Matrix Algebra and its Applications

by Avinash Sathaye, Professor of Mathematics [1]
Department of Mathematics, University of Kentucky

April 10, 2019

[1]Partially supported by NSF grant thru AMSP(Appalachian Math Science Partnership)

Introduction.

This book was designed as a study guide for students taking **MA: 322 Matrix Algebra and its Applications** at the University of Kentucky. It covers the basic material of most linear algebra textbooks.

Contents

Chapter 1

Linear Equations.

1.1 Systems of Linear Equations

1. **Def.1: A Linear Function** of variables $x_1, \cdots, x_n$ is an expression of the form $a_1x_1 + \cdots + a_nx_n - b$ where $a_1, a_2, \cdots, a_n, b$ are assumed to be numbers (or scalars) in a field F.

2. Moreover **Def.2: A Linear Equation** of variables $x_1, \cdots, x_n$ is an equation prearranged as $a_1x_1 + \cdots + a_nx_n = b$.

   A set of linear equations in the same variables is called a **System of Linear equations.**

3. In practice, the coefficients may be allowed to have some parameters or temporary variables to be replaced by convenient scalars later on.

4. **Def.3: Augmented Row** To each linear equation as above we associate an augmented row with $n+1$ entries

   $$\begin{pmatrix} a_1 & a_2 & \cdots & a_n & b \end{pmatrix} \text{ or for clarification of RHS } \begin{pmatrix} a_1 & a_2 & \cdots & a_n & | & b \end{pmatrix}$$

5. **Def.4: Augmented Matrix** To a system of Linear equations (i.e. a set of linear equations) we associate an augmented matrix by stacking their augmented rows one above the other.

6. We also add a title row of variables for convenience.

7. For example, a system of linear equations

   $$x + y + 3z = 12, y + 6z = 20, 2x + 3z = 11, -x + y - 2z = -5$$

   gives an augmented matrix:

   $$\left(\begin{array}{rrr|r} x & y & z & RHS \\ \hline 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 2 & 0 & 3 & 11 \\ -1 & 1 & -2 & -5 \end{array}\right)$$

1.2 Elementary Row Operations

1.2.1 Basics of the solution process.

1. **Def. 5. A solution of a system of linear equations.** Given one equation $f(x_1, \cdots, x_n) = b$, an n-tuple $(s_1, \cdots, s_n)$ is said to be its solution if $f(s_1, \cdots, s_n) = b$.

   We shall shorten the above notation for convenience by writing the equation as $f(X) = b$ and the solution as $s = (s_1, \cdots, s_n)$, so that we can say s is a solution of $f(X) = b$ if $f(s) = b$.

   If we have a set of equations $E = \{f_1(X) = b_1, \cdots, f_m(X) = b_m\}$, then we say that s is a solution of the system E if $f_i(X) = b_i$ for each i.

2. **Def. 6. Equivalent systems of linear equations.** Two systems of linear equations E, E^* are defined (in the book) as systems having the same set of solutions. Even though this is correct, it is not too useful since it involves solving both and verifying that the solutions match.

   We use a better test: Two systems E, E^* are equivalent if each equation in E is a linear combination of equations in E^* and conversely, each equation in E^* is a linear combination of equations in E.

3. We note that shuffling the rows of the augmented matrix describes the same system of equations. In general, we record the following elementary row operations on our matrix which are easily shown to give equivalent linear systems.

4. **Def. 7.: Elementary Row Operations.**

   (a) Swap an i-th row with a different j-th row. Notation: P_{ij}.

   (b) Multiply an i-th row by a non zero scalar k Notation: kR_i.

   (c) Multiply the j-th row by a scalar c and add it to a different i-th row. Notation: $R_i + cR_j$.

5. These notations must be learnt and memorized precisely!

1.2.2 A sample of elementary operations.

1. We now perform elementary row operations on the above augmented matrix to illustrate how the equations are simplified and hence solved.

$$\begin{array}{c} \begin{array}{cccc} x & y & z & RHS \end{array} \\ \left(\begin{array}{ccc|c} 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 2 & 0 & 3 & 11 \\ -1 & 1 & -2 & -5 \end{array}\right) \end{array} \begin{pmatrix} R_3 - 2R_1 \\ R_4 + R_1 \end{pmatrix} \begin{array}{c} \begin{array}{cccc} x & y & z & RHS \end{array} \\ \left(\begin{array}{ccc|c} 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 0 & -2 & -3 & -13 \\ 0 & 2 & 1 & 7 \end{array}\right) \end{array}$$

2.

$$\begin{array}{c} \begin{array}{cccc} x & y & z & RHS \end{array} \\ \left(\begin{array}{ccc|c} 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 0 & -2 & -3 & -13 \\ 0 & 2 & 1 & 7 \end{array}\right) \end{array} \begin{pmatrix} R_3 + 2R_2 \\ R_4 - 2R_2 \end{pmatrix} \begin{array}{c} \begin{array}{cccc} x & y & z & RHS \end{array} \\ \left(\begin{array}{ccc|c} 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 0 & 0 & 9 & 27 \\ 0 & 0 & -11 & -33 \end{array}\right) \end{array}$$

3.
$$\left(\begin{array}{ccc|c} x & y & z & RHS \\ \hline 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 0 & 0 & 9 & 27 \\ 0 & 0 & -11 & -33 \end{array}\right) \; R_4 - \frac{-11}{9}R_3 \; \left(\begin{array}{ccc|c} x & y & z & RHS \\ \hline 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 0 & 0 & 9 & 27 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

4.
   - If we read the current equations from bottom to top, we see:
     $$0 = 0, \quad 9z = 27, \quad y + 6z = 20, \quad x + y + 3z = 12.$$
   - These can be solved in order for z, y, x respectively, to deduce
     $$z = 3, y = 2, x = 1.$$
   - Recall that our equations from bottom to top, were:
     $$0 = 0, \quad 9z = 27, \quad y + 6z = 20, \quad x + y + 3z = 12.$$
   - We can ignore $0 = 0$, since it is always true!
   - Thus, $9z = 27$ gives $z = 3$.
   - Now $y + 6z = 20$ gives $y + 18 = 20$ or $y = 2$.
   - Now $x + y + 3z = 12$ gives $x + 2 + 9 = 20$ or $x = 1$.

5. **The $\mathbf{0}, 1, \infty$ principle:**
   - Sometimes, our system does not lead to any solution. For instance, a system $2x + 3y = 5, 4x + 6y = 6$ would leads to:
     $$\left(\begin{array}{cc|c} x & y & RHS \\ \hline 2 & 3 & 5 \\ 4 & 6 & 6 \end{array}\right) \; R_2 - 2R_1 \; \left(\begin{array}{cc|c} x & y & RHS \\ \hline 2 & 3 & 5 \\ 0 & 0 & -4 \end{array}\right).$$
   - The second equation is $0 = -4$ and such equations are said to be **inconsistent.** The equations have 0 solutions (i.e. no solutions).
   - If we change the above equations to $2x + 3y = 5, 4x + 6y = 10$, then it is easily seen that the system has an infinity of solutions, namely $x = 1 - 3t, y = 1 + 2t$ where t can take any value!
   - We will later show that any linear system of equations has 0, 1 or infinitely many solutions!

6. **Understanding the process:**
   - Thus, our work is M_1 to M_2 to M_3, where:
   - $$M_1 = \left(\begin{array}{ccc|c} x & y & z & RHS \\ \hline 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 2 & 0 & 3 & 11 \\ -1 & 1 & -2 & -5 \end{array}\right) \to M_2 = \left(\begin{array}{ccc|c} x & y & z & RHS \\ \hline 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 0 & 0 & 9 & 27 \\ 0 & 0 & 0 & 0 \end{array}\right)$$
     and
     $$M_3 = \left(\begin{array}{ccc|c} x & y & z & RHS \\ \hline 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

1.3 The PC list

- If a row has at least one non zero entry, **"the pivot of that row "** shall be the first nonzero entry in it.
- The book calls this as the "leading entry" and the corresponding column, the pivot column.
- **Def.8: pivot column (pc) number of a row** is defined to be the column number of the pivot of that row.
- Thus, the pc number 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 row full of zeros, there is no pivot entry and the "pc number " is defined to be ∞.
- **Def.9: pivot column (pc) list of a matrix** is the list of the pc numbers of successive rows of that matrix.

1.3.1 REF

- Thus, we see that the pc lists of the following matrices are respectively $(1, 2, 2), (1, 1, 3), (2, \infty, 1)$.

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

- We say that **Def.10: The pc list is strict** if it is an increasing sequence. Note that for this definition, we accept that a sequence of ∞ as increasing.
- For our matrices M_1, M_2 the pc lists were: $(1, 2, 1, 1)$ and $(1, 2, 3, \infty)$. For M_3 it stays the same as M_2.
- **Def.11: A matrix is said to be in REF if its pc list is strict.**
- Thus we accept $(1, 3, \infty, \infty)$ as a strict sequence also!

1.4 Gaussian Elimination.

1. By a linear system of equations we mean an augmented matrix $(A|B)$ with or without a title row specified. The separator bar is also optional.

   As we have seen, it is useful to have the system reduced to REF (i.e. with a strict pc list).

2. The big theorem is that a suitable sequence of elementary operations applied to this matrix will always produce such an REF. There is no claim (or hope) of uniqueness of the final form, however, a certain associated integer called its rank will turn out to be a very useful tool for the solution process.

3. While there are many choices of operations, we describe a certain well defined choice which is guaranteed to work. This will be described as the standard algorithm.

1.4.1 The standard algorithm.

1. Here is the set up. We assume that our augmented matrix is renamed M and it has m rows. We assume that
   - The top i rows of M are already in REF and
   - All the pivot columns in rows $i+1$ to m are strictly bigger than the pivot columns for the first i rows.
2. We describe this situation as having the top i rows inactive.
3. The Plan: We can always begin with $i = 0$ and our algorithm will push i to m. Clearly, at the end, we have REF!
4. Now we show how to make the $i+1$-th row inactive. This is the iterative step.
5. Among the rows from $i+1$ to m, we pick the one whose pivot column number is the least, say s.

   We do a row swap **only if needed** to make this row the $i+1$-th.
6. Important rule. We always choose the smallest numbered row to swap into this $(i+1)$-th place. This is the only time we use a row swap, and only if really needed.
7. Now we do a sequence of row operations to arrange the pivot column numbers of all rows from $i+2$ to m to be bigger than the pc s of the $(i+1)^{th}$ row.
   - For each $j > i+1$ we do the following well defined operation.
   - Consider the pivot entry of the $(i+1)^{th}$ row, namely $M(i+1, s)$.
   - For each $j > i+1$ we wish to arrange the pc of the j-th row to be bigger than s. We already know that it is at least s.
   - This exactly means that we need $M(j, s) = 0$ for all such $j > i+1$. We call this entry $M(j, s)$ the target - to be made zero!
   - **Important Formula:** We use the operation $R_j - cR_{i+1}$ where c is given by the formula $c = \frac{M(j,s)}{M(i+1,s)}$.
   - Note that the formula for c can be remembered as $\frac{target}{pivot}$.
   - Note that this step is carried out for each $j > i+1$ whenever $M(j, s) \neq 0$. We typically do it in sequence, but as long as $i+1$ is fixed, all these steps can be done at the same time, since they do not interfere with each other!

1.4.2 Using the algorithm.

1. As seen above, we can make all the m rows inactive and thus have REF. The pc-list is now strict and thus all rows which become zero appear only after the non zero rows.
   - At this stage, we are ready to solve the original equations.
   - **Def.12: Rank of M.** The number of pivots in the final REF is called the rank of M and is denoted by $rank(M)$.
   - Note that we have not proved the rank to be well defined. That proof will come much later.
   - Write the final form as $M^* = (A^*|B^*)$.

- We note that both A^* and M^* are in REF and these are respectively REF of A and M.

2. **Def.13: Consistency:** A system $(A|B)$ is said to be consistent if it has at least one solution.

3. **Def.14: Consistency Condition.** We note that the original system represented by $(A|B)$ is consistent if and only if $rank(A) = rank(M)$.

4. Explicitly, this means that all the pivots in M^* occur in the A^* part. In other words, if some row of A^* is zero, then it must be also the zero row of M^*.

5. Example of an inconsistent system.

   - Consider our old example with the RHS changed in the last equation.

$$\left(\begin{array}{rrr|r} x & y & z & RHS \\ \hline 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 2 & 0 & 3 & 11 \\ -1 & 1 & -2 & t \end{array}\right)$$

   - It can be shown that the same REF steps as before, produce:

$$\left(\begin{array}{rrr|r} x & y & z & RHS \\ \hline 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 2 & 0 & 3 & 11 \\ -1 & 1 & -2 & t \end{array}\right) \to \left(\begin{array}{rrr|r} x & y & z & RHS \\ \hline 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 0 & 0 & 9 & 27 \\ 0 & 0 & 0 & t+5 \end{array}\right)$$

   - Thus, our original system is consistent if and only if $t + 5 = 0$ or $t = -5$.
   - Our original system had $t = -5$ and hence was consistent! For that system, both A and M had the same rank 3.

1.5 Vector Spaces.

1. Now we present a different way of understanding our work.

   Def.15: Vectors in n-dimensions. A column of n scalars $v = \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{pmatrix}$ is said to be an n-dimensional vector.

2. **Def.16: The set of all n-dimensional vectors forms the vector space $\Re^n$.** The space $\Re^n$ has two natural operations.

   - Given $v = \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{pmatrix}$ and $w = \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{pmatrix}$ we define addition

$$v + w = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \cdots \\ a_n + b_n \end{pmatrix}.$$

- Further, for any given scalar c, we define scalar multiplication

$$cv = \begin{pmatrix} ca_1 \\ ca_2 \\ \cdots \\ ca_n \end{pmatrix}.$$

- There are some natural algebraic properties of these operations which will be formally stated later and used to define abstract vector spaces.
- **Examples.** Consider a linear system given by $2x + 3y = 5, 4x - 3y = 17$.
- Set $v = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, w = \begin{pmatrix} 3 \\ -3 \end{pmatrix}, b = \begin{pmatrix} 5 \\ 17 \end{pmatrix}$.
- Consider the vector calculation:

$$xv + yw = x\begin{pmatrix} 2 \\ 4 \end{pmatrix} + y\begin{pmatrix} 3 \\ -3 \end{pmatrix} = \begin{pmatrix} 2x + 3y \\ 4x - 3y \end{pmatrix}.$$

- Thus our linear system can be reinterpreted as a vector equation:

$$xv + yw = b.$$

- More generally, Consider a system $(A|B)$.
- Suppose that A has n columns $C_1, C_2, \cdots, C_n$ corresponding to the coefficients of the n variables $x_1, x_2, \cdots, x_n$ respectively,
- then the equation

$$x_1C_1 + x_2C_2 + \cdots + x_nC_n = B$$

has the same meaning as the original system of equations.

3. To make it more succinct, we define

Def.17: Span of a set of vectors Given any set S of vectors, we set:

$$Span\ S = \{a_1v_1 + a_2v_2 + \cdots a_mv_m\}$$

where m is any non negative integer, $v_1, v_2, \cdots, v_m$ are some m vectors in S and $a_1, a_2, \cdots, a_m$ are some scalars.

4. Note that the definition is designed to work for an infinite set S, but for a finite set with n elements, we can fix $n = m$.

5. To make our statements even simpler, we now define:

Def.18: Matrix times a vector Given a matrix A with n-columns $C_1, C_2, \cdots C_n$, and a vector $v \in \Re^n$ we set

$$Av = a_1C_1 + a_2C_2 + \cdots + a_nC_n \text{ where } v = \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{pmatrix}.$$

6. Thus, we can now rewrite the system $(A|B)$ as $AX = B$ where $X = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$, i.e.

$B \in Span\{C_1, C_2, \cdots, C_n\}$.

7. Now we define

   Def.19: Column Space of a Matrix For a matrix A with columns $C_1, C_2, \cdots, C_n$ we define $Col(A) = Span\{C_1, C_2, \cdots, C_n\}$.

8. Thus, our consistency condition can be reformulated as $(A|B)$ is consistent iff $AX = B$ has a solution iff $B \in Col(A)$.

9. In view of our earlier consistency condition, this says that B is in $Col(A)$ iff augmenting B to A does not increase its rank!

1.6 Homogeneous Equations

1. **Def. 20: Homogeneous System of Equations.** A linear system $(A|B)$ is said to be homogeneous when $B = 0$, i.e. the RHS entries in B are all 0.

   In this case, the REF of $M = (A|0)$ can be seen to be $M^* = (A^*|0)$, i.e. the column 0 can be omitted through the reduction process, since it will never change.

2. Clearly, $rank(M) = rank(A)$, so a homogeneous system is always consistent. Indeed, it is also clear that $X = 0$ is a solution to $AX = 0$ and hence consistency is directly obvious!

3. Let the common rank be r. Then there are exactly r pivot variables and $n - r$ free variables.

4. The final solution will consist of solving the pivot variables in terms of the free variables and reporting the conclusion.

Here is an example:

1. Consider a system in REF:

$$\left(\begin{array}{ccccc|c} x & y & z & w & t & RHS \\ \hline 2 & 6 & 0 & -4 & 6 & 0 \\ 0 & 0 & -1 & 3 & 2 & 0 \\ 0 & 0 & 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

2. Identify pc list as $(1, 3, 5, \infty)$, pivot variables as x, z, t and the free variables y, w.

3. The fourth equation is ignored. The third gives $t = 0$, the second gives $z = 3w + 2t = 3w$ and the first gives $x = -3y + 2w - 3t = -3y + 2w$.

4. The above solution is best reported as a vector:

$$\begin{pmatrix} x \\ y \\ z \\ w \\ t \end{pmatrix} = \begin{pmatrix} -3y + 2w \\ y \\ 3w \\ w \\ 0 \end{pmatrix} = yv_1 + wv_2$$

   where

$$v_1 = \begin{pmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 2 \\ 0 \\ 3 \\ 1 \\ 0 \end{pmatrix}.$$

Often, it is preferred to replace the original free variables by suitable parameters.

Thus, we may also write:

$$\begin{pmatrix} x \\ y \\ z \\ w \\ t \end{pmatrix} = t_1 v_1 + t_2 v_2$$

We summarize the above results:

1. Thus the solution is seen to be a member of $Span\{v_1, v_2\}$.
2. Denoting a general member of the span as v_h, we write $X = v_h$.

   It is important to remember that v_h stands for any one of an infinite collection of vectors and should not be confused with a specific single vector or with the whole span!
3. In general, if n is the number of variables and $r = rank(A)$ then the solution of $AX = 0$ is always a member of the span of $s = n - r$ vectors $\{V_1, \cdots, V_s\}$.

1.7 The general case $AX = B$

1. More generally, if we put the augmented matrix $M = (A|B)$ of a non homogeneous system into REF, say $M^* = (A^*|B^*)$. then we need to verify the consistency condition first.

   Recall that the consistency condition is: the rank of the LHS matrix A is the the same as the rank of augmented matrix $(A|B)$.

   If the matrix $(A|B)$ is put in REF $M^* = (A^*|B^*)$, then this amounts to the condition that no row in M^* has pivot in the B^* column. We will develop a "general solver" below, which can easily determine the condition even when we replace the RHS B.
2. If the condition fails, then there is no solution.
3. If the condition holds, then the solution process and reporting is just as above, except the final answer is a fixed vector plus a span of $s = (n - r)$ vectors as before.

Here is an example.

1. Consider a system (already) in REF:

$$\left(\begin{array}{ccccc|c} x & y & z & w & t & RHS \\ \hline 2 & 6 & 0 & -4 & 6 & 18 \\ 0 & 0 & -1 & 3 & 2 & -5 \\ 0 & 0 & 0 & 0 & 7 & 35 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

2. Identify pc list as $(1, 3, 5, \infty)$, pivot variables as x, z, t and the free variables y, w.
3. The fourth equation is ignored. The third gives $t = 5$, the second gives
   $z =3w + 2t + 5 = 3w + 10 + 5 = 3w + 15$ and the first gives
   $x = -3y + 2w - 3t + 9 = -3y + 2w - 15 + 9 = -3y + 2w - 6$.

4. The above solution is best reported as a vector:

$$\begin{pmatrix} x \\ y \\ z \\ w \\ t \end{pmatrix} = \begin{pmatrix} -3y+2w-6 \\ y \\ 3w+15 \\ w \\ 5 \end{pmatrix} = v_p + t_1 v_1 + t_2 v_2$$

   where

$$v_p = \begin{pmatrix} -6 \\ 0 \\ 15 \\ 0 \\ 5 \end{pmatrix}, \ v_1 = \begin{pmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \ v_2 = \begin{pmatrix} 2 \\ 0 \\ 3 \\ 1 \\ 0 \end{pmatrix}.$$

   Note that the solution forces $y = t_1$ and $w = t_2$. Sometimes, we may not introduce these new variables, but it is better to bring them in.

5. Note that $Span\{v_1, v_2\}$ is a solution of a related homogeneous equation $AX = 0$.

   Thus, $v_h = t_1 v_1 + t_2 v_2$ describes a general member of the solution of the homogeneous equation.

   It is important to remember that v_h stands for any one of an infinite collection of vectors and should not be confused with a specific single vector or with the whole span!

6. **Notations** *$rownum, colnum$*. For any matrix A, we shall define $colnum(A)$ to be the number of columns in A and $rownum(A)$ to be the number of rows in A.

7. Suppose that $r = rank(A)$ and $n = colnum(A)$. Set $s = n - r$.

   Then we have that the solution of a system $AX = B$ is of the form $X = v_p + v_h$ where v_h is a general linear combination of s solutions of the associated system $AX = 0$.

8. Def. 21: Homogeneous and Particular Solutions. We call v_p as a "particular" solution and v_h as a "homogeneous solution." Note that neither of these are unique, but with proper identifications, they exhibit all the solutions of the system in a parametric form.

9. Def. 22:A homogeneous system $AX = 0$ always has one obvious solution, namely $X = 0$. This is defined to be the "trivial solution." Moreover, as shown above, the system $AX = 0$ has a non trivial solution iff $s = colnum(A) - rank(A) > 0$, or equivalently, there is at least a free variable.

   Since the solution of a homogeneous system is of the form $X = t_1 v_1 + \cdots + t_s v_s$ we can say that $v_p = 0$ and $X = v_h$ for a homogeneous system.

1.8 RREF and its uses

1. We have illustrated how to make and use REF - the "row echelon form". We did not work much with the RREF. Roughly, it needs twice as much work as REF and hence we avoided using it, if it was not really needed. It is, however, needed for some of the later work and we include an illustration of the method to get that form.

   Unlike, REF, the form RREF is well defined and thus has theoretical merit. Typically, if we have reached RREF, then the act of "solving equations", becomes, "writing down the answers".

2. A matrix M is said to be in RREF if the following conditions hold:

- M is in REF.
- Every row is either a zero row or has pivot entry 1.
- Pivots of all rows are "lonely" in their columns. This means that the column containing a pivot entry has zero entry in all other rows.

3. An example of RREF

   Here is a worked out example of converting REF into RREF.

   (a) Consider the following augmented matrix (already in REF) and convert to RREF. Use it to write out the solution of the associated linear system.

   (b)
$$\left(\begin{array}{ccc|r} x & y & z & RHS \\ \hline 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 0 & 0 & 9 & 27 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

   (c) Notice that the pc list is $(1, 2, 3, \infty)$ and the respective pivot entries are $1, 1, 9$.

   (d) The process is to start with the last pivot, make all entries above it equal to zero and make it 1.

   Then repeat with earlier pivots.

   (e) Thus, the operations $R_1 - \frac{3}{9}R_3, R_2 - \frac{6}{9}R_3, \frac{1}{9}R_3$ give:

   (f)
$$\left(\begin{array}{ccc|r} x & y & z & RHS \\ \hline 1 & 1 & 3 & 12 \\ 0 & 1 & 6 & 20 \\ 0 & 0 & 9 & 27 \\ 0 & 0 & 0 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccc|r} x & y & z & RHS \\ \hline 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

   (g) Then the operation $R_1 - R_2$ finishes off the RREF.
$$\left(\begin{array}{ccc|r} x & y & z & RHS \\ \hline 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

   (h) The answer to the system can now be simply read off!

1.9 Linear dependence and independence

1. We have learned the alternate view that if A is a matrix with columns $C_1, \cdots, C_n$ then the equation $AX = B$ is solvable iff B is in the column space $Col(A) = Span(C_1, \cdots, C_n)$.

   We now introduce new concepts which help us decide the nature of solutions more efficiently.

2. **Def. 23: Linearly Dependent vectors** The columns $C_1, C_2, \cdots, C_n$ of A are said to be linearly dependent if the system $AX = 0$ has a non trivial solution, or equivalently $colnum(A) > rank(A)$, i.e. rank of A is less than its number of columns.

3. For future use, we restate this definition more generally thus: **Def. 24(general): Linearly Dependent vectors.** Any set S of vectors is said to be linearly dependent if there is a positive integer n such that n distinct vectors $v_1, v_2, \cdots, v_n$ of S satisfy

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0 \text{ where at least one of } c_i \text{ is non zero.}$$

   This definition is necessary since in general vector spaces, the vectors may not be columns and we may not be able to make the matrix A from them.

4. Note that this makes sense even for an infinite set S.

5. **Def. 25: Linearly Independent Vectors.** A set S of vectors is said to be linearly independent if it is not linearly dependent!

   A better way to understand this is as follows: If we take any distinct vectors $v_1, v_2, \cdots, v_n$ in S and solve the equation $c_1v_1 + \cdots + c_nv_n = 0$, then it has only the trivial solution $0 = c_1 = \cdots = c_n$.

6. Though logically clear, linear independence can be difficult to verify without better tools. We describe such a tool next.

7. **Convention:** We shall often drop the word "linearly" from the terms "linearly dependent" and "linearly independent".

Tests for dependence/independence

Suppose that we have a set of vectors $v_1, \cdots, v_n$ in some vector space V. If w is a given vector, then the vector equation $x_1v_1 + \cdots + x_nv_n = w$ is the analog of our linear system of equations. We describe the corresponding ideas for the abstract vector spaces.

1. First we recall what we know.

   - For a finite set of vectors in $\Re^m$, there is a simple criterion for dependence/independence.
   - Given vectors $v_1, v_2, \cdots, v_n$ in $\Re^m$, make a matrix A by taking these as columns and find its rank (by using REF, for example.).
   - Suppose $rank(A) = r$. It is obvious that $r \le n = colnum(A)$. Then we have:

$$v_1, v_2, \cdots, v_n \text{ are linearly dependent iff } r < n$$

     and thus, they are linearly independent iff $r = n$.

1.10 Generic Solver

1. We discuss a topic which is not in the book, but is a very efficient technique for solving Linear Algebra problems, especially if you have a good calculator handy.

   We learned how to solve a linear system $(A|B)$ for a given right hand side B. Any vector $B \in \Re^m$ can be easily seen to be a well defined combination of special elementary columns

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \cdots \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ \cdots \\ 0 \\ 0 \end{pmatrix}, e_{m-1} = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 1 \\ 0 \end{pmatrix}, e_m = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{pmatrix}.$$

2. Namely
$$B = b_1e_1 + b_2e_2 + \cdots + b_{m-1}e_{m-1} + b_me_m.$$

3. It stands to reason that if we solve each of the systems $(A|e_1), (A|e_2), \cdots, (A|e_m)$, then we can write down the complete solution of any $(A|B)$ by simply combining the answers. It would seem like a lot of work, but in reality, it is just as easy as a single system, since the necessary row operations can stay the same.

4. Thus, we set up an augmented matrix $(A|I)$ where I is the "identity matrix" with columns $e_1, e_2, \cdots, e_m$.

5. We then use the row reduction algorithm to change A to its row echelon form (REF) or, even RREF, if desired. Here is what we shall expect to see:
$$\text{The final form appears as } \left(\begin{array}{c|c} U & U^* \\ \hline 0 & G \end{array}\right)$$

6. The part U has the non zero rows of the REF of our A while 0 below it denotes all its zero rows. Suppose that U has $r = rank(A)$ rows and the last $m - r$ rows are zero.

   The part U^* is simply the transformed part of I across U and G is the important part of the answer in the last $m - r$ rows.

7. We are now ready to handle any given RHS $B = b_1e_1 + b_2e_2 + \cdots + b_{m-1}e_{m-1} + b_me_m$.

   Let $C_1, C_2, \cdots, C_m$ denote the columns of the final RHS $C = \begin{pmatrix} U^* \\ G \end{pmatrix}$.

   It is not hard to see that the REF for $(A|B)$ will have RHS equal to
   $C \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{pmatrix} = b_1C_1 + b_2C_2 + \cdots + b_mC_m.$

   Further all the entries in the last $m - r$ rows can be shown to be $G \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{pmatrix}$ and this must be zero if $(A|B)$ is consistent.

8. **Def. 26: Consistency matrix.**

   The matrix G obtained here is called the "consistency matrix for the system $(A|B)$

   It gives us a simple **Consistency test**, namely: $(A|B)$ is consistent iff $GB = 0$.

   This equation can be interpreted as a condition that B is perpendicular to all the rows of G (transposed into columns).

   Later on, we will see how a complete solution may also be deduced.

1.11 A Linear Transformation.

1. A Linear Transformation extends the idea of a function so that the domain is $\Re^n$ rather than just the field of real numbers.

   The word "Linear" also means that it has the simplest possible formula consisting of ordinary linear functions.

2. **Def.27: A Linear Transformation** is a map $L : \Re^n \to \Re^m$ satisfying two properties:
   (a) $L(v+w) = L(v) + L(w)$ for all $v, w \in \Re^n$ and
   (b) $L(cv) = cL(v)$ for all $v \in \Re^n$ and $c \in \Re$.

3. The map defined by $L(\begin{pmatrix} x \\ y \end{pmatrix}) = \begin{pmatrix} x+y \\ x-y \\ 2x+3y \end{pmatrix}$ defines a Linear Transformation from $\Re^2$ to $\Re^3$.

4. **Verification of the definition.** This is checked from definition thus: Let $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ and $w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$. Then $L(v+w)$ equals:

$$L(\begin{pmatrix} v_1+w_1 \\ v_2+w_2 \end{pmatrix}) = \begin{pmatrix} v_1+w_1+v_2+w_2 \\ v_1+w_1-v_2-w_2 \\ 2v_1+2w_1+3v_2+3w_2 \end{pmatrix}$$

   which simplifies:

$$\begin{pmatrix} v_1+v_2 \\ v_1-v_2 \\ 2v_1+3v_2 \end{pmatrix} + \begin{pmatrix} w_1+w_2 \\ w_1-w_2 \\ 2w_1+3w_2 \end{pmatrix} = L(v) + L(w).$$

5.

$$L(cv) = L(\begin{pmatrix} cv_1 \\ cv_2 \end{pmatrix}) = \begin{pmatrix} cv_1+cv_2 \\ cv_1-cv_2 \\ 2cv_1+3cv_2 \end{pmatrix} = cL(v).$$

   item**When does the definition fail?** The reason that the calculations work is that the formulas are homogeneous linear expressions in the coordinates of the vectors in $\Re^n$.

6. Thus

$$S(\begin{pmatrix} x \\ y \end{pmatrix}) = \begin{pmatrix} x+y+1 \\ x-y \\ 2x+3y \end{pmatrix} \text{ and } T(\begin{pmatrix} x \\ y \end{pmatrix}) = \begin{pmatrix} x+y \\ xy \\ 2x+3y \end{pmatrix}$$

   both fail the definition.

   This should be checked.

1.12 Matrix of a Linear Transformation.

1. Here is an example of a map guaranteed to give a Linear Transformation. Let A be a matrix with real entries having m rows and n columns.

   Def.28: The transformation T_A.

   Define the map $T_A : \Re^n \to \Re^m$ by the formula $T_A(X) = AX$.

2. Then the following calculation shows that T_A is a Linear Transformation.

$$T_A(v+w) = A(v+w) = Av + Aw = T_A(v) + T_A(w)$$

   and

$$T_A(cv) = A(cv) = cAv = cT_A(v).$$

   This can be called as the Linear Transformation defined by A.

3. **How to find the Matrix of a Linear Transformation?**
   (a) We first need some notation.
   Notation: Define a vector e_i^n to be a column with n entries which are all zero except the i-th entry is 1.
   (b) Thus for $n = 3$ we have:
   $$e_1^3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2^3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$
   While working with $\Re^n$ for a fixed n, we often drop the superscript n to simplify our display.

4. **The matrix calculated.**
   Given a map $L : \Re^n \to \Re^m$, we calculate the n columns
   $$v_1 = L(e_1^n),\ v_2 = L(e_2^n),\ \cdots,\ v_n = L(e_n^n).$$
   Let A be the matrix with n columns $v_1, v_2, \cdots, v_n$ in order.

5. Then the theorem is: L is a Linear Transformation iff $L(X) = AX$ for all $X \in \Re^n$.
   In other words, $L = T_A$.

6. **Spaces associated with a Linear Transformation..**
   Given any matrix A with m rows and n columns, we have two natural sets associated with it.
   Def.29: Column and Null spaces of a Matrix.
   (a) $Col(A) = \{AX \mid X \in \Re^n\}$.
   (b) $Nul(A) = \{X \mid AX = 0 \text{ and } X \in \Re^n\}$.

7. We now consider a Linear Transformation $L : \Re^n \to \Re^m$.
   Def.30: Kernel and Image of a Linear Transformation.
   (a) $Image(L) = \{L(X) \mid X \in \Re^n\}$.
   (b) $Ker(L) = \{X \mid L(X) = 0 \text{ and } X \in \Re^n\}$.

8. We shall later define subspaces and show that $Col\ A$ is a subspace of $\Re^m$ and $Nul\ A$ is a subspace of $\Re^n$.

1.13 Relationship with properties of a Linear Transformation.

1. **Def.31: Injective or 0ne-to-one transformation.** A function is said to be one-to-one, if it maps different elements to different elements. For a linear transformation L, it is enough to check $L(v) \neq 0$ if $v \neq 0$.
   In other words if $Ker(L)$ contains only the zero vector. In this case, we call the transformation to be injective or one-to-one.

2. **Def.32: Surjective or onto transformation.** A function is said to be onto if every element of the target space is an image of some element.
   For a linear transformation $L : \Re^n \to \Re^m$, the target space is $\Re^m$ and thus the condition reduces to $Image(L) = \Re^m$. In this case, we call the transformation to be surjective or onto.

1.14 The criteria for Injectivity and Surjectivity.

1. When $L = T_A$, we know that $Ker(L) = Nul(A)$ and so we have $L = T_A$ is injective iff $Nul(A) = 0$, or $rank(A) = colnum(A) = n$.

2. When $L = T_A$, we also know that $Image(L) = Col(A)$ and so we have $L = T_A$ is surjective iff $Col(A) = \Re^m$, or $rank(A) = rownum(A) = m$.

3. A Fundamental Fact: The rank of a matrix with m rows and n columns: It is obvious that $rank(A)$ is less than or equal to $\min(m, n)$.

4. This gives some easy but **important conclusions** which are well worth memorizing!

5. Let A be a matrix with $rownum(A) = m$ and $colnum(A) = n$. Let $L = T_A$. Then $rank(A) \leq \min(m, n)$.

6. If $n > m$, then $rank(A) < n = colnum(A)$ and hence $Nul(A)$ is non zero. Consequently, $Ker(L) \neq 0$ and L cannot be injective.

7. If $m > n$, then $rank(A) < m = rownum(A)$ and hence $Col(A)$ is smaller than $\Re^m$. Consequently, $Image(L) \neq \Re^m$ and hence L cannot be surjective.

8. If $n = m$ then $rank(A)$ may be equal to this common value or may be smaller.

   We discuss this next.

9. **Observations continued.**

   If $rank(A)$ equals this common value $n = m$, then the map L is both surjective and injective.

   Def.33: Isomorphism. The Linear transformation L is said to be an isomorphism if it is both surjective and injective. In this case, it is a one-to-one onto map from $\Re^n$ to $\Re^n$.

10. If $rank(A) < n = colnum(A)$, then $L = T_A$ is not injective.

11. If $rank(A) < m = rownum(A)$, then $L = T_A$ is not surjective.

12. **Creating a Suitable Linear Transformation.**

    We now show how to use the matrix of a transformation to create a Linear Transformation with a desired property.

13. Suppose we want $L : \Re^3 \to \Re^3$ to rotate all vectors about the z-axis.

14. Recall that the vectors e_1^3, e_1^3, e_1^3 are along the x, y, z axes respectively.

15. It is clear that we want

$$L(e_1^3) = e_2^3,\ L(e_2^3) = -e_1^3,\ L(e_3^3) = e_3^3.$$

16. Thus $L = T_A$ where $A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

17. Now, we can find the rotated image of any desired vector, say $L\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$.

Chapter 2

Matrix Algebra.

2.1 What is a Matrix.

1. A Matrix is a rectangular array A of numbers (also called scalars, in vector space terminology.) The number of rows is denoted by $rownum(A)$ and the number of columns by $colnum(A)$.

   Convention: We write $A = A_{m \times n}$ to indicate that A has m rows and n columns and may also express this by saying that A has type (or size) $m \times n$.

2. Notation: An entry in the i^{th} row and j^{th} column of a matrix A may be conveniently denoted as $A(i, j)$. Often in books this is written as A_{ij} or even a_{ij} if the author makes a convention of using corresponding small letters for entries.

3. The set of all matrices of type $m \times n$ with entries from a field K will be denoted by $M_K(m, n)$. We may drop the subscript K, if the scalars are already known, e,g, $\Re$.

2.1.1 Matrix Operations.

1. We wish to define two operations on $M_K(m, n)$.

2. Scalar Multiplication. Given a matrix A and a scalar $c \in K$, we define a new matrix cA defined by $(cA)(i, j) = c(A(i, j))$.

3. Thus if $c = 5$, then

$$\text{for } A = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 1 & -5 \\ 2 & 3 & 0 \end{pmatrix} \text{ we have } 5A = \begin{pmatrix} 5 & 0 & 15 \\ 10 & 5 & -25 \\ 10 & 15 & 0 \end{pmatrix}.$$

2.1.2 Matrix Addition.

1. Matrix Addition. Given $A, B \in M_K(m, n)$, we define $A + B \in M_K(m, n)$ by

$$(A + B)(i, j) = A(i, j) + B(i, j).$$

2. We remark that this is just like the definition in $\Re^n$. Indeed, $\Re^n$ can be thought as a special case of matrices, namely $\Re^n = M_{\Re}(n, 1)$.

3. Thus we get

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

2.1.3 Matrix Product.

1. The most complicated and most important operation is the product.

2. Matrix Multiplication. Important: Given matrices A, B, the product AB is defined only if $colnum(A) = rownum(B)$.

3. When this condition is satisfied, suppose that the common number $colnum(A) = rownum(B)$ is equal to s.

4. Then we define
$$(AB)(i,j) = \sum_{k=1}^{s} A(i,k)B(k,j).$$

5. For example, we take matrices $A_{2\times 3}$ and $B_{3\times 2}$ to calculate:
$$AB = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 5 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 4 & -10 \end{pmatrix} = \begin{pmatrix} 0 & 15 \\ 13 & 2 \end{pmatrix}.$$

2.1.4 Comments on the Product.

- We had earlier defined AX when A had type $m \times n$ and X was a column with n entries, i.e. X had type $n \times 1$. This is a special case of our general product.

- It is important to note that if we multiply AB where $A = A_{m\times s}$ and $B = B_s \times n$, then $AB = AB_{m\times n}$.

- Thus, in our example of product we can see that BA would be a matrix of type 3×3. Thus, in general, BA need not be equal to AB.

- Indeed, it is easy to make examples such that AB is defined, but BA is not! For example take, $A_{2\times 2}$ and $B_{2\times 3}$.

2.2 Some Special Matrices.

- The Zero Matrix 0 is a matrix with all zero entries. To simplify our notation, we often use the same symbol 0 for the scalar zero as well as any sized zero matrix. Its type is deduced from the equation that it fits in.

- A zero matrix serves the purpose of zero in matrix additions.

- Thus $A + 0 = 0 + A = A$ for all A, where 0 is understood to be the same type as A.

- A Square Matrix is a matrix of type $n \times n$ for some positive integer n. These are the only matrices A for which AA is defined!

- The power A^m is defined only for a square matrix A and some positive integer m. It is interpreted as a product of m copies of A.

- Later, we shall extend the definition of A^m to zero or negative integer values of m.

2.2.1 The Unit Matrix.

- The Unit Matrix I is a square matrix A which has all zero entries except for 1's down the main diagonal. We often specify the type of the unit matrix I by writing I_n, if we have an $n \times n$ matrix.
- For example:

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- I behaves like 1 for multiplication. Thus, $AI = IA$ for all matrices A.
- Important: If A is of type 2×3, then the above equation has to be interpreted as $AI_3 = I_2A$.

2.2.2 The Transpose.

- A natural flipping operation converts all members of $M_K(m,n)$ into $M_K(n,m)$. We define A^T to be the matrix obtained by turning all rows of A into columns. Thus

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

- It is easy to verify the following:

$$(A+B)^T = A^T + B^T, \quad (cA)^T = c(A^T).$$

  Also, $(A^T)^T = A$.
- With a little more work we check

$$(AB)^T = B^T A^T.$$

2.2.3 Symmetric Matrices.

- A matrix A is said to be symmetric if $A^T = A$. Also, A matrix A is said to be antisymmetric if $A^T = -A$.

  Note that A must be square for either of these definitions to hold.
- We leave it as an exercise to prove:
- Let A be any square matrix.
  1. $A + A^T$ is symmetric.
  2. $A - A^T$ is antisymmetric.
  3. Challenge: A can be written as the sum $B + C$ where B is symmetric and C is antisymmetric. Moreover, B and C are uniquely determined by A.

2.2.4 Elementary Matrices.

- Let $p \neq q$ with $1 \leq p, q \leq n$ and let $c \in K$. we define $E^n_{pq}(c)$ to be the matrix I_n where we have replaced the (p, q)-entry by c.
- Thus, for example

$$E^3_{23}(7) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 7 \\ 0 & 0 & 1 \end{pmatrix}, \quad E^4_{32}(-4) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -4 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

- Observation: We can remember $E^n_{pq}(c)$ as the matrix obtained from I_n by applying the elementary row operation $R_p + cR_q$.

2.2.5 Diagonal Matrices.

- A matrix similar to the Identity but slightly different is the diagonal matrix.
- A diagonal matrix $diag(a_1, a_2, \cdots, a_n)$ is a square matrix of type $n \times n$ which has all zero entries, except for the entries $a_1, a_2, \cdots a_n$ on the main diagonal.
- We can see that $diag(a_1, a_2, \cdots, a_n)M$ gives a matrix which is same as M, except its successive n rows are multiplied by the scalars $a_1, a_2, \cdots, a_n$.
- If we multiply the diagonal matrix on the right, then it multiplies the columns instead.

2.2.6 Permutation Matrices.

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

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

- As before, we drop the superscript n when convenient.
- Multiplication by P_{ij} on the left swaps rows i, j while multiplication on the right permutes columns i, j.

2.3 Voodoo Principle.

- An important principle of working with matrices is the voodoo principle.
- **Row operations.**

  If you wish to do something to the rows of a matrix M, make a matrix A obtained by performing it on I and then take AM.

  Naturally, choose I so that $colnum(I) = rownum(A)$.

- Example:

  Let
  $$M = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \end{pmatrix}.$$
  If you wish to swap its first two rows, then multiply it by P_{12}^3. Thus you can check
  $$P_{12}^2 M = \begin{pmatrix} 3 & 4 & 5 \\ 1 & 2 & 3 \end{pmatrix}.$$

2.3.1 More Voodoo.

- To perform $R_2 - 2R_1$ multiply by $E_{21}(-2)$.
  $$\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & -1 \end{pmatrix}.$$
- 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.** These can be performed by the same idea, except you multiply the prepared matrix on the right.
- Thus $M \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ permutes the two columns of M.

2.4 Inverse.

- Def: 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 $AM = MA = I = I_n$. Note that by definition of matrix products, M must also have type $n \times n$.
  - We shall soon prove that a matrix $A = A_{n\times n}$ is invertible iff $rank(A) = n$ iff RREF of A is I_n.

2.4.1 Some consequences of the definition.

- Note the following obvious identities:
  $$E_{ij}(c)E_{ij}(-c) = I, \ P_{ij}P_{ij} = I$$
  $$diag(a_1, \cdots, a_n)diag(\frac{1}{a_1}, \cdots, \frac{1}{a_n}) = I$$
  provided $a_1, \cdots, a_n$ are all non zero.
- If $AM = MA = I$ and $AK = KA = I$, then $M = K$. To see this, simply observe $MAK = (MA)K = K$ while at the same time $MAK = M(AK) = M$.
- Thus, it makes sense to Define: A matrix M is an inverse of a matrix A if $AM = MA = I$. Note that the definition already forces A and M to be square of the same type and we denote $M = A^{-1}$. Observe! $(A^{-1})^{-1} = A$.

2.4.2 Important obvious formula:

- We assume all matrices to be square of equal type.

$$(AB)^{-1} = B^{-1}A^{-1},\ (A_1 A_2 \cdots A_r)^{-1} = (A_r^{-1} A_{r-1}^{-1} \cdots A_1^{-1}).$$

  The formula means that the LHS exists iff all parts of the RHS exist. Note the reversal of order, it is crucial!

- A simple observation. If a matrix A is invertible, then it cannot have a zero row or a zero column.

- **Proof.** Suppose $AM = MA = I$ and A has a zero i-th row. Clearly AM would have a zero i-th row also but I has no zero rows. If A has a zero j-th column, then MA would have a zero j-th column, leading to a contradiction.

2.4.3 Criterion for Inverse discussed.

- To see if $AM = I$ has a solution, we need to try and solve n equations:
  $AX = C_1, AX = C_2, \cdots, AX = C_n$ where $C_1, C_2, \cdots, C_n$ are n columns of I.

- A little thought says we might as well solve them at once, i.e. try to reduce the augmented matrix $(A|I)$.

- The Voodoo principle says that all our row operations correspond to left multiplication by one of the following matrices.

- elementary or permutation or diagonal matrices (with non zero diagonal entries).

- Let a product of these "operational matrices" in correct order be H, so that HA is in RREF.

- Thus, we have $H(A|I) = (HA|HI) = (HA|H)$ where HA is in RREF.

2.4.4 Criterion for Inverse concluded.

- Since H is invertible, it does not have any zero row, so $HI = H$ does not have any zero rows.

- If the RREF HA has a zero row, then for consistency, the corresponding row in H would be zero - a contradiction!

- Hence, the pc list of the RREF of A must be $(1, 2, \cdots, n)$, i.e. the RREF HA must be I.

- Also, note that then columns of H are the solutions of the n equations we started with and hence $AH = I$.

- Thus, we have proved that A is invertible iff the RREF process for the augmented matrix $(A|I)$ ends in $(I|H)$ for some matrix H.

- Moreover the resulting H is the desired inverse of A.

- Also, clearly, a square $n \times n$ matrix A has RREF I iff its rank equals n. Thus, if we need to know the invertibility of A without finding A^{-1}, then $rank(A) = n$ is the criterion!

2.5 Using the Failure.

- Suppose we start with a matrix A which may or may not be square and still attempt the procedure to find an inverse. Thus, we reduce $(A|I)$ to REF, by using a matrix H so that

$$H(A|I) = (HA|H) \text{ where } HA \text{ is in REF.}$$

- Suppose that A fails our usual invertibility test which means the pc list of HA has a few ∞. We write this as

$$HA = \begin{pmatrix} A^* \\ 0 \end{pmatrix} \text{ where } A^* \text{ is in REF without zero rows.}$$

- We write the RHS part $HI = H = \begin{pmatrix} G^* \\ G \end{pmatrix}$ where G consists of the rows across the zero rows on left.

- Recall that none of the rows of G are zero, since they are rows of the invertible matrix H. This matrix G indicates the failure of invertibility and actually has many practical uses.

2.5.1 Using the matrix G.

- Recall that we have

$$H(A|I) = (HA|H) = \left(\begin{array}{c|c} A^* & G^* \\ \hline 0 & G \end{array}\right)$$

- It follows that equations $AX = B$ transform to

$$(HA)X = \left(\begin{array}{c} A^*X \\ \hline 0 \end{array}\right) = \left(\begin{array}{c} G^*B \\ \hline GB \end{array}\right) = HB$$

- The zero rows on the LHS give us $0 = GB$.

  These are the precise consistency conditions for our system $AX = B$.

2.6 The Consistency Matrix.

- Def: Row equivalence We define two matrices P, Q to be row equivalent if there is an invertible matrix H such that $HP = Q$. We have been using this concept without formally defining it. For example all matrices are row equivalent to any of their REF, since we have been using invertible operations only.

- Def: Consistency Matrix We define the matrix G to be a consistency matrix for A. It is not unique, but all consistency matrices of A are row equivalent.

- Note that if the REF of A has no zero rows, then our matrix G does not exist or is a matrix that may be called empty.

2.6.1 Use of the Consistency Matrix.

- Thus, we have Theorem: $AX = B$ is solvable iff $GB = 0$.

- **Proof.**

  The necessity of this is already evident.

  The sufficiency follows because if $GB = 0$ then it remains to solve $A^*X = G^*B$ and this is solvable since A^* is in REF with pivots in every row!

2.6.2 Understanding the Column Space.

- Thus we have a new description of the column space of any matrix A with m rows and n columns. Let its consistency matrix be G.

  Then
  $$Col(A) = \{B \mid AX = B \text{ is solvable}\} = \{B \mid GB = 0\}$$
  where B is assumed to be in $\Re^n$

- Moreover, if $rank(A) = rownum(A) = m$ then G is empty and $Col(A) = \Re^m$, because B is free of any conditions!

- Using our old definition of a Null Space, we have also established:

  $$Col(A) = Nul(G) \text{ where } G \text{ is the consistency matrix of } A.$$

Chapter 3

Determinants.

3.1 Quick Summary.

1. Given any square matrix $A = (a_{ij})$, we define its **determinant** variously denoted as $\det(A)$ or $|A|$ or $||A||$. The definition needs some auxiliary terms.

   - First for any (i, j) such that a_{ij} is defined, we define the **index of** $(\mathbf{i}, \mathbf{j})$ to be $\mathrm{ind}(i, j) = (-1)^{(i+j)}$.
   - Further we define the **minor of** $(\mathbf{i}, \mathbf{j})$ to be $\mathrm{minor}(A, i, j)$ which is the subdeterminant of A obtained by throwing away the i-th row and the j-th column.
   - We also define the **cofactor of** $(\mathbf{i}, \mathbf{j})$ to be $\mathrm{cofactor}(A, i, j) = \mathrm{ind}(i, j)\mathrm{minor}(A, i, j)$.
   - Then the determinant $|A|$ can be computed as $\sum a_{ij}\mathrm{cofactor}(A, i, j)$ where the sum is taken over all entries a_{ij} **coming from any chosen row or column.**

2. The value of the determinant gets multiplied by k if all entries in a single row or column are multiplied by k.

3. The value of the determinant gets multiplied by -1, if a single exchange of two rows or two columns is carried out. For more complicated permutations, we multiply by the sign of the permutation.

4. The value of the determinant is unchanged if we add a multiple of one row to another. Similar result holds for columns.

5. The value of a determinant is 0 if some row or column consists of zero entries only! (This follows from the definition.)

6. The value of a lower triangular determinant is equal to the product of its diagonal entries. Ditto for upper triangular. In general, this is how determinants are computed: reduce the determinant to upper or lower triangular form and then evaluate the product of the diagonal entries. If permutations are used along the way, then suitable sign is attached to the answer. Sometimes, an expansion along a suitable row/column is also used to reduce the work.

7. There is a more general expansion, the so-called Laplace expansion, which works with several rows (or columns) at once, instead of the single row (or column) as in the definition.

8. The adjoint of a matrix A is a matrix denoted by A^{adj} whose (i, j)-th entry is equal to $\mathrm{cofactor}(A, j, i)$. **Do notice the switch in the order!** The adjoint satisfies the identity

$$AA^{\mathrm{adj}} = A^{\mathrm{adj}}A = |A|I.$$

This lets us write the inverse of A as $A^{\text{adj}}/|A|$. Of course it exists iff $|A| \neq 0$.

We often use the notation $\text{adj}(A)$ in place of A^{adj}.

9. For a general matrix M **its rank** $\text{rank}(M)$ is defined to be the largest number r such that M has a nonzero subdeterminant of size r. Thus a square $n \times n$ matrix is invertible iff its rank is n. Rank of a matrix is obviously less than or equal to its rownum as well as colnum.

10. In general, **the equations** $\mathbf{AX} = \mathbf{B}$ **are solvable** iff $\text{rank}(A) = \text{rank}(A|B)$. Here $A|B$ stands for the augmented matrix. Obviously, if $\text{rank}(A) = \text{rownum}(A)$ then $AX = B$ is solvable for all B. The converse is true too!

11. **Cramer's Rule.** If we wish to solve a system $AX = B$ where A is a square $n \times n$ matrix we proceed thus.

    Define a convenient notation: $A(v, i)$ which is obtained from A by swapping the i-th column of A with the column v.

    We claim that when $AX = B$, then $\det(A)X_i = \det(A, B, i)$. The proof can be seen as follows. Consider $adj(A)AX = adj(A)B$. Since $adj(A)A = \det(A)I$ by comparing i-th row entries on both sides, we deduce that $\det(A)X_i = Row(adj(A), i)B$. If we recall the definition of the adjoint, we know that the i-th row of $adj(A)$ gives the sequence of cofactors of the i-th column of A. Thus the RHS is nothing but the $\det(A(B, i))$.

    - Assume $\det(A) \neq 0$. Then the above calculation gives a formula for the solution, namely $X_i = \frac{\det(A,B,i)}{\det(A)}$. This is the Cramer's rule.
    - If $\det(A) = 0$ but one of $\det(A(B, i))$ is not zero, then we get that the system is inconsistent.
    - If $\det(A) = 0$ and $\det(A(B, i)) = 0$ for all i, then it can be seen that one of the equations can be dropped (being dependent on the others) and one of the variables becomes free. Moving the free variables entries to the RHS (and treating them as constants) we reduce the problem to a smaller sized determinant!

      Example: Consider the system $AX = B$ where $A = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 1 & 4 \\ 4 & 3 & 10 \end{pmatrix}$ and $B = \begin{pmatrix} -2 \\ 4 \\ 0 \end{pmatrix}$.
      We check that $\det(A) = 0$ and all $\det(A(B, i)) = 0$. In fact we see that the third equation is simply the sum of the second equation with twice the first. So, we drop it and see if some pair of variables (corresponding to a pair of columns) give a nonzero 2×2 determinant. We see that $\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$ has a nonzero determinant, so the (remaining) third variable (say X_3) can be free. Thus we get a new pair of equations $PY = Q$ where $P = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$, $Y = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $Q = \begin{pmatrix} -2 - 3X_3 \\ 4 - 4X_3 \end{pmatrix}$. This can now be solved by the above rule, leaving X_3 free.

3.2 Theory behind Determinants.

After having learned how to compute a determinant of a square matrix, we now discuss the basic idea behind the notion of determinants. A determinant is a number associated with a sequence of n vectors in $\Re^n$. It is supposed to help us determine the various properties of the geometric object described by these vectors.

Thus, in $\Re^2$ consider two vectors which can be thought of as two arrows coming out from the origin. They define a parallelogram. If one of the vectors is simply a multiple of the other, then the parallelogram collapses to a line. If we take the two vectors $v_1 = \begin{pmatrix} a \\ b \end{pmatrix}$ and $v_2 = \begin{pmatrix} c \\ d \end{pmatrix}$, then we can form the determinant

$$\det(v_1, v_2) = \det(\begin{pmatrix} a & c \\ b & d \end{pmatrix} = ad - bc = \Delta \text{ say.}$$

With some geometric calculations, it is easy to establish that $|\Delta|$ gives the area of the parallelogram and moreover, the sign of Δ even gives us a measure of the angle from v_1 to v_2. Here is the calculation. Using the usual idea of polar coordinates, we can write

$$v_1 = \begin{pmatrix} a \\ b \end{pmatrix} = r_1 \begin{pmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{pmatrix} \text{ where } r_1 = \sqrt{a^2 + b^2}$$

and θ_1 is the angle made by v_1 with the x-axis measured counterclockwise.

Similarly, we write

$$v_2 = r_2 \begin{pmatrix} \cos(\theta_2) \\ \sin(\theta_2) \end{pmatrix} \text{ where } r_2 = \sqrt{c^2 + d^2}$$

and θ_2 is the angle made by v_2 with the x-axis measured counterclockwise.

Then we have

$$\Delta = r_1 r_2 \left(\cos(\theta_1)\sin(\theta_2) - \sin(\theta_1)\cos(\theta_2)\right) = r_1 r_2 \sin(\theta)$$

where $\theta = \theta_2 - \theta_1$ which is the angle measured from v_1 to v_2 counterclockwise.

Thus the determinant gives the signed area of the parallelogram constructed from v_1 towards v_2. It also lets us decide if the vectors are linearly dependent. We note that for dependent vectors, the angle θ is either 0 or π. In either case $\sin(\theta) = 0$. Thus $\Delta = \det(v_1, v_2) = 0$ iff v_1, v_2 are linearly dependent. [1]

We also note three natural properties enjoyed by these determinants:

1. If a vector is scaled by multiplying by a constant p then the determinant also gets multiplied by the same p.

2. If the vectors are swapped, the determinant gets multiplied by -1.

3. If a vector w is equal to $w_1 + w_2$, then

$$\det(w, v) \ = \ \det(w_1, v) + \det(w_2, v).$$

   This is easily checked from the formula and it can also be geometrically verified for the areas of parallelograms.

We can appeal to the geometric argument in three space $\Re^3$ as well. But in higher dimensions, we don't have a preconceived idea of "volumes". We therefore take a clue from the algebraic calculations and define the n-dimensional determinant as follows.

Axioms of determinants.

[1] We have not paid careful attention to the zero vector in this analysis. Fortunately, when one of the two vectors is zero, then the vectors are clearly linearly dependent and also the $\det(v_1, v_2) = 0$. Thus, our conclusion is easily verified!

1. Given any ordered n-tuple $v_1, v_2, \cdots, v_n$ in $\Re^n$ the expression $\det(v_1, v_2, \cdots, v_n)$ shall be a well defined real number. Thus det is a function from n-tuples of vectors to $\Re$.

2. **Alternating property.** This function shall be an alternating function. This means, if any two vectors v_i, v_j are exchanged, then the value of the determinant shall be multiplied by -1.

3. **Linearity.** The determinant shall be a linear function of each of its arguments $v_1, v_2, \cdots, v_n$. This means:

$$\det(w_1 + w_2, v_2, \cdots, v_n) = \det(w_1, v_2, \cdots, v_n) + \det(w_2, v_2, \cdots, v_n)$$

   and

$$\det(kv_1, v_2, \cdots, v_n) = k \det(v_1, v_2, \cdots, v_n) \text{ for any } k \in \Re.$$

   Similar conditions hold for each of the n arguments.[2]

4. If v_1, v_2, v_n are the unit vectors, i.e. the matrix with columns $v_1, v_2, \cdots v_n$ in order gives the identity matrix I_n, then $\det(v_1, v_2, \cdots, v_n) = 1$.

Outline of the argument.
It can be shown that subject to these conditions, there exists a unique determinant function. Moreover, it satisfies all the properties that we informally asserted.
Here is a sketch of the argument.

1. We prove the result by induction on the size n of the determinant.

   Our inductive statement is:

   Let A be the matrix formed by the vectors $v_1, v_2, \cdots, v_n$ as columns. We note that the entries of v_1 are $A(1,1), A(2,1), \cdots, A(n,1)$.

   For any $i = 1, 2, \cdots n$ consider vectors $w_2^i, w_3^i, \cdots, w_n^i$ in $\Re^{n-1}$ obtained by dropping the i-th entry from each of the $n-1$ vectors $v_2, v_3, \cdots, v_n$.

   Then $\det(v_1, v_2, \cdots, v_n) = \det(A) =$
   $A(1,1)cofactor(A,1,1) + A(2,1)cofactor(A,2,1) + \cdots + A(n,1)cofactor(A,n,1)$ where

$$cofactor(A,i,1) = (-1)^{i+1} minor(A,i,1) = (-1)^{i+1} \det(w_2^i, w_3^i, \cdots w_n^i).$$

2. Thus the starting case shall be $n = 2$ and here we know everything already.

3. Now assume the result for $n = 2$.

   For $n = 3$, our formula gives the usual expansion for the first column. The formula is easily seen to satisfy the linearity condition for the first argument v_1 and the linearity as well as the alternating property for the second and third components is easily seen from the inductive assumption.

   The only non trivial calculation is the proof that $\det(v_1, v_2, v_3) = -\det(v_2, v_1, v_3)$. This is checked by an easy but tedious calculation of expanding both determinants completely.

4. Now assume the result for all determinants of size less than or equal to $n-1$ and define $\det(v_1, v_2, \cdots, v_n)$ as in the statement, where the cofactors are defined and satisfy known properties by induction hypothesis.

   We note that the linearity of the formula is again easily seen to be true and as before, it is enough to prove that $\det(v_1, v_2, \cdots, v_n) = -\det(v_2, v_1, \cdots, v_n)$. We may either do the easy

[2] A little thought may show that in view of the alternating condition, it would be enough to assume this condition just for the argument v_1, since we can deduce it for other arguments by swapping the vectors and then swapping them back.

but tedious calculation or make the following shortcut. Note that any vector in $\Re^n$ can be written as a sum of at most n-vectors which have only one non zero entry. Using this, we can assume that our v_1 and v_2 each have only one non zero entry. In that case, the cofactor expansions by the first two vectors give an easy formula for both sides.

Suppose v_1 has only one non zero entry p in ith position and v_2 has only one non zero entry in position j. We invite the reader to verify these statements:

- If $i = j$, then $\det(v_1, v_2, \cdots, v_n) = \det(v_2, v_1, \cdots, v_n) = 0$. So the claim is true.
- If $i < j$ then $\det(v_1, v_2, \cdots, v_n) = (-1)^{i+j-1}\Delta$ where Δ is the determinant formed by vectors $w_3, \cdots w_n$ which are obtained from $v_3, \cdots, v_n$ after droppin their i-th and j-th entries. Moreover $\det(v_2, v_1, \cdots, v_n) = (-1)^{j+i}\Delta$. Thus $\det(v_1, v_2, \cdots, v_n) = -\det(v_2, v_1, \cdots, v_n)$.
- If $i > j$, then the calculation is similar, except $\det(v_1, v_2, \cdots, v_n) = (-1)^{i+j}\Delta$ and $\det(v_2, v_1, \cdots, v_n) = (-1)^{j+i-1}\Delta$.

5. Thus, the inductive step is complete and the result is proved.

Once the existence of the determinant function is proved, it is easy to prove uniqueness as well as all the properties asserted in the quick introduction.

Chapter 4

General Vector Spaces

1. We briefly discuss axioms for general vector spaces. The best way to understand them is to study examples. You should consult a separate file of notes on vector spaces on this web page.

2. Never forget the basic example $\Re^n$. Also, keep in find new examples of P (the space of all polynomials in one variables over a field - usually the field is $\Re$), function spaces with values in a field, space of matrices of a fixed size and so on.

3. We no longer presume that vectors are columns of numbers, and we no longer simply look at the rank of some matrix to answer questions, except after a careful analysis and argument.

4. We give examples of this next.

4.1 Vector Spaces

For our purposes a **field**, k, is a subset of the complex numbers that is closed under addition, multiplication, and additive and multiplicative inverses. Examples are the rationals, $\mathbf{Q}$, the reals, $\Re$, and the complex numbers, $\mathbf{C}$. In general one can assume one is working with $k = \Re$ although we will need $\mathbf{C}$ at some point.
A **vector space** over a field k is a non empty set V together with a well defined addition '+' and a scalar multiplication '$\cdot$' by elements of the field k, satisfying the following axioms.

- **Operations.** For all $u, v \in V$ and $c \in k$ we have $u + v \in V$ and $cu \in V$. Sometimes, we write $c \cdot u$ but this $\cdot$ is often omitted.

- **Addition properties.** The addition is commutative ($u + v = v + u$) and associative ($u + (v + w) = (u + v) + w$).

- **Additive identity and inverse.** There is a zero vector, denoted 0 such that $u + 0 = 0 + u = u$ for all u. Moreover, for each u there is a u^* such that $u + u^* = u^* + u = 0$. The u^* can be shown to be uniquely defined by u and is denoted as $-u$. The zero vector is also shown to be unique!

- **Distributivity and unitariness.** The two operations interact naturally:

$$c(u + v) = cu + cv \ , \ (c + d)u = cu + du \ , \ (cd)u = c(du).$$

  Moreover, $1(u) = u$ for all u.

As noted above this course, the field k is usually $\Re$ the field of real numbers. In that case, we drop the phrase "over the field $\Re$".

4.1.1 Subspaces

If V is a vector space over k and W is a non empty subset, then we say that W is a **subspace** of V if we have:

- For all $w_1, w_2 \in W$ we have $w_1 + w_2 \in W$.
- For all $c \in k$ and $w \in W$ we have $cw \in W$.

We note that the vector 0 will always belong to a subspace as soon as it is non empty, since $w \in W$ implies $0w = 0 \in W$ by the second subspace condition above.
Hence, you may replace the condition of W being non empty by the simpler condition $0 \in W$, as done in the book.

4.1.2 Examples

A challenging example. Here is an exotic example of a vector space which should be studied to verify your understanding of the above definition.
Let $V = \Re$ be made into a vector space over the usual field of real numbers $\Re$ as follows:

- We define a new addition $\oplus$ on V by the formula:

  $$v \oplus w = v + w - 1$$

  where the operations on the right hand side are the usual operations in real numbers.
- We define a new scalar multiplication $\odot$ by $\Re$ on V by the formula:

  $$c \odot v = cv + 1 - c$$

  where, as before, the operations on the right are the usual operations in real numbers.

It is instructive to verify all the axioms from these definitions. You should also identify what $-v$ means. This example should be kept in mind while analyzing all the following concepts.
You should also make an enhanced version of this example by taking V to be $\Re^n$ as the set, but using the above definitions of addition and scalar multiplication, suitably generalized. [1]

A Universal example.

Let S be any non empty set and consider

$$F_S^k = \{f : S \to k \mid \text{ where } f(s) = 0 \text{ for all except finitely many } s \in S. \}$$

It can be shown that every vector space can be described in this manner, but finding such an explicit S can be tedious and it is better to use the basic definition.
If $k = \Re$ we may drop it from the notation.
It is easy to verify how F_S^k is a vector space by defining $(f + g)(s) = f(s) + g(s)$ and $cf(s) = cf(s)$.
The extra condition on f is not necessary, but it is essential if you want to claim that every vector space has a standard structure!

[1] Can you make other examples of such weird operations? Here is a general hint and secret of all such constructions.
Let V be any vector space over a field k. Let ψ be any bijective (ii.e. injective and surjective) function from V to itself.
Define a new vector space W which uses the same V as an underlying set but defines operations as follows.

$$w_1 \oplus w_2 = \psi^{-1}(\psi(w_1) + \psi(w_2)) \text{ and } c \odot w = \psi^{-1}(c\psi(w)).$$

It can be shown that W is a vector space "isomorphic" to V which means essentially the same as V. See below for explanation of "isomorphic".
Can you guess the ψ for the example given above? Hint: try $x \to x + \alpha$ for some α.

Basic examples.

Here are some of the standard examples.

- **Euclidean spaces.** The space k^n consisting of all n-tuples of elements of k, usually written as a column.

  This can be described as F_S^k where S is the set $\{1, 2, 3, \cdots, n\}$. A typical function f in the vector space may be displayed as $\begin{pmatrix} f(1) \\ f(2) \\ \cdots \\ f(n) \end{pmatrix}$. This leads to the usual notation for k^n. For our purposes k will almost always be $\Re$ in which case we are talking about the familiar $\Re^n$.

- **The case of an infinite S.** If we take $S = \{1, 2, \cdots, n, \cdots\}$, the set of natural numbers, then we find it convenient to display $f \in F_S$ as

  $$f(1) + f(2)x + f(3)x^2 + \cdots f(n+1)x^n + \cdots .$$

  Note that the description of F_S implies that after some large enough exponent N, the coefficients are all zero and we have a set of polynomials.

  The book denotes $F_S^{\Re}$ by the symbol $\mathbb{P}$. A general notation for F_S^k is also $k[x]$ which is the ring of polynomials in x with coefficients in k, where, we have chosen to ignore the usual multiplication of polynomials!

  We now note that if we keep the same set $S = \{1, 2, \cdots, n, \cdots\}$ but drop the special condition on functions we get a much bigger set, namely

  $$H = \{f : S \to k\}.$$

  As before, any such $f \in H$ be displayed as $f(1) + f(2)x + f(3)x^2 + \cdots + f(n+1)x^n + \cdots$.

  Since there is no special condition on the function, we now get power series! The general notation for this set is $k[[x]]$, the ring of power series in x with coefficients in k, where, as before, we ignore the usual product of power series.

  It can be shown that $H = F_T^k$ for some convenient set T, but finding such a T is a daunting task!

  There is also a well known subset of H when $k = \Re$, namely the set of convergent power series. To write it as $F_T^{\Re}$ is an even more daunting task!

- **Spaces of matrices:** $Mat_{m,n}$, the set of m by n matrices with matrix addition and scalar multiplication.

- **Function spaces:** If X is any set then $F_X^{\Re} = \{f : X \to \Re\}$ is a vector space with the usual addition of functions and multiplication of functions by a scalar. (i.e. $(f + g)(x) = f(x) + g(x)$, $(\alpha f)(x) = \alpha f(x)$)

- **Direct sums:** If U and V are vector spaces and $U \oplus V = \{(u, v) \mid v \in U, v \in V\}$ then $U \oplus V$ is a vector space with addition $(u_1, v_1) + (u_2, v_2) = (u_1 + v_1, u_2 + v_2)$ and scalar multiplication $\lambda (u, v) = (\lambda u, \lambda v)$. This allows us to define vector spaces tailored to specific problems and to view more complex spaces as constructed of simpler ones. For example we can, properly interpreted, view $\Re^7$ as $\Re^2 \oplus \Re^5$.

Important Remark: A subspace of a vector space V is simply a subset of V that is closed under linear combinations. Except for the following exotic example, all of the vector spaces we will encounter in this course either fall in the above list of examples or are subspaces of members of this list. This means that once a prospective vector space has been identified as a subset of a known vector space (e.g. the above examples) then all that is required to verify that it is a vector space is to check that it is closed under linear combinations.
Example: Let $C^1(\Re)$ be the set of all differentiable functions $f : \Re \to \Re$. This is a vector space since:

- it is a subset of the vector space $F_\Re^\Re$
- If $f, g \in C^1(\Re)$ and $\alpha, \beta \in \Re$ then $\frac{d}{dx}(\alpha f + \beta g) = \alpha \frac{d}{dx} f + \beta \frac{d}{dx} g$ so $(\alpha f + \beta g) \in C^1(\Re)$

4.1.3 Basic structures in a vector space

Now let V be a k-vector space (i.e. vector space over a field k).
For any subset $A \subset V$, we define its **span**:

$$Span\ A = \{c_1 v_1, \cdots + c_m v_m \mid \text{ where } c_i \in k, v_i \in A \text{ and } m \text{ is some non negative integer.}\}.$$

Note that $Span\ A$ can be described as the set of all possible **linear combinations** of elements of A. Note that even when A is infinite, we only allow finitely many elements of it at a time! Also note that m is allowed to be zero and it gives the combination 0, by a standard convention.
We say that a set A **spans V or is a spanning set for V if** $Span\ A = V$.
A subset $A \subset V$ is said to be **linearly dependent** if there are elements $v_1, \cdots, v_m \in A$ such that $c_1 v_1 + \cdots + c_m v_m = 0$ for some $c_1, \cdots, c_m \in k$ with at least one non zero element c_i among them. In application, other convenient forms of this condition are used. One such version is:
A subset $A \subset V$ is said to be linearly dependent if there is some $v \in A$ such that v is a linear combination of some elements $w_1, \cdots, w_r \in A$ which are distinct from v. A compact way of saying this is to write that $v \in Span\ A \setminus \{v\}$.
A set $A \subset V$ is said to be **linearly independent**, if it is not linearly dependent.
We often drop the word "linearly" from these terms.
A subset $A \subset V$ is said to be **a basis** of V if

$$Span\ A = V \text{ and } A \text{ is linearly independent.}$$

Convention: If V is any vector space then $\{O\}$, the set consisting only of O, the zero vector of V is a subspace of V. It is logically true that the empty set, Φ is independent and by convention it is a basis for $\{O\}$.
We say that V is **finite dimensional** if it has a finite basis.
Important Observation: If V is a vector space and B_1 and B_2 are bases for V with $B_1 \subset B_2$ then $B_1 = B_2$.
proof: If $B_1 \neq B_2$ then there would have to be $b \in B_2$ that is not in B_1. However B_1 is a spanning set for V which means that b_2 must be a linear combination of elements in B_1. But $B_1 \subset B_2$ so b_2 is a linear combination of the other elements of B_2 which says that B_2 is dependent which is a contradiction.
Definition: The number of elements in a basis for V, is said to be **the dimension** of V. We write $\dim V$ or $\dim_k V$ if we wish to identify k.
We will soon argue that the dimension is a well defined number for any vector space, i.e. every basis of a vector space has the same number of elements.
Examples:

- The standard basis $\{e_1, e_2, \cdots, e_n\}$ is linearly independent and a spanning set for $\Re^n$, thus the dimension of $\Re^n$ is n.

- If V is a vector space, recall that by convention the empty set Φ is a basis for the zero subspace $\{O\}$. Thus the zero subspace has dimension 0 since Φ has 0 elements.

Infinite Dimensional Spaces A vector space that is not finite dimensional (i.e. does not have a finite basis is said to be infinite dimensional. It is true that even infinite dimensional spaces do have bases (linearly independent spanning sets). However we don't say that the dimension is the "number of elements in a basis". For infinite dimensional spaces we need to use the finer notion of "number of " called "cardinality[2] " which distinguishes between different infinite sets. Properly dealing with "cardinality" would take too much time so, for our purposes, we won't try to define the dimension of a vector space that is not finite dimensional but will rather simply say that all such spaces are infinite dimensional.

4.2 Homomorphisms, Isomorphisms and Automorphisms

Given k-vector spaces V, W a map $T : V \to W$ is said to be a **linear transformation** if it satisfies these two conditions:

- $T(v_1 + v_2) = T(v_1) + T(v_2)$ for all $v_1, v_2 \in V$.
- $T(cv) = cT(v)$ for all $c \in k$ and $v \in V$.

We may also use the term "homomorphism" (meaning similar structure) to denote such a map. There are two concepts associated with the notion of a linear transformation (homomorphism). First is "the image of T" which can be formally denoted as

$$T(V) = \{T(v) \mid v \in V\}.$$

Second is "the Kernel of T" which can be defined as:

$$Ker\ T = \{v \in V \mid T(v) = 0\}.$$

It is easy to verify that both the Kernel and the Image are respectively subspaces of V and W.
The homomorphism T is **injective (or "one to one ")** iff $Ker\ T = 0$ where we have used a slightly abused notation 0 in place of $\{0\}$. This abuse is routinely done!
The homomorphism T **surjective (or "onto")** if $T(V) = W$ i.e. W is its total image.
The homomorphism T is said to be an **isomorphism** if it is both injective and surjective i.e. bijective. The word "iso" denotes sameness and the concept says that the two vector spaces with an isomorphism mapping one to the other are essentially the same. They, can be treated as replacements for each other in analyzing their properties.
An isomorphism of V to itself is called an automorphism.

4.2.1 Coordinates

We now show that **every vector space is essentially of the form F_S^k for some set S.** Here "essentially" means that if we have a vector space then it is isomorphic to F_S^k for some S. In fact, S can be taken to be any basis for V and if S is finite, say S has n elements then V is isomorphic to k^n ($\Re^n$ if $k = \Re$)
Let V be a vector space with a basis A. We show that V is isomorphic to F_A^k.
Note that by the definition of the basis, every vector $w \in V$ has a unique expression
$w = c_1v_1 + \cdots + c_rv_r$ for some $v_1, \cdots, v_r \in A$ and $c_1, \cdots, c_r \in k$.

[2] Two sets A, B are said to have the same cardinality if there is a bijective map from A to B. As the example of the sets $\{1, 2, 3, \cdots, n \cdots\}$ and $\{2, 4, 6, \cdot, 2n, \cdots\}$ shows, a set can have the same cardinality as a proper subset. This suggests that one has to be very careful in dealing with infinite cardinality.

For any $v \in A$, we define the v-coordinate of w with respect to A by the formula:

$$w_{[v,A]} = c_i \text{ if } v = v_i \text{ and } 0 \text{ if } v \text{ is not among the } v_1, \cdots, v_r.$$

Now we define the map $\Phi : V \to F_A^k$ by

$$\Phi(w)(v) = w_{[v,A]}.$$

It is not hard to see that this is indeed an isomorphism!
The main trick is to note that from the definition of F_A^k, given any $f \in F_A^k$ we can see that the element $w = \sum_{v \in A} f(v)v$ is a well defined member of our vector space $Span\ A$. Our definition of Φ makes $\Phi(w) = f$. This shows surjectivity.
Injectivity is obvious.
Thus, we have shown why every vector space is essentially of the form F_S^k for some set S.

4.2.2 Connection with $\mathbf{k^n}$

Consider a finite dimensional k-vector space V and a basis B which we write as a formal vector

$$B = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$$

of vectors in V.
We construct a map $V \to k^n$ as follows. Given $v = c_1 v_1 + \cdots + c_n v_n \in V$ we shall write

$$[v]_B = \begin{pmatrix} v_1 \\ \cdots \\ v_n \end{pmatrix} \in \Re^n.$$

The resulting vector shall be called the coordinate vector of v **with respect to the ordered basis** B.
This is best remembered by a natural identity:

$$v = B[v]_B.$$

This is to be interpreted as the usual product of a row with a column, generalized for our vector entries.

4.3 Reduction to the old case.

1. Consider a real vector space V with a given basis $B = \begin{pmatrix} p & q & r & s \end{pmatrix}$. This means, every vector $v \in V$ has a unique expression

$$v = ap + bq + cr + ds \text{ where } a, b, c, d \in \Re.$$

2. We declare an important notation $[v]_B = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$, and say that $[v]_B$ is the coordinate vector of v with respect to the basis B. Remember that B is an ordered basis! A change of order changes the coordinate vector.

3. This coordinate vector satisfies a fundamental identity (to be understood and memorized):

For every $v \in V$ and a basis B of V, we have: $v = B[v]_B$.

4. This defines a linear transformation $L : V \to \Re^4$ defined by $L(v) = [v]_B$.
5. Convince yourself from basic definitions that this L is an isomorphism. So, all concepts of linear dependence, independence etc. about subsets of V can be deduced by taking the image under L.
6. For example consider the following vectors in our vector space V.
$$v_1 = 2p - q + r, v_2 = p + q - 3r, v_3 = p + q + r$$
$$v_4 = q + s, v_5 = p + q + r + s, v_6 = 3p - 2r.$$
7. Using the basis $B = \begin{pmatrix} p & q & r & s \end{pmatrix}$, we calculate the corresponding coordinate vectors:
$$[v_1]_B = \begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \end{pmatrix}, [v_2]_B = \begin{pmatrix} 1 \\ 1 \\ -3 \\ 0 \end{pmatrix}, [v_3]_B = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$
$$[v_4]_B = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, [v_5]_B = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, [v_6]_B = \begin{pmatrix} 3 \\ 0 \\ -2 \\ 0 \end{pmatrix}.$$

4.3.1 Various questions.

1. Now we can ask and answer all questions about these vectors.
2. What are the dimensions of $Span\{v_1, v_2, v_3, v_4\}$ and $Span\{v_1, v_2, v_3, v_6\}$?
3. Are the vectors v_1, v_2, v_3, v_4 independent? Do they form a basis of V?
4. Prove that $C = \begin{pmatrix} v_1 & v_2 & v_3 & v_5 \end{pmatrix}$ is a basis of V.
5. Find $[v_4]_C$, $[v_3]_C$, $[v_2 - 5v_3]_C$.
6. Find the dimension and basis for $Span\{v_1, v_2, v_6\}$.
7. Find the dimension and basis for $Span\{v_1, v_2, v_4, v_6\}$.
8. Find the dimension and basis for $Span\{v_1, v_2, v_3, v_4, v_5, v_6\}$.

4.4 Change of Basis

1. Given a basis $B = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}$ of an n-dimensional space, recall that we always have:
$$v = B[v]_B \text{ for any } v \in V.$$
2. Given another basis $C = \begin{pmatrix} w_1 & w_2 & \cdots & w_n \end{pmatrix}$ we see that $w_i = B[w_i]_B$ for all $i = 1, \cdots, n$.
3. It follows that
$$C = \begin{pmatrix} w_1 & w_2 & \cdots & w_n \end{pmatrix} = B \begin{pmatrix} [w_1]_B, [w_2]_B, \cdots, [w_n]_B \end{pmatrix}.$$
Thus we see that
$$v = C[v]_C = B \begin{pmatrix} [w_1]_B, [w_2]_B, \cdots, [w_n]_B \end{pmatrix} [v]_C = BM_B^C [v]_C$$
where $M_B^C = \begin{pmatrix} [w_1]_B, [w_2]_B, \cdots, [w_n]_B \end{pmatrix}$ is called the change of basis matrix from B to C.
4. Thus, $M_B^C [v]_C = [v]_B$.

4.4.1 Example.

1. Let V have a basis $\begin{pmatrix} u & v & w \end{pmatrix}$ and let $C = \begin{pmatrix} u+v & u+v+w & v+w \end{pmatrix}$. It can be shown that C is also a basis of V.

2. The coordinate vectors with respect to B of members of C are
$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

3. Verify that
$$M_B^C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

4. This says that $C = BM_B^C$. Show that $C(M_B^C)^{-1} = B$ or $M_C^B = (M_B^C)^{-1}$.

5. Using it or directly to calculate $[u]_C, [v]_C, [w]_C$.

6. The inverse comes out to be:
$$M_C^B = \begin{pmatrix} 0 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 0 \end{pmatrix}.$$

7. It follows that for any vector h with $[h]_B = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ we get
$$[h]_C = M_C^B \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

8. In particular:
$$[u]_C = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, [v]_C = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, [w]_C = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

4.5 Matrix of a Transformation.

1. Consider a vectors space V with basis B and vector space W with basis C.

   Explicitly, assume that $B = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}$ and $C = \begin{pmatrix} w_1 & w_2 & \cdots & w_m \end{pmatrix}$.

2. Let $L : V \to W$ be a linear transformation. We define the matrix of transformation of L with respect to bases B, C to be the matrix $M = \begin{pmatrix} [L(v_1)]_C & [L(v_2)]_C & \cdots & [L(v_1)]_C \end{pmatrix}$.

3. The matrix can also be defined by the property that for all $v \in V$, we have $[L(v)]_C = M[v]_B$.

4.5.1 Example.

1. Let $V = P_3$ with basis $B = \begin{pmatrix} 1 & x & x^2 & x^3 \end{pmatrix}$ and $W = P_2$ with basis $C = \begin{pmatrix} 1 & x & x^2 \end{pmatrix}$. Let $L : P_3 \to P_2$ defined by $L(p(x)) = p'(x) - xp''(x)$.
2. Note that $L(1) = 0, L(x) = 1, L(x^2) = 0, L(x^3) = -3x^2$.
3. Then the matrix of L with respect to the bases B, C is:
$$M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}.$$
4. Next we show how to use this matrix to calculate properties of L.
5. As above, assume that $L : V \to W$ be a linear transformation and let M be its matrix with suitable bases B, C for V, W respectively.
6. We see that L is injective iff $Ker(L) = 0$ iff $Nul(M) = 0$.
7. We see that L is surjective iff $Im(L) = W$ iff $Col(M) = \Re^m$ where $m = \dim(W) =$ the number of elements in the basis C.
8. Given vectors $u_1, u_2, \cdots u_s \in V$, we can test if their images $L(u_1), L(u_2), \cdots, L(u_r)$ are linearly independent iff the columns $M[u_1]B, M[u_2]_B, \cdots, M[u_r]_B$ are independent.
9. Given a vector $w \in W$ we can find a vector $v \in V$ with $L(v) = w$ thus:

   Solve the equation $MX = [w]_C$. Then $v = BX$ is the answer. Of course, the number of solutions can be $0, 1, \infty$ by the usual theory of equations.

4.5.2 Example of using the matrix.

1. Recall the map $L : P_3 \to P_2$ studied earlier. Recall the matrix $M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}$.

   Now we decide its properties.
2. Note that $rank(M) = 2$ and thus its columns are dependent and $Col(M)$ is a 2-dimensional space.

   Since $\dim(P_3) = 4 > 2 = rank(M)$ L is not injective. Moreover $Ker(L) = BNul(M)$.

   Since $Nul(M)$ has a basis $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$,

   the $Ker(L)$ has basis $1, x^2$.

4.6 A Fundamental Theorem (existence of bases).

This is the fundamental theorem in the theory of vector spaces.

Theorem 4.1. *Let V be a vector space with a spanning set A. Then there is a subset B of A such that B is a basis of V. That is B is also a spanning set for V and B is independent.*

We shall give a complete proof in case A is finite and a partial proof in case A is infinite.
Proof. If A is independent, then it is a basis and we are done. If A is dependent, then there is some vector $v \in A$ which is a linear combination of vectors in $A_1 = A \setminus \{v\}$.
We now claim that $Span\ A = Span\ A_1$ by the following argument.
Proof of claim. Note that any $w \in Span\ A$ can be written as: $w = cv + w_1$ where $c \in k$ and $w_1 \in Span\ A_1$.
By assumption, $v \in Span\ A_1$, so cv and hence w belongs to $Span\ A_1$. Thus $Span\ A \subset Span\ A_1$.
Clearly $Span\ A_1 \subset Span\ A$ since $A_1 \subset A$.
This shows $Span\ A = Span\ A_1$.
Thus, if $V = Span\ A$ and A is dependent, then we get a proper subset A_1 of A such that $V = Span\ A_1$. We can now apply the same argument to A_1 and either get a basis or a smaller spanning set A_2.
In case A is a finite set to begin with, this cannot continue indefinitely and we must get a basis at some stage.
We remark that we may run into an empty subset of A, in case the vector space is the zero space $\{0\}$. However, in this case the whole set A can only be $\{0\}$ or empty and we have nothing to argue!
It is also possible to do the above proof in a reversed process. We can start with independent subsets of A and enlarge them as much as possible. We argue that eventually we should get a basis.
In case the set A is infinite requires a more general inductive principle called Zorn's Lemma. This lemma allows us (we omit the details here) to prove that there are maximal independent subsets in any vector space. Such sets are easily seen to be spanning sets.
The next important result in vector spaces is that **any two bases of a vector space have exactly the same number of elements.** This number is the **dimension** of the vector space.
We only describe the argument when we have one finite basis.

Theorem 4.2. *Suppose V is a vector space and $F = \{F_1, F_2, \cdots, F_s\}$ is a finite subset of V. Suppose further that the set of vectors $B = \{B_1, B_2, \cdots, B_s, B_{s+1}\}$ is contained in the linear span of F. Then B is linearly dependent set.*

Proof: For each $i = 1 \cdots s+1$ there is a linear combination

$$a_{i,1}F1 + a_{i,2}F_2 + \cdots + a_{i,s}F_s = B_i$$

Let A be the $s+1 \times s$ matrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,s} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,s} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i,1} & a_{i,2} & \cdots & a_{i,s} \\ \vdots & \vdots & \vdots & \vdots \\ a_{s,1} & a_{s,2} & \cdots & a_{s,s} \\ a_{s+1,1} & a_{s+1,2} & \cdots & a_{s+1,s} \end{pmatrix}$$

Then letting $F = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_i \\ \vdots \\ F_s \end{pmatrix}$ and $B = \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_i \\ \vdots \\ B_s \\ B_{s+1} \end{pmatrix}$ we write

$$AF = B$$

Now we augment A by the $s+1 \times s+1$ identity matrix to get $< A|I >$ which we then row reduce to get $< R|Q >$ where R is an REF of A and Q is an invertible matrix such that $QA = R$.
Since R has more rows than pivots its last row must be all zeros and therefore the last row of Q is a row of the consistency matrix *which cannot be all zero since it is a row of an invertible matrix!*.
Let this last row be $C = [c_1, c_2, \cdots, c_{s+1}]$ then $CB = 0$ because $AF = B$ is consistent. Hence $CB = c_1B_1 + c_2B_2 + \cdots + c_sB_s + c_{s+1}B_{s+1} = 0$ where not all of the c_i are 0. Thus, we have that B is linearly dependent.
Example: Let P_3 be the vector space of polynomials of degree at most 3 and let $F_1 = x, F_2 = x^2, F_3 = x^3$ and $B_1 = x + x^2$,$B_2 = x - 4x^3$, $B_3 = x + 3x^2 + x^3$, $B_4 = 2x - x^2 + x^3$ then according to the theorem the four elements of $B = \{B_1, B_2, B_3, B_4\}$ are linearly dependent since they are linear combinations of the three members of $F = \{F_1, F_2, F_3\}$. Moreover, we can calculate a dependence relation. Here

$$1\,F_1 + 1\,F_2 + 0\,F_3 = B_1$$

$$1\,F_1 + 0\,F_2 + -4\,F_3 = B_2$$

$$1\,F_1 + 3\,F_2 + 1\,F_3 = B_3$$

$$2\,F_1 - 1\,F_2 + 1\,F_3 = B_4$$

so setting

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -4 \\ 1 & 3 & 1 \\ 2 & -1 & 1 \end{pmatrix}$$

We have $A \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix}$

and $REF(< A|I >) = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & -4 & -1 & 1 & 0 & 0 \\ 0 & 0 & -7 & -3 & 2 & 1 & 0 \\ 0 & 0 & 0 & -\frac{32}{7} & \frac{5}{7} & \frac{13}{7} & 1 \end{pmatrix}$

According to the theorem it must be true that

$$\begin{pmatrix} -\frac{32}{7} & \frac{5}{7} & \frac{13}{7} & 1 \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{pmatrix} = 0$$

In fact $\frac{-32}{7}(x + x^2) + \frac{5}{7}(x - 4x^3) + \frac{13}{7}(x + 3x^2 + x^3) + (1)(2x - x^2 + x^3) = 0$

Theorem 4.3. *Any two bases of a vector space have the same number of elements.*

Proof: We prove the theorem for finite dimensional vector spaces. The theorem is true for infinite dimensional spaces but requires the idea of *cardinality* to replace "number of" and that would take too much time for an introductory course.
Suppose the vector space has two linearly independent spanning sets, B and C and that B has m elements and C has n elements. We can assume $m \leq n$. If $m < n$ then $c_1, c_2, \cdots, c_m, c_{m+1}$ are all in the linear span of the m elements of B and are therefore linearly dependent, a contradiction.
Thus $m = n$.

Definition: If V is a vector space then the **dimension** of V is the number of elements in any basis of V.
The above theorem says that dimension is well-defined since all bases have the same number of elements.

Theorem 4.4. *Any $n+1$ elements in a vector space of dimension n is linearly dependent.*

Proof: This follows immediately from Theorem A.

Corollary 4.5. *If V is a vector space of dimension n and W is a subspace of V then $dim(W) \leq n$.*

Corollary 4.6. *If V is a vector space of dimension n then any n linearly independent elements of V is a basis of V.*

Proof: Let $B = \{b_1, b_2, \cdots, b_n\}$ be a set of n linearly independent elements of V. We need only show that V is in the linear span of B since we know it is linearly independent. Let v be any element of V. Then $\{b_1, b_2, \cdots, b_n, v\}$ is a set with $n+1$ elements in an $n-$dimensional vector space so it is dependent. This means that there are scalars α_i, not all zero, such that

$$\alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n + \alpha_{n+1} v = O$$

It must be true that $\alpha_{n+1} \neq 0$ since otherwise

$$\alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n = O$$

but that is not possible since If $\alpha_{n+1} \neq 0$ then $v = -\frac{\alpha_1}{\alpha_{n+1}} b_1 + -\frac{\alpha_2}{\alpha_{n+1}} b_2 + \cdots + -\frac{\alpha_n}{\alpha_{n+1}} b_n$ and we are done.

Corollary 4.7. *If V is a vector space of dimension n and W is a subspace of V of dimension n then $W = V$.*

Proof: The basis for W has n independent elements so it must be a basis for V.

4.6.1 Extension of independent sets to Bases

Theorem 4.8. *Suppose V is a vector space and A is a linearly independent subset of V then there is a basis B of V which contains A.*

Proof: Although the theorem is true in general, the proof for infinite dimensional spaces requires extra tools and we will restrict our attention to the case when V is a subspace of some finite dimensional space W. Let W have dimension N.
Suppose L is the linear span of A and $A = \{a_1, a_2, \cdots, a_s\}$. A is a basis for L so if $L = V$ then V has a basis containing A.
If $L \neq V$ then there is $a_{s+1} \in V$ that is not in L. It follows then that $\{a_1, a_2, \cdots, a_s, a_{s+1}\}$ is linearly independent. If this were not the case then there would be

$$\alpha_1 a_1 + \alpha_2 a_2 + \cdots + \alpha_s a_s + \alpha_{s+1} a_{s+1} = O$$

with not all of the $\alpha_i = 0$. But α_{s+1} must be 0 or a_{s+1} is in the span of A. However if this is the case then A is linearly dependent.
If $\{a_1, a_2, \cdots, a_s, a_{s+1}\}$ does not span V then the process can be repeated to find $a_{s+2}, \cdots, a_{s+k}$. Since $s + k \leq N$ and N is the maximum number of elements in an independent subset of V the process must stop at which time $\{a_1, a_2, \cdots, a_s, a_{s+k}\}$ is a spanning set for V. But since it is independent it must be a bais and $s = k = N$.

4.6.2 A matrix approach to extension of independent sets to bases

The following is a method for implementing the Theorem 4.8 in the case where V is $\Re^n$ and A is a linearly independent subset of $\Re^n$.
Suppose $v_1, v_2, ..., v_r$ are independent vectors in $\Re^n$. Then the above theorem says that one can find vectors $v_{r+1}, \cdots, v_n$ such that the vectors $v_1, \cdots, v_n$ are linearly independent and are therefore a basis for $\Re^n$. If A_1 is the matrix with columns $v_1, v_2, ..., v_r$ then this is equivalent to saying that we can add $n-r$ columns to A_1 to get a matrix A with n linearly independent columns. The following construction does this.
Form $(A_1|I)$ by augmenting A_1 by the n by n identity and let R be an REF of $(A_1|I)$. Let C be the consistency matrix of A_1. Since $rank(A_1) = r$ and has n rows, the REF of A_1 has exactly $n-r$ zero rows and hence the consistency matrix C has $n-r$ rows. If C^t is the transpose of C then its columns are independent since they are columns of an invertible matrix. Then $A = A_1|C^t$, is n by n and we claim its columns are linearly independent. We will prove this later as an application of orthogonality.

Example: Let $A_1 = \begin{pmatrix} 1 & 3 \\ 0 & 1 \\ 2 & 4 \\ 3 & -1 \end{pmatrix}$ then the REF of $< A|I >$ is $\begin{pmatrix} 1 & 3 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 2 & 1 & 0 \\ 0 & 0 & 0 & 7 & \frac{-3}{2} & 1 \end{pmatrix}$. The consistency matrix is $C = \begin{pmatrix} -2 & 2 & 1 & 0 \\ 0 & 7 & \frac{-3}{2} & 1 \end{pmatrix}$. So $A =< A_1|C^t >= \begin{pmatrix} 1 & 3 & -2 & 0 \\ 0 & 1 & 2 & 7 \\ 2 & 4 & 1 & \frac{-3}{2} \\ 3 & -1 & 0 & 1 \end{pmatrix}$

To see that A has independent rows we can, for instance calculate its REF and get
$\begin{pmatrix} 1 & 3 & -2 & 0 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 9 & \frac{25}{2} \\ 0 & 0 & 0 & \frac{314}{9} \end{pmatrix}$

4.7 Dimension theorem for vector spaces

Now we come to the most useful theorem in dimension theory of vector spaces.

Theorem 4.9. *Let $T : V \to W$ be a surjective linear transformation of vector spaces. Then*

$$\dim V = \dim W + \dim Ker\ T.$$

Proof. Select a basis A_1 of $Ker\ T$ and a set A_2 such that $T(A_2)$ gives a basis of W. (This is done using the surjectivity of T.)
Now it is possible to argue that $A = A_1 \bigcup A_2$ is a basis of V. Here is the outline.
First show that $Span\ A = V$. Let $v \in V$. Then $T(v) \in Span\ T(A_2)$. Thus we have $u_2 \in Span\ A_2$ such that $T(v) = T(u_2)$ or $v - u_2 = u_1 \in Ker\ T$.
Then $u_1 \in Span\ A_1$ and thus $v = u_1 + u_2$ is in $Span\ A_1 \bigcup A_2$.
The independence of the set A is argued thus. Suppose some non trivial combination of elements of A is zero. Write the combination as $u_1 + u_2 = 0$ where $u_1 \in Span\ A_1$ and $u_2 \in Span\ A_2$.
Taking the image by T we see that

$$T(u_1 + u_2) = T(u_1) + T(u_2) = 0 + T(u_2).$$

Since $T(A_2)$ is a basis of W, we see that $T(u_2)$ is a trivial combination and hence u_2 is also a trivial combination.
Thus $u_2 = 0$.

From independence of A_1 it follows that u_1 is also the trivial combination.
Now note that A_1, A_2 are disjoint sets, since otherwise a basis of W will contain a zero vector (the image of a common element).
Hence, for finite dimensional V we get that the number of elements of A equals the sum of the number of elements of A_1 and A_2.
This proves the formula.
Note: Our proof did not use the finite dimensionality of the vector spaces. It is needed only to count the elements of A as those of A_1 and those of A_2.

4.7.1 A matrix theory explanation of the dimension formula

Let V have dimension n and W have dimension m. Then there are isomorphisms $\phi_1 : \Re^n \to V$ and $\phi_2 : W \to \Re^m$ which means that we have the composite linear transformation

$$T^* : \Re^n \to V \to W \to \Re^m$$

Since ϕ_1 and ϕ_2 are isomorphisms, it follows that $kernel(T) = \phi_1(kernel(T^*)$ and $image(T^*) = \phi_2(image(T))$.
This means that we can assume that $V = \Re^n$ and $W = \Re^m$ in proving the formula.
Let A be the standard matrix for T then $T(X) = AX$ for $X \in \Re^n$ and the kernel of T is the null space of A while the image of T is the column space of A. This means we are reduced to showing the following:

Theorem 4.10. *Let A be an m by n matrix and R an REF of A. If r is the number of pivot columns of R and $A_{p_1}, \cdots, A_{p_r}$ are the corresponding columns of A then $A_{p_1}, \cdots, A_{p_r}$ is a basis for the column space of A and $n - r$, the number of free variables, is the dimension of the null space of A.*

Proof The construction of the complete parametric solution to the homogeneous linear system $AX = O$ represents each element of the null space in the from $x_{f_1}V_{f_1} + \cdots + x_{n-r}V_{f_{n-r}}$ where the x_{f_i} are the free variables. The corresponding V_{f_i} ar linearly independent because V_{f_i} contains as a sub-vector the corresponding i^{th} column of the $n - r$ by $n - r$ identity matrix. This shows that the dimension of the null space is $n - r$ by exhibiting a basis with $n - r$ elements.
To see that the pivot columns of A (*not the pivot columns of* R) form a basis for the column space we first note that the pivot columns of R are linearly independent since the sub-matrix they form is in REF with a pivot in every column. If the pivot columns of A were dependent then there is a non-trivial vector $X \in \Re^n$ such that the pivot variable components of X are 0 and $AX = 0$.
However A and R have the same null space so $RX = O$ which means that the pivot columns of R are dependent, a contradiction.
Finally, if R_u is any non-pivot variable of the R then Ru is a linear combination of the pivot columns of R. This means that there is a vector X with 1 in the position corresponding to R_u and such that $RX = O$. Again, since A and R have the same null space, $AX = O$ which means that the column A_u of A is a linear combination of the pivot columns of A. This means that all of the columns of A are in the span of the pivot columns of A and since they are independent, they are a basis for the column space.

Corollary 4.11. *The dimension of the column space of a matrix A is the rank of A. This means that the rank is well-defined and that all calculations of the rank of a matrix must yield the same value.*

4.7.2 An Important Consequence of the Dimension Formula

If U and W are subspaces of the vector space V then $U + W = \{u + w \mid u \in U, w \in W\}$ and $U \cap W = \{v \in V \mid v \in U \text{ and } v \in W\}$ are subspaces of V. It is very easy to check that both are closed under linear combinations.

Theorem 4.12. *If U and W are subspaces of the vector space V then*

$$dim(V + W) + dim(V \cap W) = dim(V) + dim(W).$$

Proof: Let $M = V \oplus W$. If B_V is a basis for V and B_W is a basis for W then $\{(b_V, O) \mid b_V \in B_V\} \cup \{(O, b_W) \mid b_W \in B_W\}$ is a basis for M so

$$dim(M) = dim(V) + dim(W)$$

Let $T : M \to V + W$ be defined by $T(v, w) = v - w$. Since $(v, -w) \in M$ for every $v \in V$, $w \in W$ and $T(v, -w) = v + w$ we see that T is surjective. So by the dimension formula

$$\dim U \oplus W = \dim U + W + \dim Ker\ T.$$

or

$$\dim U + \dim W = \dim U + W + \dim Ker\ T.$$

So to complete the theorem we must prove that $\dim Ker\ T = \dim(U \cap W)$.
To see this we note that $kernel(T) = \{(u, w) \mid T((u, w)) = u - w = O\}$ that is $kernel(T) = \{(u, v) \mid u \in U, v \in V, u = v\}$. With this in mind we define $\phi : kernel(T) :\to U \cap V$ by $\phi((u, w)) = u(= w)$
The mapping ϕ is easily checked to be an isomorphism which means that $\dim(kernel(T)) = \dim(U \cap W)$.
Example: Suppose A is a 7 dimensional subspace of $\Re^{10}$ and B is a 6 dimensional subspace of $\Re^{10}$. Show that $A \cap B$ contains a non-zero vector.
Solution: By the above theorem $\dim(A \cap B) = dim(A) + dim(B) - dim(A + B)$. Since $A + B$ is a subspace of $\Re^{10}$ it has dimension at most 10 so $\dim(A \cap B) \geq dim(A) + dim(B) - 10$ or

$$\dim(A \cap B) \geq 7 + 6 - 10 = 3.$$

Therefore $\dim(A \cap B)$ contains a linearly independent set with 3 elements, no one of which can be O.

Chapter 5

Eigenvalues and Eigenvectors.

5.1 Eigenvectors and Eigenvalues in abstract spaces

5.1.1 Introduction

In these notes, we start with the definition of eigenvectors in abstract vector spaces and follow with the more common definition of eigenvectors of a square matrix.
Then we discuss the diagonalization problem for a linear transformation.
Finally, we discuss all cases eigenvectors of 2×2 matrices.

5.1.2 Eigenvectors and Eigenvalues.

Let V be a vector space over a field K. Let L be a linear transformation from V to itself.
A scalar $\lambda \in K$ is said to be an **eigenvalue** for L if there is a **non zero vector** v such that $L(v) = \lambda v$.
A vector $v \in V$ is said to be an **eigenvector for L** if it satisfies two conditions:

1. $v \neq 0$.

2. $L(v) = \lambda v$ for some $\lambda \in K$.

When the above conditions are satisfied, we get that λ is an eigenvalue for L and we will describe this by saying v belongs to the eigenvalue λ.

Examples.

1. Let $A = \begin{pmatrix} 5 & 2 & 1 \\ 0 & 3 & -1 \\ 0 & 0 & 1 \end{pmatrix}$. Define a transformation L from $\Re^3$ to $\Re^3$ by $L(v) = Av$. Note that

$$\begin{pmatrix} 5 & 2 & 1 \\ 0 & 3 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \\ 0 \end{pmatrix} = 5 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

This shows that the vector $v = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is an eigenvector of L belonging to the eigenvalue 5.

It is also clear that any non zero multiple of v also has the same property.

It is possible to show that $\lambda = 3$ and $\lambda = 1$ are also eigenvalues for L. What are eigenvectors belonging to them?

2. Consider a linear transformation $T : P_3 \to P_3$ defined by $T(p(x)) = xp'(x) - 3p'(x)$.

   Verify that the polynomials $1, (x-3), (x-3)^2, (x-3)^3$ are all eigenvectors belonging to different eigenvalues.

   For example

$$T((x-3)^2) = x(2(x-3)) - 3(2(x-3)) = 2(x-3)(x-3) = 2(x-3)^2.$$

   Thus, $(x-3)^2$ is an eigenvector belonging to eigenvalue 2.

3. Consider a linear transformation S on the space of twice differentiable real functions given by $S(y) = y''$.

   Verify that the eigenvectors for S are the exponential functions $\exp(rx)$ for various values of r. What eigenvalue does $\exp(rx)$ belong to?

   For example $\exp(3x)$ belongs to the eigenvalue 9.

5.1.3 Eigenspaces.

Let V be a vector space with a linear transformation L from V to itself. Let λ be any scalar in K. We define the subspace V_λ to be the space of all vectors v such that $L(v) = \lambda v$.
Note that we are not requiring the vector v to be non zero here!

- Clearly, every non zero vector in V_λ is an eigenvector for L belonging to the eigenvalue λ.
- Note that λ is an eigenvalue for L iff V_λ is not just the zero vector space.
- We define V_λ to be the eigenspace belonging to λ if it is not just the zero vector space (i.e. λ is an eigenvalue.)
- If $V = \Re^n$ and L is defined by an $n \times n$ matrix A as $L(X) = AX$, then the space $V_\lambda = Nul(A - \lambda I)$.

  Thus, for such transformations, all eigenvalues λ can be identified as scalars λ for which $A - \lambda I$ is singular, i.e. $\det(A - \lambda I) = 0$.

Examples of eigenspaces.

- For the transformation L in example (a) above, there are three non zero eigenspaces. For the eigenvalue 5, the space V_5 is the one dimensional space spanned by $w_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. It is also described as $Nul((A - 5I))$.

  Similarly, for the eigenvalue 3 we calculate the $V_3 = Nul((A - 3I))$ and find its basis to be $w_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$. Finally for the eigenvalue 1 we get V_1 with basis $\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}$.

  Later on we shall see that the resulting set of 3 vectors $w_1, w_2, w_3\}$ is linearly independent and hence a basis for the space $\Re^3$.

- In the example (b) above, we find the eigenspaces by solving the equation $S(p) = xp' - 3p' = \lambda p$ for the various eigenvalues. Unlike example (a) we don't have just the luxury of finding Null spaces of a suitable matrix.

  The solution to $xp' - 3p' = \lambda p$ can be seen to be $p = (x-3)^\lambda$ and thus we get the eigenvalues $0, 1, 2, 3$ and the resulting four polynomials $1, (x-3), (x-3)^2, (x-3)^3$ are independent and form a basis of their respective eigenspaces. The four vectors together form a basis of P_3.

- For the example (c) above, we have to work like example (b) and solve $S(y) = y'' = \lambda y$.

  You should verify that for $y = \exp(rx)$ we get $y'' - \lambda y = (r^2 - \lambda)\exp(rx)$, so for $\lambda = 0, 1, 4, 9$ we get respective eigenvectors $\exp(0) = 1, \exp(x), \exp(2x), \exp(3x)$ respectively.

  Even though these are independent and form a basis of their respective eigenspaces, they do not give a basis of our vector space since the vector space itself is infinite dimensional.

  Note: Before long, we will learn how to compute the eigenspaces by evaluating the Null space of a suitable square matrix.

- **Notations.** Let $L : V \to V$ be a linear transformation. Let $B = (v_1 \quad v_2 \quad \cdots \quad v_r)$ be a sequence of vectors in V which we treat as if it is a **generalized row vector.**

  If we have a column vector $\begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_r \end{pmatrix}$ then we shall give the natural meaning to

  $B\begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_r \end{pmatrix} = a_1v_1 + a_2v_2 + \cdots + a_rv_r..$ Also, if M is any matrix with r rows, then BM similarly makes sense.

  Then, we will define $L(B) = (L(v_1) \quad L(v_2) \quad \cdots \quad L(v_r))$.as the image of the generalized vector.

  Important Observation. In this notation, suppose that $B = (v_1 \quad v_2 \quad \cdots \quad v_r)$ is a sequence of eigenvectors of L belonging to the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_r$ then we get a natural equation:

  $$L(B) = (\lambda_1v_1 \quad \lambda_2v_2 \quad \cdots \quad \lambda_rv_r) = (v_1 \quad v_2 \quad \cdots \quad v_r)\begin{pmatrix} \lambda_1 & 0\cdots & 0 & \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \cdots & \lambda_r \end{pmatrix}.$$

  If our vector space is the usual $\Re^m$, and if our linear transformation is multiplication by a matrix A, then this takes a more familiar form as follows. Let P be the matrix formed by the vectors of B as columns and D the diagonal matrix of type $r \times r$ with $\lambda_1, \lambda_2, \cdots, \lambda_r$ along the diagonal. Then the above equation takes on the form:

  $$AP = PD.$$

  We will be interested in the special case of this when P itself is a square and invertible matrix. Then the equation will be more conveniently rewritten as $P^{-1}AP = D$ and this will be the so-called diagonalization process described below.

5.2 Eigenvectors of square matrices.

As we saw above, for more abstract vector spaces we have to adopt a different strategy to find eigenvalues and eigenvectors. Now we show how we can reduce the problem of finding eigenvectors to that of finding a Null space of a matrix.

5.2.1 The matrix of a linear transformation.

We describe how to calculate the matrix of a linear transformation of a finite dimensional vector space with respect to a given basis.
Here are the steps.

- Let L be a linear transformation from a vector space V to itself and assume that V has a basis $B = \begin{pmatrix} w_1 & w_2 & \cdots & w_n \end{pmatrix}$.
- Next we calculate the vectors $L(w_1), L(w_2), \cdots, L(w_n)$.
- Let $v_1 = [L(w_1)]_B, v_2 = [L(w_2)]_B, \cdots, v_n = [L(w_n)]_B$ be their respective coordinate vectors in the basis B.
- Make a matrix A whose columns are these vectors $v_1, v_2, \cdots, v_n$.

This matrix is the so-called matrix of the linear transformation in the basis B.
Consider $F(\lambda) = \det(A - \lambda I)$ which is easily seen to be a polynomial in λ of degree n.
Definition. Given a square $n \times n$ matrix A, the polynomial $F(\lambda) = \det(A - \lambda I)$ is called **its characteristic polynomial** and the equation $F(\lambda) = 0$ is called **its characteristic equation.**
The roots of the characteristic polynomial of A are called **the eigenvalues of A** and for any such eigenvalue λ, the space $Nul(A - \lambda I)$ is **its eigenspace**.
Another way to explain this is as follows: If we define a linear transformation $T_A : \Re^n \to \Re^n$ to be $T_A(X) = AX$. Then these eigenvalues and eigenspaces correspond to the eigenvalue and eigenspaces of T_A.

5.2.2 How to use the characteristic polynomial.

We now show that the eigenvalues of L are simply the roots of the polynomial $F(\lambda)$, i.e. the solutions of the polynomial equation $F(\lambda) = 0$. Moreover, if λ is such a root, an eigenvector belonging to it is a vector whose coordinate vector is a non zero member of the $Nul(A - \lambda I)$.
Idea of the proof. Let v be an eigenvector for L belonging to an eigenvalue λ.
We then have $v \neq 0$ and $L(v) = \lambda v$.
But we note that $v = B \cdot [v]_B$ and hence $\lambda v = \lambda B([v]_B)$. So:

$$L(v) = L(B[v]_B) = BA[v]_B = B\lambda[v]_B = \lambda B[v]_B = \lambda v.$$

Thus v is an eigenvector belonging to λ iff its coordinate vector $[v]_B$ is a non zero vector in $Nul(A - \lambda I)$.
We illustrate this on our example (b) above.
Choose the basis $B = \begin{pmatrix} 1 & x & x^2 & x^3 \end{pmatrix}$. Then calculation of $L(v)$ for each of the basis vectors gives

$$L(B) = \begin{pmatrix} 0 & (x-3) & (x-3)(2x) & (x-3)(3x^2) \end{pmatrix}.$$

This gives the matrix of the transformation:

$$A = \begin{pmatrix} 0 & -3 & 0 & 0 \\ 0 & 1 & -6 & 0 \\ 0 & 0 & 2 & -9 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

It is now clear that its characteristic polynomial is $\lambda(\lambda - 1)(\lambda - 2)(\lambda - 3)$.
The eigenvalues are $0, 1, 2, 3$ and the respective eigenvectors can be seen to be the columns of the matrix

$$M = \begin{pmatrix} 1 & -3 & 9 & -27 \\ 0 & 1 & -6 & 27 \\ 0 & 0 & 1 & -9 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

When you look for the polynomials with these coordinate vectors, you get the given answer.

5.3 Diagonalization.

Let V be a vector space over a field K and let L be a linear transformation from V to itself.
We say that L is diagonalizable **if V has a basis consisting of eigenvectors of L.** Then the matrix of L in such a basis is a diagonal matrix. This is the reason for the term.
Given a square $n \times n$ matrix A, we say it is diagonalizable if the corresponding transformation T_A is diagonalizable.
This can be made more explicit thus. Suppose B is a basis which diagonalizes T_A. Form a matrix P with coordinate vectors of vectors in B as columns.
Note: In this case, the vectors of B are their own coordinate vectors in the standard basis.
Then we see that $AP = PD$ where D is a diagonal matrix and we get $A = PDP^{-1}$.
Thus, **for a square matrix** A, we can make **a simpler definition** which says: A is diagonalizable iff $A = PDP^{-1}$ for some $n \times n$ matrix P and a diagonal matrix D.
Example. For the example (a) above, we take the three eigenvectors w_1, w_2, w_3 and form the matrix

$$P = \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix} \text{ and note that } AP = P \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It follows that $A = PDP^{-1}$ and we say that P diagonalizes A.
The most important theorem about diagonalization is this:
Theorem of Independence of eigenvectors. Suppose that $v_1, \cdots, v_r$ are eigenvectors for a linear transformation L belonging to **different** eigenvalues $\lambda_1, \cdots, \lambda_r$.
Then $v_1, \cdots, v_r$ are linearly independent.
Proof.
We use induction on r.
Induction step 1. If $r = 1$, then the result is trivially true since v_1 being an eigenvector is non zero and hence is an independent vector.
Induction step 2. Suppose the induction result is true for $r - 1$. Then we prove it for r.
Suppose if possible the result is false and $v_1, \cdots, v_r$ are linearly dependent. Since $v_1, \cdots, v_{r-1}$ are linearly independent by induction hypothesis, we must have

$$EQ1 \qquad v_r = a_1v_1 + c_2v_2 + \cdots + a_{r-1}v_{r-1}.$$

for some scalars $a_1, \cdots, a_{r-1}$.
Applying L to both sides, we see:

$$EQ2 \qquad \lambda_r v_r = \lambda_1 a_1 v_1 + \cdots \lambda_{r-1} a_{r-1} v_{r-1}.$$

Then we calculate $EQ2 - \lambda_r EQ1$ to get:

$$EQ3 \qquad 0 = a_1(\lambda_1 - \lambda_r)v_1 + \cdots + a_{r-1}(\lambda_{r-1} - \lambda_r)v_{r-1}.$$

By induction hypothesis $v_1, \cdots, v_{r-1}$ are linearly independent and we see that

$$a_1(\lambda_1 - \lambda_r) = \cdots = a_{r-1}(\lambda_{r-1} - \lambda_r)v_{r-1} = 0.$$

Since all the λ_i are distinct, we must have: $a_1 = \cdots = a_{r-1} = 0$. It follows from EQ1 that $v_r = 0$ which is a contradiction!
Hence the theorem is true.
Corollary 1. We can deduce the main criterion for diagonalization based on this theorem. It is:
Suppose that V has dimension n and $V_{\lambda_1}, \cdots, V_{\lambda_r}$ are all the eigenspaces for L with distinct eigenvalues $\lambda_1, \cdots, \lambda_r$ with corresponding dimensions $d_1, \cdots, d_r$.

Then the transformation L is diagonalizable iff $d_1 + \cdots + d_r = n$.
Idea of Proof.
If the condition is satisfied then the n vectors obtained by taking the union of the basses of all these eigenspaces give n independent eigenvectors in V and hence form a basis of V consisting of eigenvectors. This gives the diagonalization.
Conversely, if V has a basis consisting of n eigenvectors then it is easy to show that these n vectors are simply obtained by taking a union of bases for eigenspaces.
Corollary 2. If a linear transformation L of a vector space V of dimension has n distinct eigenvalues, then the set of n eigenvectors, one for each eigenvalue form a basis of V and hence L is diagonalizable.

5.3.1 Examples.

1. The matrix $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ as well as the corresponding transformation T_A are not diagonalizable. The reason is that the characteristic polynomial $(\lambda - 1)^2$ has only one root 1 and the corresponding eigenspace has a basis of a single vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Thus, we cannot have two independent eigenvectors and the diagonalization fails.

2. If A is a square matrix which is upper triangular with distinct entries on its diagonal, then it is diagonalizable. The reason is that the diagonal entries are seen to be its eigenvalues and since these are distinct, the Corollary 2 applies.

   The result holds for a lower triangular matrix for the same reason.

3. Diagonalize the matrix
$$A = \begin{pmatrix} 3 & 1 & 2 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$
   if possible. Explain the reason if this is not possible.

   Answer.

   First we compute the characteristic polynomial
$$\begin{vmatrix} 3-\lambda & 1 & 2 \\ 0 & 1-\lambda & 0 \\ 1 & 0 & 2-\lambda \end{vmatrix} = (1-\lambda)((3-\lambda)(2-\lambda) - (2)(1))$$
   where we have expanded the determinant by the second row. The determinant further factors as
$$(1-\lambda)(\lambda^2 - 5\lambda + 4) = (1-\lambda)(\lambda - 1)(\lambda - 4).$$
   Thus the eigenvalues are 1 and 4 where 1 is a double root.

   To calculate V_1 we find:
$$Nul(\begin{pmatrix} 2 & 1 & 2 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}).$$
   It is not hard to see that the matrix has rank 2 and hence there is only one free variable in the Null space calculation. This means the eigenspace V_1 will have dimension 1.

The eigenspace V_4 requires the solution of

$$Nul(\begin{pmatrix} -1 & 1 & 2 \\ 0 & -3 & 0 \\ 1 & 0 & -2 \end{pmatrix}).$$

This matrix has rank 2 as well and thus the dimension of the Null space is again 1.

Thus, there are at most two independent eigenvectors and it is not diagonalizable.

4. Diagonalize the matrix

$$A = \begin{pmatrix} 3 & 1 & 2 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

if possible. Explain the reason if this is not possible.

Answer. This is similar to above but has three different eigenvalues $2, 1, 4$. Thus, by Corollary 2 it is diagonalizable.

It is not hard to see that the respective eigenvectors are the columns of the matrix:

$$P = \begin{pmatrix} 0 & -1 & 2 \\ 2 & 0 & 0 \\ -1 & 1 & 1 \end{pmatrix}.$$

Thus we have:

$$A = P \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} P^{-1}.$$

5.4 The 2×2 real matrix.

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a 2×2 matrix. We analyze the eigenspaces and diagonalization for A completely.
First we note that the characteristic polynomial is easily seen to be
$F(\lambda) = \lambda^2 - (a + d)\lambda + (ad - bc)$. This is easy to remember by noting that the coefficient of λ is the sum of the diagonal entries or the trace of the matrix and the constant term is its determinant.
Assumption. Here we assume that our matrix has real coefficients.
There are three cases for the roots which we analyze next.

1. **Distinct real roots.** We assume that $F(\lambda)$ has two distinct real roots, p, q. By corollary 2, we know it is diagonalizable and we simply need to find the matrix of the two eigenvectors.

   For $\lambda = p$, we solve:

$$\left(\begin{array}{cc|c} a-p & b & 0 \\ c & d-p & 0 \end{array}\right).$$

   If the first row is not zero, then we solve its equation by inspection: $\begin{pmatrix} b \\ p-a \end{pmatrix}$ and know that this is an eigenvector belonging to p. We note that the second row must give a dependent row and hence we can ignore it!

   Could the first row be a zero vector? Yes, it is possible when $a = p$ and $b = 0$. In that case, we solve the second equation to get $\begin{pmatrix} p-d \\ c \end{pmatrix}$.

What if this one is zero too?

We see that we must have $b = c = 0, a = d$ so our matrix is aI and already diagonalized!

Similarly, for $\lambda = q$, we deduce an eigenvector $\begin{pmatrix} b \\ a-q \end{pmatrix}$, if the first row is non zero, $\begin{pmatrix} q-d \\ c \end{pmatrix}$ if the first row is zero but the second is not and note that the matrix is aI and is already diagonalized in the remaining case.

Thus the matrix

$$P \text{ equals } \begin{pmatrix} b & b \\ p-a & q-a \end{pmatrix} \text{ or } \begin{pmatrix} p-d & q-d \\ c & c \end{pmatrix} \text{ or } I.$$

2. **Double root case.** We assume that the characteristic polynomial factors as $(\lambda - p)^2$.

   The eigenspace V_p is $Nul((A - pI))$ and it is diagonalizable iff this space has dimension 2. But that means the matrix $A - pI$ must have rank zero, i.e. it must be the zero matrix. This means $A = pI$ or A is already diagonal!

   Thus A is diagonalizable iff it is already diagonal!

3. **Two complex roots.** In this case, there are no real eigenvalues so the matrix cannot be diagonalizable over the reals. However, we can put the matrix in a certain form which helps us understand the nature of the transformation.

   Worked Example: distinct complex roots.

   Let $A = \begin{pmatrix} 1 & 4 \\ -2 & 5 \end{pmatrix}$.

   The characteristic polynomial is $\lambda^2 - 6\lambda + 13$. So the eigenvalues are $3 \pm 2i$.

   Suppose we were to find the eigenvector for the complex eigenvalue $3 - 2i$ as before. We would have to solve the equations in Complex numbers represented by this augmented matrix:
$$\left(\begin{array}{cc|c} 1-(3-2i) & 4 & 0 \\ -2 & 5-(3-2i) & 0 \end{array}\right) = \left(\begin{array}{cc|c} -2+2i & 4 & 0 \\ -2 & 2+2i & 0 \end{array}\right).$$

   As before, the second equation is a multiple of the first (by $\frac{1}{1-i}$) and we solve the first equation by inspection as $\begin{pmatrix} 4 \\ 2-2i \end{pmatrix}$. Suppose we write this vector into its real and complex components as:
$\begin{pmatrix} 4 \\ 2-2i \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} + i \begin{pmatrix} 0 \\ -2 \end{pmatrix} = v_1 + iv_2.$
   Then we know $Av = (3 - 2i)v$, i.e.

$$A(v_1 + iv_2) = (3 - 2i)(v_1 + iv_2) = (3v_1 + 2v_2) + i(-2v_1 + 3v_2).$$

   Splitting this into real and complex parts, we get:

$$Av_1 = (3v_1 + 2v_2) \text{ and } Av_2 = (-2v_1 + 3v_2).$$

   If we form a matrix P with columns v_1, v_2, then we see that

$$AP = P \begin{pmatrix} 3 & 2 \\ -2 & 3 \end{pmatrix}.$$

Thus if we change our basis to columns of P, the new matrix is $\begin{pmatrix} 3 & 2 \\ -2 & 3 \end{pmatrix}$.

Conclusion. Thus we conclude that in general if we have a complex eigenvalue $a - bi$ with $b \neq 0$ and $v = v_1 + iv_2$ is an eigenvector belonging to it, then we can set P to be the matrix with columns v_1, v_2 and get the equation:

$$A = P \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Comment: If one interprets the points of the plane as complex numbers, then this corresponds to multiplication by the complex number $(a - bi)$. Moreover, if we write $a - bi = r\exp(i\theta)$ using the usual polar representation, then this can be described as a rotation by the angle θ followed by expansion of scale by a factor r.

Chapter 6

Inner Products.

6.1 Inner product, length, and angle between two vectors

Given a real vector space V, an inner product is defined to be a **bilinear** map $F : V \times V \to \Re$ such that the following holds:

- Commutativity: For all $v_1, v_2 \in V$, we have $F(v_1, v_2) = F(v_2, v_1)$.
- Distributivity:For all $v_1, v_2, v_3 \in V$, we have $F(v_1, v_2 + v_3) = F(v_1, v_2) + F(v_1, v_3)$.
- Scalar multiplicativity: For all $v_1, v_2 \in V$ and $c \in \Re$ we have $F(cv_1, v_2) = F(v_1, cv_2) = cF(v_1, v_2)$.
- **Positivity:**For all $v \in V$, we have $F(v, v) \geq 0$. Moreover $F(v, v) = 0$ iff $v = 0$.

Notation. We usually do not use a name like F, but write $< v, w >$ in place of $F(v, w)$. Often, we also just write $v \cdot w$ and call it a "dot" product.
Warning. Many books will define a more general inner product where the last property of positivity is not assumed in the beginning but imposed later. The positivity is essential for definitions of angles and lengths.

6.1.1 Norm, angle

We now use the shortened notation $< \ , \ >$ for an inner product and define

- $||v||^2 =< v, v >$ or $||v|| = \sqrt{< v, v >}$. This $||v||$ is the length of the vector v for the chosen inner product, so strictly speaking, it should carry a marker indicating the inner product. Here, using a function name F helps us put such a marker and write $||v||_F$.
- It can be proved that for any two vectors v, w, we have

  $$| < v, w > | \leq ||v|| ||w|| \textbf{Cauchy Schwartz Inequality.}.$$

  Moreover, we get equality iff v, w are linearly dependent.

  Further, if v, w are non zero vectors, then $| < v, w > | = ||v|| ||w||$ implies that one of the following two things happens.

  Either we have: $< v, w >= ||v|| ||w||$ in case v, w are positive multiples of each other (or can be considered to be in the same direction) or $< v, w >= -||v|| ||w||$ in case v, w are negative multiples of each other (or can be considered to be in the opposite direction).

- We define the angle between non zero vectors v, w by

$$\angle(v, w) = \arccos(\frac{< v, w >}{||v||||w||}.$$

The Cauchy Schwartz inequality guarantees that we get a meaningful angle between 0 and 180 degrees.

Warning: One should not lose sight of the fact that this is dependent on the chosen inner product and as before, a marker F can be attached if necessary.

6.1.2 Examples

Here are some examples of inner products in known vector spaces.

- The most common example is in $\Re^n$. We define $< v, w >= v^T w$. This gives the usual dot product. It is obvious that $||v||$ corresponds to the usual length of a vector and for $n = 2, 3$, direct calculations can verify the angles to be consistent with usual convention.

- Still in $\Re^n$ a more general inner product can be defined by **a symmetric matrix** $A = A_{n\times n}$ by defining:

$$F(v, w) = v^T Aw.$$

We may write $< v, w >_A$ as a shortened notation, or as an alternative drop all special references to A if no confusion follows.

A random choice of A will not satisfy the positivity condition. It can be shown that a necessary and sufficient condition for a symmetric matrix A to define an inner product is that all its principle minors be positive. This means all the determinants using first few entries of the main diagonal are positive.

- If we go to the space of polynomials P_n or even P, the infinite dimensional space, then we can define an inner product:

$$F(p(t), q(t)) = \int_0^1 p(t)q(t)dt.$$

Clearly, the interval can be changed to other finite intervals leading to different inner products.

- The above example can be generalized to define an inner product on the space $C[a, b]$ which is the space of continuous functions on the interval $[a, b]$. The inner product is defined as

$$F(f(t), g(t)) = \int_a^b f(t)g(t)dt.$$

- In the space of polynomials P_n, define an inner product thus:

Choose a set of distinct numbers $a_0, a_1, \cdots, a_n$ and define

$$< p(t), q(t) >= p(a_0)q(a_0) + p(a_1)q(a_1) + \cdots + p(a_n)q(a_n).$$

This defines an inner product. A little thought shows that the map

$$p(t) \to \begin{bmatrix} p(a_0) \\ p(a_1) \\ \cdots \\ p(a_n) \end{bmatrix}$$

is an isomorphism of P_n onto $\Re^{(n+1)}$ and all we are doing is using the usual inner product in the target space $\Re^{(n+1)}$ to define our inner product.

This is a usual method of building new inner products.

6.2 Orthogonal sets

Given an inner product $< \ , \ >$ on a vector space V we say that a set of vectors $v_1, \cdots, v_r$ is orthogonal, if for any $i \neq j$ we have $< v_j, v_j >= 0$.
It is easily seen that a set of non zero orthogonal vectors are linearly independent.
Proof. Suppose $v_1, \cdots, v_r$ are non zero orthogonal vectors and $c_1 v_1 + \cdots + c_r v_r = 0$. Take the inner product of both sides with some v_i to get:

$$c_1 < v_i, v_1 > + \cdots + c_i < v_i, v_i > + \cdots + c_r < v_i, v_r >=< v_i, 0 >= 0.$$

Clearly all but the term $c_i < v_i, v_i >$ are zero. Moreover, $< v_i, v_i >\neq 0$, so $c_i = 0$. Thus each c_i is zero and we have proved independence of our vectors.
This is the most important reason to study and use the inner product!
The set of vectors $v_1, \cdots, v_r$ is said to be **orthonormal** if it is orthogonal and also $< v_i, v_i >= 1$ for all i. This last condition means that $||v_i|| = 1$ for each $i = 1, \cdots r$.
Vectors with norm (length) equal to 1 are said to be unit vectors. **Note** that given any non zero vector v, the vector $\pm \frac{v}{||v||}$ is always a unit vector. Moreover, if we take the plus sign, then it is in the same direction as v and is in the opposite direction if we use the minus sign.
This gives a simple but useful observation:
Every nonzero vector v is of the form cu where u is a unit vector and $c = \pm||v||$.

6.3 Coordinate vectors

If we have a set of n non zero orthogonal vectors, $v_1, \cdots, v_n$ in an n-dimensional vector space V, then, in view of the above result, they clearly form a basis $B = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ of V.
Moreover, for any vector $v \in V$, it is easy to find its coordinate vector $[v]_B$ as follows.
Suppose we write $v = c_1 v_1 + \cdots c_n v_n$. By taking inner product with v_i and using the same reasoning as above, we see that $< v, v_i >= c_i < v_i, v_i >$ and thus $c_i = \frac{< v, v_i >}{< v_i, v_i >}$. This defines the coordinate vector:

$$[v]_B = \begin{bmatrix} c_1 \\ \cdots \\ c_n \end{bmatrix} = \begin{bmatrix} \frac{< v, v_1 >}{< v_1, v_1 >} \\ \cdots \\ \frac{< v, v_n >}{< v_n, v_n >} \end{bmatrix}.$$

6.4 Projections

One of the main goals of Linear Algebra is to give efficient methods to solve linear equations $AX = B$. In general, if there are more equations than variables (i.e. A has more rows than columns), then the solutions may not exist. However, in many Scientific and Statistical applications, it makes sense to ask for an answer which makes the equation close to true as much as possible.
If we have an inner product in our vector space, then we can reformulate the problem of solution of $AX = B$ as "find a vector w such that $||B - Aw||$ is as small as possible.
This can be shown to be equivalent to finding a w such that $B - Aw$ is orthogonal to each column of A. If we are using the usual inner product in $\Re^n$, then this is easily seen to be guaranteed by:

Normal Equations. $$A^T A w = A^T B$$

From the properties of the inner product, we can show that if the columns of A are independent, then the matrix $A^T A$ is invertible. (See proof below). Using this, we get a formal solution:

$$w = (A^T A)^{-1} A^T B.$$

The vector Aw so obtained is geometrically the projection of the vector B into the space $Col\ A$.
Proof that $\mathbf{A^T A}$ is invertible. Suppose if possible, $A^T A$ is singular. Then there is a non zero vector u such that $A^T Au = 0$. Then

$$< Au, Au >= u^T A^T Au = u^T (A^T Au) = 0.$$

Hence $Au = 0$. But since columns of A are independent, this implies $u = 0$, a contradiction!

6.4.1 Associated Spaces

Given an $m \times n$ matrix A, we know the two associated spaces $Col(A)$ and $Nul(A)$ which are respectively subspaces of $\Re^m$ and $\Re^n$.
If we use the transpose A^T instead, then we get two other spaces: $Col(A^T)$ which we call $Row(A)$ or the row space of A and also $Nul(A^T)$ or sometimes called the left null space of A.
Note that $Row(A)$ is a subspace of $\Re^n$ and consists of rows of A **transposed into column vectors.**
Similarly, $Nul(A^T)$ is a subspace of $\Re^m$ consisting of all column vectors X such that $A^T X = 0$. Taking transpose, we see that these correspond to row vectors X such that $X^T A = 0$. Hence the name of "left null space."
The concept of inner product gives another meaning to these. Thus, the left null space $Nul(A^T)$ can be thought of all vectors orthogonal to all vectors of $Col(A)$.
In general, we define **an orthogonal subspace** to a given space W as $\{v \mid < v, w >= 0 \text{ for all } w \in W\}$. We denote this as $W^\perp$.
It is not hard to see that $(W^\perp)^\perp = W$ for any subspace W. Thus, we note that $Col(A) = (Nul(A^T))^\perp$. This expresses the starting space $Col(A)$ as a null space of some other matrix. This was the basis of our results on writing a column space as a null space or conversely, writing a null space as a column space.
Note We already know another method to find this left null space. Recall the consistency matrix G obtained by finding an REF of $(A|I)$ and taking the part of the transformed I in front of zero rows in REF of A. We know that vectors $v \in Col(A)$ are characterized by $Gv = 0$. This means $v^T G^T = 0$ and thus $Nul(A^T) = Col(G^T))$ as desired.
Similarly, we can describe $Row(A)$ as $(Nul(A)))^\perp$.
It is easy to see that for any subspace W of V we have $\dim(W) + \dim(W^\perp) = \dim(V)$. This is another formulation of the fundamental dimension theorem.
Proof. Write $W = Col(A)$ for some $m \times n$ matrix A, so that W is a subspace of $\Re^m$. We know that $\dim(W) = rank(A)$.
Then

$$W^\perp = \{Y \mid < w, Y >= 0 \text{ for all } w \in W\}.$$

Since $< w, Y >= w^T Y$, we see that $W^\perp = Nul(A^T)$ and we know that its dimension is $m - rank(A^T) = m - rank(A)$. Thus, we have proved that

$$\dim(W) + \dim(W^\perp) = rank(A) + m - rank(A) = m.$$

6.5 Orthonormal Bases.

Suppose that we have a vector space V with an inner product and a given subspace W.

The above results make it clear that we would greatly benefit if given any basis (or even a spanning set) of the subspace W, we can find a suitable orthogonal (or even orthonormal) basis for W from the given set.
This can be accomplished by a slight modification of our row reduction algorithm. This is a way of codifying the Gram-Schmidt process discussed in the book. We show the method below, which is not in the book.

6.5.1 The Inner Product Matrix

Suppose that $v_1, \cdots v_r$ is a spanning set for W. First step is to make a matrix M^* such that $M^*_{ij} =< v_i, v_j >$ for all $i, j = 1, \cdots r$.
Note that M^* is a symmetric $r \times r$ matrix and we can think of M^* as $< B, B >$ where B is the row of vectors $\begin{bmatrix} v_1 & v_2 & \cdots & v_r \end{bmatrix}$. This is said to be the I.P. (Inner Product) matrix of the spanning set B.
If we replace B by linear combinations of $v_1, \cdots v_r$ then we can think of the new set of vectors as BP where P is the matrix describing the combinations.
If P is invertible, then vectors of BP form a new spanning set for the same space W and its I.P. matrix is $P^T M^* P$. We shall show that there is an invertible matrix R such that $R^T M^* R$ is a diagonal matrix.
It follows that the new generating set BR consists of orthogonal vectors.
If the original vectors of B were independent, then the new vectors BR will indeed be an orthogonal basis. Moreover, in this case, the matrix R can be chosen to be (unit) upper triangular. This is known as the Gram-Schmidt theorem.

6.5.2 The Gram-Schmidt Algorithm

We present a different Gram-Schmidt algorithm which is lot easier to implement than the one in the book. It is based on just the usual row reduction algorithm and does not involve complicated expressions.
Start with the I.P. matrix M^* of a spanning set $v_1, \cdots, v_r$ for W. If we are working with the "usual inner product" in $\Re^n$, then we simply have $M^* = A^T A$ where A is the matrix whose columns are $v_1, \cdots, v_r$. In either case, M^* is a square $r \times r$ square matrix.
Let I_r be the usual identity matrix. Set $M = (M^*|I_r)$ the augmented matrix as usual.
Perform the usual row reductions on M to try and convert it to REF to get a matrix $G = (N|S)$ where N is upper triangular and S is the lower triangular, invertible matrix such that $G = SM^*$.
(Note that for N to be upper triangular one can only do row operations in which a row is modified by adding to it a multiple of row above it and in which a row with a zero in the intended pivot position is interchanged with one below it. Scaling rows is permitted.)
Let $R = S^T$. Since R and $R^T M^* = SM^*$ are both square, upper triangular, so is their product $R^T M^* R$.
In case $M^* = A^T A$ this expression is $R^T A^T AR =< AR, AR >$.
In the more general case with $M^* =< B, B >$ we get that $R^T M^* R =< BR, BR >$.
Finally, we note that this matrix $R^T M^* R$ is upper triangular and equal to its own transpose, hence it must be a diaagonal matrix!
It follows immediately that the columns of AR (or vectors in BR)are mutually orthogonal.
If columns of A are linearly dependent, then we will simply have some columns in AR as zero columns and the non zero columns will give a basis of $Col(A)$.
Similarly, if $v_1, \cdots, v_r$ are linearly dependent, then some of the vectors in BR will be zero and we get a basis for $Span(B)$ after dropping them.

Example 6.1. Let $A = \begin{bmatrix} -2 & 1 \\ 2 & -2 \\ 1 & 1 \end{bmatrix}$. Then the inner product matrix is

$$IP = A^t A = \begin{bmatrix} 9 & -5 \\ -5 & 6 \end{bmatrix}.$$

We augment IP by the identity to get

$$M = \left[\begin{array}{cc|cc} 9 & -5 & 1 & 0 \\ -5 & 6 & 0 & 1 \end{array}\right]$$

The row operation $R_2 \to R_2 + \frac{5}{9}R_1$ produces $\left[\begin{array}{cc|cc} 9 & -5 & 1 & 0 \\ 0 & \frac{29}{9} & \frac{5}{9} & 1 \end{array}\right]$
We stop at this point since the IP matrix is now in REF.
Now we take R to be the transpose of the matrix derived from the identity. $R = \begin{bmatrix} 1 & \frac{5}{9} \\ 0 & 1 \end{bmatrix}$
Then the row operations have multiplied A^TA by R^T on the left and the column operations have multiplied it by R so we have $R^T(A^TA)R = \begin{bmatrix} 9 & 0 \\ 0 & \frac{29}{9} \end{bmatrix}$.
This says that $(AR)^t(AR) = \begin{bmatrix} 9 & 0 \\ 0 & \frac{29}{9} \end{bmatrix}$
Since R is invertible, AR has the same linear span as the columns of A. So the columns of AR are spanning set for $col(A)$. Since they are mutually independent they are then an independent spanning set for $Col(A)$ so the columns of

$$AR = \begin{bmatrix} -2 & 1 \\ 2 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{5}{9} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -\frac{1}{9} \\ 2 & -\frac{8}{9} \\ 1 & \frac{14}{9} \end{bmatrix}.$$

are an orthogonal basis for the column space of A.
If we want an orthonormal basis then since $(AR)^T(AR) = \begin{bmatrix} 9 & 0 \\ 0 & \frac{29}{9} \end{bmatrix}$
we have that 9 is the square of the length of column 1 of AR and $\frac{9}{29}$ is the square of the length of column 2. To convert an orthogonal set to an orthonormal set all we have to do is divide each element by its length. That is we divide column 1 by 3 and column 2 by $\sqrt{\frac{29}{9}}$ to get
$S = \begin{bmatrix} -\frac{2}{3} & -\frac{\sqrt{29}}{87} \\ \frac{2}{3} & -8\frac{\sqrt{29}}{87} \\ \frac{1}{3} & 14\frac{\sqrt{29}}{87} \end{bmatrix}$. which is a matrix whose columns are an orthonormal basis for $Col(A)$. We can check the orthonormality by calculating $S^TS = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

In the above example the matrix A had independent columns. Even if A has dependent columns then the matrix AR will still have mutually orthogonal columns- only some of them would be the zero vector (which is orthogonal to everything). Then all one has to do is delete the zero columns to have an orthogonal basis.

Example 6.2. Here is an example where the starting vectors are not independent. To emphasize the basic details we give only the IP matrix.
Suppose the matrix A has columns v_1, v_2, v_3 and the following I.P. matrix:

$$A^TA = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 5 & 6 \\ 3 & 6 & 9 \end{bmatrix}.$$

As before, we make the augmented matrix $< A^TA|I >$:

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 3 & 1 & 0 & 0 \\ 1 & 5 & 6 & 0 & 1 & 0 \\ 3 & 6 & 9 & 0 & 0 & 1 \end{array}\right].$$

- $$\left[\begin{array}{ccc|ccc} 2 & 1 & 3 & 1 & 0 & 0 \\ 0 & 9/2 & 9/2 & -1/2 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{array}\right].$$

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

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

$$R^TA^TAR = (AR)^T(AR) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 9/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- Note that the third new vector (column 3 of AR) has norm zero and hence it is zero! This is $A\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$ so the process indicates that the third vector is $w_3 = v_3 - v_2 - v_1 = 0$ and thus it identifies the linear dependence relation too!

- We can now conclude that our vector space $Span\{v_1, v_2, v_3\}$ is actually two dimensional with $w_1 = v_1$ and $w_2 = v_2 - (\frac{1}{2})v_1$ as an orthogonal basis. The lengths of w_1, w_2 are $\sqrt{2}$, $\sqrt{\frac{9}{2}}$ respectively.