrix} 0 & 0 & 2 & 0 & -3 \end{pmatrix}$ is 3 whereas it is ∞ for $\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \end{pmatrix}$.
     For a matrix, we define its pivot positions as a sequence of the pivot positions of its rows. Thus, we see that the pivot positions of the following matrices

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 5 \\ 0 & 2 & 7 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 7 \end{pmatrix}, \quad \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 0 \\ 2 & 1 & 7 \end{pmatrix}$$

     are respectively $(1, 2, 2), (1, 1, 3), (2, \infty, 1)$.
   - **Pivots or leading entries.** For a non zero row, its pivot or the leading entry is said to the entry in its pivot position. For a zero row, we say that the pivot is missing.
   - We shall say that a sequence of pivot positions is **strict** if it consists of a strictly increasing sequence of positive integers which may be followed by ∞ repeated some number of times. Note that the ∞ part may be absent, if there are no zero rows and the integer part may be absent for a zero matrix.
   - **Echelon forms.** A matrix is said to be in **a row echelon form (REF)** provided the sequence of its pivot positions is strict.

     It is said to be in a **reduced row echelon form (RREF)** if every pivot is 1 and **moreover** it is the only non zero entry in its column.
   - **Elementary Row operations.** We shall use the following notations. In the following i, j are different row numbers of a matrix M and c is a number.
     - The operation $R_i \leftrightarrow R_j$ swaps the i-th row R_i with the j-th row R_j. Sometimes, this may be shortened as P_{ij} where P reminds of the word permute. All other rows stay the same.
     - The operation $R_i + cR_j$ modifies the row R_i by adding c times R_j to it. All other rows stay the same.
     - If $c \neq 0$, then the operation cR_i replaces the i-the row by c times itself. All other rows stay the same.

4. The main use of elementary operations is that if M is an augmented matrix of a system of linear equations, then the set of solutions of the system transformed by elementary row operations is the same as the original.

   Thus, the main idea of solving systems of linear equations is to transform the associated augmented matrix into a form which lets us simply read off the solutions. This is exactly the RREF.

   As an intermediate step, we also discuss a form from which the final solutions can be quickly deduced (rather than read off). This is the REF.

   We shall discuss the algorithm to derive these forms next.

5. **The Gauss Elimination.** Here are the steps of the Gauss Elimination algorithm. **Convention and notation.** We will be working with the active part of a given matrix M and reduce it to to REF in a sequence of steps. At the beginning, the whole matrix is active and as we proceed, we shall declare certain rows and columns inactive. This means then that they are not used in further pivot calculations. Also, any further row transformations will be such that they do not change the inactive rows.

   Main Lemma. Consider non zero rows R_i, R_j with $i \neq j$. Suppose that R_i, R_j have the same pivot position and that their pivot entries are respectively c_i, c_j. Then the operation $R_j - \frac{c_j}{c_i} R_i$ produces a new jth row whose pivot position is bigger than that of R_i.

   Moreover, suppose that the pivot position of R_i is less than or equal to that of R_j for each $j > i$, then repeated application of the Lemma gives that the pivot position of R_i is strictly less than the pivot position of all the new rows R_j with $j > i$.

   This transformation leaves all the first i rows unchanged.

   - **Start:** Assume that the whole matrix is active.
   - **Cleanup.** If there is no active row left, then go to the end.

     Otherwise, suppose that R_i is the first active row. By swapping it, if necessary, with some R_j with $j > i$; arrange that the pivot position of R_i is less than or equal to all the pivot positions of R_j for $j > i$.
   - If R_i is the zero row, then clearly all the rows R_j with $j > i$ are also zero. Declare all rows inactive and go to end.

     Otherwise, apply the Main Lemma to arrange that the pivot positions of all R_j with $j > i$ are bigger than the pivot position of R_i. Note that this does not disturb the inactive part.
   - Set the R_i as inactive. Go to the Cleanup step.
   - **End.**

   Note that the pivot positions of the inactive matrix always stay strict and hence we end up with an REF when the whole matrix becomes inactive.

6. When the matrix is in REF, the pivot positions is an increasing sequence of positive integers followed by a sequence of ∞. (Of course, either of these can be empty.)

   The number of finite pivot positions in the REF is called the rank of the matrix. It does not depend on the process used to get the REF. This invariance will be proved later.

7. Starting with REF, it is possible to get the reduced form RREF as follows:

   - First make all pivot entries equal to 1 by dividing the corresponding pivot rows by the pivot entries. Explicitly, if c_i is the pivot entry in R_i then use the operation $\frac{1}{c_i} R_i$.
   - Note that REF implies that all entries below a pivot are already 0. Use the cleanup operations to make all entries above each pivot using the pivot entry 1. This may be described as cleaning the pivot column.

     This cleaning is best done in reverse order, i.e. cleaning the columns with pivots from last to the first.

8. **Vectors** We define $\Re^n$ to be the set of columns of n-tuples of real numbers.

   This $\Re^n$ can be also given a geometric interpretation by thinking of the n entries as the n-coordinates of points in an n-dimensional space.

   For $v \in \Re^n$ we shall make the convention that v_i is the i-th entry of v for $i = 1, \cdots n$. Thus, if $n = 2$ and $v = \begin{pmatrix} -5 \\ 6 \end{pmatrix}$, then $v_1 = -5$ and $v_2 = 6$.

   The set $\Re^n$ has a natural addition defined by componentwise addition. Thus $(v + w)_i = v_i + w_i$. There is also a scalar multiplication so that for any $c \in \Re$ and $v \in \Re^n$ we have cv defined as a vector with $(cv)_i = cv_i$.

   There is a natural subtraction $v - w$ which is realized to be the same as $v + (-1)w$. [1]

[1] For those who know these terms, we state: The addition is commutative and asoociative. the scalar multiplication is associative with respect to scalars $c(dv) = (cd)v$. Also the addition is distributive over the scalar multiplication. Addition has an identity, namely the zero vector with all zero entries and thus the addition operation makes an additive group. The scalar multiplication also has a multiplicative unit 1 such that $1v = v$ for all v.

Later, any set with operations having such properties will be declared as an abstract vector space.

Members of $\Re^n$ are called vectors and geometrically may be thought of as an arrow pointing from the origin towards the point whose coordinates are the n entries of the column.

For example, when $n = 2$, a column $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ describes an arrow from the origin $(0, 0)$ towards the point $(1, 2)$.

In general an arrow from a point $P(a, b)$ to a point $Q(c, d)$ can be identified with the vector corresponding to $Q - P$, namely $\begin{pmatrix} c-a \\ d-b \end{pmatrix}$.

For convenience, we may say that the point $P(a, b)$ is associated with a vector $\overline{OP} = \begin{pmatrix} a \\ b \end{pmatrix}$.

Then the addition can be given a geometric interpretation thus: Given vectors v, w associated with the points P, Q, let O denote the origin. Then the vector $v + w$ is associated witb a point R such that $OPRQ$ is a parallelogram. A vector cv is associated with a point P_1 on the line joining O and P such that the distance OP_1 between O and P_1 is $|c|OP$. Moreover, if $c > 0$ then P, P_1 lie on the same side of O, while they lie on the opposite sides if $c < 0$.

9. **New Perspective on Linear Equations.** Given vectors $v_1, \cdots, v_r$ in $\Re^n$ we define a linear combination of them to be a vector of the form $c_1v_1 + c_2v_2 + \cdots + c_rv_r$.

   Collectively, we define the span

   $$Span\{v_1, \cdots, v_r\} = \{c_1v_1 + c_2v_2 + \cdots + c_rv_r \mid c_1, c_2, \cdots, c_r \in \Re\}.$$

   In this notation, it is easy to see that given an augmented matrix $(A|b)$ belonging to a linear system of equations in $x_1, \cdot, x_r$, we can say that the system of equation is :

   BasicEquation. $$x_1A_1 + x_2A_2 + \cdots x_rA_r = b$$

   where $A_1, A_2, \cdots, A_r$ are the columns of the matrix A. They are members of $\Re^n$ if A has n rows.

   In other words, solving the system of euuations is the same as expressing b as a linear combination of $A_1, \cdots, A_r$. In other words, the system has a solution iff b is a member of the $Span\{A_1, \cdots, A_r\}$.

   If X denotes the column of r numbers $x_1, \cdots, x_r$, then it is a member of $\Re^r$.

   We define AX to be the left hand side of the above basic equation.

   Thus, given a matrix with n rows and r columns, we say that for a column X the product AX is defined iff X has exactly r entries. In other words, as a matrix, X has r rows and 1 column.

   More generally, if B is any matrix with r rows and s columns $B_1, B_2, \cdots B_s$, then we define the product

   $$AB = \text{ The matrix with columns } AB_1, AB_2, \cdots, AB_s.$$

   Thus, the resulting matrix AB has n rows and s columns. Note that for AB to make sense the number of columns of A must be equal to the number of rows of B.

10. **Applications to solutions of systems of equations.** We know that given a system $(A|b)$ where A has n rows and r columns, we can say that:

    - **Consistency condition.** The system has a solution iff when put in REF, every pivot is in one of the first r columns. This condition is the same as saying that the first non zero entry in a row does not occur on the right hand side of our equation.
    - **Uniqueness condition.** Assuming that the system is consistent, it has a unique solution iff there are no free variables. Equivalently, the number of pivots in REF equals the number r.

    Using this, we have two very simple yet powerful conclusions.

    In expanded form, we write our system thus:

    $$A \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} = b \quad \text{or} \quad x_1A_1 + \cdots + x_rA_r = b$$

where $A_1, \cdots, A_r$ are the columns of A in order.

Suppose d is the number of pivots in an REF of A. Note that we are not including b in the calculation of this number. We shall write $rank(A) = d$. This will be proved to be independent of our choice of row transformations later.

If $d = n$, then the consistency condition is satisfied regardless of the right hand side. In other words every vector in $\Re^n$ is in the span of columns of A.

Proof. Reduce $(A|b)$ to REF. We have $d = n$ pivots, so every row ofthe REF has a pivot on the left hand side of the equation! Thus, the system is consistent, regardless of what b is.

Hence evary $b \in \Re^n$ is in the span of columns of A.

If $d = r$, then there are no free variables. Thus, either the system is inconsistent or it has a unique solution.

Proof. Evident!

Now assume $r = n$, i.e. the matrix A is square. Then we have two possible cases. Either $d = r = n$ or $d < r = n$.

In the first situation, we get to claim that $(A|b)$ has a unique solution for all b. Naturally, such matrices A are important. They are called non singular or invertible. We can make a formula for the solution of $(A|b)$ for such matrices.

In the second situation, depending on b, either the system has no solution or infinitely many. The matrix A is said to be singular in such cases.

11. **Matrices as Linear Transformations.**

Given any map $T : \Re^n \to \Re^r$ we say T is a linear transformation if the following conditions hold.

$$T(X + Y) = T(X) + T(Y) \quad \text{and} \quad T(kX) = kT(X)$$

where X, Y are any vectors in $\Re^r$ and k is any real number.

The simplest example of a linear transformation is obtained when we take a matrix A with n rows and r columns.

Define $T(X) = AX$ for any $X \in \Re^n$. the two properties are easily verified by definition of the matrix multiplication.

Conversely, it is possible to prove the following:

Theorem. Given any linear transformation $T : \Re^n \to \Re^r$ there is a well defined matrix A such that $T(X) = AX$ for all X.

Proof.

First we prepare **some useful notation.**

We define a sequence of vectors $e_1^n, \cdots e_r^n$ as follows. For each $i = 1, 2, \cdots, n$ we define e_i^n to be the vector which has only zero entries except for one 1 in the i-th place.

Thus, for example, when $n = 3$ we have:

$$e_1^3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2^3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Similarly, when $n = 2$ we have:

$$e_1^2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e_2^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Important shortened notation. Often, it is tedious to keep on writing the superscript for these.

So, we may simply write e_1, e_2, e_3 in place of e_1^3, e_2^3, e_3^3, if we know that we are working in $\Re^3$.

Similarly, we will simply use e_1, e_2 in place of e_1^2, e_2^2, if we are working in $\Re^2$.

If we are working with two different spaces at the same time, then we find it convenient to give new names to these.

We will say that $(e_1^n, e_2^n, \cdots, e_n^n)$ is a **standard basis of $\mathbf{\Re^n}$.**

Important Observation. Every $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ can be **uniquely** expressed as $x_1 e_1^n + \cdots x_n e_n^n$.

Thus for example, when $n = 3$, **using shortened notation** we get:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 e_1 + x_2 e_2 + x_3 e_3.$$

Now consider the standard basis $e_1, \cdots, e_n$ of $\Re^n$ in shortened notation.

Construct a matrix A whose columns are $T(e_1), \cdots, T(e_n)$ in order.

We claim that $T(X) = AX$ for all $X \in \Re^n$.

Note that

$$X = x_1 e_1 + \cdots + x_n e_n$$

and by the definition of a linear transformation, we see that

$$T(X) = x_1 T(e_1) + \cdots + x_n T(e_n).$$

But the right hand side of the last equation is exactly

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = AX.$$

This proves our claim.

It is easy to see that if matrices A, B satisfy $AX = BX$ for each $X \in \Re^n$, then $A = B$. **Hint:** Take X to be successively $e_1, \cdots, e_n$ and deduce that A and B have identical columns.

Thus, it makes sense to make: **Corollary/Definition:**

The matrix A such that $T(X) = AX$ is uniquely defined by T and is defined to be the **standard matrix for the linear transformation T.**

We use the notation M_T to denote the matrix A.

Example: Suppose $T : \Re^2 \to \Re^3$ is defined by

$$T(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}) = \begin{pmatrix} 2x_1 + x_2 \\ x_1 - x_2 \\ 3x_1 + 5x_2 \end{pmatrix}.$$

Then

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and hence} \quad T(e_1) = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}.$$

Similarly

$$e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{and hence} \quad T(e_2) = \begin{pmatrix} 1 \\ -1 \\ 5 \end{pmatrix}.$$

Thus,

$$M_T = \begin{pmatrix} 2 & 1 \\ 1 & -1 \\ 3 & 5 \end{pmatrix}.$$

It is easy to verify that:

$$T(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}) = \begin{pmatrix} 2 & 1 \\ 1 & -1 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{or} \quad T(X) = \begin{pmatrix} 2 & 1 \\ 1 & -1 \\ 3 & 5 \end{pmatrix} X.$$

12. **Properties of Linear Transformation.**

Let T be a linear transformation from $\Re^n$ to $\Re^r$ as before.

We shall **define** T to be **surjective** if it is an onto map. This means that every $b \in \Re^r$ can be written as $T(X)$ for some X. Thus, if A is the standard matrix M_T, then this is equivalent to the condition that $(A|b)$ is consistent for all b.

From what we already observed, **the condition for surjectivity is thus** $d = r$ where $d = rank(A)$.

Note that the number of pivots in an REF of any matrix has to be less than or equal its number of rows as well as its number of columns, since different pivots are in different rows and diffent columns.

Hence:

$$rank(A) = d \leq \min(n, r).$$

In particular, we note that if $n < r$ then $d \leq \min(n, r) < r$ and the map cannot be surjective! Note that we need to do extra calculations if $n = r$.

We also **define** T to be **injective** if it is one to one as a map. This means if $T(X) = T(Y)$ then $X = Y$. As before, using the standard matrix $A = M_T$ we see that

$$T(X) = T(Y) \quad \text{iff} \quad AX = AY \quad \text{iff} \quad A(X - Y) = 0$$

where the matrix 0, by convention is the matrix with all zero entries or the so called zero matrix.

Thus setting $X - Y = Z$ we can reformulate the condition of injectivity as $AZ = 0$ implies $Z = 0$.

This means that the system $(A|0)$ always has a unique solution 0, or, as discussed before, in REF, there are no free variables!

From what we already observed, **the condition for injectivity is thus** $d = n$ where $d = rank(A)$.

As before, we can deduce that if $n > r$, then $d \leq \min n, r < n$ and thus the map T is not injective. As before, the case $n = r$ needs further evidence!

13. **Independence of vectors.** Given r vectors $v_1, \cdots, v_r$ in $\Re^n$ we define them to be **linearly dependent** if $x_1 v_1 + \cdots + x_n v_n = 0$ for some $x_1, \cdots, x_n$ such that at least one of the x_i is non zero.

If we build a matrix A with columns $v_1, \cdots, v_r$, then this condition is seen to mean that the linear system described by $(A|0)$ has a non zero solution. Thus, the system must have more than one solutions, since setting all $x_i = 0$ is an obvious solution.

From what we already know, this is equivalent to $rank(A) < r$ or the linear transformation defined by $T(X) = AX$ is not injective.

We may simply write $\{v_1, \cdots, v_r\}$ is a linearly dependent set if this condition is satisfied.

It can be shown that the condition can also be expressed by saying that one of the v_i is a linear combination of the rest.

The idea is to note that if $x_1 v_1 + \cdots + x_i v_i + \cdots x_r v_r = 0$ and $x_i \neq 0$ then we can write:

$$v_i = -\frac{x_1}{x_i} v_1 - \cdots - \frac{x_r}{x_i} v_r$$

where v_i does not appear on the right hand side.

Moreover, by choosing the last i with $x_i \neq 0$, we may even make a stronger looking claim that v_i is a linear combination of $v_1, \cdots, v_{i-1}$.

These descriptions are useful when we are trying prove properties of linearly dependent vectors.

The set of vectors $\{v_1, \cdots, v_r\}$ is **defined to be linearly independent** if it is not linearly dependent. Equivalently, the rank of the corresponding matrix A with columns v_i equals r or that the associated transformation T is injective.

Here are some simple observations that can be made from these definitions.

Let S be the set of vectors $\{v_1, \cdots, v_r\}$ in $\Re^n$.

- If one of the vectors $v_1, \cdots, v_r$ is the zero vector then the set S is linearly dependent.

- If one of the v_i is a multiple of another v_j, then the set S is linearly dependent.
- If one of the v_i is a linear combination of some other members of S, then the set S is linearly dependent.
- If a subset of S is linearly dependent, then S is linearly dependent.
- If $r > n$ than S is linearly dependent.
- If S is contained in a linearly independent set of vectors S_1, then it is linearly independent.