Preamble

These lecture notes are a first draft prepared this past summer (2023) after teaching the course for the first time in Fall 2022. I expect a lot of mistakes/typos/errors etc. in it, which we will correct in class as we go along in this offering in Fall 2023. When you are reading these notes if you find any errors not corrected in class, please let me know so I can correct them going forward.

Chapter 1

Applied Algebra

1.1 Determinants of matrices

1. The determinant Δ of a square matrix $A_{n\times n}$ with elements $[a_{i,j}]_{i=1,...,n,j=1,...,n}$ which may be real or complex numbers or polynomials or rational or irrational functions etc. of size n is defined as:
$$\Delta = \sum_{\sigma} \mathrm{sgn}(\sigma)\Pi_{i=1}^{i=n} a_{i,\sigma(i)} \tag{1.1}$$
where the sum is over all permutations σ of the n integers.

2. A permutation σ of integers $1, ..., n$ (alternatively a countable finite set) is a bijective mapping (function) from the set to itself.

3. Examples of permutation and ways of representing it follow:

   (a) Consider $n = 3$, then one permutation is: σ:
   $$\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1 \tag{1.2}$$
   $$\sigma : \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \text{Cauchy's two line representation} \tag{1.3}$$
   $$\sigma \qquad (1,2,3) = (2,3,1) \ \text{ i.e } \ 1 \to 2 \to 3 \to 1 \quad \text{cycle representations} \tag{1.4}$$

   (b) The identity permutation does not have a cycle decomposition.

   (c) With $n = 5$, a permutation τ is
   $$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 1 & 2 \end{pmatrix} \tag{1.5}$$
   τ can be written in terms of (product ie composition of)smaller cycle represented permutations (read right to left) in a variety of ways:
   $$\tau = (1,3,4)(2,5) = (2,5)(1,3,4) \qquad \text{disjoint cycles,commutative} \tag{1.6}$$
   $$\tau = (1,4)(1,3)(2,5) = (2,5)(1,4)(1,3) \ \text{transpositions,noncommutative} \tag{1.7}$$

Transpositions are cycles of length 2.

4. A permutation is even if it can be decomposed into cycles with an even number of even cycles (i.e cycles whose length is even. Note: 0 is an even number) else the permuation is odd. The sign of a permutation is 1 if even, else it is -1.

5. σ in 3(a) is even and has sgn(σ) = 1(it has 0 cycles of even length or if decomposed as product of transpositions $\sigma = (1,3)(1,2)$ it has **two** cycles of length 2).

6. τ from 3(c) is an odd permutation as it has **one** cycle of length=2 from (6) or **three)** cycles of length 2 from (7) and so sgn(τ) = -1.

7. The identity permutation is even as it has 0 cycles of 0 length!

8. For the set $1, 2, ..., n$ there are $n!$ permutations and for $n > 1$ half of them are even and the other half are odd.

9. Going back to our determinant definition in 1 above, for $n = 3$, the six permutations in cycle representation are: Even: (1,2,3),(1,3,2) and identity; Odd: (1,2),(1,3),(2,3); A third order matrix then has the expansion

$$\Delta = a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{3,2}a_{2,1} + a_{1,1}a_{2,2}a_{3,3} - a_{1,2}a_{2,1}a_{3,3} - a_{1,3}a_{3,1}a_{2,2} - a_{2,3}a_{3,2}a_{1,1} \tag{1.8}$$

1.1.1 Other Results on Determinants

From the above main section other properties of Determinants and the Laplace Expansion of determinants in terms of minors can be obtained.

1. The minor $M_{i,j}$ of an element $a_{i,j}$ of a (square) matrix(A) is the determinant of the residual matrix formed from A by eliminating its i^{th} row and j^{th} column. Laplace expansion of a determinant by row i or column j follows below using this concept of minor and can be derived from the definition in (1) (For a proof see Wikipedia article on Laplace Expansion of Determinant).

$$\Delta = \det(A) = \sum_{j=1}^{n}(-1)^{i+j}M_{i,j}a_{i,j} = \sum_{i=1}^{n}(-1)^{i+j}M_{i,j}a_{i,j} \tag{1.9}$$

2. The Laplace definition is often what is taught in elementary courses on matrices and determinants etc. However it is not the fundamental definition. The permutation definition is. Why ? – When seeking curl of a 3-D vector in orthogonal coordinates, determinants come in. How does one define curl (tensor products) of vectors in higher dimensions than 3, permutations are useful in such definitions and are used there. Hence the definition in terms of permutations generalizes concepts from 3-D to n-D.

3. The cofactor ($C_{i,j}$) of the element $a_{i,j}$ is defined as $C_{i,j} = (-1)^{i+j}M_{i,j}$. This shortens the writing of the Laplace expansion.

4. The scalar product s is defined as the multiplication of a row vector which is a matrix $A_{1\times n}$ with a column vector which is a matrix $B_{n\times 1}$

$$s = \sum_{j=1}^{n} a_{1,j} b_{j,1} \tag{1.10}$$

5. The product $S_{m\times p}$ of matrix $A_{m\times n}$ with matrix $B_{n\times p}$ written as $S = AB$ is defined as the matrix with elements

$$s_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j}, \qquad i = 1..m, j = 1..p \tag{1.11}$$

Note: that this is a generalization of the scalar product in that you take the i^{th} row of A and the j^{th} column of B and compute $s_{i,j}$ using the scalar multiplication of row and column vector matrices from (10). Note further matrix multiplication may not be always definable, as to do it, the number of columns in the row source matrix A must be equal to the number of rows in the column source matrix B. If matrix multiplication can be defined in a specific order between two matrices, then the matrices are said to be compatible for multiplication in that order. When $m = p = n$ (i.e. square matrices A, B) then matrix multiplication in general is non- commutative ie $AB \neq BA$ (in general). It is easy to see this in Matlab/Octave using the following commands:

```
>> diary Eg1.txt
>> A=rand(5);B=rand(5); AB=A*B;BA=B*A;
>> AB
AB =

1.5752   1.3403   1.1492   1.0748   1.1682
2.3035   2.1484   2.0218   1.7140   1.7325
1.4070   1.5901   1.3881   1.0410   1.1578
0.9699   0.9905   0.8467   0.7621   0.8454
1.9288   2.0011   1.4140   1.4319   1.5831

>> BA
BA =

1.0346   1.0486   1.4553   1.5607   1.7773
1.0335   1.0270   1.8369   1.6651   1.8435
0.9914   1.0009   1.8422   1.7344   1.8191
0.7406   0.9035   1.5311   1.2996   1.3929
1.1120   1.4757   1.7336   1.7874   2.2535

>> diary off
```

6. The adjoint of a square matrix A is another matrix $\text{Adj}(A)$ formed as the transpose of the matrix of cofactors of the elements of A. The inverse of a square matrix A, if it exists, is defined as $A^{-1} = \text{Adj}(A)/\det(A)$. It exists if and only if $\det(A) \neq 0$. The inverse of A has the familiar property that A (matrix multiplied) by its inverse i.e $AA^{-1} = A^{-1}A = E$ where E is the identity matrix with 1 for the diagonal elements and 0 on the off diagonal elements. In Matlab/Octave the det and inv functions evaluate determinants and inverse of a square matrix respectively. I illustrate the inverse product identity above in the symbolic processing package WxMaxima below:

(% i1) A:matrix([a11,a12],[a21,a22]);

$$\begin{pmatrix} \text{a11} & \text{a12} \\ \text{a21} & \text{a22} \end{pmatrix} \qquad (\%\ \text{o1})$$

(% i3) Ainverse:invert(A);

$$\begin{pmatrix} \frac{\text{a22}}{\text{a11 a22}-\text{a12 a21}} & -\frac{\text{a12}}{\text{a11 a22}-\text{a12 a21}} \\ -\frac{\text{a21}}{\text{a11 a22}-\text{a12 a21}} & \frac{\text{a11}}{\text{a11 a22}-\text{a12 a21}} \end{pmatrix} \qquad (\%\ \text{o3})$$

(% i6) ratsimp(Ainverse.A);

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad (\%\ \text{o6})$$

(% i7) ratsimp(A.Ainverse);

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad (\%\ \text{o7})$$

1.2 Assignment

Assuming A, B, C, D are compatible for multiplication (Block matrix multiplication if need be) and inverses exist for any square matrix formed from operations on them (E is the identity matrix and O the zero matrix of appropriate dimension), show that

1. $(AB)^T = B^T A^T$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. Prove the matrix inversion lemma (very useful to convert batch form of least squares to recursive form and in the Kalman filter): $(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}$

4. Prove the generalized matrix inversion lemma: $M = (A - BD^{-1}C)^{-1}$ is defined as Schur Complement (after Issai Schur) of the matrix on the left side of the identity below.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{bmatrix} \tag{1.12}$$

5. Show that (I call this manipulation the Generalized Gabriel Kron Reduction The result is needed when we do conditional Gaussian distributions, we will use it there. The Kron reduction is extensively used in power systems and networks for forming the Z Bus and Y Bus matrices of a power system and in the corresponding software packages.)

$$\begin{bmatrix} E & -BD^{-1} \\ O & E \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E & O \\ -D^{-1}C & E \end{bmatrix} = \begin{bmatrix} A - BD^{-1}C & O \\ O & D \end{bmatrix} \tag{1.13}$$

6. Derive 4. from 5.

7. Show that determinant of product of square matrices is the product of the determinant of each of the square matrix.

8. Using 7. and 5., show that

$$\det\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix}\right) = \det\left(A - BD^{-1}C\right)\det(D) \tag{1.14}$$

9. Given permutations of $n = 5$ as $\sigma = (1, 4, 3, 5)$ and $\tau = (1, 3, 4, 2, 5)$ find the permutations $\sigma\tau$ and $\tau\sigma$. Are the two resultant permutations the same ? What is the sign of the two resultant permutations ?

10. If a square matrix has two rows or two columns identical or a multiple of the other the prove that the determinant of the matrix is 0.

1.3 Solutions of Linear Equations

The set of equations $3x + 5y = 1; 5x + 6y = -1$ is linear (why?). The set of equations $e^{3x+5y} = 100; xy = 6$ is nonlinear. The set of equations $3x^2 + 4y = 4; 5x^2 + 6y = 1$ appears nonlinear but if x^2 is called z, then it appears linear in form. However, there is no guarantee that solutions to this variable transformation exist in real numbers as we can't apriori guarantee $z > 0$. However solutions may exist in the complex numbers or the equations may be inconsistent ie no solutions exist or many solutions exist. The diode equation $I = I_0 e^{V/V_T}$ when transformed with logarithms reads $\log(I) = \log(I_0) + V/V_T$ and appears linear in form with variable transformation to $\log(I)$,$\log(I_0)$.

m linear equations in unknown variables $x_i, i = 1..., n$ arranged in a column vector X can be written as a matrix equation $A_{m\times n}X = b_{m\times 1}$ where $b_{m\times 1}$ is a column vector of the right hand

side of the equations. The equations may be inconsistent (ie no solution X) can be found, or consistent (solution(s) X exists). If the equations are consistent, the solution X may be unique or many solutions for X exist. Additional constraints on for example the size of X can be imposed to isolate a solution from the many solutions. In this Section the elementary treatment of classifying and exposing these solution(s) if any, by (elementary) row operations is given. These elementary row operations are carried out on the augmented matrix Aug $= \left[\begin{array}{cc} A & b \end{array}\right]_{m\times n+1}$ to reduce it to an appropriate row reduced echelon form from which conclusions can be drawn. I will restrict us to real number and/or complex numbers as the elements of the various matrices in the following notes of this section albeit the ideas can be carried over to polynomial matrices and rational matrices over (stable Laplace domain) polynomials with extensions (See for example: M. Newman – Integral Matrices, T. Kailath - Linear Systems or M. Vidyasagar - Control Systems Synthesis– a factorization approach.)

The notation I follow below is from Towers - Guide to Linear Algebra. Elementary row operations are done with matrices $E_{r,s_{m\times m}}$ acting on the left of a given matrix $A_{m\times p}$ by matrix product rule. Three types of elementary row matrices are useful $E_r(\alpha), E_{r,s}(\alpha), E_{rs}$ which stand respectively for multiply row r by α, to row r add row s multiplied by α, interchange row r and s respectively.

Example: with $m = 3$

$$E_2(3) = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{array}\right], E_{2,1}(-5+3j) = \left[\begin{array}{ccc} 1 & 0 & 0 \\ -5+3j & 1 & 0 \\ 0 & 0 & 1 \end{array}\right], E_{2,3} = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{array}\right] \tag{1.15}$$

I illustrate the process by 5 examples (Examples 2, 3 4 are from Towers Guide to Linear Algebra Pg 59- Exercise 3.2.1(a),(b),(d)). Showing the process by examples is far easier than to write in general. You may wish to consult Towers Chapter 3 for generic notation etc for reduced row echelon form. Finally this method may be familiar to you as Gaussian Elimination in any numerical methods course you may have done. There you would have considered pivoting. As done here we are not addressing pivoting.

1.3.1 Example 1

Consider the equations $3x + 5y + 6z + 4v = 1, x + y + z + 2v = 2, x - y - 3z - 2v = 3$. With $x_1 = x, x_2 = y, x_3 = z, x_4 = v$ these set of equations can be written as $AX = b$ where

$$A = \left[\begin{array}{cccc} 3 & 5 & 6 & 4 \\ 1 & 1 & 1 & 2 \\ 1 & -1 & -3 & -2 \end{array}\right], b = \left[\begin{array}{c} 1 \\ 2 \\ 3 \end{array}\right], X = \left[\begin{array}{c} x \\ y \\ z \\ v \end{array}\right] \tag{1.16}$$

```
>> A=[3 5 6 4;1 1 1 2;1 -1 -3 -2],b=[1;2;3]
A =
```

```
3   5   6   4
1   1   1   2
1  -1  -3  -2

b =

1
2
3

>> Aug=[A b]
Aug =

3   5   6   4   1
1   1   1   2   2
1  -1  -3  -2   3

>> E1=[1 0 0;-1/3 1 0;0 0 1]
E1 =

 1.0000         0         0
-0.3333    1.0000         0
 0         0              1.0000

>> Aug=E1*Aug
Aug =

3.0000   5.0000   6.0000   4.0000   1.0000
0       -0.6667  -1.0000   0.6667   1.6667
1.0000  -1.0000  -3.0000  -2.0000   3.0000

>> E2=[1 0 0;0 1 0;-1/3 0 1]
E2 =

1.0000         0         0
0              1.0000    0
-0.3333        0         1.0000

>> Aug=E2*Aug
Aug =

3.0000   5.0000   6.0000   4.0000   1.0000
0       -0.6667  -1.0000   0.6667   1.6667
```

```
0       -2.6667  -5.0000  -3.3333   2.6667

>> E3=[1 0 0;0 -1/Aug(2,2) 0;0 0 1]
E3 =

1.0000        0             0
0             1.5000        0
0             0             1.0000

>> Aug=E3*Aug
Aug =

3.0000   5.0000   6.0000   4.0000   1.0000
0       -1.0000  -1.5000   1.0000   2.5000
0       -2.6667  -5.0000  -3.3333   2.6667

>> E4=[1 0 0;0 -1 0;0 0 1]
E4 =

1   0   0
0  -1   0
0   0   1

>> Aug=E4*Aug
Aug =

3.0000   5.0000   6.0000   4.0000   1.0000
0        1.0000   1.5000  -1.0000  -2.5000
0       -2.6667  -5.0000  -3.3333   2.6667

>> E5=[1 0 0;0 1 0;0 -Aug(3,2) 1]
E5 =

1.0000        0             0
0             1.0000        0
0             2.6667   1.0000

>> Aug=E5*Aug
Aug =

3.0000   5.0000   6.0000   4.0000   1.0000
0        1.0000   1.5000  -1.0000  -2.5000
0        0       -1.0000  -6.0000  -4.0000
```

```
>> diary off
```

The final form of Aug above is the elementary row reduced form of the system of equations. From this one can see that infinite solutions exist as $z = 4 - 6v, y = -2.5 + v - 1.5z, x = \frac{(1-4v-6z-5y)}{3}$. Recursively substitution in WxMaxima, we get x, y, z in terms of v and constants below:

(% i3) z:4-6*v;y:-2.5+v-1.5*z;x:(1-4*v-6*z-5*y)/3;

$$4 - 6v \qquad (\%\ o1)$$

$$v - 1.5\,(4 - 6v) - 2.5 \qquad (\%\ o2)$$

$$\frac{-5\,(v - 1.5\,(4 - 6v) - 2.5) - 4v - 6\,(4 - 6v) + 1}{3} \qquad (\%\ o3)$$

(% i4) ratsimp(y);

$$rat : replaced - 2.5by - 5/2 = -2.5rat : replaced - 1.5by - 3/2 = -1.5$$

$$\frac{20v - 17}{2} \qquad (\%\ o4)$$

(% i6) ratsimp(x);

$$rat : replaced - 2.5by - 5/2 = -2.5rat : replaced - 1.5by - 3/2 = -1.5$$

$$-\frac{12v - 13}{2} \qquad (\%\ o6)$$

->

1.3.2 Example 2

Consider $2x_1 + 3x_2 - x_3 = -1, -x_1 - 4x_2 + 5x_3 = 3, x_1 - 2x_2 - 3x_3 = 3$. Forming the augmented matrix and reducing in Matlab/Octave as below we arrive at the final row reduced echelon form for this set of equations:

```
>> Aug=[2 3 -1 -1;-1 -4 5 3;1 -2 -3 3]
Aug =

2   3  -1  -1
-1  -4   5   3
1  -2  -3   3

>> E1=[1/Aug(1,1),0 0;0 1 0;0 0 1];
```

```
>> Aug=E1*Aug
Aug =

1.0000   1.5000  -0.5000  -0.5000
   -1.0000  -4.0000   5.0000   3.0000
1.0000  -2.0000  -3.0000   3.0000

>> E2=[1 0 0;1 1 0;-1 0 1]
E2 =

1   0   0
1   1   0
-1   0   1

>> Aug=E2*Aug
Aug =

1.0000   1.5000  -0.5000  -0.5000
0       -2.5000   4.5000   2.5000
0       -3.5000  -2.5000   3.5000

>> E3=[1 0 0;0 1/Aug(2,2) 0;0 0 1]
E3 =

1.0000        0             0
0            -0.4000        0
0             0             1.0000

>> Aug=E3*Aug
Aug =

1.0000   1.5000  -0.5000  -0.5000
0        1.0000  -1.8000  -1.0000
0       -3.5000  -2.5000   3.5000

>> E4=[1 0 0;0 1 0;0 -Aug(3,2) 1]
E4 =

1.0000        0             0
0             1.0000        0
0             3.5000   1.0000

>> Aug=E4*Aug
```

```
Aug =

1.0000   1.5000  -0.5000  -0.5000
0        1.0000  -1.8000  -1.0000
0        0       -8.8000        0

>> diary off
```

We see from the final Aug matrix, that the equations are consistent and the solutions are unique: $x_3 = 0, x_2 = -1, x_1 = -0.5 + 1.5 = 1$

1.3.3 Example 3

Consider $x_1 + 2x_2 - x_3 = 1, 3x_1 - 5x_2 + 2x_3 = 6, -x_1 + 9x_2 - 4x_3 = -4$ Forming the augmented matrix and reducing in Matlab/Octave as below we arrive at the final row reduced echelon form for this set of equations:

```
>> Aug=[1 2 -1 1;3 -5 2 6;-1 9 -4 -4]
Aug =

1   2  -1   1
3  -5   2   6
-1   9  -4  -4

>> E1=[1 0 0;-3 1 0;1 0 1]
E1 =

1   0   0
-3   1   0
1   0   1

>> Aug=E1*Aug
Aug =

1    2   -1    1
0  -11    5    3
0   11   -5   -3

>> E2=[1 0 0;0 1 0;0 1 1]
E2 =

1   0   0
0   1   0
```

```
0   1   1

>> Aug=E2*Aug
Aug =

1    2   -1    1
0  -11    5    3
0    0    0    0

>> diary off
```

From the final form of the Aug matrix, we see that $0x_3 = 0$ so any value of x_3 satisfies this equation. Now the remaining two equations (first two rows of Aug) can be solved for in terms of constants and the free variable x_3 like in Example 1 above (Complete this). So the original equations are consistent with many solutions. In 3-D with $x_1 = x, x_2 = y, x_3 = z$, the solutions with various $x_3 = \lambda$ is a line. This is best seen by expressing the coordinates as the vector equation $\vec{r} = \vec{p} + \lambda\vec{d}$ and finding numerical values for the point p and the direction vector d etc.. (Try it.)

1.3.4 Example 4

Consider $x_1 + 3x_2 + x_3 = 0, 2x_1 - x_2 - x_3 = 1, x_1 - 4x_2 - 2x_3 = 2$. Forming the augmented matrix and reducing in Matlab/Octave as below we arrive at the final row reduced echelon form for this set of equations:

```
>> Aug=[1 3 1 0;2 -1 -1 1;1 -4 -2 2]
Aug =

1   3   1   0
2  -1  -1   1
1  -4  -2   2

>> E1=[1 0 0;-2 1 0;-1 0 1]
E1 =

1   0   0
-2   1   0
-1   0   1

>> Aug=E1*Aug
Aug =

1   3   1   0
0  -7  -3   1
0  -7  -3   2
```

```
>> E2=[1 0 0;0 1 0;0 -1 1]
E2 =

1   0   0
0   1   0
0  -1   1

>> Aug=E2*Aug
Aug =

1   3   1   0
0  -7  -3   1
0   0   0   1

>> diary off
```

From the last row of the final Aug matrix we see $0x_3 = 1$. No value of x_3 can satisfy this so the original set of equations is inconsistent.

1.3.5 Example 5

Consider Example 2 above and add to those original equations, an additional fourth equation $2x_1 + 3x_2 + x_3 = -1$. We now have 4 equations in 3 variables. Upon carrying out the reduction, the final Aug matrix as below:

```
Aug =

2.0000   3.0000  -1.0000  -1.0000
0       -2.5000   4.5000   2.5000
0        0       -8.8000        0
0        0        0              0
```

From this we can see that the set of 4 equations are consistent with the same unique solution as in Example 2.

1.4 Assignment

1. Consider the fourth equation to be $2x_1 + 3x_2 + x_3 = 1$ in Example 5. Investigate what happens to the augmented matrix after reduction and draw conclusions.

2. Investigate how this method of reducing to row echelon form can be used to find the inverse of a square matrix if it exists and how it reveals the non existence of the inverse of a matrix. Take some examples of A matrix 3 by 3 which are singular and non-singular and carry out the process and verify.

1.5 Algebraic Structures

1.5.1 Groups

Good references are Abstract Algebra- The Basic Graduate Year by Robert Ash Chapter 1 which I follow here in its development as it gets us to where we want quickly or J.B. Fraleigh et. al. A first course in Abstract Algebra (Chapters 1 to 11). This book provides a gentle introduction to all this.

1. A group G is a nonempty set of elements on which a binary operation . is defined with the following properties.
   (a) Closure: $a.b \, \epsilon \, G \, \forall \, a, b \, \epsilon \, G$.
   (b) Associativity: $(a.b).c = a.(b.c) \, \forall \, a, b, c \, \epsilon \, G$
   (c) Existence of an Identity element in G for this operation: There is an element in G denoted 1 such that $a.1 = 1.a = a \, \forall \, a \, \epsilon \, G$
   (d) Existence of an inverse element for every element in G for this operation: For each element $a \, \epsilon \, G \, \exists \, a^{-1} \, \epsilon \, G$ such that $a.a^{-1} = a^{-1}.a = 1$.

   Example: The set of all permutations of n letters forms a group under the binary operation of composition. Verify this for say $n = 5$.
   The identity element and inverse element for each $a \, \epsilon \, G$ are unique.

2. The order of a group G is the number of elements of G and is denoted $|G|$.

3. A group G is an abelian group if . is commutative i.e. $a.b = b.a \, \forall \, a, b \, \epsilon \, G$.
   Example: The set of integers modulo m, $(\mathcal{Z}_m)$ where m is a (positive) integer, integers $\mathcal{Z}$, Real numbers $\mathcal{R}$, Complex numbers $\mathcal{C}$ under the operation of addition as normally defined forms respective groups with the number 0 as identity element.

4. H is a subgroup of G if (and only if) it is a nonempty subset of G which forms a group under the same binary operation . of the group G.
   Example: Consider the Group $\mathcal{Z}_4 = \{0, 1, 2, 3\}$ with addition modulo 4 as the binary operation. (Verify it is a group). $H = \{0, 2\}$ is a subgroup of G. Consider $\mathcal{Z}_5$; its only subgroups are $\{0\}$ and itself.

5. The subgroup generated by A a subset of elements of the group G is the smallest subgroup of G which contains A. This subgroup generated by A is denoted $\langle A \rangle$.

6. A group generated by a single element is called cyclic and is therefore always abelian and is denoted $\langle a \rangle$ if the binary operation is understood or explicitly as $\langle a, . \rangle$ where the . refers to the binary operation of the group and a denotes the generating element. The groups elements can then be listed as $\{a, a^2, ..., a^n\}$ or $\{1, a, a^2, ...a^{n-1}\}$ with the understanding that $a^k.a^l = a^{k+l}$.
   Example: $\mathcal{Z}_4 = \langle 1 \rangle = \langle 3 \rangle$. $\mathcal{Z}_5 = \langle 1, +_5 \rangle$.

7. The order of an element a in a group G is denoted $|a|$ is the least positive integer n such that $a^n = 1$. If no such integer n exists, the order of a is infinite. Thus if $|a| = n$, then the cyclic subgroup $\langle a \rangle$ has exactly n elements, and $a^k = 1$ if and only if k is a multiple of n.

8. If G is a finite cyclic group of order n i.e G has n elements, then G has exactly one (necessarily cyclic) subgroup of order n/d for each positive divisor d of n, and G has no other subgroups.
Example: Instead of a formal proof of this result – consider our earlier $\mathcal{Z}_4$ (In point 4 above) which has 4 elements. The factorization of 4 is 1×4 or 2×2 and so the only nontrivial subgroup is H as in item 4. For $\mathcal{Z}_5$ the subgroups are only those indicated in item 4 as 5 is a prime number and has only the factorization 1×5.
Example: What are the non trivial subgroups of $\mathcal{Z}_{24}$ a cyclic group generated by $\langle 1, +_{24}\rangle$? Factorizing $24 = 2 \times 12 = 3 \times 8 = 4 \times 6$. So the nontrivial subgroups will have 2 elements, 3 elements, 4 elements, 6 elements, 8 elements and 12 elements. The trivial ones are $0, \mathcal{Z}_{24}$. What are these nontrivial subgroups: $\{12, 0\}, \{8, 16, 0\}, \{6, 12, 18, 0\}, \{4, 8, 12, 16, 20, 0\}, ...$ Hopefully you see the pattern by which the subgroups are generated. These are the only nontrivial subgroups.
Example: What if we consider generating the elements of $\mathcal{Z}_{24}$ using the element 5, we will get the entire group as 5 does not factor 24!. In our notation the group generated by 5 is $\langle 1^5, +_{24}\rangle$. We further can say $|1^5| = 24$. All positive integers which are relatively coprime to 24 (i.e. greatest common divisor (GCD i.e the largest positive integer that divides the two numbers) between the integer being considered and 24 is 1) will generate the group just as well. Before casting these results in a general form we detour to the Extended Euclid's algorithm in the next item. Why are we studying all this ? The basis of modern cryptography requires all these ideas from (applied) Algebra.

9. **Extended Euclid's Algorithm for GCD of two integers:** The extended Euclid's algorithm (Wikipedia has a decent article on the details) solves for the GCD of two integers p, q by computing integers x, y to satisfy the Bezout/Aryabhatta's identity as:

$$px + qy = gcd(p, q) \tag{1.17}$$

If p, q are relatively prime then $px + qy = 1$. This can be stated as $px \bmod q = 1$ i.e if p, q are relatively prime then there exists an integer x such that $px = 1 \bmod q$ or alternatively p is a unit $\bmod q$. or p has a multiplicative inverse $\bmod q$.
Example: Since integers 5 and 24 are relative prime. So the Bezout identity and gcd are obtained as 5*5-24*1=1 with $p = 5, q = 24, x = 5, y = -1$. Likewise 7 and 24 are relatively prime and 7*7-24*2=1 etc.. So we get the following set of relative prime numbers to 24 (which are less than 24): $P = \{1, 5, 7, 11, 13, 17, 19, 23\}$. The number of elements of this set is 8 integers. The Set P can be converted to a group by defining a binary operation as multiplication of integers with the result reduced modulo 24. Verify this! Note we are changing the binary operation from addition to multiplication followed by the modulo operation to make the set P into a group. In this form this new group becomes very useful in cryptography.

10. Couched in the language of GCDs item 8 above can now be stated as Ash does in his text-book/lectures (Abstract Algebra - The Basic Graduate Year, Page 3, Chapter 1) as:
If G is a cyclic group of order n generated by the element a, the following conditions are equivalent:

    (a) $|a^r| = n$.

(b) r and n are relatively prime.

(c) r is a unit mod n, in other words, r has a multiplicative inverse mod n (ie an integer s such that $rs = 1 \mod n$).

Furthermore, the set U_n of units mod n forms a group under multiplication. The order of this group is $\phi(n)$, the number of positive integers $\leq n$ that are relatively prime to n. ϕ is the Euler's totient function.

11. The set P in item 9 above has 8 elements and is $\phi(24)$. Can we find a general expression for $\phi(n)$. To do this we need to understand the inclusion-exclusion principle for elements of the union of two or more sets.

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2| \tag{1.18}$$

$$|A_1 \cup A_2 \cup A_3| = \sum_{1 \leq i \leq 3} |A_i| - \sum_{1 \leq i < j \leq 3} |A_i \cup A_j| + |A_1 \cap A_2 \cap A_3| \tag{1.19}$$

$$\left| \bigcup_{i=1}^{n} A_i \right| = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \leq i_1 < \cdots < i_k \leq n} |A_{i_1} \cap \cdots \cap A_{i_k}| \right) \tag{1.20}$$

This result is useful in probability theory later and is proved by induction

Now we apply this to compute $\phi(n)$ as follows. Let $n = \Pi_{i=1}^{k} p_i^{e_i}$. (Example $24 = 2^3 * 3^1$.) The number of elements in $\mathcal{Z}_n$ divisible by p_i $= n/p_i$. The number of elements divisible by both p_i, p_j is $n/(p_i p_j)$ etc..

Consider $k = 2$ ie n is factorizable into the product of two primes ($24 = 2^3 * 3^1$)The number of elements left over in the set of n numbers after removing the numbers divisible by p_1, p_2 will be $n - [n/p_1 + n/p_2 - n/(p_1 p_2)] = n(1 - 1/p_1)(1 - 1/p_2)$ by the inclusion exclusion principle.

Generalizing to arbitrary (again by induction on) k we get:

$$\phi(n) = n \Pi_{i=1}^{k} (1 - 1/p_i) \tag{1.21}$$

For $\mathcal{Z}_{24}$ this gives $\phi(24) = 24(1 - 1/2)(1 - 1/3) = 8$ as we got before in item 9.

For a prime number p, $\phi(p) = p - 1$.

12. Cosets of a subgroup H of an element g of a group G are defined as left coset $gH = \{g.h \, \epsilon \, G | \, \forall \, h \, \epsilon \, H\}$ and right coset $Hg = \{h.g \, \epsilon \, G | \, \forall \, h \, \epsilon \, H\}$. If the operation . of the group is commutative i.e. we are dealing with an abelian group then left and right cosets generated by an element g are the same. Let us work with abelian groups when dealing with cosets.

Two cosets generated by two distinct elements a, b of G acting on a subgroup H are equal if and only if $a.b^{-1} \, \epsilon \, H$.

Example: Consider $\mathcal{Z}_4 = \langle 1, +_4 \rangle$, with the subgroup $H = \{0, 2\}$ The cosets generated by $0, 1, 2, 3$ (elements of G) acting on H are respectively (remember the group operation is $+_4$): $\{0, 2\}, \{1, 3\}, \{0, 2\}, \{1, 3\}$. It is seen that the elements $0, 2$ produce the same coset and trivially

$0 +_4 2^{-1} = 0 +_4 2 = 2 \,\epsilon\, H$ and likewise the elements $1, 3$ produce the same coset $\{1, 3\}$ and $1 +_4 3^{-1} = 1 +_4 1 = 2 \,\epsilon\, H$. Note that each coset has the same number of elements as H.

So if we define two elements a, b of G to be equivalent wrt the subgroup H if $a.b^{-1} \,\epsilon\, H$ then for the above example the group G above can be thought to have just two equivalent classes represented as $0, 1$ wrt H as $0 \cong \{0, 2\}$, $1 \cong \{1, 3\}$.

The (integer) number of equivalent classes of G wrt a subgroup H is called the index and is denoted $[G : H]$.

13. Lagrange's theorem (one of many under his name) connects the number of elements in G,subgroup H and index as $|G| = |H|\,[G : H]$. If G is finite then $|H|$ divides $|G|$.

14. Trivial as Lagrange's theorem looks above it has 4 very important corollaries summarized from Ash Page 9, Chapter 1) for finite Groups G with order $|G| = n$ all of which have a lot of applications within the broad field of ECE and Computer Science:

    (a) For $a \,\epsilon\, G$ the subgroup generated by a in G has $|a|$ elements which divides n. (In $\mathcal{Z}_{24}$, the element 2 has order 12 and divides 24.). Hence, $a^n = 1$ (for element 2 in $\mathcal{Z}_{24}$ this says $2 + 2 + 2...24 \mod 24 = 0$. Note here 1 is the identity element in the group G and a^n denotes $a^{n-1}.a$ where . is the group operation. The result follows as we can factorize n by Lagrange's theorem into $|a|\,[G : \text{Subgroup generated by } a]$. Thus n is a multiple of the order of each of its elements, so if we define the exponent of G to be the least common multiple of $\{|a| : a \,\epsilon\, G\}$, then n is a multiple of the exponent.

    (b) If n is prime integer, then G is cyclic i.e the entire G is generated by an element (any element except the identity element).

    (c) Euler's theorem (The Basis of RSA algorithm discussed later): If a and n are relatively prime positive integers with $n \geq 2$ then $a^{\phi(n)} = 1 \mod n$. Note that the set P was converted to a group with the standard integer multiplication modulo n and so to cerify this for $\mathcal{Z}_{24}$ we can compute in Matlab/Octave using the set P of item 9 and standard integer power operation as follows:

    ```
    >> mod([1 5 7 11 13 17 19 23].^8,24)
    ```

    (d) Fermat's little theorem (We will use this later when we work on Galois Fields): If n is a prime integer p and a is a positive integer, then $a^{p-1} = 1 \mod p$. Again to verify this consider the integers $1, 2, 3, 4$ to power 4 modulo 5 and verify they are all 1 in Octave/Matlab.

    Assignment problem: Show that for any positive integer n, $n^{33} - n$ is divisible by 3 and 5 and so is also divisible by 15.

    (e) **Example:** The use of Euler's theorem in RSA encryption (Rivest, Shamir, Adlemann algorithm):

        i. Consider two large primes p, q and generate $n = p.q$ (ordinary multiplication generates n). Messages and their encryptions are considered as integer elements of $\mathcal{Z}_n$.
        ii. Compute $\phi(n) = (p - 1)(q - 1)$.

iii. Select a number (exponent of the algorithm e) in the range between $\{1, 2, ...\phi(n)-1\}$. which is relatively prime to $\phi(n)$.

iv. Compute the private key $k_{private}$ by $k_{private}.e = 1 \bmod \phi(n)$. The private key always exists by Euler's theorem above.

v. Transmit he public key as the pair (n, e) to whoever wants to communicate with you.

vi. Given the message x that some one wants to send you (note this is a integer in the range $0, 1, ..., n-1$, they will encrypt the message as $y = x^e \bmod n$ and transmit y to you.

vii. You decrypt the message as $x = y^{k_{private}} \bmod n$.
Why does this work? Since $e.k_{private} = 1 \bmod \phi(n)$ and Euler's theorem takes over.

viii. To illustrate with numbers Consider numerical values of $p = 5, q = 11, n = 55, phi(55) = 40$,Choose $e = 7$ relatively prime to 40, then $k_{private} = 23$ from $k_{private}.e \bmod 40 = 1$. (I find this solution using Matlab/Octave by computing mod(7*[1:39],40) and finding the index of the integer at which this is 1.
Let the message be $x = 13$. Then $y = x^e \bmod n = 13^7 \bmod 55 = 7$. At the receiving end computing mod(7^{23}, 55) directly in Matlab/Octave will give 0 as the mod function in Octave/Matlab fails on large integers!. Instead we compute this as follows recursively:

```
>> z=1;for i=1:23 z=mod(z*7,55);end;z
z = 13
```

ix. What makes RSA hard to crack. First given n in the public key one must factorize it into the two prime factors p, q. With current length of RSA at 2048 bits for n this is a hard problem If that could be done quickly then p, q are known and the private key can be calculated etc. from the public key e as $\phi(n)$ can be computed. This factorization of n into p, q is estimated to take a long time (One estimate seen by me on this is trillion years on normal desktop computers circa 2019 but with a quantum computer powerful enough it may take just 10 seconds(estimate)!). If two certificates where public keys are given out with two different n, have a common prime factor, then the extended Euclidean algorithm given before can find this common prime factor on desktop machines in seconds and both certificates can now be broken and the private keys worked out in a matter of seconds! This has been shown in research and on TLS (transport layer security) certificates on the net!. The two primes are selected by random search in for example SSH and there is no guarantee that two or more TLS certificates won't have the same common factor in their n given the large number of TLS certificates that are generated now that practically every internet site is https. For an introduction of some practical details on SSL(Secure Socket layer)/TLS certificates etc. see for example https://kulkarniamit.github.io/whatwhyhow/howto/verify-ssl-tls-certificate-signature.html

1.5.2 Rings

So far groups had just one binary operation. But we have been using integers on which two binary operations were defined (addition and multiplication). So we would like to study structures in which

two binary operations are defined appropriately like in integers. These algebraic structures are Rings.

1. A ring $\langle R, +, . \rangle$ is a set R, with two binary operations $+$ and . defined on R such that

   (a) $\langle R, + \rangle$ is an abelian group

   (b) . operation is associative

   (c) $a.(b+c) = a.b + a.c$ and $(b+c).a = b.a + c.a \; \forall \; a, b, c \; \epsilon \; R$

   As defined the operation . may not be commutative. The identity element of the group in the ring is denoted 0 (zero) under the operation $+$. The identity element for multiplication operation . in the ring may or may not exist. If it exists, it will be denoted 1 (unity).

2. **Example:** Consider the set of integers Z with the usual operation of $+$ and . (multiplication). This is a ring with additive identity 0 and multiplicative identity 1 and the operation . is commutative. Consider the set of even integers $2Z$ with the usual operation of $+$ and .. This is also a ring but does not have a multiplicative identity (i.e has no unity element).

   Consider the set of all $n \times n$ matrices M_n over integers (or complex numbers or integers) with addition and multiplication defined as matrix addition and matrix multiplication. This is a ring with the multiplication operation being non-commutative for $n \geq 2$. Identity elements for both $+$ and . exist. (However the very important issue with rings is that there is no guarantee of multiplicative inverse as the definition does not require that even when multiplicative identity exists in the ring). In the 3 examples provided so far multiplicative inverses for each element does not exist in the ring for all elements (even after excluding 0). If a multiplicative inverse of an element exists in the ring (necessarily for this to occur the multiplicative identity must exist in the ring) then this element is classified as a unit in the ring.

   Assignment Problem: Prove that the units in the ring of matrices M_2 over the integers are matrices whose determinant is ± 1. Another name for such matrices are unimodular matrices in this ring. The result generalizes to the ring of matrices of order n over integers.

   Consider $\mathcal{Z}_{24}$ which was a group under the $+_{24}$. We can convert this to a ring by defining usual operation of multiplication as also modulo 24. We will denote this ring as $\langle Z, +_{24}, .{}_{24} \rangle$

3. Consider $\langle Z, +_{24}, .{}_{24} \rangle$. Solve the quadratic $x^2 - 5x + 6 = 0$ polynomial equation for solutions in this ring if any. One pair of solution is $x = 2$ or $x = 3$ and both "roots" represented as a ordered pair $(2, 3)$) are in the ring being considered. Are there more ? Since under multiplication modulo 24 all of the following: $a_1 \times a_2 = 2 \times 12, 3 \times 8, 4 \times 6, 6 \times 4, 8 \times 312 \times 2 = 24$ we can generate other ordered pair of roots by using $(a_1 + 2, a_2 + 3)$ and all are valid roots! Are there more- how about $a_1 \times a_2 = 9 \times 8, 10 \times 12, 8 \times 9, 12 \times 10$. Yes these will also give ordered pair solutions to the polynomial as roots in the ring. Every pair of integers whose product modulo 24 is a zero from the ring contributes to a solution! a_1, a_2 such that $a_1 .{}_{24} a_2 = 0$ are the divisors of zero. (You should have noticed that the integers from which we generate multiple solutions which are the divisors of zero are those which are relatively not prime to 24.)

   To make a solution of a polynomial equation, if it exists, in the ring $\langle Z, +_n, .{}_n \rangle$ unique, what is required is that for all non-zero elements of the ring a_1,a_2 there is no solution to $a_1 .{}_n a_2 = 0$. i.e the non-zero elements of the ring. If n is a prime then this condition is satisfied.

4. Based on the above discussion and making the multiplication operation in a ring also commutative, we define a commutative ring with no (non-zero) elements as divisors of zero to be an integral domain. (The name comes from the properties of the integers that if $a.b$=0 and if $a \neq 0$ then $b = 0$ holds in the integers - i.e. cancellation law applies.)

5. A polynomial in the indeterminate x over a commutative (integral domain) ring R with unity is an infinite series given below **with only a finite number of coefficients** a_i, from the ring R, being non zero.

$$f(x) = \sum_{i=0}^{\infty} a_i x^i, \quad a_i \, \epsilon \, R \tag{1.22}$$

   The highest integer i for which a_i is non zero is the degree of the polynomial.

6. The set of all polynomial $R[x]$ as defined above forms a ring with the usual operations of addition and multiplication of polynomials. Note: when manipulating the coefficients of the polynomial, the operations of the underlying ring R from which the coefficients emanate apply. A polynomial $f(x)$ in the ring $R[x]$ is irreducible if it cannot be written as $f(x) = g(x)h(x)$ with $g(x), h(x)$ having degree strictly less than that of $f(x)$.

   Example: Consider the ring $R = \langle Z_3, +_3, .3 \rangle$ and the polynomial ring $R[x]$, then the sum of the two polynomials $p(x) = 2x + x^2$ and $q(x) = 1 + 2x + x^2 + 2x^3$ is $1 + x + 2x^2 + 2x^3$ and $-p(x)$ is the polynomial is $x + 2x^2$ as $p(x) + (-p(x)) = 0$ (the rhs in this equation is the zero polynomial). The product $p(x)q(x)$ is $2x^5 + 2x^4 + x^3 + 2x^2 + 2x$. Note computations can be done in Octave/Matlab where polynomials are entered as coefficients in descending powers of x. conv function in Octave/Matlab will do the multiplication and all operations are done mod 3.

```
>> p=[1 2 0],q=[2 1 2 1]
p =

1   2   0

q =

2   1   2   1

>> p=[0 p]
p =

0   1   2   0
>> mod(p+q,3)
ans =

2   2   1   1
```

```
>> mod(-p,3)
ans =

0   2   1   0

>> mod(conv(p,q),3)
ans =

0   2   2   1   2   2   0

>> mod(q-p,3)
ans =

2   0   0   1

>> mod(p-q,3)
ans =

0   2   1   0
```

7. **Assignment Problem:** Show that $M_n(R[x])$ (matrices of order n over the commutative ring of polynomials over the integral domain R) form a ring with unimodular matrices having determinant ± 1.

1.5.3 Fields

Commutative rings have elements which may not have multiplicative inverses. Yet the rational numbers, real numbers and complex numbers have multiplicative inverses for every non-zero element. So the next logical abstract algebra structure is Fields and we define this in general as below (I follow Walter Rudin - Principles of Mathematical Analysis - Chapter 1, Pg. 5. Rudin's definition of the field does not use the concepts of rings or groups but includes their definitions in that of the field.)

1. A field F is a set with two operations called addition (+) and multiplication (.) which satisfy the following axioms:

   A1 If $x, y \epsilon F$ then $x + y \epsilon F$.

   A2 $x + y = y + x \; \forall \; x, y \epsilon F$.

   A3 $(x + y) + z = x + (y + z) \; \forall \; x, y, z \epsilon F$.

   A4 F contains an element called 0 such that $x + 0 = 0 + x = x \; \forall \; x \epsilon F$.

   A5 To every $x \epsilon F \; \exists \; -x \epsilon F$ such that $-x + x = 0$.

   M1 If $x, y \epsilon F$, then $x.y \epsilon F$.

   M2 $x.y = y.x \; \forall \; x, y \epsilon F$.

M3 $(x.y).z = x.(y.z) \forall\, x, y, z \epsilon F$.

M4 F contains a multiplicative identity denoted $1 \neq 0$ such that $1.x = x \forall\, x \epsilon F$.

M5 If $x \epsilon\, , x \neq 0$, $\exists\, x^{-1} \epsilon F$ denoted often as $\frac{1}{x}$ such that $x.x^{-1} = 1$.

D $x.(y + z) = x.y + x.z \forall\, x, y, z \epsilon F$.

So a field F is a commutative ring with operations + and . and having a multiplicative identity 1 such that for every non-zero element in it a multiplicative inverse exists.

2. Examples: The set of real numbers $\mathbb{R}$, the set of complex numbers $\mathbb{C}$, the set of rational numbers $\mathbb{Q}$ with the usual operations of + and . are fields. The set $\mathbb{Z}$ of integers is not a field with the usual operations. The Ring $\langle Z, +_n, .n \rangle$ is a field if $n = p$, a prime integer and this field is denoted $GF(p)$ (GF for Galois Field). The Boolean set $\{0, 1\}$ under the operation of exor ($\oplus$) for addition with the and operation (.) as multiplication is a field and is denoted $GF(2)$. Note the exor and the and operations can be thought of as addition and multiplication modulo 2 here and so the notation fits with that introduced earlier $GF(p)$. We remind ourselves of Fermat's little theorem which states that in $GF(p)$, $x^{p-1} = 1 \bmod p$ or equivalently $x^p = x \bmod p$ or equivalently the polynomial $x^p - x = 0 \bmod p$. Note in the last form 0 on the right hand side is the zero polynomial in the ring of polynomials of degree less than or equal to $p - 1$ as we will see soon.

3. Let us now look at $GF(2^l)[x]/p(x)$ which stands for the Galois field of polynomials with coefficients in $GF(2)$ which are of degree $< l$ and in which multiplication and addition are defined in the usual manner for multiplying and adding polynomials with $GF(2)$ rules for coefficients and the (multiplication) results are reduced (modulo) the prime polynomial $p(x)$. The prime polynomial $p(x)$ reduces the degree of the product of two polynomials to lie in the field (and makes the ring into a field) What constitutes a prime polynomial? It is a polynomial which cannot be factorized into lower power polynomials with coefficients in $GF(2)$ and its degree is l (and this degree term has coefficient 1- a redundant statement).

4. **Examples:** Let us consider creating the field $GF(2^2)[x]/p(x)$. We recognize $l = 2$ and the degree of $p(x) = 2$. Since there are 4 elements in this field (2^2) we consider from Fermat's theorem the polynomial $x^4 - x$ and factorize this to get $p(x)$ a 2nd order polynomial out of it which is irreducible. $x^4 - x = x(x - 1)(x^2 + x + 1) = x(x + 1)(x^2 + x + 1)$. Consider all 2nd degree polynomials of x with leading coefficient of 1: They are: $x^2, x^2 + x, x^2 + 1, x^2 + x + 1$. Of these 4 the only one which does not have zeroes (roots) 0 or 1 is $x^2 + x + 1$. This is also a factor in Fermat's equation. Let us see if it meets our requirements of being irreducible.

   Let us start with $\{0, x, x^2, x^3\}$ as the polynomials in $GF(2^2)[x]$ and then reduce with $p(x) = x^2 + x + 1$ the two elements x^2, x^3. Doing the long division we get $x + 1, 1$. The field is complete with elements $\{0, x, x + 1, 1\}$ and is closed under exor addition of coefficients of polynomials and on multiplication modulo $p(x)$. Alternative representations include coding the field elements with the coefficients of the polynomial elements and we get the elements as two bit representation for the elements $\{00, 10, 11, 01\}$ with exor addition and bitwise multiplication and addition (no carry) and reduction by $p(x)$ coefficients ($100 = 011$) produces the same results

as the original polynomial field. A third form is to define α as the zero(root) of $p(x)$ and the elements as $\{0, \alpha, \alpha^2, \alpha^3 = 1\}$. Since α is a root of $p(x)$, $\alpha^2 = \alpha + 1$.

Let us consider now $GF(2^3)[x]/p(x)$. To find $p(x)$, factorize $x^8 - x = x(x+1)r(x)$, $r(x) = x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$. We need to factorize the $r(x)$ polynomial into irreducible factors (Remember coefficients are $\{0, 1\}$ and exor addition etc.) Consider all 2nd degree polynomials with leading coeff of 1 and ending with constant of 1 (constant of 1 at the end is needed to get the 1 as the x^0 coefficient of the $r(x)$ polynomial), these are: $x^2 + 1, x^2 + x + 1$. We reject $x^2 + 1$, as its roots are 1. Dividing $r(x)$ by $x^2 + x + 1$ we get a remainder of 1 (verify this). So we do not have 2nd order factors for $r(x)$. Consider third order factors: The only ones we need to consider are $x^3 + x^2 + x + 1, x^3 + x + 1, x^3 + x^2 + 1$. Of these $x^3 + x^2 + x + 1$ is not a factor as division of $r(x)$ gives a remainder $x^2 + x + 1$. The other two are factors of $r(x)$.

We now have $x^8 - x = x(x+1)(x^3 + x + 1)(x^3 + x^2 + 1)$ and I will choose $p(x) = x^3 + x + 1$. The elements of $GF(2^3)[x]/p(x) = \{0, x, x^2, x^3, x^4, x^5, x^6, x^7\} = \{0, x, x^2, x+1, x^2+x, x^2+x+1, x^2+1, 1\} = \{000, 010, 100, 011, 110, 111, 101, 001\} = \{0, \alpha, \alpha^2, \alpha + 1, \alpha^2 + \alpha, \alpha^2 + \alpha + 1, \alpha^2 + 1, 1\}$ where α is a root of $p(x)$. One can verify that the constructed $GF(2^3)$ is a field under exor addition and the and coefficient multiplication rules modulo $p(x)$.

5. **Generating Reed Solomon Codes is an application of GF theory**: Consider the same $GF(2^3)[x]/p(x)$ as the underlying field with $p(x) = x^3 + x + 1$. Consider a code generating polynomial $P(x) = b_0 + b_1 x + b_2 x^2$. (b_0, b_1, b_2) is the message in blocks each of 3 bits for a block size of 9 bits (and therefore can code upto 8^3 different messages). The 9 bit message is mapped to $[P(0), P(\alpha), P(\alpha^2), ...P(\alpha^7)]$. (Following Reed-Solomon's original paper-"Polynomial Codes Over Certain Finite Fields", I. S. Reed and G. Solomon, Journal of the Society for Industrial and Applied Mathematics, SIAM, Jun., 1960, Vol. 8, No. 2, pp. 300-304. Get it from the library (**Assignment**)).

   Example Message: $(b_0 = 000, b_1 = 010, b_2 = 011)$ ie $(b_0 = 0, b_1 = \alpha, b_2 = \alpha^3 = \alpha + 1)$. The corresponding code word for this 9 bit message is computed as: $P(0) = 000, P(\alpha) = b_1\alpha + b_2\alpha^2 = \alpha^2 + \alpha^3 + \alpha^2 = \alpha^3 = \alpha + 1 = 011, P(\alpha^2) = b_1\alpha^2 + b_2\alpha^4 = \alpha^3 + \alpha^5 + \alpha^4 = \alpha + 1 + \alpha^2 + \alpha + 1 + \alpha^2 + \alpha = \alpha = 010, P(\alpha^3) = \alpha = 010, P(\alpha^4) = 1 = 001, P(\alpha^5) = 0 = 000, P(\alpha^6) = \alpha + 1, P(\alpha^7) = 1 = 001$. So the 9 bit message $[000, 010, 011]$ is coded as a 24 bit word in this $[000, 011, 010, 010, 001, 000, 011, 001]$. This is the same result as the example in the paper referred above but they use a different coding of binary 3 bits. (**Assignment:** study how they suggest decoding the received message and understand their process of decoding. The paper is a classic powerful example of concise clear and precise writing. Reed Solomon codes are one of the means by which the 1977 Voyager Mission launched by NASA communicate with earth along with convolutional coding techniques. (See IEEE Press book -Reed Solomon Codes and their Applications Edited by S.B. Wicker and V.K. Bhargava.)

6. **Assignment:** Choose $p(x) = x^3 + x^2 + 1$ and set up $GF(2^3)[x]/p(x)$ for this $p(x)$ and generate RS code for 9 bits to 24 bits with the same coding polynomial and same message as above example.

7. One can now consider matrices over $GF(2^l)[x]/p(x)$ and as before study their properties. For example solve the following problem from Midterm 2022 for this course:

In the field $GF(2^3)[x]/p(x) = x^3 + x + 1$, the following matrix A is written in terms of the elements of this field with α the root of $x^3 + x + 1$). Find the determinant of A in this field in its simplest form as a polynomial in α with degree ≤ 2. (Numbers for m, n, p see **Data** below)

$$A = \begin{bmatrix} 1 & \alpha^m & \alpha^{2m} \\ 1 & \alpha^n & \alpha^{2n} \\ 1 & \alpha^p & \alpha^{2p} \end{bmatrix} \tag{1.23}$$

Data: $m = 2, n = 3, p = 5$

1.6 Vector Spaces

1. A vector space over a field F is a set V whose elements are called vectors with an addition operation + and an operation . for (scalar) multiplication between elements of the field F of with elements of the space V such that the space is closed under these operations and the following axioms are satisfied:

   V1 the addition operation is commutative.

   V2 the addition operation is associative

   V3 the exists an identity element for the operation of addition $0 \epsilon V$ called zero such that $0 + v = v \; \forall \; v \epsilon V$

   V4 for each element of V there exists an additive inverse element in V

   V5 for $1 \epsilon F$, $1.v = v \; \forall \; v \epsilon V$

   V6 $(\lambda.\mu)v = \lambda.(\mu.v) \; \forall \; \lambda, \mu \epsilon F$ and $v \epsilon V$

   V7 $\lambda.(v + w) = \lambda.v + \lambda.w \; \forall \; \lambda \epsilon F$ and $v, w \epsilon V$

   V8 $(\lambda + \mu) \, .v = \lambda.v + \mu.v \; \forall \; v \epsilon V$ and $\lambda, \mu \epsilon F$.

2. **Examples:** The traditional collection of 2-D, 3-D n-D vectors over the field $\mathbb{R}$ or *mathbbC* is a classic example of vector spaces with addition being defined as scalar addition of the 'coordinates' of the vectors and scalar multiplication defined as scaling each 'coordinate' of the vector by the scalar from the field.

   The set of Matrices $M_n(\mathbb{R})$ with addition defined as per addition of matrices and scalar multiplication defined as multiply each element of the matrix by the element of the field constitutes a vector space.

   The set of polynomials of degree $< n$ in the indeterminate x with coefficients from $\mathbb{R}$ with usual addition of coefficients of same powers and scalar multiplication forms a vector space. Each polynomial of this vector space can be put into a one-one mapping onto n-D vectors from the first example above (Matlab/Octave do this with descending powers of the powers and exploit this internally). Matlab/Octave intrinsically are geared to operate on n-D vectors.

Conider the set denoted $\mathbb{Q}\sqrt{2} = \{a+b\sqrt{2}, a, b \,\epsilon\, \mathbb{Q}\}$ over the field $\mathbb{Q}$ of rational numbers. Define the addition operation as $(a+b\sqrt{2})+(c+d\sqrt{2}) = (a+c)+(b+d)\sqrt{2}$ and scalar multiplication $\alpha(a+b\sqrt{2}) = a\alpha + b\alpha\sqrt{2}$. This set now becomes a vector space and can be put into a one to one mapping onto 2-D vectors.

The set of all functions over the real numbers or complex numbers with addition and scalar multiplication defined in the usual manner constitute a vector space.

3. A subspace W of V is defined as a non-empty $W \subset V$ such that W by itself is a vector space (i.e W is closed under the operations of + and . scalar multiplication of the space V).

   Examples: In n-D vector space, the 0 vector alone forms a trivial subspace.

   In 3-D vector space whose typical element may be $[x, y, z]'$, the 2-D vectors $[x, y, 0]'$ form a subspace.

4. If $\{v_i, i = 1, 2, ..., n\}$ are a set of vectors in V and $\exists\, \alpha_i \; i = 1, 2, ..., n$ not all 0 such that $\sum_{i=1}^{n} \alpha_i v_i = 0$ the the set $\{v_i, i = 1, 2, ..., n\}$ are a set of dependent vectors. If $\sum_{i=1}^{n} \alpha_i v_i = 0 \implies \alpha_i = 0 \,\forall\, i = 1, 2, ...n$ the the vectors $v_i,\; i = 1, 2, ...n$ are independent.

   Examples: The set of vectors $\{[1, 0, 0]', [0, 1, 0]', [0, 0, 1]'\}$ are independent vectors in 3-D vector space and so as a corollary $1, x, x^2$ are independent vectors in the polynomial vector space of degree < 3. So also are the vectors $\{[1, 0, 0]', [1, 1, 0]', [1, 1, 1]'\}$ independent in 3-D vector space and as a corollary so are the polynomials $1, 1 + x, 1 + x + x^2$ for polynomial vector space of degree < 3.

5. **Assignment Problem:** Show that each linearly independent set of vectors $\{v_i, i = 1, 2, ...n\}$ defines a subspace by generating vectors as (linear combination) $\sum_{i=1}^{n} \alpha_i v_i,\; \alpha_i \,\epsilon\, F$.

6. The number of vectors in the smallest set of independent vectors which generates the entire space V is called the dimension of V often written $\dim(V)$. The dimension of a vector space V is unique. Any smallest independent set of vectors which generates the vector space V by linear combination is a (set of) basis (vectors) of the vector space. For a n-dimensional vector space, I will use B_n to denote a basis and when needed spell out its elements if needed later. Note many B_n exist for a vector space.

7. The vector space of polynomials of any degree has dimension ∞ as e^x an infinite series in x belongs to it.

8. A transformation T between two vector spaces V, W over the same field F is a mapping of vectors from V to W (a rule (one-one or many-one) which takes each $v \,\epsilon\, V$ to a vector $w \,\epsilon\, W$). Transformations are classified as linear or nonlinear.

   Examples: V, W are 2-D (x,y) and 3-D spaces (x,y,z) respectively. Consider T_1 defined as $T_1([x, y]') = [x+y, x-y, 2x+3y]'$ and $T_2([x, y]') = [x^2, x-y, 2x+3y]'$. Both are transformations between V, W.

9. A linear transformation T between V, W is a mapping such that:

   (a) $T(v_1 + v_2) = T(v_1) + T(v_2) \,\forall\, v_1, v_2 \,\epsilon\, V$

(b) $T(\alpha v) = \alpha T(v) \; \forall \; \alpha \; \epsilon \; F, v \; \epsilon \; V$

From the two examples on transformations above T_1 is linear while T_2 is nonlinear.

10. if B_n and B_m are a basis of V with $\dim(V) = n$ and W with $\dim(W) = m$, then a linear transformation T has a matrix representation to transform a vector v with coordinates $\alpha_i, i = 1, 2, ..., n$ in the basis B_n to a vector w with coordinates $\beta_j, j = 1, 2, ..., m$. This is how matrices arise and while the transformation T is unique, since V, W have many different basis, the matrices obtained in each pair of basis for the same T can be different!.

Examples: Consider T_1 in item 8 above. The way the rule was defined a tacit basis (canonical) was assumed viz $B_2 = \{v_1 = [1, 0]', v_2 = [0, 1]'\}$ and $B_3 = \{w_1 = [1, 0, 0]', w_2 = [0, 1, 0]', w_3 = [0, 0, 1]'\}$.

The vector $[x, y]' = xv_1 + yv_2$ in the input space V.

$T_1(v_1) = [1, 1, 2]' = 1w_1 + 1w_2 + 2w_3$ and $T_1(v_2) = [1, -1, 3]' = 1w_1 - 1w_2 + 3w_3$. Using (9) $T_1([x, y]') = xT_1(v_1) + yT_1(v_2)$. We now arrive at the matrix representation for T_1 between these two basis as

$$\begin{bmatrix} x+y \\ x-y \\ 2x+3y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{1.24}$$

So T_1 has the matrix representation with numbers of how the basis vectors of the input space gets transformed into the basis vectors in the output space.

We will consider the linear transformation defined by the derivative operator $D = \frac{d}{dx}$ operating on polynomial space with basis $[v_1 = 1, v_2 = x, v_3 = x^2]$ and producing an output in the polynomial space with basis $[w_1 = 1, w_2 = 1 + x]$. (note: that the output space basis is not the usual canonical basis!). Applying the D operator to the basis $[1, x, x^2]$ we get $[0, 1, 2x]$ which in the output basis is $[0w_1 + 0w_2, 1w_1 + 0w_2, -2w_1 + 2w_2]$. If the input polynomial is represented as $a + bx + cx^2 = av_1 + bv_2 + cv_3$ then the matrix of numbers associated with D in these two basis for V, W is given by:

$$\begin{bmatrix} b - 2c \\ 2c \end{bmatrix} = \begin{bmatrix} 0 & 1 & -2 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{1.25}$$

Note that $(b-2c)1+2c(1+x) = b+2cx$ which correctly describes what D does to the polynomial $a + bx + cx^2$.

What if the output basis was canonical as well ie the output basis is $[w_1 = 1, w_2 = x]$ then the D operator is given by the matrix

$$D \cong \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{1.26}$$

What happens if the output space basis is in canonical form $[w_1 = 1, w_2 = x]$ but the input space basis $=[v_1 = 1, v_2 = 1 + x, v_3 = 1 + x + x^2]$? D operating on these vectors produces $Dv_1 = 0w1 + 0w2, Dv_2 = 1w_1 + 0w_2, Dv_3 = 1w_1 + 2w_2$ and the matrix of D is now given by:

$$D \cong \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix} \tag{1.27}$$

The input polynomial is represented as $\sum_{i=1}^{3} \alpha_i v_i$ ie the given generic polynomial in the old canonical basis $a + bx + cx^2 = av_{1_{old}} + bv_{2_{old}} + cv_{3_{old}} = \sum_1^3 \alpha_i v_i$. While we can work this out to find α_i in terms of a, b, c, let us look at this as:

$$v_1 = v_{1_{old}} \tag{1.28}$$
$$v_2 = v_{1_{old}} + v_{2_{old}} \tag{1.29}$$
$$v_3 = v_{1_{old}} + v_{2_{old}} + v_{3_{old}} \tag{1.30}$$

and so the a, b, c coordinates in the old canonical basis are related to the α_i in the new coordinates as:

$$\left(A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \tag{1.31}$$

In (26) we have the D representation when the input basis was canonical. So therefore if we multiply that matrix representation by A from (31) on the right then we should get the representation of D of (27) in the new input basis. Verify that this is correct.

Finally from (26) verify that the representation of D in (25) is obtained by $Y^{-1}D$ where the Y matrix describes coordinate transformation from canonical to the basis $[w_1 = 1, w_2 = 1 + x]$ used in deriving the representation of D in (25).

$$Y = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{1.32}$$

Through the study of the matrix representation of the derivative operator on polynomials, what we have shown is that the general matrix representation of the operator D can be written as $Y^{-1}D_{canonical}A$ where A, Y represent coordinate transformations and $D_{canonical}$ is the matrix representation of D in (an arbitrary) canonical basis.

11. We presumed in the previous item Y^{-1} exists. For it to exist, we have in general: If T is a bijective transformation from V to W (one to one and onto) then $\exists\, T^{-1}$ such that $T^{-1}Tv = v \,\forall\, V \,\epsilon\, V$ and $TT^{-1}w = w \,\forall\, w \,\epsilon\, W$ (Note: this also says for a linear transformation $\dim(V) = \dim(W)$.)

12. The kernel of a transformation T from V to W is the set of all vectors v_i, such that $Tv_i = 0$ and is denoted $\ker(T)$. The kernel of a linear transformation is a subspace. (Prove this.)

13. $\dim(\ker(T))$ of a linear transformation is called the nullity of the transformation denoted n_T.

14. Let $\mathrm{Im}(T)$ of a linear transformation T from V to W be the subspace in W of vectors such that $w \,\epsilon\, \mathrm{Im}(T) \implies \exists\, v \,\epsilon\, V$ such that $Tv = w$. (Show that $\mathrm{Im}(T)$ is a subspace as claimed.) The rank of the transformation T is $\dim(\mathrm{Im}(T))$ and is denoted r_T.

15. For a linear transformation T from V to W, $r_T + n_T = \dim(\mathrm{V})$.

16. Therefore for a linear matrix equation $T_{n\times p}x = b_{p\times 1}$

    (a) The equations are consistent if $r_{[T\ b]} = r_T$.

    (b) If $p = n$, the equations have a unique solution if $r_T = n$

    (c) If the equations are consistent, the number of free variables in the set of unknowns is $n_T = p - r_T$

    Matlab/Octave have the rank function built into them and compute it using a decomposition called singular value decomposition. $[U, S, V] = \mathrm{svd}(A) \implies USV' = A$. If A is real or complex the S matrix (non-square if A is non-square) is a diagonal matrix with real elements (≥ 0) along its diagonal called the singular values and the number of non-zero elements in this diagonal is the rank of the matrix A. The matrices U, V in the above are normalized by Matlab/Octave and their columns are called singular vectors. The normalization is done such that the sum of squares of the singular vector elements is 1. The singular vectors are such that the matrix product of two different singular vectors in U or V written as $u_i'u_j, v_i'v_j = 0$ if $i \neq j$.

17. **Assignment Problem:** Show that the determinant of the Vandermonde Matrix of order 3 is given by $(x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$. For what is the Vandermonde matrix and this problem see Wikipedia article on this and understand the Linear Mapping proof in it. Find the basis transformation T that transforms the Vandermonde matrix into the matrix L shown there. (Do not do row column operations etc. Understand the linear transformation (mapping) proof there and fill the details for a 3×3 Vandermonde matrix) Can you apply this result to calculating the determinant for the problem in item 7 in Section 5.3 (Fields (Equation(23))).

18. **Example on linear equations and rank condition testing:** Consider $Ax = b_i, i = 1, 2$ given by:

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 2 & -1 & -1 \\ 1 & -4 & -2 \end{bmatrix}, b_1 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, b_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \tag{1.33}$$

    Applying the rank test to $[A, b_i], i = 1, 2$ we get:

```
>> A=[1 3 1;2 -1 -1;1 -4 -2]
A =
```

```
1   3   1
2  -1  -1
1  -4  -2

>> b1=[0 1 2]',b2=[0 1 1]'
b1 =

0
1
2

b2 =

0
1
1

>> rank([A b1])
ans = 3
>> rank([A,b2])
ans = 2
>> rank(A)
ans = 2
```

So $Ax = b_1$ is inconsistent. While $Ax = b_2$ is consistent and possible non-unique solutions exist. The input space for the A matrix is 3-D but its rank is 2. Therefore the null space of A has dimension 1. One free variable exists in the solution of $Ax = b_2$. Let us choose arbitrarily x_3 the third component of the vector $x = 1$ and consider obtaining a null space vector as the solution of $Ax = 0$. Using x_3=1, the new matrix equation $Ax = 0$ is equivalent to

$$A = \begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \Longrightarrow \tag{1.34}$$

$$\begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \Longrightarrow \tag{1.35}$$

$$x_1 = 0.2857, x_2 = -0.4286 \quad \text{and} \quad x_1 - 4x_2 = 2 \text{ is automatically satisfied} \tag{1.36}$$

Since the null space of A is 1 dimensional the vector $[0.2857, -0.4286, 1]'$ generates this null space. To generate a particular solution to $Ax = b_2$ consider arbitrarily $x_1 = 0$ and solve for x_2, x_3 as 0.5 and -1.5 respectively. So all solutions for $Ax = b_2$ are now given by $x = [0, 0.5, -1.5]' + \alpha[0.2857, -0.4276, 1]'$, where α is a free variable real number. If you now impose

restrictions such as $\sum_1^3 x_i^2 = 1$ then α can be found numerically as $\alpha = 2.1612$ or 0.5485 (Check this). Alternatively if we seek to minimize $\sum_1^3 x_i^2$, then $\alpha = 1.3458$ (Check this too).

19. **Eigenvalues (λ_i) and Eigenvectors (u_i) of a linear transformation T from U to V both spaces of same dimension** (i.e the matrices associated with T are square) are defined for a matrix of T (with appropriate basis selected in the spaces U, V) as solutions of the corresponding matrix equation $Tu_i = \lambda_i u_i$ where λ_i (the eigenvalues) is a scalar in $\mathbb{C}$ in general. For the matrix associated with T this is equivalent to finding all null space vectors of the transformation $T - \lambda_i E$ (remember E is the identity matrix.) So in general, n eigenvalues exist and at least one eigenvector exists for each unique value of λ_i.

    Example: Consider the D operator between vector spaces of polynomials of degree <=2 with both output space and input space having canonical basis (See item 10 above). The D operator has a matrix representation

$$D = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \tag{1.37}$$

    The eigenvalues of D are 0 repeated thrice. (Matlab/Octave eig(D) will give the eigenvalues.) What about the eigenvectors? To find them purportedly with Octave/Matlab we can call the function eig again as follows and the V below contains if eigenvectors exist for each eigenvalue as column vectors while the diagonal matrix has the eigenvalues on its diagonal.

```
>> [V,Lambda]=eig(D)
V =

1.0000  -1.0000   1.0000
0        0.0000  -0.0000
0        0        0

Lambda =

Diagonal Matrix

0   0   0
0   0   0
0   0   0

>>
```

    Are the eigenvectors correct as obtained ? One suspects not given how the columns in V look. How can we determine how many eigenvectors exist for the eigenvalue 0. We need to look at rank of $D - 0E = D$. The rank of D is 2. So the null space dimension of $D - 0E$ is 1. So only one eigenvector exists for the eigenvalue 0.

Example: Consider A in Equation (33).

```
[V,D]=eig(A)
V =

   -0.3792   0.8043   0.2540
    0.4366   0.5239  -0.3810
0.8158  -0.2804   0.8890

D =

Diagonal Matrix

   -4.6056e+00              0              0
0                       2.6056e+00     0
0                       0             -6.4351e-17
>> eps
ans = 2.2204e-16
```

The (3,3) element of the diagonal matrix should be treated as 0 as it is smaller than the $|\epsilon|$ precision to which Octave/Matlab computes. (This can be found by asking for *eps* in the command line and is given above as well.) Since the three eigenvalues are unique the corresponding columns in the V matrix above are eigenvectors. The null space eigenvector is column 3. When we compare this vector with the null space vector we found before when solving equations in item 18, we notice that we assumed there $x_3 = 1$ while now we have $x_3 = 0.8890$. Scaling this column 3 by $1/0.8890$ we get the new equivalent null space vector as

```
>> format long e
>> V(:,3)*(1/V(3,3))
ans =

 2.857142857142859e-01
-4.285714285714287e-01
 1.000000000000000e+00
```

which matches what we obtained earlier as the null space vector. Note that when Matlab/Octave calculate eigenvectors, they always try to normalize them such that for each eigenvector $\sum_1^n x_i^2 = 1$.

20. Some applications of Eigenvalues and Eigenvectors

    (a) A $n \times n$ square matrix with n eigenvectors can be diagonalised if eigenvectors exist by using the eigenvector matrix V from $[V, D] = \text{eig}(A)$ as $D = V^{-1}AV$ or equivalently

$A = VDV^{-1}$. Going either way this can be remembered as $AV = VD$ where D is the diagonal matrix. This diagonalization (when it exists) can be used compute $A^k = VD^kV{-}1$. Since D is a diagonal matrix, D^k is the diagonal matrix with elements to power k from the original matrix D. This leads to one of the 17 dubious ways of computing $e^{At} = I + At + A^2t^2/2!... + A^nt^n/n!... = Ve^{Dt}V^{-1}$.

(b) The Cayley-Hamilton theorem for Matrices. States that if the eigenvalues are found using the characteristic equation $\det(A - \lambda E) = \sum_{i=0}^{n} \alpha_i \lambda^i = 0$ then the characteristic equation is satisfied by the matrix A as well i.e. $\sum_{i=0}^{n} \alpha_i A^i = 0$ matrix . (Prove this assuming diagonalization is possible for A.)

(c) For square matrix A over $\mathbb{C}$, if $A^{*\prime} = A$ then A is a hermitian matrix (if the matrix is over $\mathbb{R}$ then the matrix is symmetric). $*$ denotes complex conjugate operation and many times A' automatically is assumed to be $A^{*\prime}$ for example in Octave/Matlab. if $-A^{*\prime} = A$ then A is skew-symmetric (if A is over $\mathbb{R}$, then anti-symmetric.)

(d) If A_n is symmetric, then n real eigenvalues and n real eigenvectors exist and A is diagonalizable. Further more, with the eigenvector matrix V normalized (and orthogonalized/orthonormalized for repeated eigenvalues) as per Matlab/Octave $V^{-1} = V'$. If A_n is hermitian, then the same result holds with the eigenvalues real and eigenvectors having complex numbers. For an anti-symmetric matrix, eigenvalues are either 0 or imaginary. (Prove these results.)

(e) Consider the quadratic form $f(x_{3\times1}) = 3x_1^2 + 2x_1x_2 + 5x_1x_3 + 0x_2x_3 + 2x_2^2 + 7x_3^2$. This can be represented by the matrix form:

$$x'Ax = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 3 & 1 & 2.5 \\ 1 & 2 & 0 \\ 2.5 & 0 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = z'Dz \tag{1.38}$$

where D is the eigenvalue diagonal matrix associated with A and $z = V'x$ where V is the normalized eigenvector matrix associated with A. Now the quadratic form $f(x)$ can be classified as positive definite, positive semidefinite, indefinite, negative semidefinite or negative definite depending on the signs of the eigenvalues in D.

(f) We discussed the svd decomposition superficially in item 16 above. With the properties there show that the SVD of a matrix A that the non-zero eigenvalues are the square root of the eigenvalues of $A'A$.

(g) Least squares problems: Consider the linear matrix equation: $A_{n\times m}x_{m\times1} = b_{n\times1}$ with $n > m$. The least squares solution which finds x such that the sum of squares of $(Ax{-}b)'(Ax{-}b)$ is minimized (least) is given by $x = (A'A)^{-1}A'b$. (Show this is the solution.) The $A'A$ matrix and its inversion can be done by diagonalization (or by SVD which is often more efficient.)

1.7 Inner Products on Vector Spaces and some applications

1. The hermitian inner product $\langle u, v\rangle$ is a map of $u, v \,\epsilon\, U$ to a scalar $\epsilon\, \mathbb{C}$ with the following properties:

   (a) $\langle u, v\rangle = \langle v, u\rangle^*$

   (b) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle \;\forall\; u, v, w \,\epsilon\, U$

   (c) $\langle \alpha u, v\rangle = \alpha^* \langle u, v\rangle$ where $\alpha \,\epsilon\, \mathbb{C}$.

   (d) $\langle u, u\rangle \geq 0$ and is 0 if and only if $u = 0$.

   Vector spaces with inner products defined on them are called inner product spaces.

   Examples:

```
>> u=[1+j*1,1-j*2].'
u =

1 + 1i
1 - 2i
>> v=[2+j*3,3-j*2].'
v =

2 + 3i
3 - 2i
>> u'
ans =

1 - 1i   1 + 2i
>> u'*v
ans =  12 +  5i
>> v'*u
ans =  12 -  5i
>> u'*u
ans = 7
```

   Consider the space of continuous complex functions ($C[a, b]$) over the (closed) interval $[a, b]$ in $\mathbb{R}$ and define the inner product (verify the 4 properties listed above for this definition only then it can be accepted as an inner product) by $\langle u, v\rangle = \int_a^b u^*(t)v(t)dt$.

2. A norm function is a map of $u \,\epsilon\, U$ to $\mathbb{R}^+ \cup 0$ (Real numbers ≥ 0) which measures the "length" of a vector u satisfies the following properties:

   (a) $||u|| \geq 0 \;\forall\; u \,\epsilon\, U$ and is zero if and only if $u = 0$.

(b) $||u+v|| \leq ||u|| + ||v|| \; \forall \; u, v \; \epsilon \; U$. This is the triangle inequality

(c) $||\alpha u|| = |\alpha| ||u||$ where $\alpha \; \epsilon \; \mathbb{C}$ or $\mathbb{R}$ the underlying field of the vector space.

3. In an inner product space $\sqrt{\langle u, u \rangle}$ is a measure of the length of the vector and it is a norm. Verify the 3 properties above to show that the square root of an inner product is a norm on u when both input vectors are the same u.

4. The Cauchy-Schwarz inequality states: $|\langle u, v \rangle| < ||u|| ||v||$. An outline of a proof follows: consider unit vectors a, b and the inner product expansion of $||a - \alpha b||^2 \geq 0$ where α is a complex number taken as $\langle b, a \rangle$. This will reduce to $1 - \alpha\alpha^* = 1 - |\alpha|^2 \geq 0$. Hence $|\langle a, b \rangle| < 1$. Replace a, b with $\frac{u}{||u||}, \frac{v}{||v||}$ then we get the Cauchy-Schwarz inequality.

5. Two vectors u, v are defined to be orthogonal if $\langle u, v \rangle = 0$. The vectors are orthonormal if $||u||, ||v||$ are 1 and their inner product is 0.

6. Given independent vectors (usually a basis) $\{v_i, i = 1, 2, ..., n\}$ one can extract a set of orthonormal vectors $\{u_i, i = 1, 2, ..., n\}$ by the following Gram-Schmidt orthogonalization procedure:

$$u_1 = \frac{v_1}{||v_1||} \tag{1.39}$$

$$x_2 = v_2 - \langle u_1, v_2 \rangle u_1, \quad u_2 = \frac{x_2}{||x_2||} \tag{1.40}$$

$$x_3 = v_3 - \sum_{i=1}^{2} \langle u_i, v_3 \rangle, \quad u_3 = \frac{x_3}{||x_3||} \tag{1.41}$$

$$x_n = v_n - \sum_{i=1}^{n-1} \langle u_i, v_n \rangle, \quad u_n = \frac{x_n}{||x_n||} \tag{1.42}$$

From each vector v_i we retain only the new "perpendicular" information ion v_i which is not contained int the unit vectors $\{u_k, k = 1, 2..., i-1\}$ in x_i and normalize x_i to generate u_i. The $x_i, i = 2, ..., n$ are called the innovations sequence of the set $\{v_i\}$.

7. An easy way of computing Gram-Schmidt orthonormal vectors is v_i are n-dimensional vectors over $\mathbb{R}$ or $\mathbb{C}$ is to put them into a matrix A as columns and compute in Matlab/Octave using the qr routine. (The qr routine is extensively used in the internals of Octave/Matlab and is part of BLAS and LAPACK and used in finding eigenvalues and eigenvectors). The columns of matrix Q obtained from the qr routine is the Gram-Schmidt orthogonalization of the vectors in A. Verify this by applying the Gram-Schmidt orthogonalization above to the 3×3 matrix A below column by column and incidentally explain the matrix R as well.

```
>> A=rand(3)
A =

7.6732e-01   9.9081e-01   5.2790e-01
8.7911e-03   8.9559e-01   2.0659e-01
```

```
5.9712e-01   6.6045e-01   4.4932e-01

>> [Q,R]=qr(A)
Q =

-7.8916e-01  -5.3184e-02  -6.1188e-01
-9.0413e-03  -9.9513e-01   9.8157e-02
-6.1412e-01   8.2994e-02   7.8484e-01

R =

-0.9723  -1.1956  -0.6944
 0       -0.8891  -0.1964
 0        0        0.0499

>> Q'*Q
ans =

1.0000         0         0.0000
0              1.0000   -0.0000
0.0000        -0.0000    1.0000

>> Q*R-A
ans =

 1.1102e-16  -3.3307e-16   1.1102e-16
-1.7347e-18   1.1102e-16   1.1102e-16
-1.1102e-16  -2.2204e-16            0
```

8. **Assignment:** The gain of a linear transformation matrix A is defined as $\max \frac{||Ax||}{||x||}$. Characterize this gain definition in terms of the singular values of A.

Chapter 2

Applied Probability

Appendix A of B.D.O Anderson and J.B. Moore "Optimal Filtering" is a brief review of Probability Theory. We will follow that appendix. I am not duplicating that material here. However, I will solve problems in the following to illustrate the details of the condensed theory there and as needed provide some sketch of proofs of the material there.

2.1 Problems on Bayes Rule:

Bayes Rule states $P(A|B) = P(A \cap B)/P(B) = P(B|A)P(A)/P(B)$. While elementary as it appears, it has profound uses.

1. Consider the binary variable - a person has a specific disease $x = 1$, or does not have the disease $x = 0$. The probability of a person having the disease is 2%.

   Consider a binary variable test for the disease $t = 1$ if the person tested has a positive outcome on testing and $t = 0$ if the person tested does not have a positive outcome on testing. The test is claimed to be 95% reliable.

   A person took the test and the outcome is positive. What is the probability the person has the disease?

   To "confirm", the person duplicated the test again and it came out positive again. What is the probability now that the person has the disease?

   The data: $P(x = 1) = 0.02, P(x = 0) = 0.98, P(t = 1|x = 1) = 0.95, P(t = 0|x = 0) = 0.95 \implies P(t = 0|x = 1) = 0.05, P(t = 1|x = 0) = 0.05$

   We are interested in computing for the first part $P(x = 1|t = 1)$. By Bayes rule this is given by $P(x = 1|t = 1) = P(t = 1|x = 1)P(x = 1)/P(t = 1) = 0.95 \times 0.02/P(t = 1)$ We can obtain $P(t = 1) = P(t = 1|x = 1)P(x = 1) + P(t = 1|x = 0)P(x = 0) = 0.95 \times 0.02 + 0.05 \times 0.98 = 0.068$. So $P(x = 1|t = 1) = 0.2794$.

   For the confirmation part by Bayes rule: $P(x = 1|t_1 = 1, t_2 = 1) = P(t_1 = 1, t_2 = 1|x = 1)P(x = 1)/P(t_1 = 1, t_2 = 1)$. Now assuming the events $t_1 = 1|x = 1 or 0$ is independent of $t_2 = 1|x = 1 or 0$ we have, $P(t_1 = 1, t_2 = 1) = [P(t = 1|x = 1)]^2 P(x = 1) + [P(t = 1|x =$

$0)]^2 P(x=0) = 0.95^2 * 0.02 + 0.05^2 * 0.98$. So $P(x=1|t_1=1,t_2=1) = 0.95^2 * 0.02/(0.95^2 * 0.02 + 0.05^2 * 0.98) = 0.8805$. If one were to test 4 times and all 4 results come out positive then the repeated testing confirms the disease at a Probability value = 0.9996.

Assignment: Look at

`https://covid19-sciencetable.ca/wp-content/uploads/2022/02/Use-of-Rapid-Antigen-Tests-during-the-Omicron-Wave_published_20220211.pdf`

Now for the omicron covid 19 rapid antigen test (this is test available in for example Shoppers Drug Mart.), estimate if the test outcome is positive in the first test what is the probability a person has omicron covid. Assume an estimate of 3% probability of omicron covid in the population in Ontario as of now. What happens if this probability is 10 or 20%?

In the same article the delta virus rapid antigen detection sensitivity is available. Calculate using that how many times should one be tested to achieve a probability of having the disease given positive tests to be greater than 0.95.

Note: Your results does not mean that the rapid antigen test is useless for omicron covid as a graph in the paper Fig. 3 indicates that the tests give 80% sensitivity for example on day 3 and 100% on days 4 and 7. It is the early detection that is the issue!

2. A fraction p of memory chips are faulty with $p = 0.1$ ie. the probability of a good memory chip is $1-p$ and a faulty chip is p. The probability that a good chip will work after testing to time t from time (t,∞) is $e^{-\alpha t}$. the probability that a bad chip (which may work initially but will quickly fail after testing from (t,∞) is $e^{-1000\alpha t}$ where $\alpha = 2\times 10^{-5}$ and t is in hours.

   It is desired that we test for a time t the chips so that the probability that the chip shipped is good is 0.99. Find the time t for testing to ensure this?

   Let T the event that the chip is functioning after t hours of tests. Let G and B denote the events chip is good or chip is bad.

   We want $P(G|T) = 0.99$. We are given $P(T|G) = e^{-\alpha t}, P(G) = 0.9, P(T|B) = e^{-1000\alpha t}, P(B) = 0.1$. So applying Bayes rule we have, $P(G|T) = 0.99 = P(T|G)P(G)/P(T)$
   $= P(T|G)P(G)/(P(T|G)P(G) + P(T|B)P(B)) = \frac{0.9e^{-\alpha t}}{0.9e^{-\alpha t}+0.1e^{-1000\alpha t}}$. Solving this equation for t we get 120 hours (approx). ie. 5 days of testing of each chip!

2.2 Distribution and density functions

1. For the chip problem above, if the chip is a good chip the probability of the random variable which is the lifetime of a chip is $\leq t$ is given by $F_X(0 \leq X \leq t) = 1-e^{-\alpha t}$. For negative lifetimes the probability is zero and I am not stating that here. Therefore its pdf is $f_X(t) = \alpha e^{-\alpha t}\ \forall\ t \geq 0$.

   The average lifetime of a good chip then is given by $\int_0^\infty t f_X(t) dt$. We compute this integral in wxMaxima as:

(% i1) assume(α>0);

$$[\alpha > 0] \qquad (\%\ o1)$$

(% i3) integrate(t*α*%e^ (-α*t),t,0,inf);

$$\frac{1}{\alpha} \qquad (\%\ o3)$$

So for a good chip the average (mean) lifetime in hours is 1/2e-5 = 50000 hours. While for a bad chip the average lifetime in hours is 1/(2e-5*1000) =50 hours (50000 hours is slightly less than 6 years for a good chip and 50 hours is just about 2 days for a bad chip). (To assure ourselves that the chip after testing is a good one with probability of 0.99 (confidence level) we need to test for 5 days from the results above.)

The variance (σ^2) of the lifetime of good chips is $\int_0^\infty \left(t - \frac{1}{\alpha}\right)^2 f_X(t)dt$. From wxMaxima we get this variance as: $\sigma^2 = \frac{1}{\alpha^2} = 2.5 \times 10^9$ hours^2. Note the standard deviation $\sigma = \frac{1}{\alpha}$ hours. While for the bad chips the variance is 2.5×10^3 hours^2.

2. Chebyschev's inequality (Pg 313 of Appendix A of Anderson and Moore) states that $P(|t - \frac{1}{\alpha}| > \frac{3}{\alpha}) \leq 1/9 = 0.111$ no matter what the underlying distribution is - we just used the average and variance values in Chebyschev's inequality to estimate this probability.

   Knowing that the distribution of lifetimes is an exponential distribution, the above probability value can be calculated as: $\int_{\frac{4}{\alpha}}^\infty \alpha e^{-\alpha * t} = e^{-4} \approx 0.0183$.

   Chebyschev's inequality is exactly that an inequality. Its power is independence of the distribution.

3. Consider the stock price (closing) data on Hydro One (quotemedia.com) for a set of days. The average and standard-deviation of the return on each day can be calculated quickly and from that by Chebyschev we can answer that the price average + 3*std deviation or average - 3*std deviation can only exceed by 0.1111 no matter how the price is distributed.

   What one has to be careful on is that we are using this to forecast but there is no concept of forecasting in this probability measure.

4. **Markov's Inequality** states Given a RV $X \epsilon [0, \infty)$ then $\Pr(X \geq a) \leq \frac{E(X)}{a}$. The proof is fairly simple: For any given $a > 0$ let RV I be defined by $I = 1$ if $X \geq a$ else 0 then $I \leq \frac{X}{a}$ and $E(I) = \Pr(X \geq a) \leq \frac{E(X)}{a}$. This bound is a weak bound but it leads to another bound which is stronger viz. Chernoff bounds.

5. **Chernoff Bounds** are derived from Markov's inequality by applying the inequality to e^{tX} and leads to (**Assignment:** Show the results below:)

$$\Pr(X \geq a) = \Pr\left(e^{tX} \geq e^{ta}\right) \leq M_X(t)e^{-ta} \qquad (t > 0) \qquad M_X(t) = E(e^{tX}) \qquad (2.1)$$

$$\Pr(X \leq a) = \Pr\left(e^{tX} \geq e^{ta}\right) \leq M_X(t)e^{-ta} \qquad (t < 0) \qquad (2.2)$$

Note: $M_X(t)$ is the moment generating function (MGF) and we will look at this again later in Section 2.7 and tie it with Laplace transform of the pdf and Fourier transform of pdf etc. as well as why it is called Moment generating function.

Since t is selectable in the above, the inequalities are written as:

$$\Pr(X \geq a) \leq \inf_{t>0} M(t)e^{-ta} \text{ and } \Pr(X \leq a) \leq \inf_{t<0} M(t)e^{-ta} \tag{2.3}$$

where inf is the infimum of the expression to the right and is the lowest limit of it as t is varied to satisy the respective requirements of $t > 0$ or $t < 0$. Chernoff bounds are Equation 2.3 above.

6. We apply Chernoff bounds to the random variable X which is the sum of n independent, identically distributed (iid) Poisson (binary) random variabless $Y_i, i = 1, 2, ..., n$ defined by $\Pr(Y_i = 1) = p$ else $\Pr(Y_i = 0) = 1 - p$. The moment generating function $M_{Y_i}(t) = e^{t\times 0}(1 - p) + e^{t\times 1}p = 1 - p + e^t p = p(e^t - 1) + 1 < e^{p(e^t-1)}$. (To prove this consider: $p(e^t - 1) = \beta$ then $p(e^t - 1) + 1 = 1 + \beta < e^\beta$. The following Octave commands and the generated plot (You can change β range in it) should convince you of the result and help visualize this. The result shows up in a surprising number of places (Information theory is one such place) and is useful in itself.

```
beta=[-1:0.01:1];plot(beta,1+beta,beta,exp(beta));grid;
```

(For a formal proof proceed by defining $f(y) = e^y - (1 + y)$. Find the minimum of this function and show that the min is unique and the second derivative of $f(y) > 0 \,\forall\, y$.)

Since X is the sum of the iid Y_i its average is $E(X) = \mu = np$ (**Assignment:** Show this) and its MGF is given by $M_X(t) = \Pi_{i=1}^n M_{Y_i}(t) < \Pi_{i=1}^n e^{p(e^t-1)} = e^{\mu(e^t-1)}$.

Now consider computing a bound for $\Pr(X \geq (1 + \delta)\mu)$ with $\delta > 0$. Applying Chernoff bound we can obtain an inequality for this as:

$$\Pr(X \geq a) \leq \inf_{t>0} M_X(t)e^{-ta} \text{ with } a = (1 + \delta)\mu \quad \leq \inf_{t>0} e^{\mu(e^t - t(1+\delta)-1)} \tag{2.4}$$

The minimum that we seek through the inf will be attained when $e^t - t(1 + \delta) - 1$ is minimum as $\mu > 0$ (We will have to check if we use the calculus that a) $t > 0$ at this minimum and that b) this is a minimum (second derivative > 0). We do this in wxMaxima.

(% i3) y:%e^ t-(1+δ)*t-1;

$$-t\ (\delta + 1) + \%e^t - 1 \tag{\% o3}$$

(% i4) solve(diff(y,t),t);

$$[t = \log(\delta + 1)] \tag{\% o4}$$

(% i5) subst(%o4,y);

$$\delta - (\delta + 1)\log(\delta + 1) \qquad (\%\ o5)$$

(% i6) subst(%o4,diff(y,t,2));

$$\delta + 1 \qquad (\%\ o6)$$

The result above in (%o5) from wxMaxima then says that the bound is now given in Equation (2.5) and both conditions a,b referred above are satisfied as $\delta > 0$.

$$\Pr(X \geq (1+\delta)\mu) \leq \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu} \qquad (2.5)$$

A similar but different bound is available for $\Pr(X \leq (1-\delta)\mu)$ and for its details see M. Mitzenmacher and E. Upfal - "Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis", 2nd Edition, Theorem 4.5 Page 71.

Let us apply what has been derived above for $\delta = 2$ and so if we have 6 Poisson RVs (Coin tosses with $p = 0.5$) then $\mu = 3$ and the bound $= 0.2737^{\mu} = 0.0205$.

The variance of a Poisson random variable (Binary variable) is $p(1-p)$ and the variance of their iid sum is $np(1-p)$. For the case of 6 trials with $p = 0.5$ the variance is 1.5. The comparable bound from Chebyschev's inequality is with K in Anderson & Moore Appendix A pg 313 set at 2μ we get this probability estimate to be 0.0208 (assuming symmetric distributon). The Chernoff bound is marginally tighter in this case. In general my experience is Chernoff bounds require a lot of work to apply correctly and Chebyschev often is quicker although not as tight.

2.3 Needle Problem

The problem: An infinite grid of parallel (thin) wires has been strung with a uniform distance of unit length between them. A needle of length $l \geq 2$ units is dropped to land on these wires (both ends land on the grid simultaneously and the needle is rigid and massless etc.). What is the probability that the needle will not fall between two wires and slip through the sieve. Verify by simulations that the probability calculated is correct using a "good" uniform random number generator.

Solution: The geometry and coordinates of the red needle falling on the black wires is shown in Fig. 1. The centre of the needle of is marked x and it is a uniform random variable which can take values in the interval [0 1). Since the grid is infinite, we need to consider only this interval as the centre can be reduced modulo the unit length and brought to within this interval across a pair of wires. The angle marked θ is the orientation fo the needle on the wires and is another uniform random variable with values in the interval [0 pi). The random variable θ is independent of the random variable x. The probability density function of the joint distribution of the random variables is given by

$$p(x,\theta) = p(x)p(\theta) = \frac{1}{1}\frac{1}{\pi} = \frac{1}{\pi} \qquad (2.6)$$

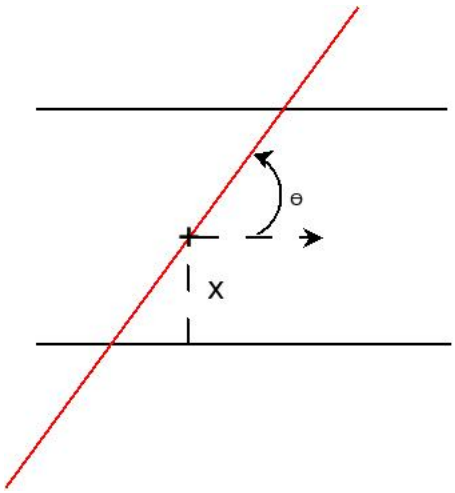

Figure 2.1: Needle and a pair of parallel wires - definition of random variables

The needle will not slip through between the wires if

$$x - \frac{l}{2}\sin(\theta) < 0 \text{ and } x + \frac{l}{2}\sin(\theta) > 1 \tag{2.7}$$

The first and second inequalities are satisfied if

$$\sin^{-1}\left(\frac{2x}{l}\right) < \theta < \pi - \sin^{-1}\left(\frac{2x}{l}\right) \text{ and } \sin^{-1}\left(\frac{2(1-x)}{l}\right) < \theta < \pi - \sin^{-1}\left(\frac{2(1-x)}{l}\right) \tag{2.8}$$

To visualize these conditions let us consider two cases: random variable $x < 0.5$ and random variable $x > 0.5$ viz. consider $x = 0.3$ and $x = 0.7$ with $l = 3$. Let us solve for θ_m that just satisfies these inequalities by plotting $\sin(\theta)$ and $2(1-x)/l$ and $2x/l$ respectively for these cases. With $x = 0.7$ the first inequality limits possible θ while in the case $x = 0.3$ the second (right side) inequality limits. In both cases the same bound operates viz from $0.4855 < \theta < \pi - 0.4855$. What does this imply if we go from $0 \leq x \leq 0.5$ then the right side inequality in (2) is bounding while for $0.5 < x < 1$ the lower side inequality in (1) is limiting but otherwise the region of the angle for which there is satisfaction of the inequality is the same and so we can calculate the required probability as:

$$2\int_0^{\frac{1}{2}} \int_{\sin^{-1}\left(\frac{2(1-x)}{l}\right)}^{\pi-\sin^{-1}\left(\frac{2(1-x)}{l}\right)} p(x,\theta)d\theta dx = \frac{2}{\pi}\int_0^{\frac{1}{2}} \left[\pi - 2\sin^{-1}\left(\frac{2(1-x)}{l}\right)\right] dx \tag{2.9}$$

Now from wxmaxima (commands are: assume($l > 0$);integrate(asin((1-x)*2/l), x, 0, 1/2);)

$$\int_0^{\frac{1}{2}} \sin^{-1}\left(\frac{2(1-x)}{l}\right) dx = \frac{\sqrt{l^2-4} + 2\sin^{-1}\left(\frac{2}{l}\right)}{2} - \frac{\sqrt{l^2-1} + \sin^{-1}\left(\frac{1}{l}\right)}{2} \tag{2.10}$$

The overall probability is now given by

$$\frac{2}{\pi}\left[\frac{\pi}{2} + \sqrt{l^2-1} + \sin^{-1}\left(\frac{1}{l}\right) - \{\sqrt{l^2-4} + 2\sin^{-1}\left(\frac{2}{l}\right)\}\right] \tag{2.11}$$

Simulation code for this problem in Octave/Matlab:

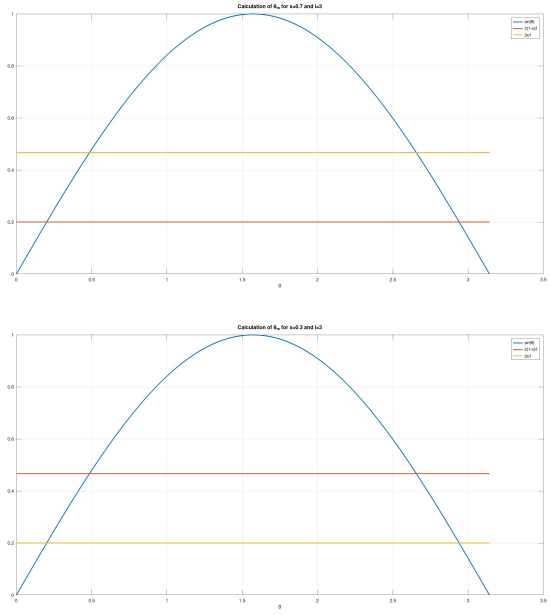

Figure 2.2: Solutions for the θ for inequalities in (2)

```
function [p,pt]=needle(N,l)
p=0;
for i=1:N
x=rand();
theta=rand()*pi;
if (x+0.5*l*sin(theta)>1) && (x-0.5*l*sin(theta) < 0) p=p+1;end;
end
p=p/N;
pt=(2/pi)*(pi/2+sqrt(l^2-1)+asin(1/l)-sqrt(l^2-4)-2*asin(2/l));
```

Some runs of the code:

```
>> [p,pt]=needle(1000000,2)
p = 0.4365
pt = 0.4360
>> [p,pt]=needle(1000000,3)
p = 0.6645
pt = 0.6643
>> [p,pt]=needle(1000000,4)
p = 0.7546
pt = 0.7545
>> [p,pt]=needle(1000000,5)
p = 0.8057
pt = 0.8057
```

```
>> [p,pt]=needle(1000000,4.5)
p = 0.7834
pt = 0.7831
>> [p,pt]=needle(1000000,10)
p = 0.9038
pt = 0.9041
```

2.4 Some Discrete Probability Problems

1. Factory A produces 70% of umbrellas with a defective rate of 2% (irrespective of color). A produces 25% of all blue umbrellas. Factory B produces the remaining umbrellas all of blue color with a defective rate of 1%. What is the defective rate of an umbrella purchased from the collective set of umbrellas produced by these two factories? If an umbrella is defective what is the probability that it was produced by Factory B ? Given that it is a blue umbrella that is defective what is the probability it was produced by Factory B ? (An example of Bayes Rule)
   Soln: Let Defective umbrella event be D and Defective and Blue umbrella event be DB. Then:

$$P(D) = P(D|A)P(A) + P(D|B)P(B) = 0.02 \times 0.7 + 0.01 \times 0.3 = 0.017 \tag{2.12}$$

$$\begin{aligned} P(DB) &= P(DB|A)P(A) + P(DB|B)P(B) = P(D|B)P(B|A)P(A) + P(D)P(B) & (2.13)\\ &= 0.02 \times 0.25 \times 0.7 + 0.01 \times 0.03 = 0.0065 & (2.14) \end{aligned}$$

$$\begin{aligned} P(B|D) &= \frac{P(D|B)P(B)}{P(D)} = \frac{0.01 \times 0.3}{0.017} = 0.1765 & (2.15)\\ P(B|DB) &= \frac{P(DB|B)P(B)}{P(DB)} = \frac{0.01 \times 0.3}{0.0065} = 0.4615 & (2.16) \end{aligned}$$

2. A tournament of two players is won if one player wins n games. Player A has probability p of winning a game over player B. (Either A or B must win a game - no draw) What is the probability that player A looses the tournament after winning k games ?
   Soln: The probability of Player B winning a game is $(1-p)$. The tournament ends when B wins the last game out of $n+k$ games. Therefore in a total of $n+k-1$ games prior to this last game A has won k games, B has won $n-1$ games and the final game is won by B and so the probability of A loosing after winning k games is

$$\left(\begin{array}{c} n-1+k \\ k \end{array} \right) p^k (1-p)^{n-1}(1-p) \tag{2.17}$$

3. A bit x (0 or 1) is transmitted through N relaying stations to you. Each relaying station may flip the bit with a probability p. What is the probability that the bit received by you is x.
   Soln: Let the probability that the bit is $\bar{x}$ at relaying station $n-1$ be denoted by P_{n-1} (the

probability that the bit is x at station $n-1$ is 1-P_{n-1}), then the probability that after going through the relaying station n you get x is given by:

$$1 - P_n = P_{n-1}p + (1 - P_{n-1})(1 - p) \Rightarrow P_n = (1 - 2p)P_{n-1} + p \tag{2.18}$$

The recursive equation has boundary condition $P_0 = 0$ and holds for $n \geq 1$. To solve for a closed form solution, we unroll the recursive equations and sum to get

$$P_N = (1 - 2p)^N P_0 + p \sum_{0}^{N-1} (1 - 2p)^i = \frac{1}{2}(1 - (1 - 2p)^N) \tag{2.19}$$

The probability that we want after going through n relaying stations is $1-P_N = \frac{1}{2}(1+(1-2p)^N)$.

4. We flip a coin which has a probability of landing heads = p. We count the number of flips before we get the first head. What is the probability of the number of flips (the random variable X for this problem) = n ? What is the average value of this random variable ie. $E(X)$?
Soln: $P(X = n) = (1 - p)^{n-1}p$. This is called a geometric random variable in the literature. Its expectation is given by:

$$E(X) = \sum_{n=1}^{\infty} n(1 - p)^{n-1}p = p \sum_{n=1}^{\infty} n(1 - p)^{n-1} = -p\frac{d}{dp} \sum_{n=1}^{\infty} (1 - p)^n = \frac{1}{p} \tag{2.20}$$

5. The probability of two series resistors A, B failing are 0.2 and 0.3 respectively. The failure events are mutually independent. What is the probability that no current flows in the series circuit made up of these two resistors?

   $P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B) = 0.5 - 0.06 = 0.44$

   The probability that current flows in the series circuit is 1-0.44 = 0.56. Note the inclusion-exclusion principle used in calculating the failure of current flow in the series circuit.

   If the resistors are in parallel- No current flow in the circuit requires both resistor must have failed and the probability is 0.06. The probability of current flow in the parallel circuit therefore 0.94.

   Can you try to compute the probability of current flow in either circuits directly instead of using 1-P(no current flow).

2.5 Joint, Marginal and Conditional pdf problem:

1. The random variables X, Y have pdf $f_{X,Y}(x, y) = 6xy\ 0 < x < 1, 0 < y < \sqrt{(x)}$ and 0 elsewhere . Find after checking if this is a valid pdf:

   (a) $f_X(x)$ and $f_Y(y)$.

   (b) Are X and Y independent ?

   (c) The conditional probability $f_{X|Y}(x|y)$.

(d) $E\{X|Y=y\}$ and $E\{X|Y=0.5\}$.

(e) The Var$\{X|Y=y\}$ and Var$\{X|Y=0.5\}$.

Validity: $\int_{x=0_+}^{x=1_-}\int_{y=0_-}^{y=\sqrt{x}} 6xydxdy = \int_{x=0_+}^{x=1_-} 3xy^2|_0^{\sqrt{x}}dx = \int_{x=0_+}^{x=1_-} 3x^2dx = 1.$

To visualize the area for the pdf so the various limits for each of the items to be found is clear, I have the following sketch generated by Octave code below:

```
x=[0:0.001:1];y=sqrt(x);area(x,y);
```

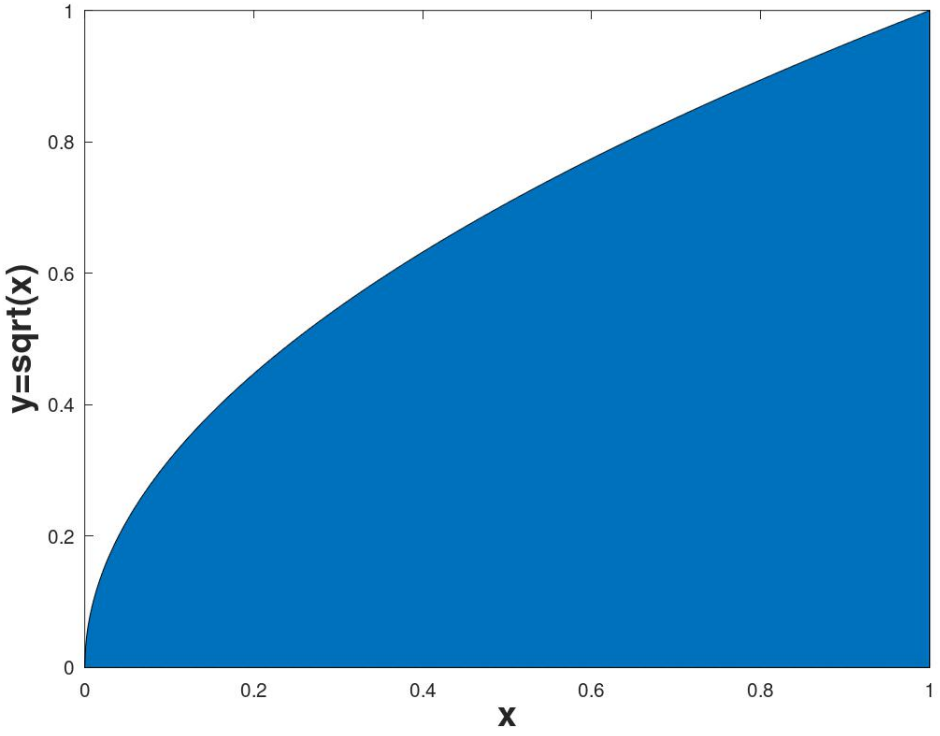

Figure 2.3: Shaded area is under consideration as the domain of the pdf

To find $f_X(x)$, x is fixed and y can vary in the blue region of the domain vertically between 0 and $\sqrt{x}$. Hence $f_X(x) = 6x\int_{y=0}^{y=\sqrt{x}} ydy = 3x^2$. To find $f_Y(y)$, y is fixed and in the blue area of the domain, x varies from y^2 to 1. Hence $f_Y(y) = 6y\int_{x=y^2}^{x=1} xdx = 3y(1-y^4)$. To check these results (only a confirmatory test that integrations have not been done wrong we can compute $\int_{y=0}^{y=1} f_Y(y)dy$ and verify it is 1. Performing this we get the integral as $(1.5y^2 - 0.5y^6)|_0^1 = 1$. You should check $f_X(x)$ for this confirmatory test.

if X, Y are independent then $f_{X,Y}(x,y)(= 6xy)$ should be equal to $f_X(x)f_Y(y)(= 9x^2y(1-y^4))$ in the domain and they are not equal. So X, Y are not independent.

The conditional probability $f_{X|Y}(x|y) = f_{X,Y}(x,y)/f_Y(y) = 2x/(1-y^4)$. Applying our confirmatory test we have: $\int_{x=y^2}^{1} f_X(x|y)dx = 1$. Note the lower limit on x is y^2 from the domain.

$E\{X|Y=y\} = \frac{\int_{x=y^2}^{1}(2x^2)dx}{1-y^4} = \frac{2(1-y^6)}{3(1-y^4)}$. To obtain $E\{X|Y=0.5\}$ substitute $y=0.5$ in the expression to get 0.7.

To compute the variance ($\mathrm{Var}\{X|Y=y\}$) One strategy could be find $E\{X^2|Y=y\}$ and then use this to compute variance as the difference between this and the square of the mean (average) found earlier. Now $E\{X^2|Y=y\} = \frac{\int_{x=y^2}^{1}(2x^3)dx}{1-y^4} = \frac{1-y^8}{2(1-y^4)}$. The variance is now given from wxMaxima as (and its numerical value when y=0.5 is also found from wxMaxima):

```
(% i1)  (1/2)*(1-y^ 8)/(1-y^ 4) - (2*(1-y^ 6)/(3*(1-y^ 4)))^ 2;
```

$$\frac{1-y^8}{2\left(1-y^4\right)} - \frac{4(1-y^6)^2}{9(1-y^4)^2} \qquad (\%\ o1)$$

```
(% i2)  ratsimp(%o1);
```

$$\frac{y^8+2y^6-6y^4+2y^2+1}{18y^4+36y^2+18} \qquad (\%\ o2)$$

```
(% i3)  subst(y=0.5,%o2);
```

$$0.04125 \qquad (\%\ o3)$$

2.6 Probability density function of a function of a random variable

Given the pdf of the two sided exponential distributions (in the memory chip testing problem we considered the one sided exponential distribution) $f_X(x) = 0.5\alpha e^{-\alpha|x|}$, find the pdf of the random variable $Y = X^2$.

The required pdf is given by

$$f_Y(y) = \frac{f_X(x_1 = +\sqrt{y})}{|dy/dx|_{x_1=+\sqrt{y}}} + \frac{f_X(x_2 = -\sqrt{y})}{|dy/dx|_{x_2=-\sqrt{y}}} = \frac{\alpha e^{-\alpha\sqrt{y}}}{2\sqrt{y}} \qquad (2.21)$$

Assignment: Given the normal distribution of the RV X as $f_X(x) = N(0,1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, find the distribution of the chi-square distribution (with 1 degree of freedom) $Y = X^2$. Plot the resultant pdf and compare with the results of the pdf on the web.

2.7 Moment generating function (MGF) & characteristic function of a random variable

The MGF of a random variable X is defined as $M_X(t) = E\{e^{tX}\}$. The Characteristic function of Anderson and Moore Appendix A is the MGF with $t = js$ where j is $\sqrt{-1}$. With t of MGF as $-s$ we are computing the Laplace transform (2 sided) of the pdf of X. While the Characteristic function is the Fourier transform of the pdf.

For independent random variables, the MGF/Characteristic function of their joint distribution is the product of their respective MGFs/Characteristic functions. (**Assignment:** Prove this result.)

The main result for MGF is that all the moments of a random variable (k^{th} moment $m_k = E\{X^k\}$) can be computed from the MGF as $m_k = \frac{d^k M(t)}{dx^k}|_{t=0}$. This follows from the standard expansion of the exponential in an infinite series.

Example: Find the MGF of the double-sided exponential random variable X, with pdf $= \frac{\alpha}{2}e^{\alpha|x|}$ and hence obtain its kurtosis.

The double-sided exponential random variable is zero mean. Therefore its kurtosis is given by the ratio of the 4th (central) moment to the square of the 2nd (central) moment (square of its variance.)

The MGF is by definition given by

$$\int_{-\infty}^{\infty} e^{tx}\frac{\alpha}{2}e^{\alpha|x|}\frac{\alpha}{2}e^{\alpha|x|}dx = \frac{\alpha}{2}\left[\int_{0}^{\infty} e^{(t-\alpha)x}dx + \int_{-\infty}^{0-} e^{(t+\alpha)x}dx\right] = \frac{\alpha^2}{\alpha^2 - t^2} \quad -\alpha < \text{real}(t) < \alpha \tag{2.22}$$

Now kurtosis can be found in wxMaxima as below:

```
(% i2)  subst(t=0,diff(α^ 2/(α^ 2-t^ 2),t,2));
```

$$\frac{2}{\alpha^2} \tag{\% o2}$$

```
(% i3)  subst(t=0,diff(α^ 2/(α^ 2-t^ 2),t,4));
```

$$\frac{24}{\alpha^4} \tag{\% o3}$$

```
(% i4)  kurtosis:ratsimp(%o3/%o2);
```

$$\frac{12}{\alpha^2} \tag{\% o4}$$

Assignment: Look up the MGF (or derive using wxMaxima) of the Gaussian (Normal), zero mean, unit variance normal random variable and use it to compute the kurtosis of this Gaussian RV.

2.8 Markov Chains

We will consider Markov Chains with discrete(integer instants k in 2.23 below) state transitions and the defining characteristic of a Markov chain is

$$\Pr(X_k|\{X_{k-1}, X_{k-2}, ...X_0\}) = \Pr(X_k|X_{k-1}) \tag{2.23}$$

The state transition probabilities of a finite Markov Chain can now be captured in a matrix. The X_k is the state to which the chain can advance from its previous state X_{k-1}. Inherently we consider only a finite number of states (ie. a finite Markov chain).

Let $P = [p_{ij}]_{n\times n} = \Pr(X_j|X_i), i = 0, 2, ..., n-1 \ \& \ j = 0, 2, ..., n-1$ i.e the row number i of the matrix is the current state and the column number j is the state to which transition might occur with probability value$= p_{ij}$. It should be obvious that $\sum_j p_{ij} = 1$ as p_{ii} measures the probability of staying in state i.

The dual structure to the above formulation is to define the transpose of the above matrix as the state transition probability matrix. Albeit rare, our earlier study of eigenvalues/eigenvectors of matrices which are very useful for Markov chains is more natural in this form.

Very often we want to know that if we started in state say 1 and want to go to state 3 in two (2) transitions what is the probability given we are dealing with a Markov chain. This is denoted by $p_{13}^{(2)}$ and is given by $p_{11}p_{13} + p_{12}p_{23} + p_{13}p_{33}$. Generalizing this such probabilities can be obtained by first computing the matrix product P^2 and reading p_{13} from the resultant matrix. Generalizing this, we have,The i, j entry of $P^n = P(P^{n-1})$ gives the probability $p_{i,j}^{(n)}$ that the Markov chain, starting in state i, will be in state j after n steps.

Before going further let us consider the coupon collector problem:

One of 4 distinct coupons are found in boxes (of cereal) with equal probability of any of them being in a box. When one collects all 4 distinct coupons, the manufacturer of the box will give a prize to the client who has all 4 coupons. Typically the manufacturer would like to know what is the average number of boxes a client who wins has to buy ?

The probability of finding one of the four coupons in a box is 0.25. The state transistion diagram in which the state numbers are the number of distinct coupons a customer currently has for this is fairly straight forward and is shown below. The black edges give the probability of transition out of the state while the red self loops give the probability of staying in the current state upon buying a box.

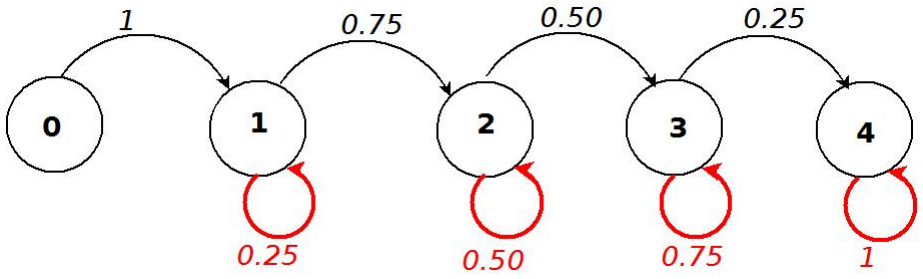

Figure 2.4: Transistion diagram for coupon collector problem with equal probability for coupons.

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.24}$$

Let us find given that the coupon collector started in state 0 that after 5 buying 5 boxes the collector will have all 4 distinct coupons. We are asking for $p_{04}^{(5)}$. To find this we compute as follows:

```
>> P=[0 1 0 0 0;0 0.25 0.75 0 0;0 0 0.5 0.5 0;0 0 0 0.75 0.25;0 0 0 0 1]
P =

0   1.0000         0         0         0
0   0.2500    0.7500         0         0
0         0    0.5000    0.5000         0
0         0         0    0.7500    0.2500
0         0         0         0    1.0000

>> P5=P*P*P*P*P
P5 =

0   0.0039    0.1758    0.5859    0.2344
0   0.0010    0.0908    0.5273    0.3809
0         0    0.0312    0.4121    0.5566
0         0         0    0.2373    0.7627
0         0         0         0    1.0000
```

If you start in state 0, then after 5 boxes bought, the probability that the collector has 4 distinct coupons is row 1, column 5 element of the matrix = 0.2344. The probability that after 5 boxes were bought that the collector has only 1 distinct coupon (i.e has 5 duplicates of one coupon) is 0.0039 etc..

So with powerful computers accessible to all, computing on Markov chains become a problem of clear drawing of transition diagrams with probabilities marked correctly and setting up the transition probability matrix and computing on it.

Assuming that the starting probabilities of the states are described by a row vector u i.e the elements of this vector give the probabilities that one is in state i for each possible state, then the probability vector $u^{(n)}$ of the states after n transition is given by $u^{(n)} = uP^n$.

We follow C.M. Grinstead and J.L. Snell's - "Introduction to Probability" - The Chance Project July 2006 which is distributed in pdf format under GNU license further below on Markov Chains (Chapter 11). we define a Markov chain to have absorbing state if there is no probability of leaving that state ie. $p_{ii} = 1$ for some i defines that state as an absorbing state. For the coupon collector problem above, the state 4 is an absorbing state. A Markov chain is absorbing if it has at lease one state which is an absorbing state and it is possible to go to it from any state. From out state transition diagram for the coupon collector problem we see it is an absorbing Markov chain. In a absorbing Markov chain, a state which is not absorbing is a transient state.

The transistion probability matrix for an absorbing Markov chain by numbering the states properly can be brought to the canonical block matrix form with $E, \mathcal{O}$ being a square identity matrix and possibly rectangular matrix of zeros:

$$P = \begin{bmatrix} \mathcal{Q} & \mathcal{R} \\ \mathcal{O} & E \end{bmatrix} \tag{2.25}$$

For the coupon collector problem the transition probability matrix is already in a canonical form.

We next consider the properties of the canonical form of the transition probability matrix (again from Grinstead and Snell) for an absorbing Markov Chain.

1. $\lim_{n\to\infty} \mathcal{Q}^n = 0$. Since $\mathcal{Q}$ is the transition probability from transient states and the Chain is absorbing, we will eventually leave the transient states and reach the absorbing states. For example consider in the numerics above P^5 has the upper 4×4, $\mathcal{Q}^5$ matrix whose elements are < 1 and if we continue then in the linit $\mathcal{Q}^n = 0$ as $n \to \infty$.

2. Consider the matrix $N = (E - \mathcal{Q})^{-1}$ and the claim is (bfAssignment: Prove this:) $N = I + \mathcal{Q} + \mathcal{Q}^2 + \mathcal{Q}^3 + ...$

3. The elements of N in each row (ie transient state) i give the average number of steps to the transient state j column before entering the absorbing state.

4. The sum of the elements of each row i of N gives the number of steps to reach an absorbing state from state i.

5. The i^{th} row elements of the matrix $N\mathcal{R}$ is the probability of which absorbing state the system started in state i will reach.

Applying the above to our coupon collector problem, we have the following numerical results

```
>> P
P =

0   1.0000         0         0         0
0   0.2500    0.7500         0         0
0         0    0.5000    0.5000         0
0         0         0    0.7500    0.2500
0         0         0         0    1.0000

>> Q=P(1:4,1:4)
Q =

0   1.0000         0         0
0   0.2500    0.7500         0
0         0    0.5000    0.5000
0         0         0    0.7500
>> R=P(1:4,5)
R =

0
0
0
0.2500
```

```
>> N=inv(eye(4)-Q)
N =

1.0000   1.3333   2.0000   4.0000
0   1.3333   2.0000   4.0000
0        0   2.0000   4.0000
0        0        0   4.0000

>> N*[1;1;1;1] % summing the rows of N with column vector of all 1's.
ans =

8.3333
7.3333
6.0000
4.0000
>> N*R
ans =

1
1
1
1
```

Starting at state 0 then it will take 8.3333 cereal boxes that a person has to buy before they will collect all 4 distinct coupons on an average.

We now consider a small variation of the coupon collector problem (from the Final Exam - 2022):

A company organizes a coupon collection program to improve sales with each item bought by customers containing one of 5 coupons. A collector who collects all 5 different coupons will be given a prize. However, the company has organized the coupons in such a way that the probability of coupons 1 to 4 in an item are the same while that of coupon 5 is $\frac{1}{8}$. What is the expected number of items that a collector must buy before the customer has all 5 different type of coupons to claim the prize.

A state transition diagram for this problem with state numbers is shown in Figure below (code: 1000,0 denotes coupon type 1 has been found and all other coupons (3 successive zeros) with equal probability are not yet obtained and ,0 denotes the special coupon 5 which has not been obtained. This coupon 5 has probability of being obtained $= p_5 = \frac{1}{8}$ while the other 4 coupons have equal probability of being obtained $= p = \frac{1-\frac{1}{8}}{4} = \frac{7}{32}$): With this state diagram, the transition probability matrix P can be filled as follows:

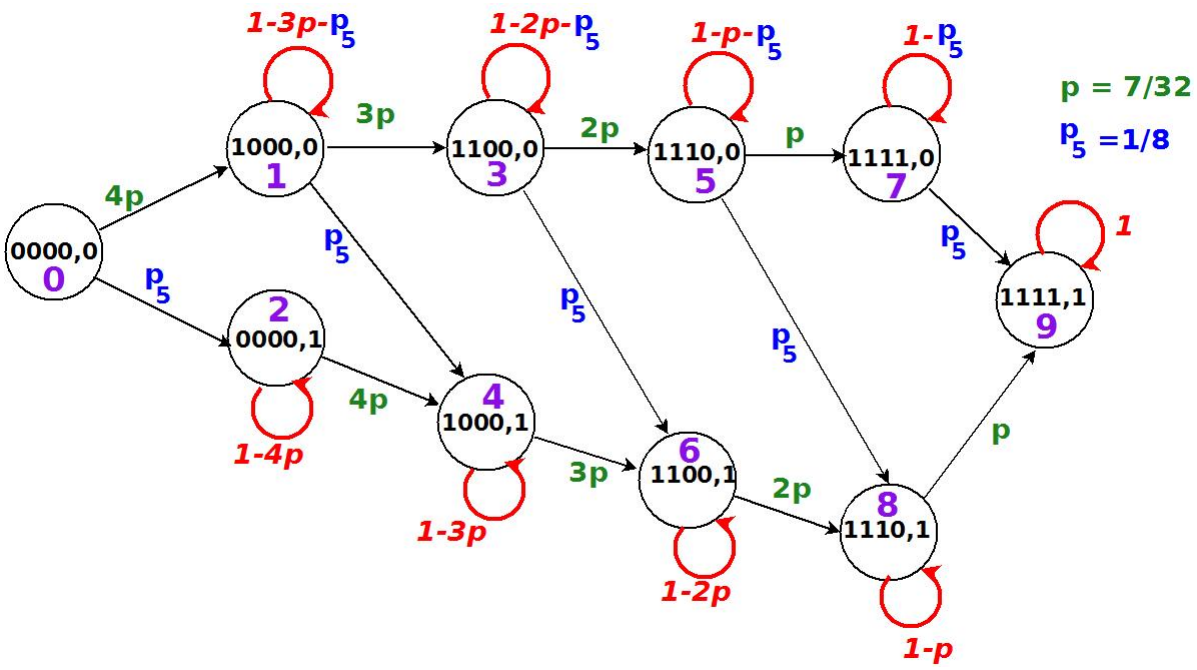

Figure 2.5: Transistion diagram for coupon collector problem with different probability for a coupon.

State Number

State Number	0	1	2	3	4	5	6	7	8	9
0	0	$4p$	p_5	0	0	0	0	0	0	0
1	0	$1-3p-p_5$	0	$3p$	p_5	0	0	0	0	0
2	0	0	$1-4p$	0	$4p$	0	0	0	0	0
3	0	0	0	$1-2p-p_5$	0	$2p$	p_5	0	0	0
4	0	0	0	0	$1-3p$	0	$3p$	0	0	0
5	0	0	0	0	0	$1-p-p_5$	0	p	p_5	0
6	0	0	0	0	0	0	$1-2p$	0	$2p$	0
7	0	0	0	0	0	0	0	$1-p_5$	0	p_5
8	0	0	0	0	0	0	0	0	$1-p$	p
9	0	0	0	0	0	0	0	0	0	1

```
>> P
P =

0    0.8750    0.1250         0         0         0         0         0         0         0
0    0.2188         0    0.6562    0.1250         0         0         0         0         0
```

```
0         0    0.1250         0    0.8750         0         0         0         0         0
0         0         0    0.4375         0    0.4375    0.1250         0         0         0
0         0         0         0    0.3438         0    0.6562         0         0         0
0         0         0         0         0    0.6562         0    0.2188    0.1250         0
0         0         0         0         0         0    0.5625         0    0.4375         0
0         0         0         0         0         0         0    0.8750         0    0.1250
0         0         0         0         0         0         0         0    0.7812    0.2188
0         0         0         0         0         0         0         0         0    1.0000

>> Q=P(1:9,1:9);N=inv(eye(9)-Q);R=P(1:9,10);
>> N*ones(9,1)
ans =

12.4341
11.7070
9.5238
10.8167
8.3810
9.6623
6.8571
8.0000
4.5714
>> N*R
ans =

1.0000
1.0000
1.0000
1.0000
1.0000
1.0000
1.0000
1.0000
1.0000
```

Starting in state 0, a collector will buy an average of 12.4341 items.

The hard part of such problems is getting the state transition diagram correct for any given finite markov chain problem.

Assignment: Solve the following problem from the Final Exam 2022: A gambler starts with \$X in his pocket ($0 < X < 13$ and X is an integer). In each gamble taken, the gambler either wins \$1 or looses \$1 with probability 0.457 or 0.543. The gambler quits if he reaches \$13 or can't continue if all money is lost else continues gambling. What is the initial amount of money X the gambler should have so that the probability of getting to \$13 and quitting is greater than 0.5 ? Note: the gambling house makes sure that the winning probability is slightly less than loosing probability!!.

Let us consider a Markov Chain problem which is a simplified version of the google search and ranking of pages (This is a solved problem (Example 11.30 Pg 671) in Alberto Leon Garcia- "Probability, Statistics and Random Processes for Electrical Engineering", 3rd Edition. I am correcting a small error in this edition of the problem in these notes as well and hopefully not introducing others.)

The web pages universe is a 5 page universe given in the diagram below. A websurfer selects from the outgoing links on any page at random (equal probability) to go to another page. Page 2 is a problem as it has no outgoing links and so any of the 5 pages may be selected at random from it. The Google page rank algorithm in an abbreviated form generates the probability transition matrix P from the diagram and then modifies this in ad-hoc manner as shown below to $P_{modified}$. The eigenvector (with sum =1) corresponding the eigenvalue of 1 of $P_{modified}$ gives the ordering of the listing of web pages by probabilities.

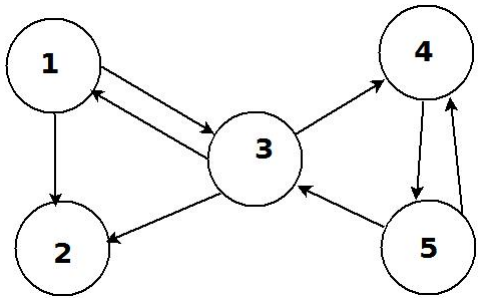

Figure 2.6: Web Page link diagrams from A. L. Garcia "Probability, Statistics and Random Processes for Electrical Engineering".

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \tag{2.26}$$

$$P_{modified} = (\alpha = 0.85)P + (1-\alpha)\frac{1}{5}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{2.27}$$

```
>> alpha=0.85,P=[0 0.5 0.5 0 0;0.2*ones(1,5);1/3 1/3 0 1/3 0;0 0 0 0 1;0 0 0.5 0.5 0]
alpha = 0.8500
```

```
P =

0   0.5000   0.5000        0        0
0.2000   0.2000   0.2000   0.2000   0.2000
0.3333   0.3333        0   0.3333        0
0        0        0        0   1.0000
0        0   0.5000   0.5000        0

>> Pm=alpha*P+(1-alpha)*0.2*ones(5,5)
Pm =

0.030000   0.455000   0.455000   0.030000   0.030000
0.200000   0.200000   0.200000   0.200000   0.200000
0.313333   0.313333   0.030000   0.313333   0.030000
0.030000   0.030000   0.030000   0.030000   0.880000
0.030000   0.030000   0.455000   0.455000   0.030000
>> [V,D]=eig(Pm')
V =

Columns 1 through 4:

-0.2649 +      0i  -0.3028 +      0i  -0.2334 +      0i  -0.2879 - 0.1407i
-0.3774 +      0i  -0.6598 +      0i   0.8524 +      0i  -0.1129 + 0.0582i
-0.4779 +      0i   0.0107 +      0i  -0.4362 +      0i   0.6480 +      0i
-0.5008 +      0i   0.3761 +      0i  -0.1690 +      0i   0.1433 - 0.3362i
-0.5551 +      0i   0.5757 +      0i  -0.0138 +      0i  -0.3905 + 0.4187i

Column 5:

-0.2879 + 0.1407i
-0.1129 - 0.0582i
0.6480 -      0i
0.1433 + 0.3362i
-0.3905 - 0.4187i

D =

Diagonal Matrix
 Columns 1 through 4:

1.0000 +      0i                  0                  0                  0
0   0.3604 +      0i                  0                  0
0                  0  -0.0913 +      0i                  0
```

```
0                   0                   0  -0.4745 + 0.1976i
0                   0                   0                   0

Column 5:

0
0
0
0
-0.4745 - 0.1976i

>> Pr = V(:,1)/sum(V(:,1))
Pr =

0.1217
0.1734
0.2196
0.2301
0.2551
```

So page 5 ,4 ,3,2 1 is the ranking order of the pages in Google search in this simplified algorithm.

2.9 Generating random numbers on a computer to follow a distribution function or pdf function.

This section is useful when simulations are to be done on a computer and the computer has a "good" uniform random number generator. To illustrate the ideas, consider the one-sided exponential distribution function whose pdf is $f_X(x) = e^{-x}$ and cumulative distribution ie $F_X(x) = \Pr(0 < X \leq x) = 1 - e^{-x}$. A plot of the cumulative distribution function is shown below: Now to generate random numbers that follow this distribution, we first select a uniform random number u between (0,1) and then using the CDF compute the corresponding x as follows for this case as $x = -\ln(1-u)$. The x values so obtained will follow the exponential distribution as shown in the histogram below. This method requires the CDF of the distribution for it to work.

For Gaussian random variables $N(0,1)$ is generated from the uniform random variables by using two uniform random variables $u_1, u_2 \, \epsilon \, (0,1)$ and generating the corresponding Normal $N(0,1)$ variables as $z_1 = \sqrt{-2\ln(u_1)}\cos(2\pi u_2), z_2 = \sqrt{-2\ln(u_1)}\sin(2\pi u_2)$. This approach uses the 2-D Gaussian distribution and is known as the Box-Muller transformation. The method used by Octave/Matlab for normal random number generation is a better method than the Box-Muller transformation (numerically).

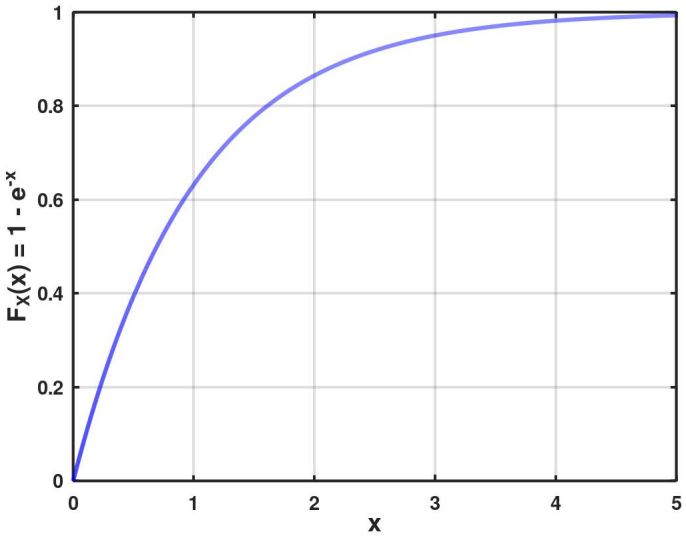

Figure 2.7: CDF of one-sided exponential distribution function with $\alpha = 1$

2.10 The Gaussian/Normal distribution

The joint Gaussian probability density function of RV $X_{N\times 1}$ with mean (vector) μ and covariance matrix $\Sigma\,[= E(\{X-\mu\}\{X-\mu\}')]$ is defined as:

$$p_{X_{N\times 1}}(x) = N(\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det(\Sigma)}} e^{-\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2}} \tag{2.28}$$

The scalar case distribution with σ = standard deviation is

$$p_{X_{1\times 1}}(x) = N(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.29}$$

with $Z = \frac{(X-\mu)}{\sigma}$ the scalar density function is

$$p_Z(z) = N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{2.30}$$

The CDF ($F_Z(z)$) is tabulated in many books/net for $N(0,1)$.

The CDF of $N(0,1)$ is closely related to the error function which is defined as

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-}t^2 dt \quad x\,\epsilon\,[0,\infty] \tag{2.31}$$

This relationship is

$$F_Z(z) = \frac{1 + \text{erf}\left(\frac{z}{\sqrt{2}}\right)}{2} \tag{2.32}$$

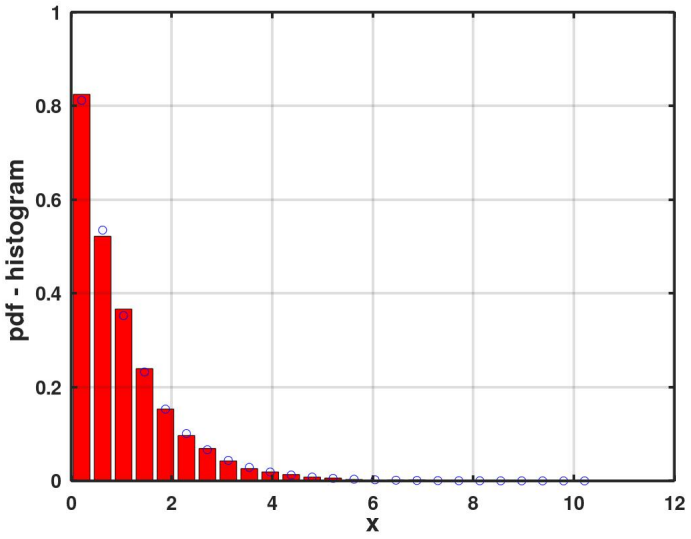

Figure 2.8: Histogram and pdf('o') plot of generated rvs x

Octave does not have the CDF function built into it without the statistics package. However it does have the erf function and so equation 2.32 is useful.

At this stage let us look at the results in Anderson & Moore Appendix A pg 321 - 322 and apply them to some problems.

1. Let $X = [X_1, X_2]'$ be a normal random vector with mean and covariance matrices μ, Σ. Let $Y = AX + b$ where A is a matrix and b a vector and Y a transformed RV vector.

$$\mu = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \Sigma = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}, A = \begin{bmatrix} 2 & 1 \\ -1 & 1 \\ 1 & 3 \end{bmatrix}, b = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{2.33}$$

   Find

   (a) the joint pdf of X_1, X_2
   (b) the pdf of X_1
   (c) the $\Pr(X_1 < 0)$
   (d) the $\Pr(Y_1 < 0)$
   (e) the covariance matrix of Y
   (f) the covariance between Y_1, Y_2
   (g) $E\{X_1 X_2\}$
   (h) $E\{Y_1 Y_2\}$
   (i) the covariance between X_1, Y_1

(j) $E\{X_1 Y_1\}$

The joint pdf of X_1, X_2 and the pdf of X_1 are from wxMaxima. Note the pdf of X_1 can be written by inspection using the results on page 321-322 of Anderson and Moore, though this is done very formally below in wxMaxima:

(% i3) Sigma:matrix([4,1],[1,4]);μ:matrix([1],[2]);X:matrix([x_1],[x_2]);

$$\begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix} \qquad (\%\ o1)$$

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad (\%\ o2)$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad (\%\ o3)$$

–>

(% i4) f_X:ratsimp((1/(2*%pi*sqrt(determinant(Sigma))))*%e^ (-transpose(X-μ).invert(Sigm μ)/2));

$$\frac{\%e^{-\frac{2{x_2}^2}{15}+\frac{x_1 x_2}{15}+\frac{7x_2}{15}-\frac{2{x_1}^2}{15}+\frac{2x_1}{15}-\frac{8}{15}}}{2\sqrt{15}\pi} \qquad (\%\ o4)$$

(% i5) f_X_1:integrate(f_X,x_2,-inf,inf);

$$\frac{\%e^{-\frac{{x_1}^2-2x_1+1}{8}}}{2^{\frac{3}{2}}\sqrt{\pi}} \qquad (\%\ o5)$$

(% i6) integrate(f_X_1,x_1,-inf,inf);

$$1 \qquad (\%\ o6)$$

Defining $Z_1 = (X_1 - 1)/2 \implies X_1 = 2Z_1 + 1$. The condition $X_1 < 0$ is satisfied by $Z_1 < -0.5$. So we need the CDF of -0.5 which using octave is:

```
>> (1+erf(-0.5/sqrt(2)))/2
ans = 0.3085
```

Y is an affine transformation of X and so is a Gaussian RV vector with mean given by $\mu_Y = A\mu + b$ and covariance matrix $\Sigma_Y = A\Sigma A'$. Computing these in Octave we have numerical values:

```
>> Sigma=[4 1;1 4],mu=[1;2],A=[2 1;-1 1;1 3],b=[-1;0;1]
Sigma =

4   1
1   4

mu =

1
2

A =

2   1
   -1   1
1   3
b =

-1
0
1

>> mu_y=A*mu+b
mu_y =

3
1
8

>> Sigma_Y=A*Sigma*A'
Sigma_Y =

24   -3   27
-3    6    6
27    6   46
```

The pdf of Y_1 is now given by $\frac{1}{\sqrt{2\pi}\sqrt{24}}e^{-\frac{(y_1-3)^2}{2\times 24}}$. To find $\Pr Y_1 < 0$ we now form $z_1 = \frac{y_1-3}{\sqrt{24}} \implies y_1 = 3 + \sqrt{24}z_1$. Imposing $y_1 < 0 \implies z_1 < -\frac{3}{\sqrt{24}}$ and computing in Octave as before using erf function we get the $\Pr Y_1 < 0 = 0.2701$

The covariance matrix of Y has already been obtained above as Σ_Y in Octave.

The covariance between Y_1, Y_2 is -3 from the 1,2 element (or 2,1 element of) of the matrix Σ_Y

To find $E\{X_1X_2\}$ we use, $E\{(X_1-\mu_1)(X_2-\mu_2)\} = \Sigma_{1,2} = 1 \implies E\{X_1X_2\} - \mu_1\mu_2 = 1 \implies E\{X_1X_2\} = 3$.

$E\{Y_1Y_2\} = -3+3 = 0$.

To solve the last two parts of the problem define a linear transformation $V = [Y_1, X_1]' = A_1X + b_1$ with

$$A_1 = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}, b_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \tag{2.34}$$

Computing the mean and covariance of V in Octave $\mu_V = A_1\mu + b_1, \Sigma_V = A_1\Sigma A_1'$ we have:

```
>> A_1=[2 1;1 0],b_1=[-1;0]
A_1 =

2   1
1   0

b_1 =

-1
0

>> mu_V=A_1*mu+b_1
mu_V =

3
1

>> Sigma_V=A_1*Sigma*A_1'
Sigma_V =

24    9
9    4
```

Hence covariance between $Y_1, X_1 = 9$ and $E\{X_1Y_1\} = 9+3 = 12$

2. Consider the Gaussian RVs X_1, X_2 as in the last problem. Let $Y = [2,1]X$ be a measurement equation (ie. you can only measure the combination of X_1, X_2 as defined by this equation). What is the $E(X|Y=y)$ and Covariance of $X|Y=y$.

Define $Z = [X_1, X_2, Y]'$ through the equation $Z = AX$ with A given by

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \tag{2.35}$$

Then $mu_Z, Sigma_Z$ is obtained as

```
>> mu_Z=A*mu,Sigma_Z=A*Sigma*A'
mu_Z =

1
2
4

Sigma_Z =

4    1    9
1    4    6
9    6   24
```

The required conditional mean and covariances computations are now given by:

```
>> Sigma_12=Sigma_Z(1:2,3),Sigma_22=Sigma_Z(3,3)
Sigma_12 =

9
6

Sigma_22 = 24
>> mu-Sigma_12/Sigma_22
ans =

0.6250
1.7500

>> Sigma_Conditional=Sigma-Sigma_12*Sigma_12'/Sigma_22
ans =

 0.6250  -1.2500
-1.2500   2.5000
```

```
>> Sigma_12/Sigma_22
ans =

0.3750
0.2500
```

So the conditional mean is now

$$E(X|Y=y) = \begin{bmatrix} 0.6250 \\ 1.7500 \end{bmatrix} - \begin{bmatrix} 0.3750 \\ 0.2500 \end{bmatrix} y \tag{2.36}$$

Note that the conditional covariance has decreases the variances of both states from the unconditional variance and therefore the measurement and conditioning on it has improved the estimates of X.