Dimensionality Reduction

What is Dimensionality Reduction?

- In machine learning classification problems, there are often too many factors on the basis of which the final classification is done.
- These factors are basically variables called features.
- The higher the number of features, the harder it gets to visualize the training set and then work on it.
- Sometimes, most of these features are correlated, and hence redundant. This is where dimensionality reduction algorithms come into play.
- Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principal variables.
- It can be divided into feature selection and feature extraction.

Methods of Dimensionality Reduction

Various methods used for dimensionality reduction include:

- Principal Component Analysis (PCA)
- Factor Analysis (FA)
- Independent Component Analysis (ICA)
- Linear Discriminant Analysis (LDA)
- Generalized Discriminant Analysis (GDA) etc…

Curse of Dimensionality

Large number of dataset makes the model extremely slow and difficult to find a solution. This is referred to as curse of dimensionality.

Advantages of Dimensionality Reduction

- It helps in data compression, and hence reduced storage space.
- It reduces computation time.
- It also helps remove redundant features, if any.
- It gives better visualization to gain important insights by detecting patterns.

Disadvantages of Dimensionality Reduction

- It may lead to some amount of data loss.
- PCA tends to find linear correlations between variables, which is sometimes undesirable.
- PCA fails in cases where mean and covariance are not enough to define datasets.
- We may not know how many principal components to keep- in practice, some thumb rules are applied

Principal Component Analysis (PCA)

- **Principal component analysis** (**PCA**) is a statistical procedure that is used to reduce the dimensionality.
- It uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

Steps Involved in the PCA

Step 1: Standardize the dataset.

Step 2: Calculate the covariance matrix for the features in the dataset.

Step 3: Calculate the eigenvalues and eigenvectors for the covariance matrix.

Step 4: Sort eigenvalues and their corresponding eigenvectors.

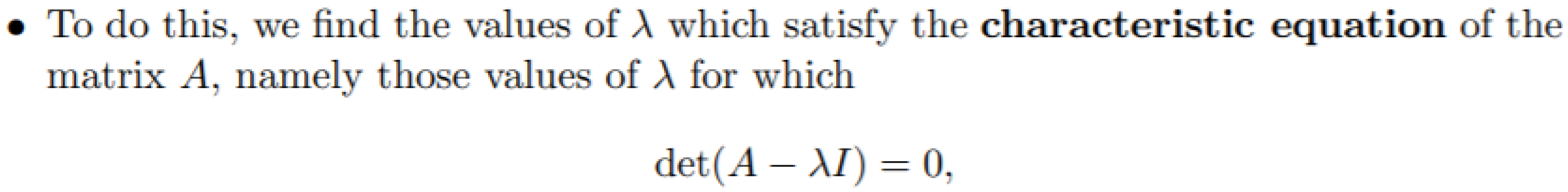
Step 5: Pick k eigenvalues and form a matrix of eigenvectors.

Step 6: Transform the original matrix.

Eigen values and vectors (prerequisite)

- An **eigenvector** is a nonzero vector that changes at most by a scalar factor when that linear transformation is applied to it.

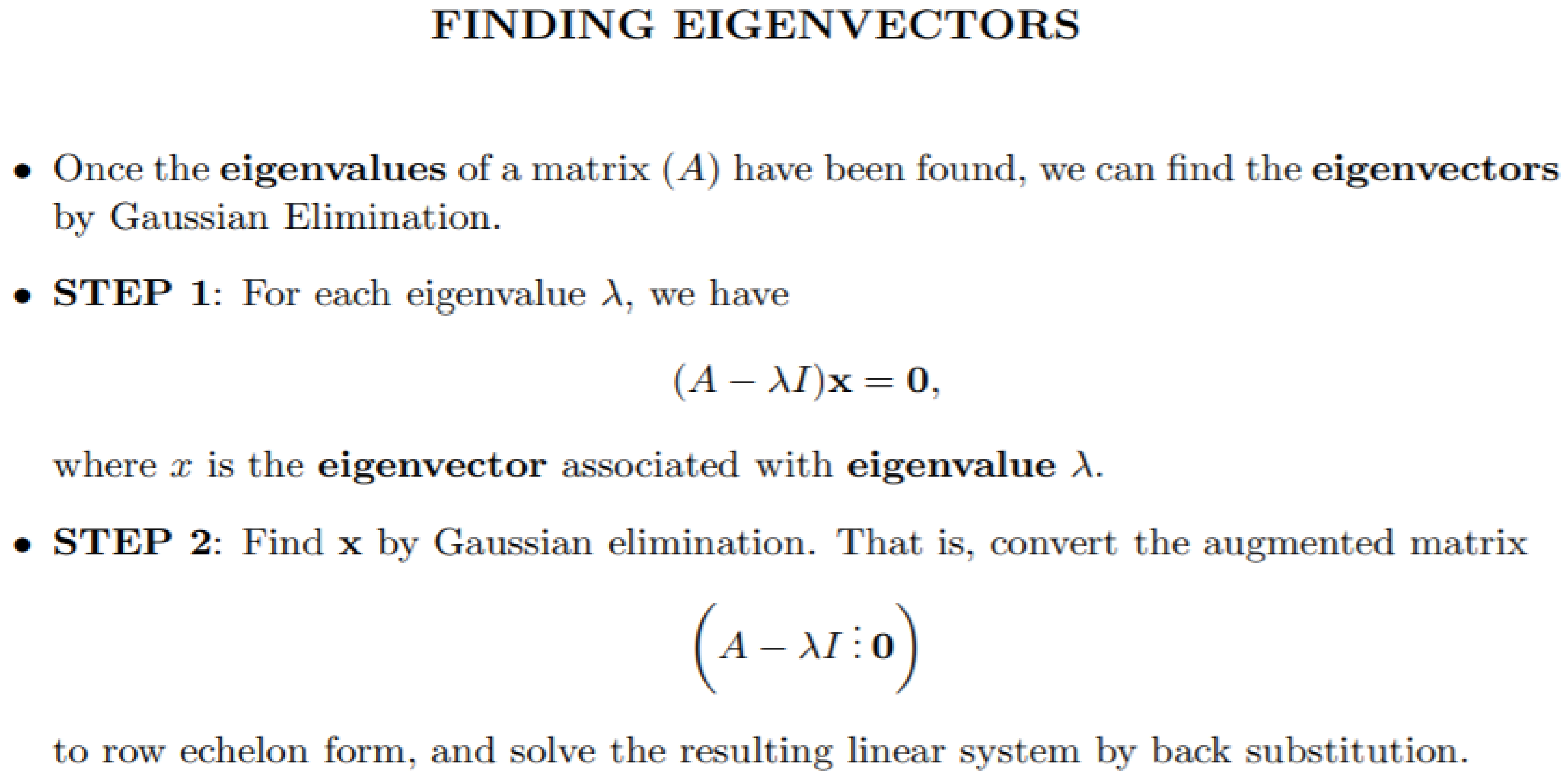
- The corresponding **eigenvalue** is the factor by which the eigenvector is scaled.



- To do this, we find the values of λ which satisfy the **characteristic equation** of the matrix A, namely those values of λ for which

$$\det(A - \lambda I) = 0,$$

Eigen values and vectors (prerequisite)



FINDING EIGENVECTORS

- Once the **eigenvalues** of a matrix (A) have been found, we can find the **eigenvectors** by Gaussian Elimination.
- **STEP 1**: For each eigenvalue λ, we have

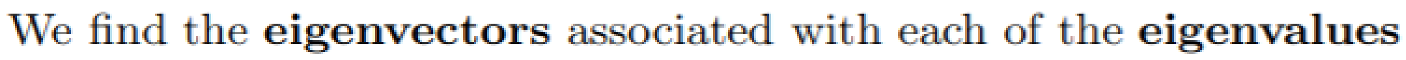
$$(A - \lambda I)\mathbf{x} = \mathbf{0},$$

where x is the **eigenvector** associated with **eigenvalue** λ.

- **STEP 2**: Find $\mathbf{x}$ by Gaussian elimination. That is, convert the augmented matrix

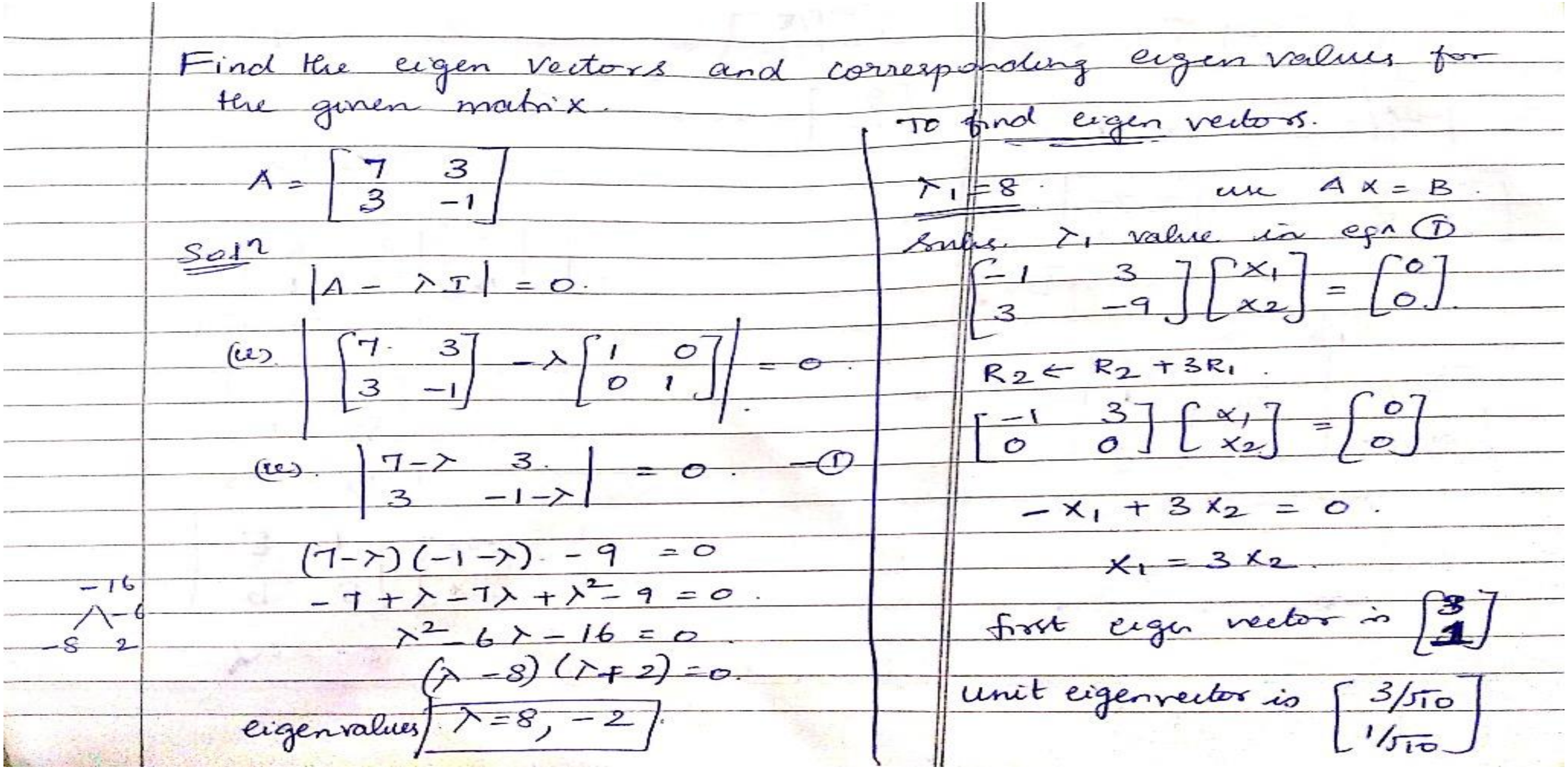
$$\left(A - \lambda I \vdots \mathbf{0}\right)$$

to row echelon form, and solve the resulting linear system by back substitution.



We find the **eigenvectors** associated with each of the **eigenvalues**

Eigen values and vectors – Example 1



Find the eigen vectors and corresponding eigen values for the given matrix.

$$A = \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$$

Soln

$$|A - \lambda I| = 0.$$

(ie) $$\left| \begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0.$$

(ie) $$\begin{vmatrix} 7-\lambda & 3 \\ 3 & -1-\lambda \end{vmatrix} = 0. \quad —①$$

$$(7-\lambda)(-1-\lambda) - 9 = 0$$

$$-7 + \lambda - 7\lambda + \lambda^2 - 9 = 0.$$

$$\lambda^2 - 6\lambda - 16 = 0.$$

$$(\lambda - 8)(\lambda + 2) = 0.$$

eigenvalues $\boxed{\lambda = 8, -2}$

-16

-6

$-8 \quad 2$

To find eigen vectors.

$\lambda_1 = 8$. we $AX = B$.

Subs. λ_1 value in eqn ①

$$\begin{bmatrix} -1 & 3 \\ 3 & -9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

$R_2 \leftarrow R_2 + 3R_1$.

$$\begin{bmatrix} -1 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

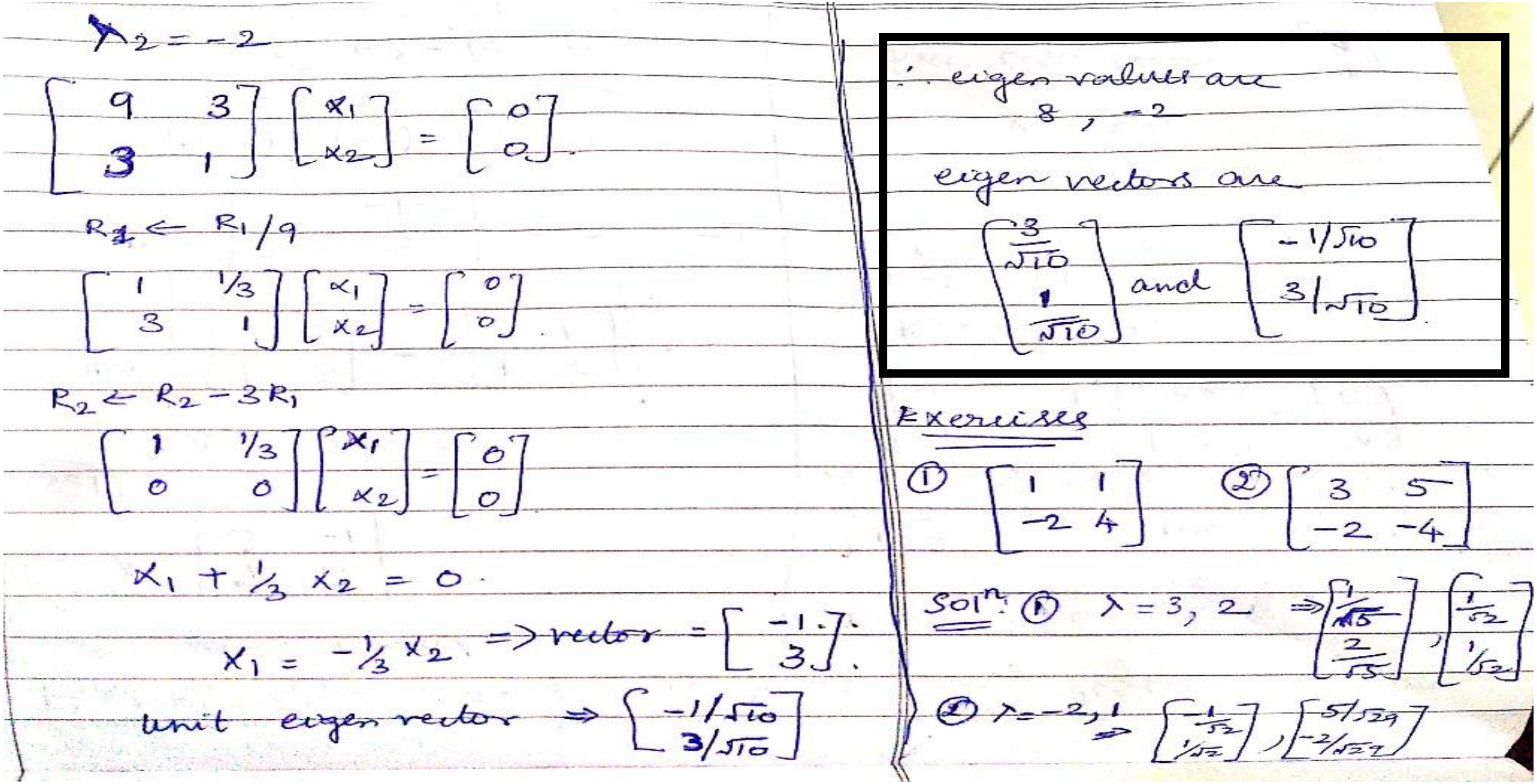
$$-x_1 + 3x_2 = 0.$$

$$x_1 = 3x_2.$$

first eigen vector is $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$

unit eigenvector is $\begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$

Eigen values and vectors – Example 1



$\lambda_2 = -2$

$$\begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

$R_1 \leftarrow R_1/9$

$$\begin{bmatrix} 1 & 1/3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

$R_2 \leftarrow R_2 - 3R_1$

$$\begin{bmatrix} 1 & 1/3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$x_1 + \frac{1}{3} x_2 = 0.$

$x_1 = -\frac{1}{3} x_2 \Rightarrow$ vector $= \begin{bmatrix} -1 \\ 3 \end{bmatrix}.$

unit eigen vector $\Rightarrow \begin{bmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \end{bmatrix}$

$\therefore$ eigen values are
8 , −2

eigen vectors are

$$\begin{bmatrix} \frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{bmatrix} \text{ and } \begin{bmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \end{bmatrix}.$$

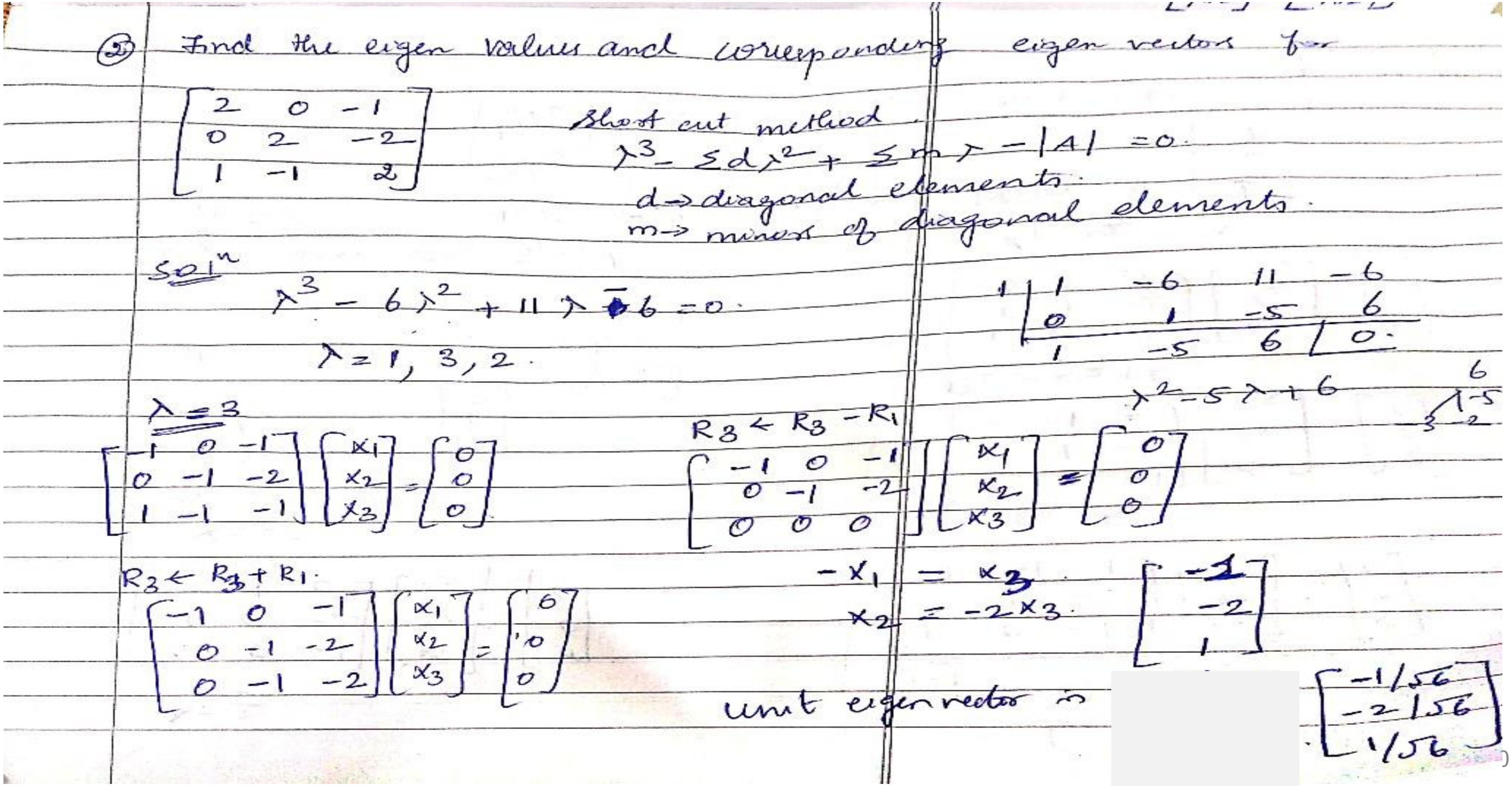
Exercises

① $\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$ ② $\begin{bmatrix} 3 & 5 \\ -2 & -4 \end{bmatrix}$

Sol^n: ① $\lambda = 3, 2 \Rightarrow \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 1/\sqrt{2} \end{bmatrix}$

② $\lambda = -2, 1 \Rightarrow \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} 5/\sqrt{29} \\ -2/\sqrt{29} \end{bmatrix}$

Eigen values and vectors – Example 2



(2) Find the eigen values and corresponding eigen vectors for

$$\begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -2 \\ 1 & -1 & 2 \end{bmatrix}$$

Short cut method.

$\lambda^3 - \Sigma d \lambda^2 + \Sigma m \lambda - |A| = 0.$

d → diagonal elements.

m → minors of diagonal elements.

Sol

$\lambda^3 - 6\lambda^2 + 11\lambda - 6 = 0.$

$\lambda = 1, 3, 2.$

$$\begin{array}{r|rrr|r} 1 & 1 & -6 & 11 & -6 \\ & 0 & 1 & -5 & 6 \\ \hline & 1 & -5 & 6 & 0. \end{array}$$

$\lambda^2 - 5\lambda + 6$

$$\begin{array}{c} 6 \\ -5 \\ -3 \quad -2 \end{array}$$

$\lambda = 3$

$$\begin{bmatrix} -1 & 0 & -1 \\ 0 & -1 & -2 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$R_3 \leftarrow R_3 + R_1.$

$$\begin{bmatrix} -1 & 0 & -1 \\ 0 & -1 & -2 \\ 0 & -1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$R_3 \leftarrow R_3 - R_1$

$$\begin{bmatrix} -1 & 0 & -1 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$-x_1 = x_3.$

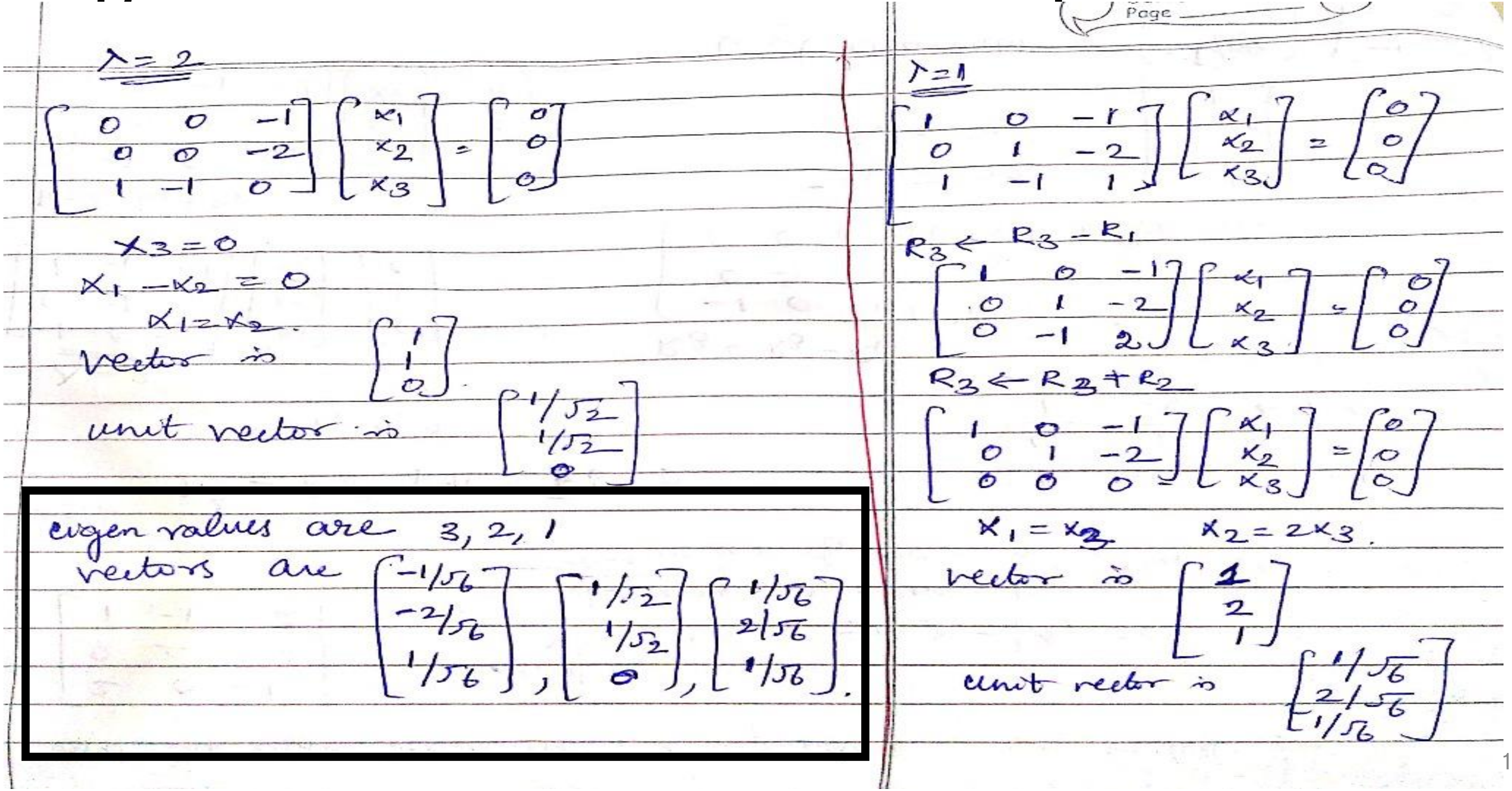
$x_2 = -2x_3.$

$$\begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}$$

unit eigen vector is

$$\begin{bmatrix} -1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$$

Eigen values and vectors – Example 2



Page

$\lambda = 2$

$$\begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & -2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$x_3 = 0$

$x_1 - x_2 = 0$

$x_1 = x_2$.

vector is $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.

unit vector is $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$

eigen values are 3, 2, 1

vectors are $\begin{bmatrix} -1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$.

$\lambda = 1$

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$R_3 \leftarrow R_3 - R_1$

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$R_3 \leftarrow R_3 + R_2$

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$x_1 = x_3 \qquad x_2 = 2x_3$.

vector is $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

unit vector is $\begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$

Exercise

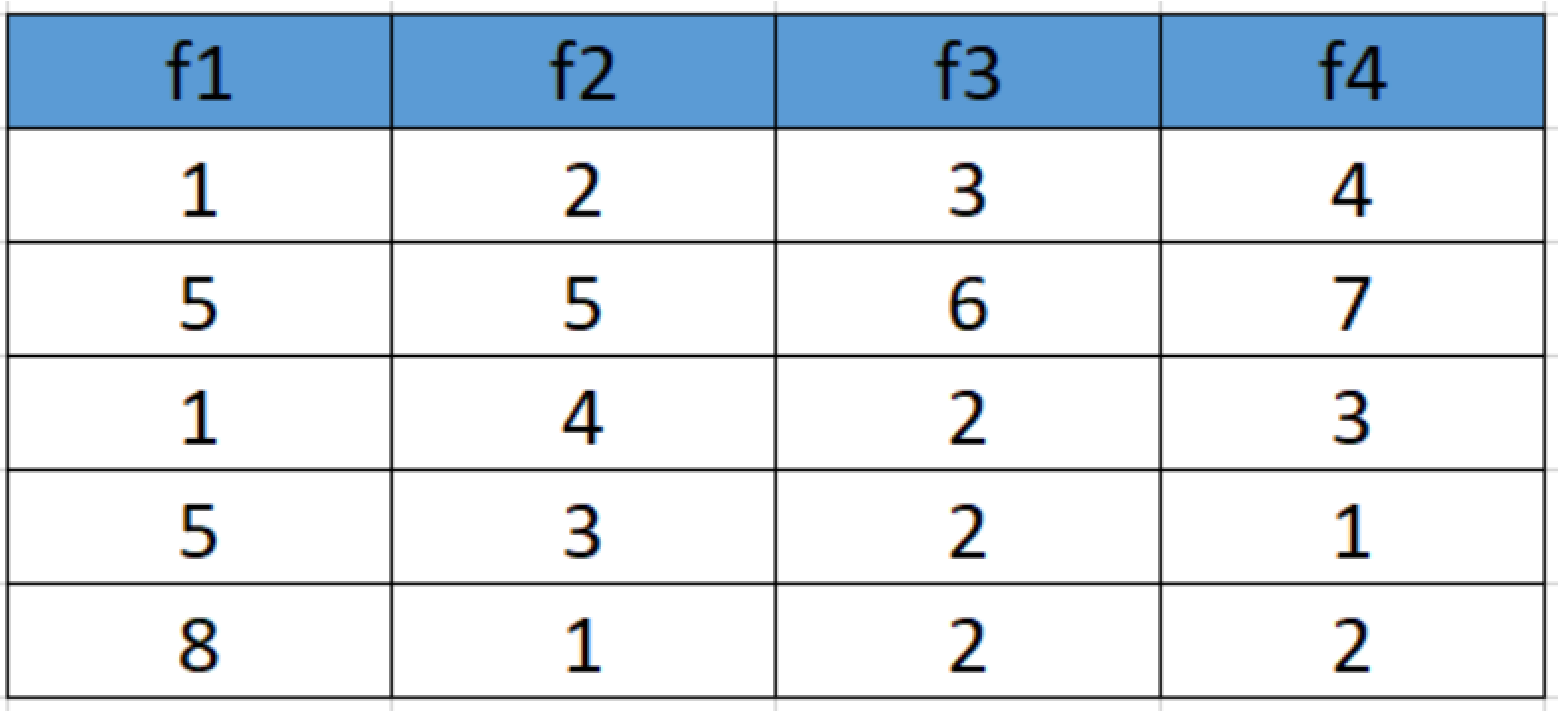
Find the eigen values and corresponding eigen vectors for the following

1) $\begin{bmatrix} 0 & 1 & 3 \\ 0 & 6 & 0 \\ -6 & 2 & 9 \end{bmatrix}$

2) $\begin{bmatrix} 5 & 0 & -1 \\ 0 & 8 & 0 \\ -3 & 0 & 7 \end{bmatrix}$

Principal Component Analysis (PCA) – Example 1

- Consider the dataset given. Identify the principal components and the transformed data.



f1	f2	f3	f4
1	2	3	4
5	5	6	7
1	4	2	3
5	3	2	1
8	1	2	2

Dataset matrix

PCA Steps

Steps Involved in the PCA

Step 1: Standardize the dataset.

Step 2: Calculate the covariance matrix for the features in the dataset.

Step 3: Calculate the eigenvalues and eigenvectors for the covariance matrix.

Step 4: Sort eigenvalues and their corresponding eigenvectors.

Step 5: Pick k eigenvalues and form a matrix of eigenvectors.

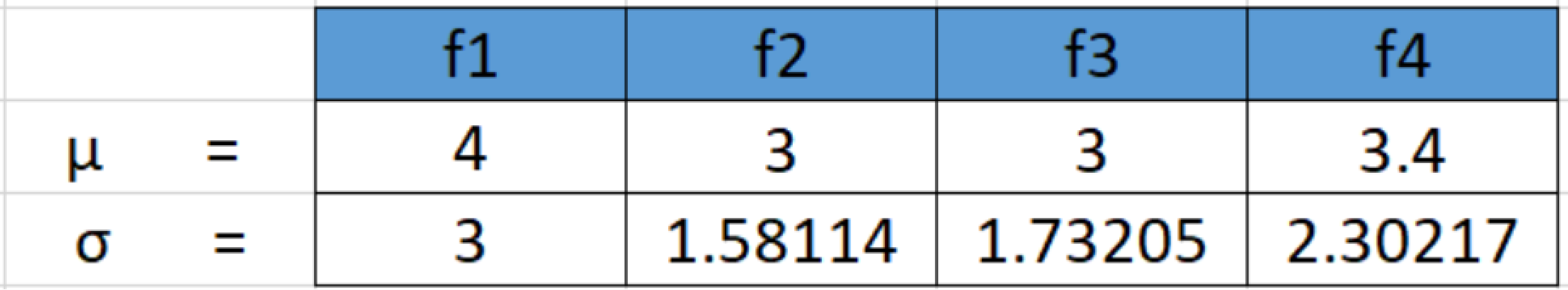
Step 6: Transform the original matrix.

Example 1

1. Standardize the Dataset

$$x_{new} = \frac{x - \mu}{\sigma}$$

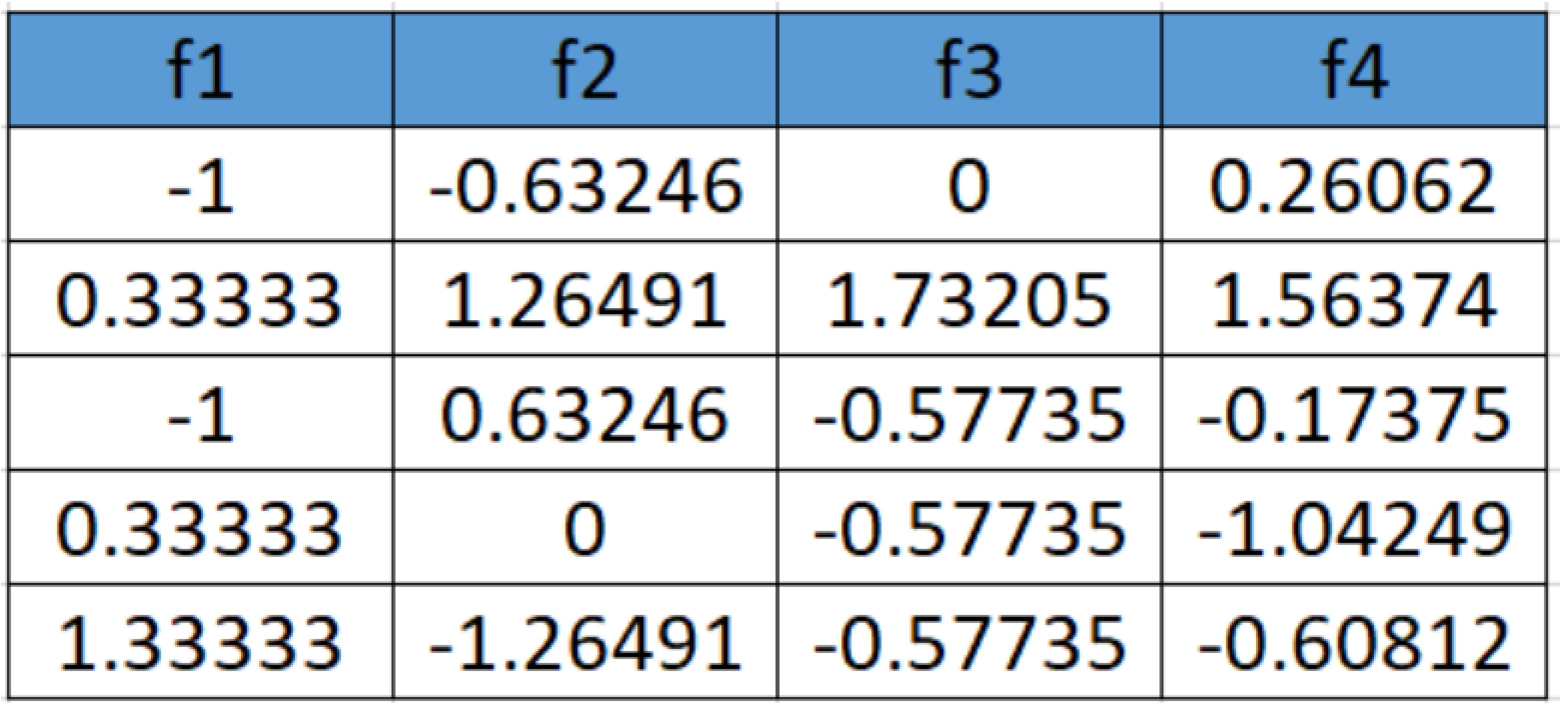
Standardization formula



	f1	f2	f3	f4
μ =	4	3	3	3.4
σ =	3	1.58114	1.73205	2.30217

Example 1

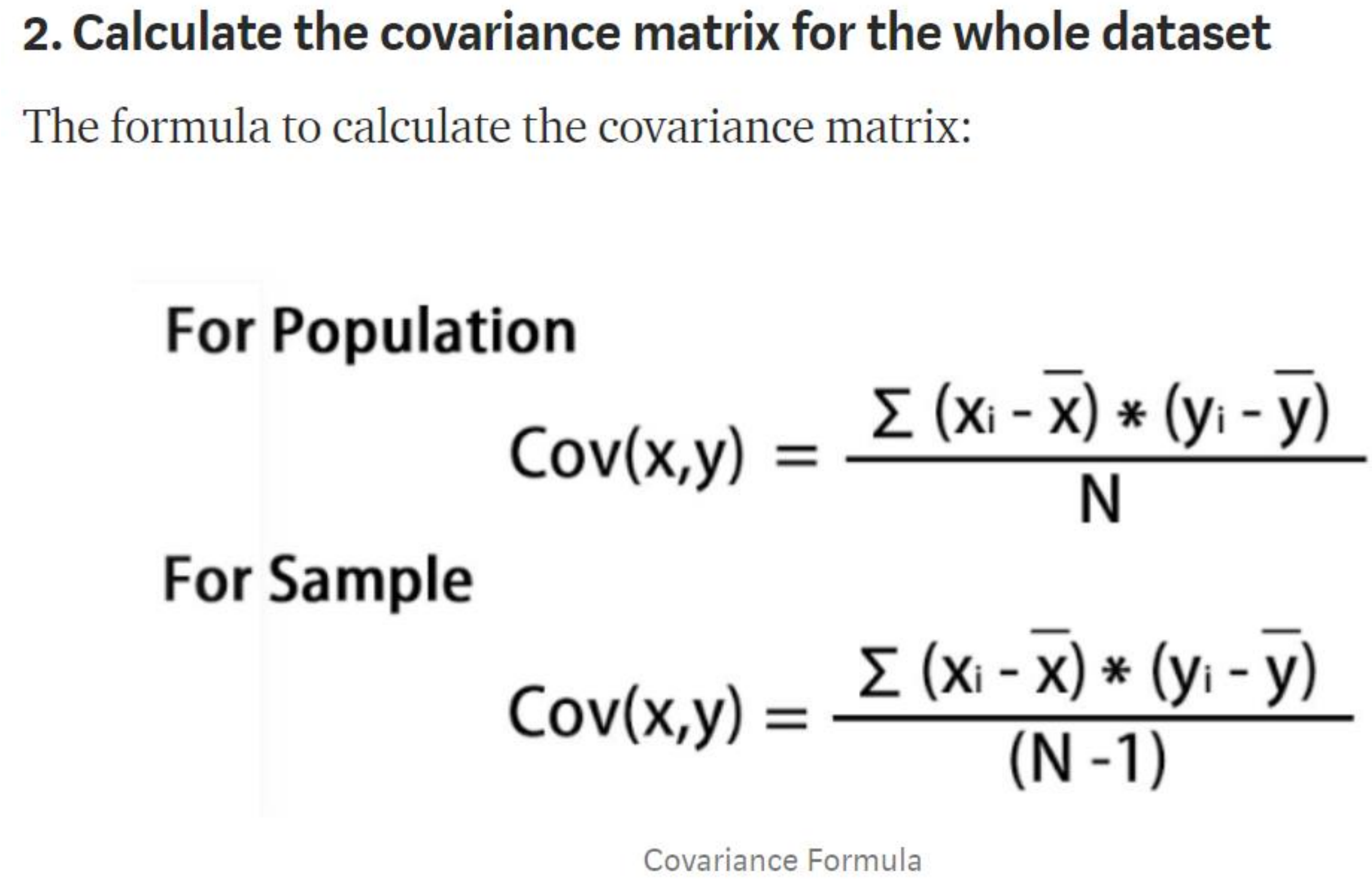
- Standardized dataset



f1	f2	f3	f4
-1	-0.63246	0	0.26062
0.33333	1.26491	1.73205	1.56374
-1	0.63246	-0.57735	-0.17375
0.33333	0	-0.57735	-1.04249
1.33333	-1.26491	-0.57735	-0.60812

Standardized Dataset

Example 1



2. Calculate the covariance matrix for the whole dataset

The formula to calculate the covariance matrix:

For Population

$$\mathrm{Cov}(x,y) = \frac{\Sigma\,(x_i - \overline{x}) * (y_i - \overline{y})}{N}$$

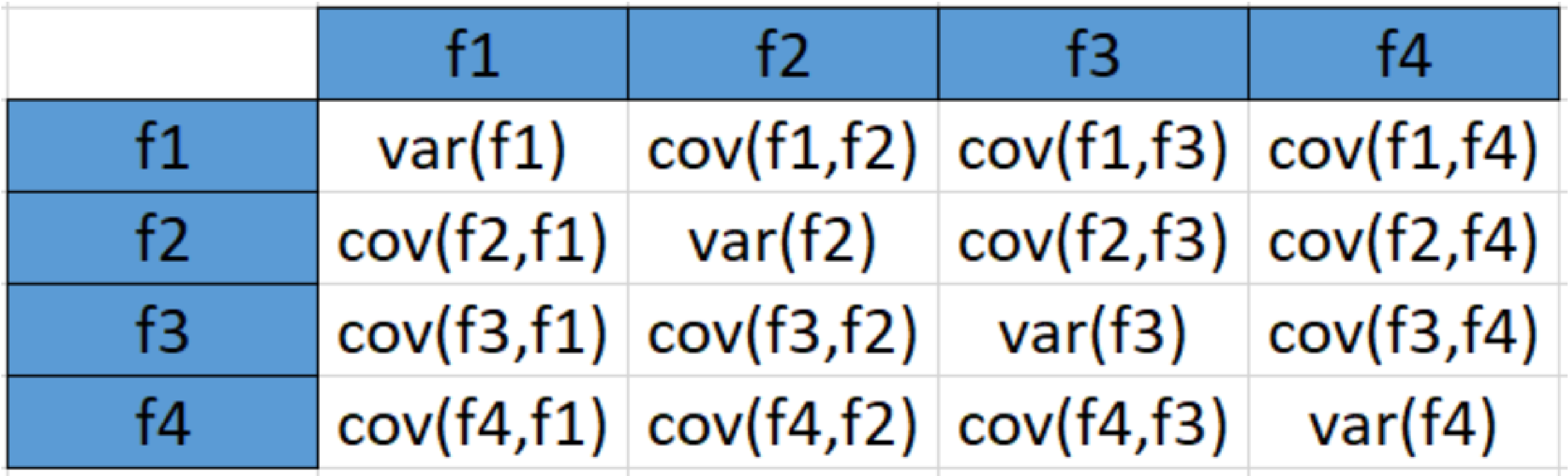
For Sample

$$\mathrm{Cov}(x,y) = \frac{\Sigma\,(x_i - \overline{x}) * (y_i - \overline{y})}{(N - 1)}$$

Covariance Formula

Example 1

the covariance matrix for the given dataset will be calculated as below

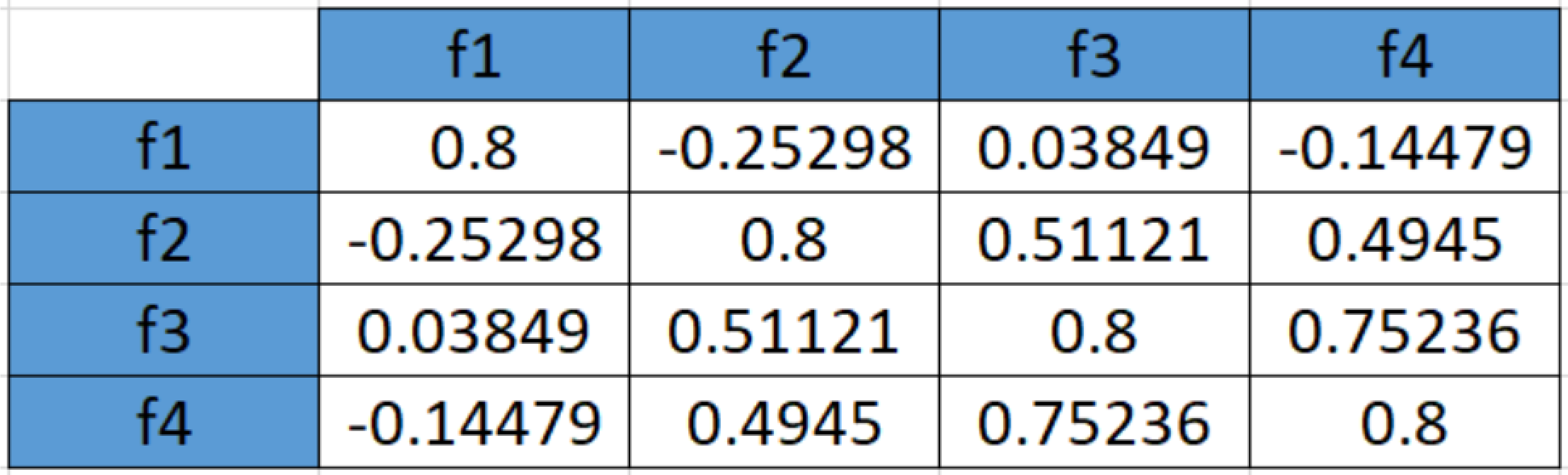


	f1	f2	f3	f4
f1	var(f1)	cov(f1,f2)	cov(f1,f3)	cov(f1,f4)
f2	cov(f2,f1)	var(f2)	cov(f2,f3)	cov(f2,f4)
f3	cov(f3,f1)	cov(f3,f2)	var(f3)	cov(f3,f4)
f4	cov(f4,f1)	cov(f4,f2)	cov(f4,f3)	var(f4)

Since we have standardized the dataset, so the **mean for each feature is 0** and the standard deviation is 1.

Example 1

- Covariance matrix



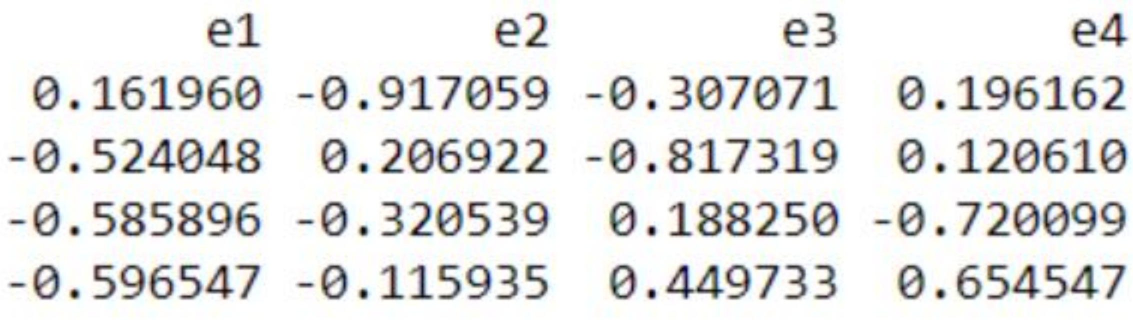
	f1	f2	f3	f4
f1	0.8	-0.25298	0.03849	-0.14479
f2	-0.25298	0.8	0.51121	0.4945
f3	0.03849	0.51121	0.8	0.75236
f4	-0.14479	0.4945	0.75236	0.8

covariance matrix (population formula)

Example 1

3. Calculate eigenvalues and eigen vectors.

$$\lambda = 2.51579324\,,\ 1.0652885\,,\ 0.39388704\,,\ 0.02503121$$



```
       e1        e2        e3        e4
 0.161960 -0.917059 -0.307071  0.196162
-0.524048  0.206922 -0.817319  0.120610
-0.585896 -0.320539  0.188250 -0.720099
-0.596547 -0.115935  0.449733  0.654547
```

eigenvectors(4 * 4 matrix)

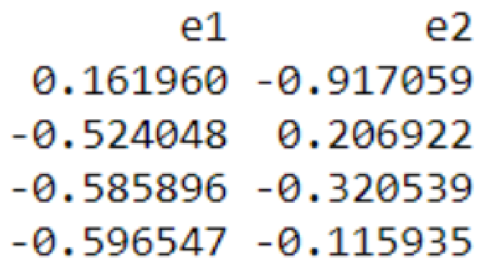
Example 1

4. Sort eigenvalues and their corresponding eigenvectors.

Since eigenvalues are already sorted in this case so no need to sort them again.

5. Pick k eigenvalues and form a matrix of eigenvectors

If we choose the top 2 eigenvectors, the matrix will look like this:



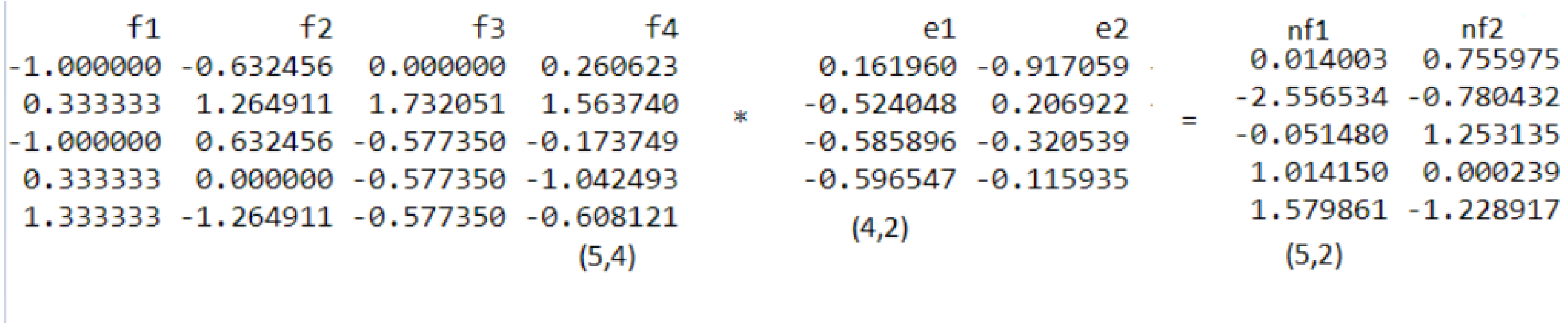
```
       e1        e2
 0.161960 -0.917059
-0.524048  0.206922
-0.585896 -0.320539
-0.596547 -0.115935
```

Top 2 eigenvectors(4*2 matrix)

Example 1

6. Transform the original matrix.

Feature matrix * top k eigenvectors = Transformed Data



```
       f1        f2        f3        f4                      e1        e2             nf1       nf2
-1.000000 -0.632456  0.000000  0.260623                0.161960 -0.917059        0.014003  0.755975
 0.333333  1.264911  1.732051  1.563740       *       -0.524048  0.206922    =  -2.556534 -0.780432
-1.000000  0.632456 -0.577350 -0.173749               -0.585896 -0.320539       -0.051480  1.253135
 0.333333  0.000000 -0.577350 -1.042493               -0.596547 -0.115935        1.014150  0.000239
 1.333333 -1.264911 -0.577350 -0.608121                                          1.579861 -1.228917
                                                          (4,2)
                                 (5,4)                                            (5,2)
```

Data Transformation

PCA – Example 2

- Consider the dataset given. Identify the principal components and the transformed data.

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3
2.3	2.7
2	1.6
1	1.8
1.5	1.6
1.1	0.9

PCA – Example 2

	x	y	Std. x	Std. y	x - x'	(x - x')^2	y - y'	(y - y')^2
	2.5	2.4	0.926	0.554	0.9259	0.857291	0.554	0.306916
	0.5	0.7	-1.759	-1.687	-1.7591	3.094433	-1.687	2.845969
	2.2	2.9	0.524	1.213	0.5239	0.274471	1.213	1.471369
	1.9	2.2	0.121	0.29	0.1209	0.014617	0.29	0.0841
	3.1	3	1.732	1.344	1.7319	2.999478	1.344	1.806336
	2.3	2.7	0.658	0.949	0.6579	0.432832	0.949	0.900601
	2	1.6	0.255	-0.501	0.2549	0.064974	-0.501	0.251001
	1	1.8	-1.087	-0.237	-1.0871	1.181786	-0.237	0.056169
	1.5	1.6	-0.416	-0.501	-0.4161	0.173139	-0.501	0.251001
	1.1	0.9	-0.953	-1.424	-0.9531	0.9084	-1.424	2.027776
Mean	**1.81**	**1.98**	**0.0001**	**0**				
Std. Dev.	**0.744916**	**0.758683**	**1.000071**	**1.000062**				

PCA – Example 2

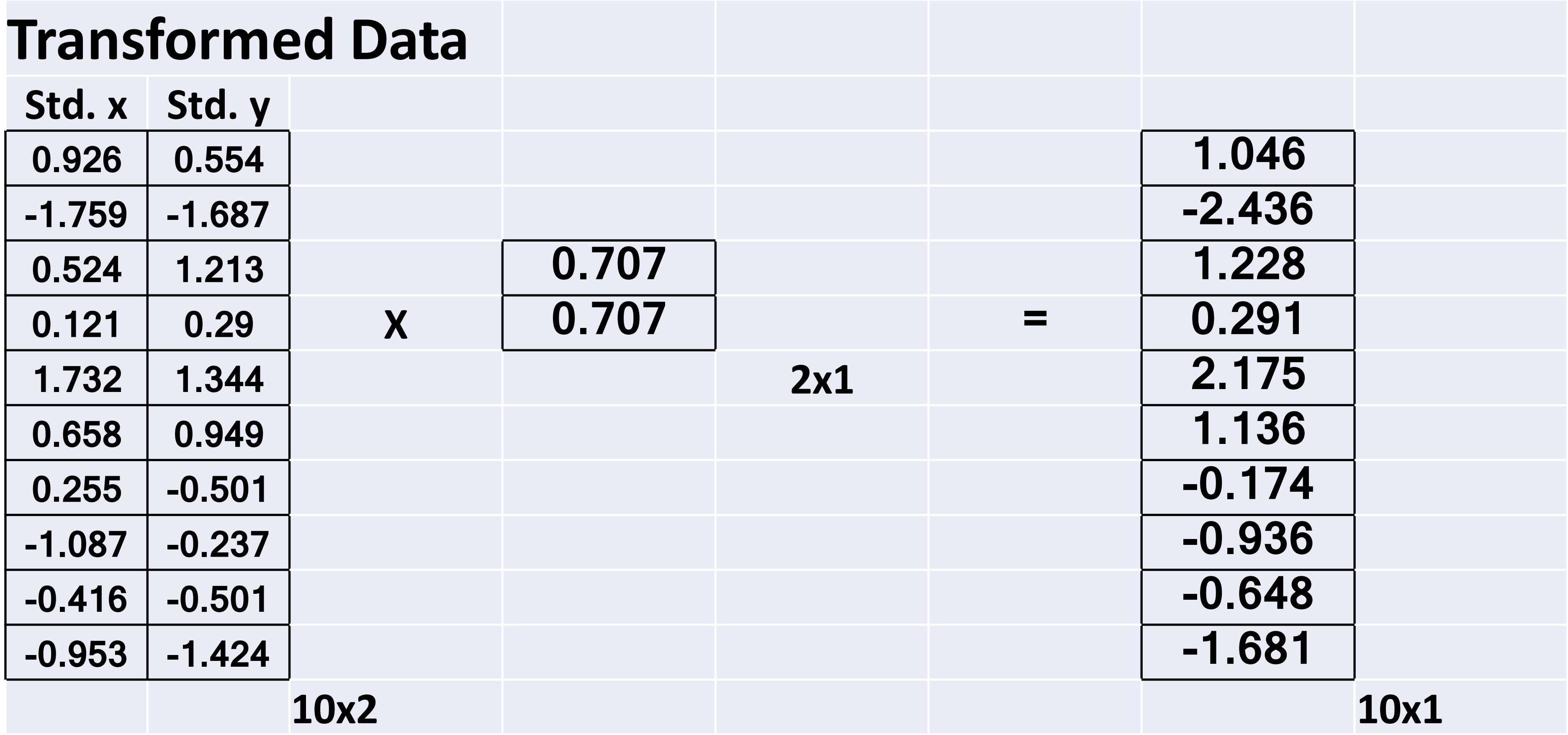
	covariance	
	x	**y**
x	1.111	0.978
y	0.978	1.111

Calculate the eigen values and the eigen vector	
E1	0.133
E2	2.089

Contribution of E1	**0.059856**	**6%**
Contribution of E2	**0.940144**	**94%**

			Unit Vector
Eigen vector for E1		**-1**	**-0.707**
		1	**0.707**
			Unit Vector
Eigen vector for E2		**1**	**0.707**
		1	**0.707**

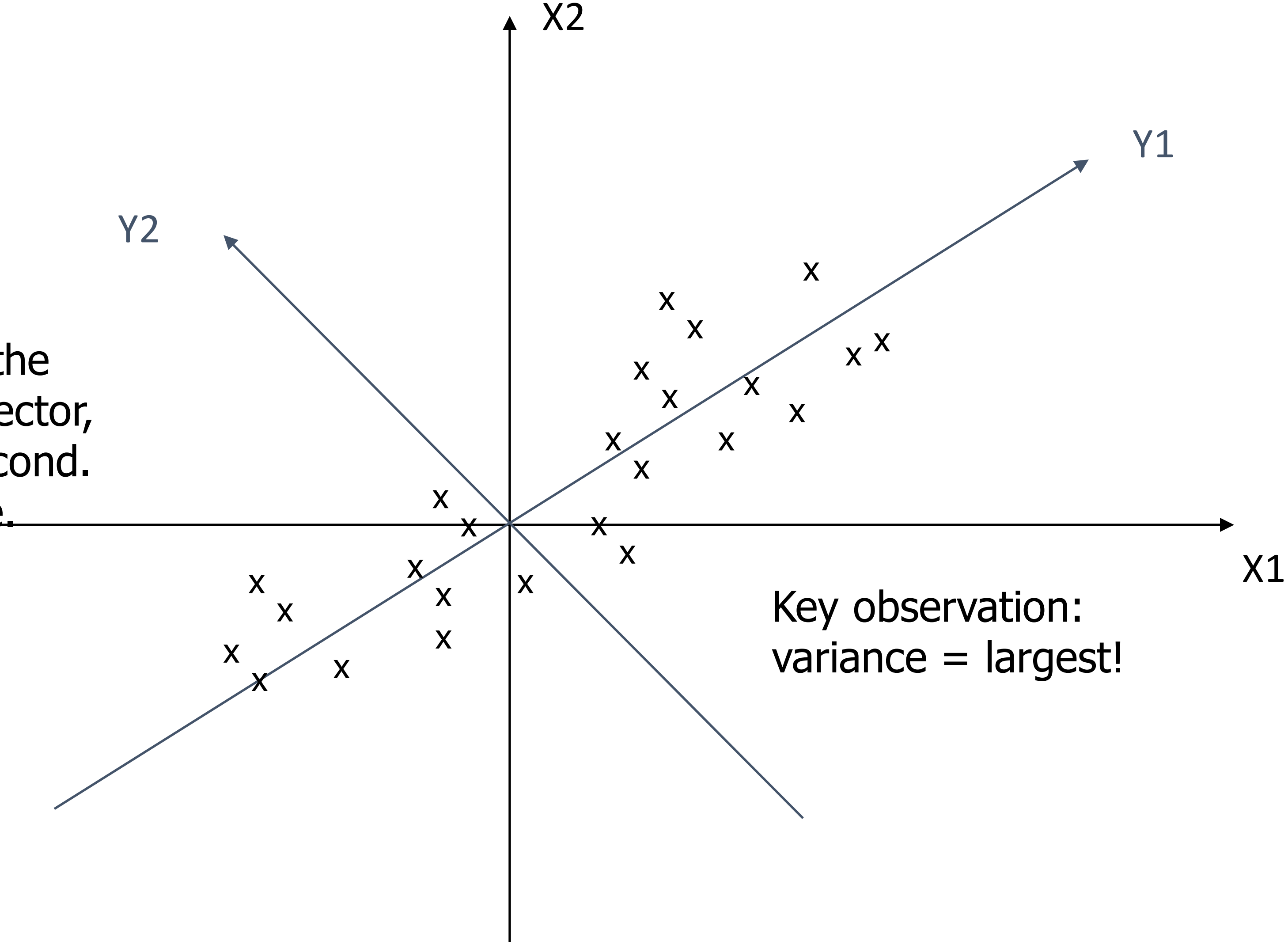
PCA – Example 2



Transformed Data

Std. x	Std. y						
0.926	0.554					1.046	
-1.759	-1.687					-2.436	
0.524	1.213		0.707			1.228	
0.121	0.29	X	0.707		=	0.291	
1.732	1.344			2x1		2.175	
0.658	0.949					1.136	
0.255	-0.501					-0.174	
-1.087	-0.237					-0.936	
-0.416	-0.501					-0.648	
-0.953	-1.424					-1.681	
		10x2					10x1

Principal Component Analysis



Note: Y1 is the first eigen vector, Y2 is the second. Y2 ignorable.

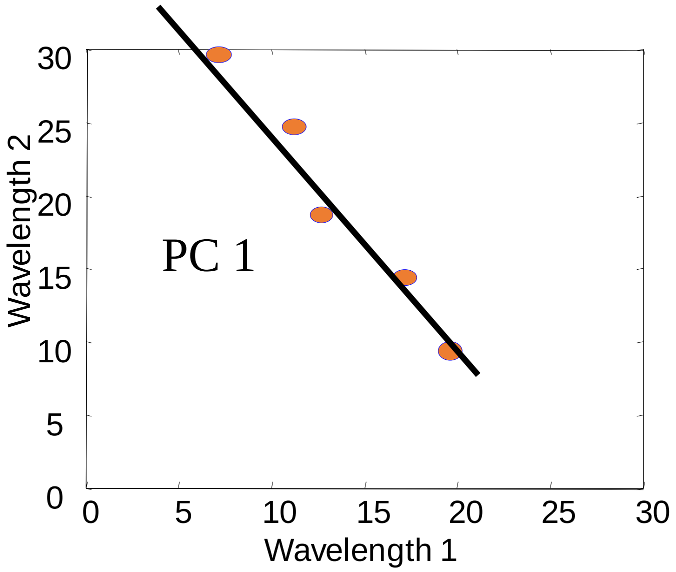
Key observation: variance = largest!

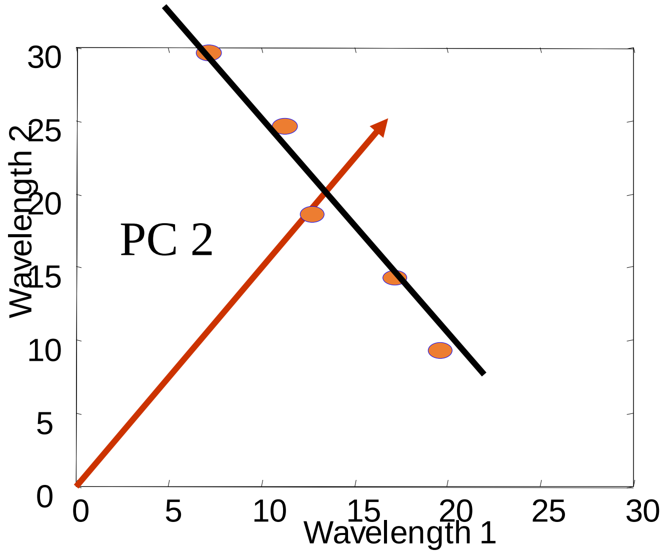
Data Presentation

- Better presentation than ordinate axes?
- Do we need a **n** dimension space to view data? Say if n=50?
- How to find the ‘best’ low dimension space that conveys maximum useful information?
- One answer: Find “Principal Components”

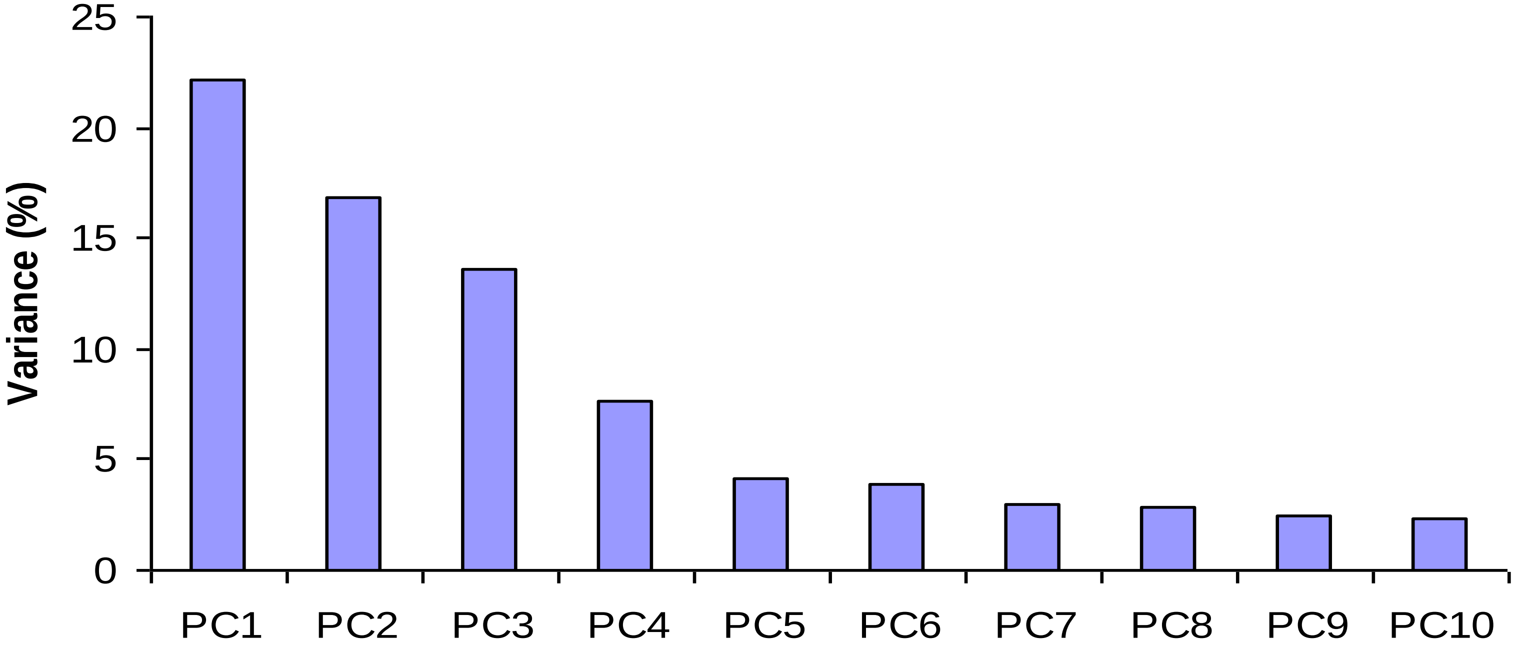
Principal Components

- All principal components (PCs) start at the origin of the ordinate axes.
- First PC is direction of maximum variance from origin
- Subsequent PCs are orthogonal to 1st PC and describe maximum residual variance





Principal components - Variance



PCA - Summary

- An exploratory technique used to reduce the dimensionality of the data set to 2D or 3D
- Can be used to:
  - Reduce number of dimensions in data
  - Find patterns in high-dimensional data
  - Visualize data of high dimensionality
- Example applications:
  - Face recognition
  - Image compression
  - Gene expression analysis

Exercise

- The following measurements are made on seven individuals in a random sample from a population. Identify the principal components of the sample and hence derive the new data set. Reduce the dimension of the sample data set if the contribution of the least significant component is less than 15%.

X	Y
1.0	1.0
1.5	2.0
3.0	4.0
5.0	7.0
3.5	5.0
4.5	5.0
3.5	4.5

Factor Analysis (FA)

- Factor analysis is a statistical tool to examine the inter-relationships among various variables
- It investigates several variables simultaneously and tries to locate them into a small number of dimensions called factors
- It helps in summarization of the data and identifies the latent dimensions or factors in the dataset
- The primary aim is data reduction as the variables that are highly co-related to each other are identified and represented as single factor
- The variables that do not correlate are identified as independent factors (orthogonal factors)
- In general, FA identifies a small number of factors which are orthogonal to each other (i.e.) they lie at right angles to each other when graphed

FA - Purpose

- To identify the underlying structure of relationships among the variables and classify them into homogenous group of clusters referred to as factors
- Data reduction/dimensionality reduction
- To remove redundancy from a set of correlated variables
- To understand the data by identifying the relationships among the variables
  - To remove duplicate variables from correlated variables
  - To identify orthogonal factors that are independent of each other
- We start from a large set of variables and identify the corelated and non-corelated variables in it

FA - Assumptions

- The dataset should be interval in nature (i.e.) the variables of the dataset should be represented as intervals
- FA can also be done if the variables are represented in ordinal scale with scores represented in Likert scale form
- Variables subjected to FA should be linearly correlated to each other
- Variables should also exhibit moderate to high correlation among each other
- FA is used in development of psychometric tests (say, have to find the verbal and the mathematical intelligence of candidates based on the aptitude questionnaire answered by them)

FA - Types

- FA is of two types: Exploratory FA (EFA) and Confirmatory FA (CFA)
- EFA –
  - This is used to identify the total number of factors that exist for the given correlated variables
  - It also helps in determining the correlation between the variables and the factors in the dataset
- CFA –
  - This is used to confirm or validate the priori theorized or hypothesized factor structure and its underlying variables
  - Here, the researcher already knows the variables and their factors

FA - Methods

Major methods are

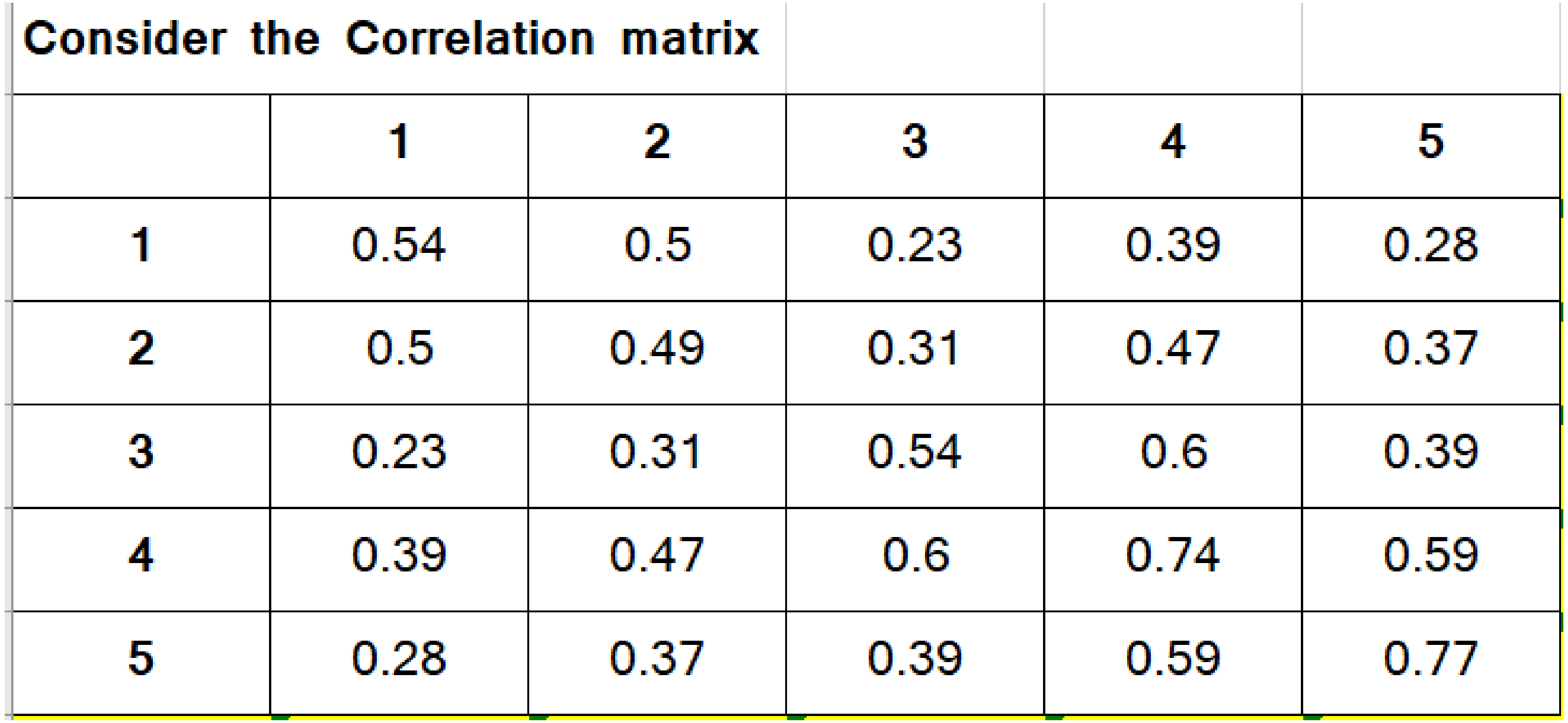
- Centroid method
- Summation method
- Principal axes method
- Principal components method

FA – diagonal values

- The basic data required for extraction of factors is the correlation matrix
- For any correlation matrix derived from the dataset, the diagonal cell values are empty
  - These cells are filled with 1.00 in PCA method
  - In centroid method, the communalities are placed in the diagonal cells (The **communality** is the sum of the squared component loadings up to the number of components you extract)

  This is done to calculate the factors
- The choice of the diagonal cells, effects the number of factors extracted and the loadings of each factor on each test

FA – Example 1

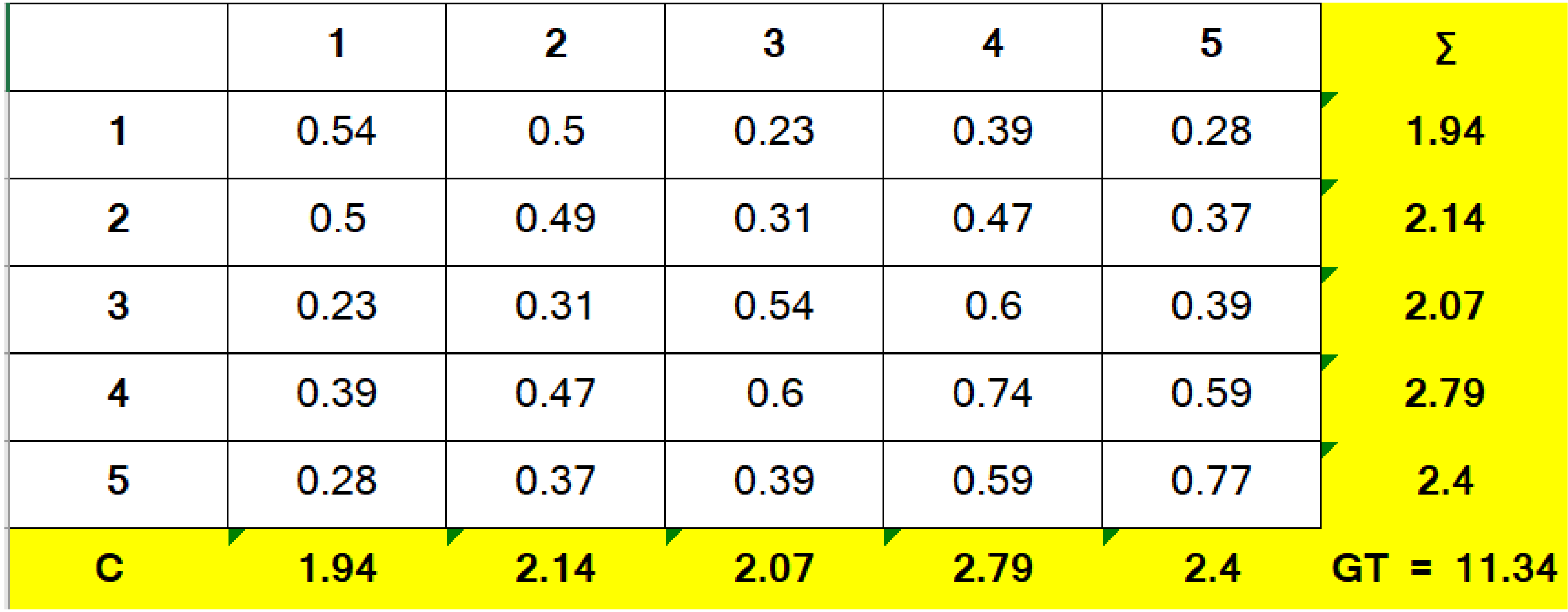


Consider the Correlation matrix

	1	2	3	4	5
1	0.54	0.5	0.23	0.39	0.28
2	0.5	0.49	0.31	0.47	0.37
3	0.23	0.31	0.54	0.6	0.39
4	0.39	0.47	0.6	0.74	0.59
5	0.28	0.37	0.39	0.59	0.77

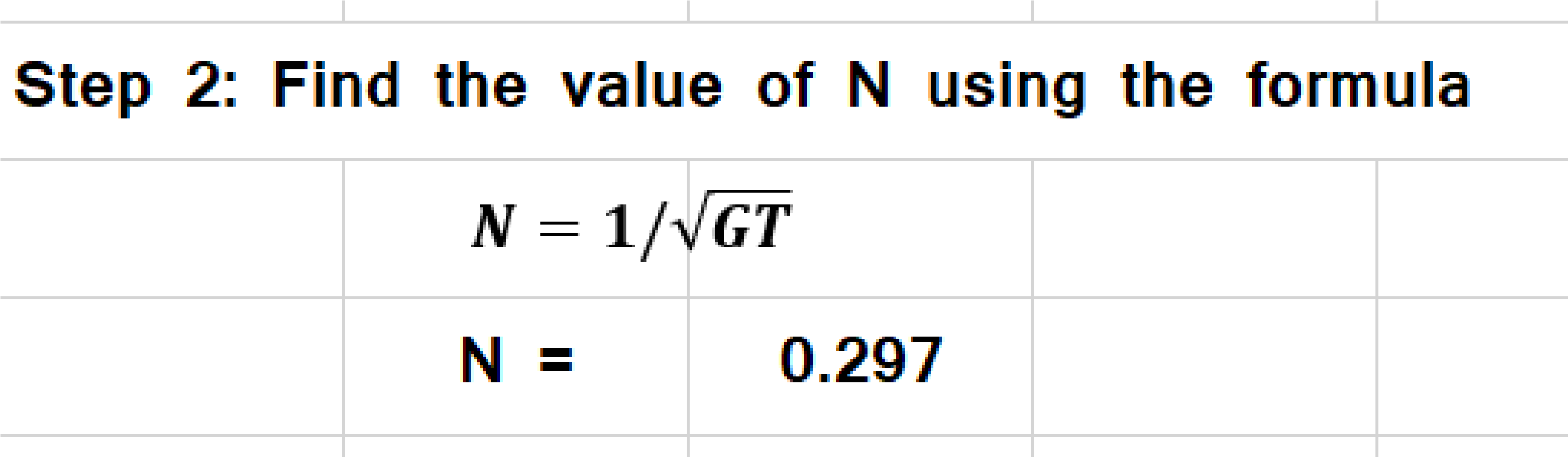
FA – Example 1

Step 1: Add the values column wise and row wise and name it as ∑ and C respectively



	1	2	3	4	5	∑
1	0.54	0.5	0.23	0.39	0.28	**1.94**
2	0.5	0.49	0.31	0.47	0.37	**2.14**
3	0.23	0.31	0.54	0.6	0.39	**2.07**
4	0.39	0.47	0.6	0.74	0.59	**2.79**
5	0.28	0.37	0.39	0.59	0.77	**2.4**
C	**1.94**	**2.14**	**2.07**	**2.79**	**2.4**	**GT = 11.34**

FA – Example 1

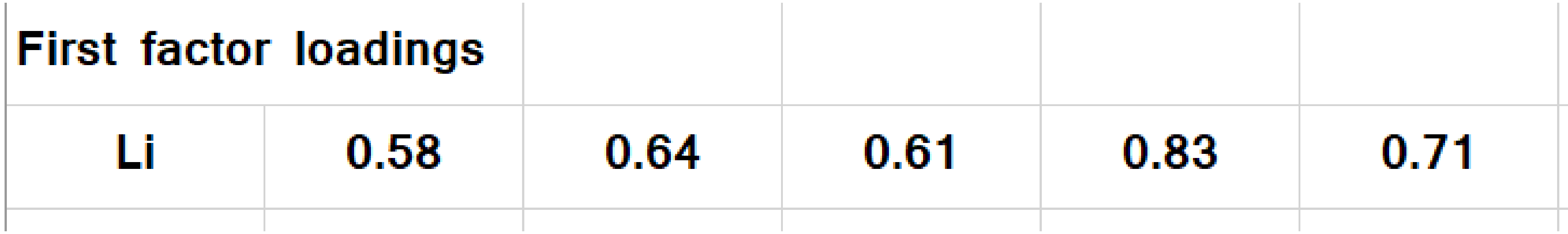


Step 2: Find the value of N using the formula

$$N = 1/\sqrt{GT}$$

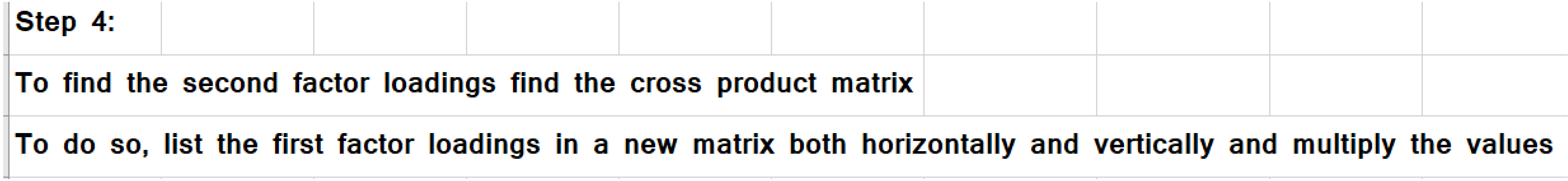
	N =	0.297

Step 3: Multiply each column sum with N to get the first factor loadings (i.e.) Li = C*N



First factor loadings					
Li	0.58	0.64	0.61	0.83	0.71

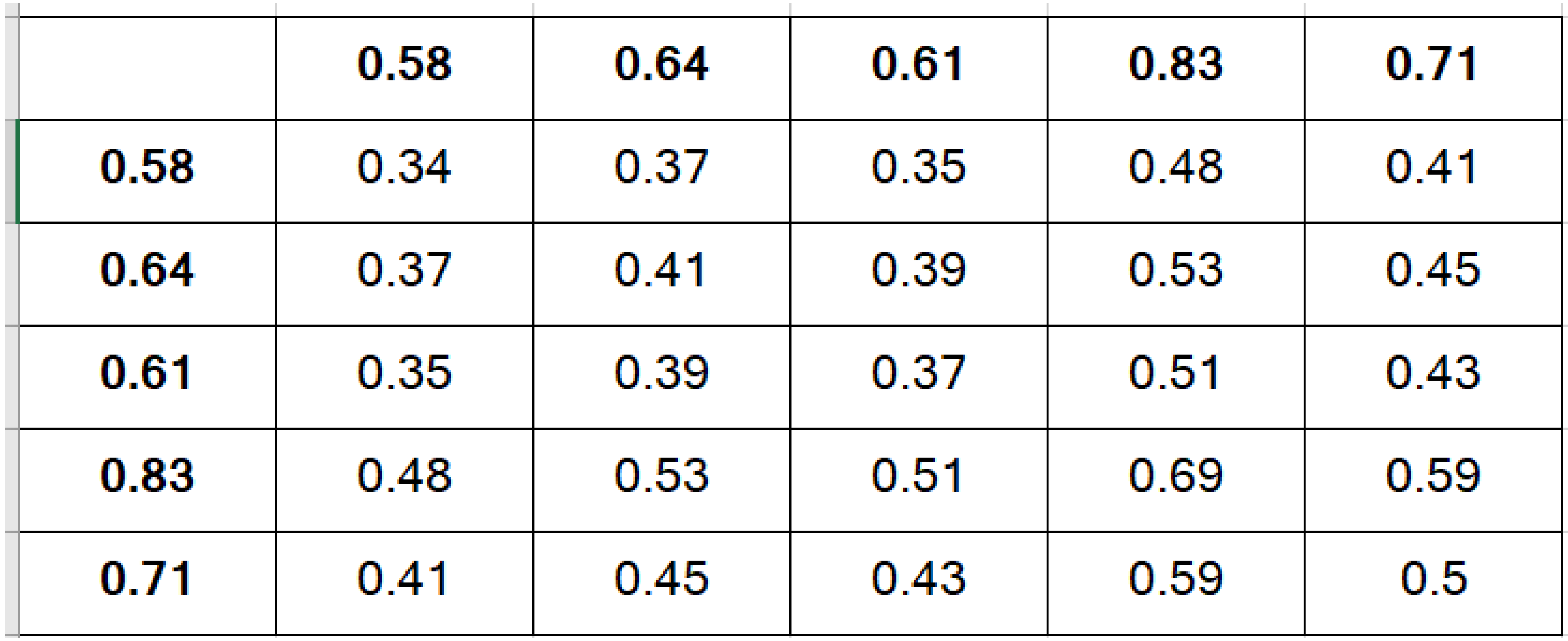
FA – Example 1



Step 4:

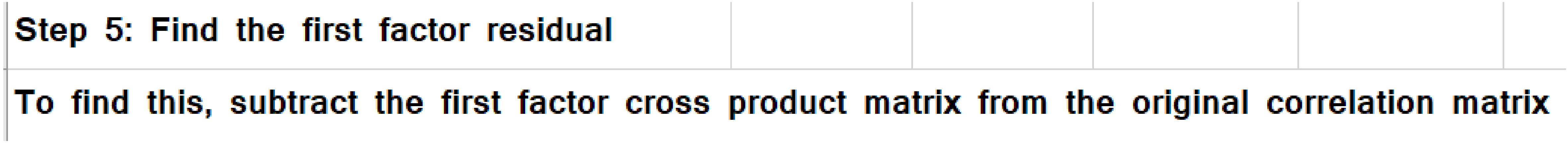
To find the second factor loadings find the cross product matrix

To do so, list the first factor loadings in a new matrix both horizontally and vertically and multiply the values



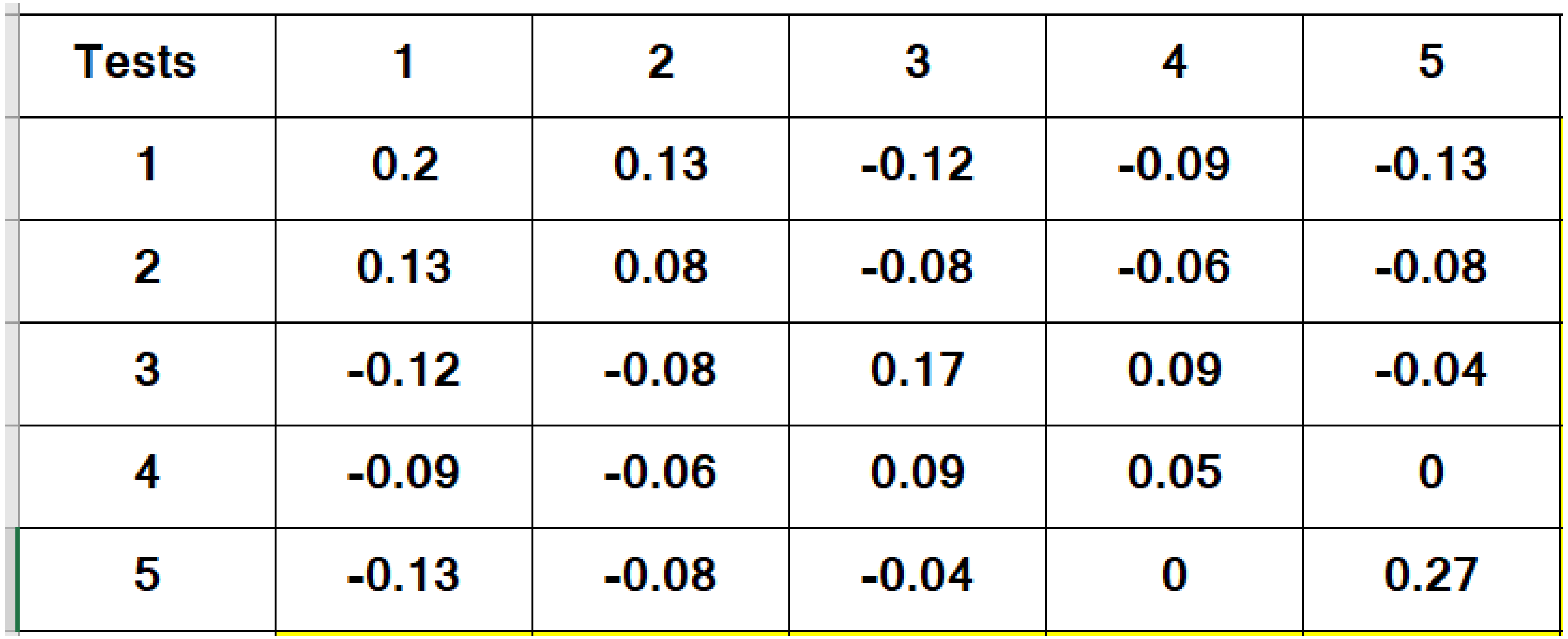
	0.58	**0.64**	**0.61**	**0.83**	**0.71**
0.58	0.34	0.37	0.35	0.48	0.41
0.64	0.37	0.41	0.39	0.53	0.45
0.61	0.35	0.39	0.37	0.51	0.43
0.83	0.48	0.53	0.51	0.69	0.59
0.71	0.41	0.45	0.43	0.59	0.5

FA – Example 1



Step 5: Find the first factor residual

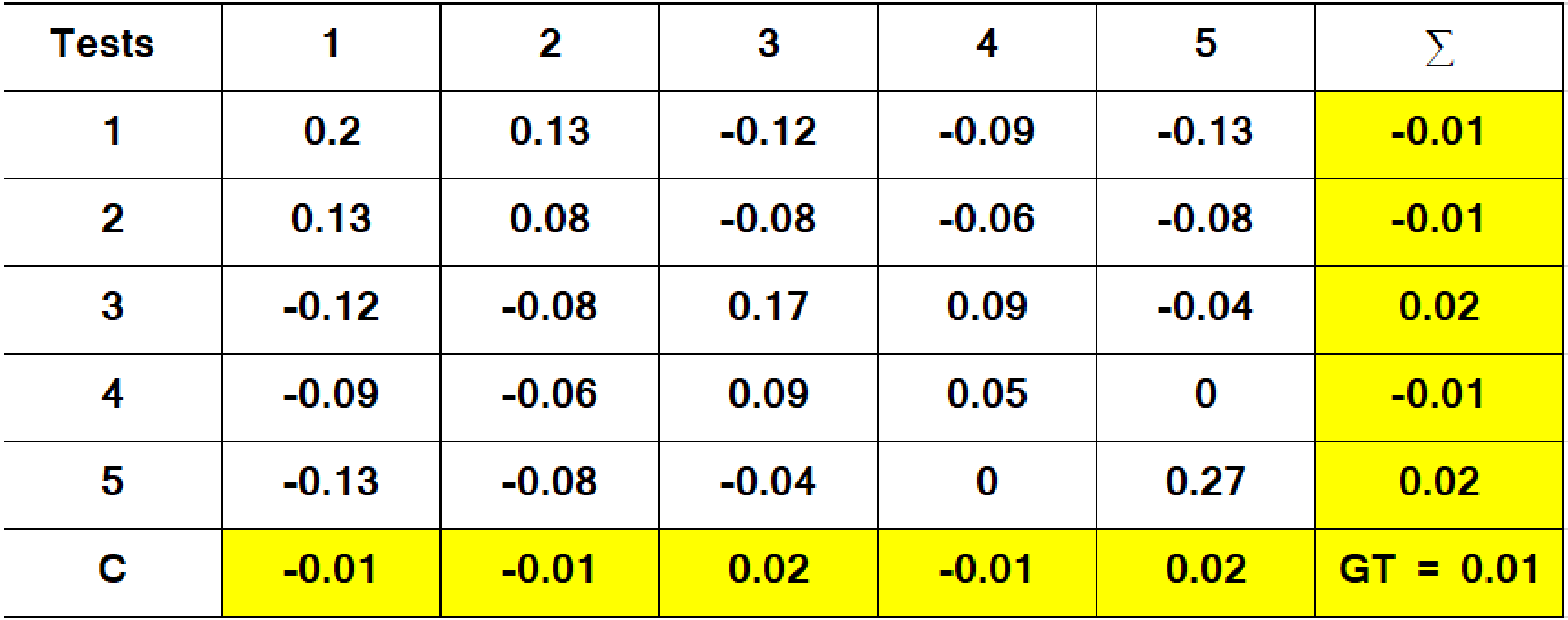
To find this, subtract the first factor cross product matrix from the original correlation matrix



Tests	1	2	3	4	5
1	0.2	0.13	-0.12	-0.09	-0.13
2	0.13	0.08	-0.08	-0.06	-0.08
3	-0.12	-0.08	0.17	0.09	-0.04
4	-0.09	-0.06	0.09	0.05	0
5	-0.13	-0.08	-0.04	0	0.27

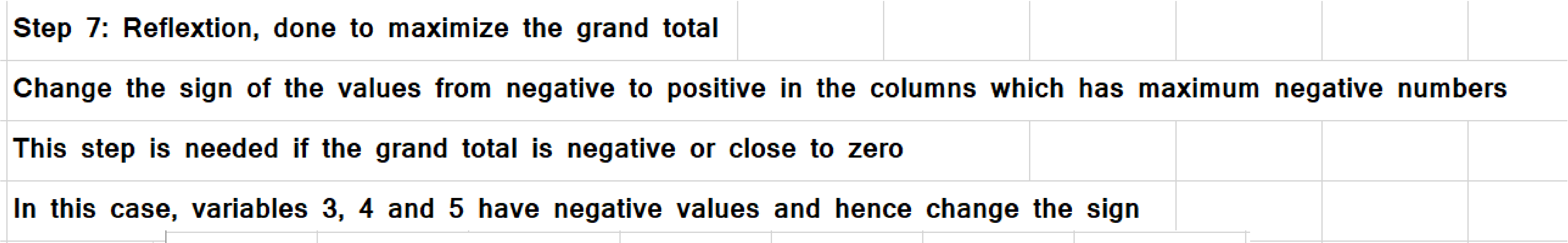
FA – Example 1

Step 6: Add values column wise and row wise and denote by symbols ∑ and C respectively and calculate the grand total



Tests	1	2	3	4	5	∑
1	0.2	0.13	-0.12	-0.09	-0.13	-0.01
2	0.13	0.08	-0.08	-0.06	-0.08	-0.01
3	-0.12	-0.08	0.17	0.09	-0.04	0.02
4	-0.09	-0.06	0.09	0.05	0	-0.01
5	-0.13	-0.08	-0.04	0	0.27	0.02
C	-0.01	-0.01	0.02	-0.01	0.02	GT = 0.01

FA – Example 1

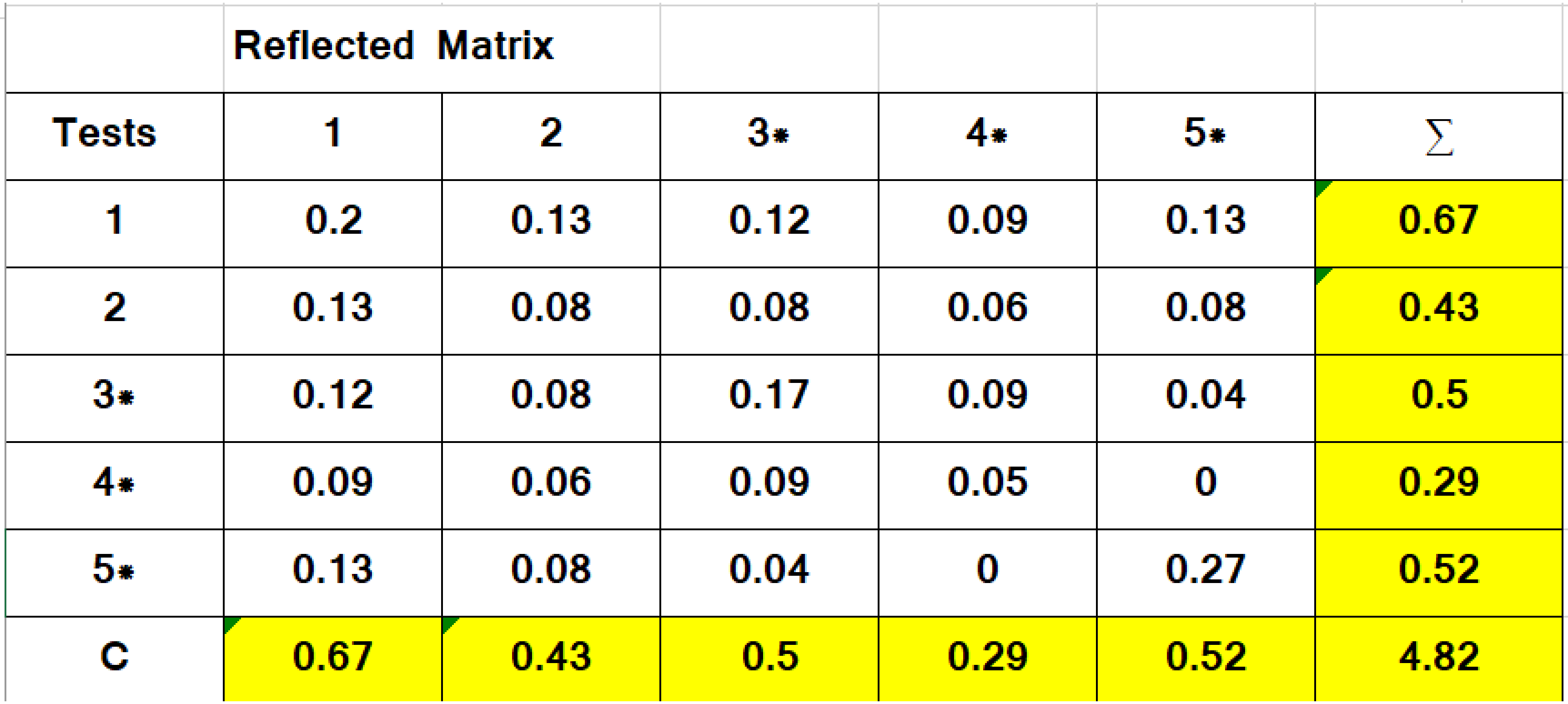


Step 7: Reflextion, done to maximize the grand total

Change the sign of the values from negative to positive in the columns which has maximum negative numbers

This step is needed if the grand total is negative or close to zero

In this case, variables 3, 4 and 5 have negative values and hence change the sign



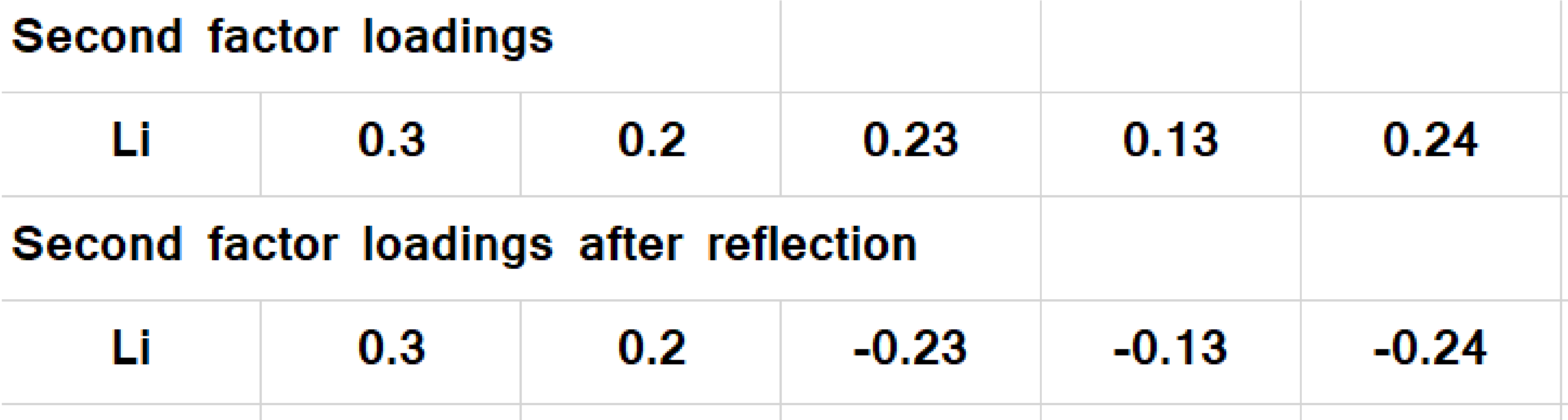
	Reflected Matrix					
Tests	1	2	3*	4*	5*	∑
1	0.2	0.13	0.12	0.09	0.13	0.67
2	0.13	0.08	0.08	0.06	0.08	0.43
3*	0.12	0.08	0.17	0.09	0.04	0.5
4*	0.09	0.06	0.09	0.05	0	0.29
5*	0.13	0.08	0.04	0	0.27	0.52
C	0.67	0.43	0.5	0.29	0.52	4.82

Reflection is also referred to as Rotation

FA – Example 1

Step 8: Find the second factor loadings same as described above (Repeat steps 1 to 3)

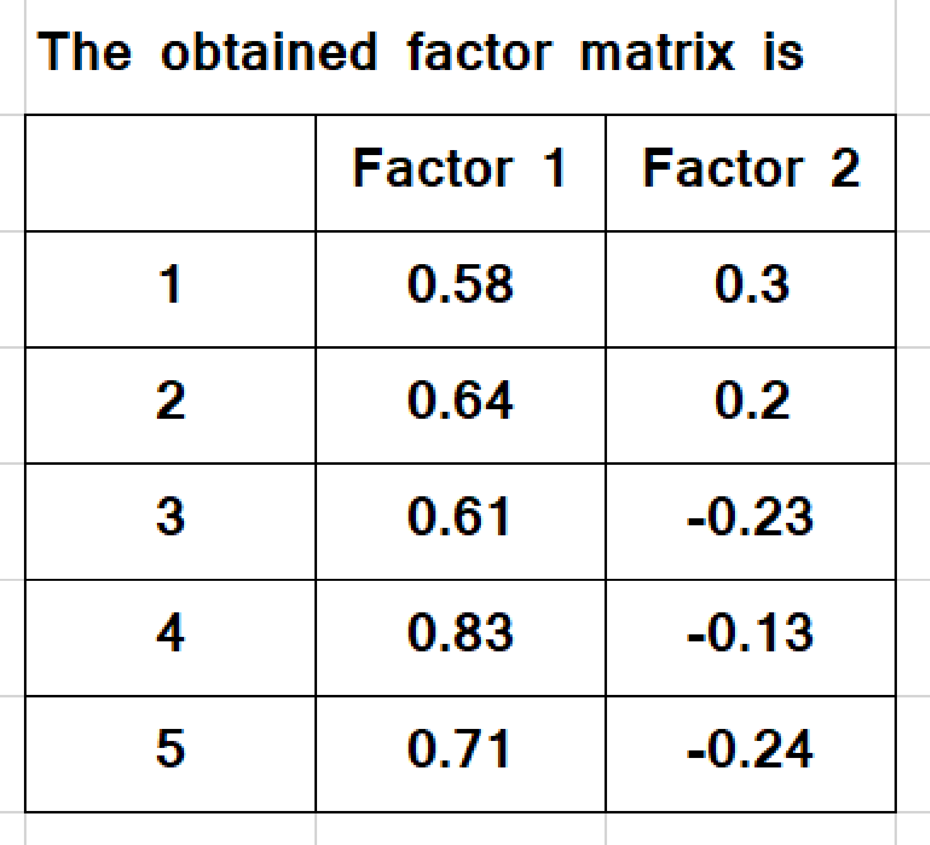
N = 0.455



Second factor loadings					
Li	0.3	0.2	0.23	0.13	0.24
Second factor loadings after reflection					
Li	0.3	0.2	-0.23	-0.13	-0.24

FA – Example 1

- **Step 9: This process can be continued to find the 3rd, 4th etc. factory loadings**



The obtained factor matrix is

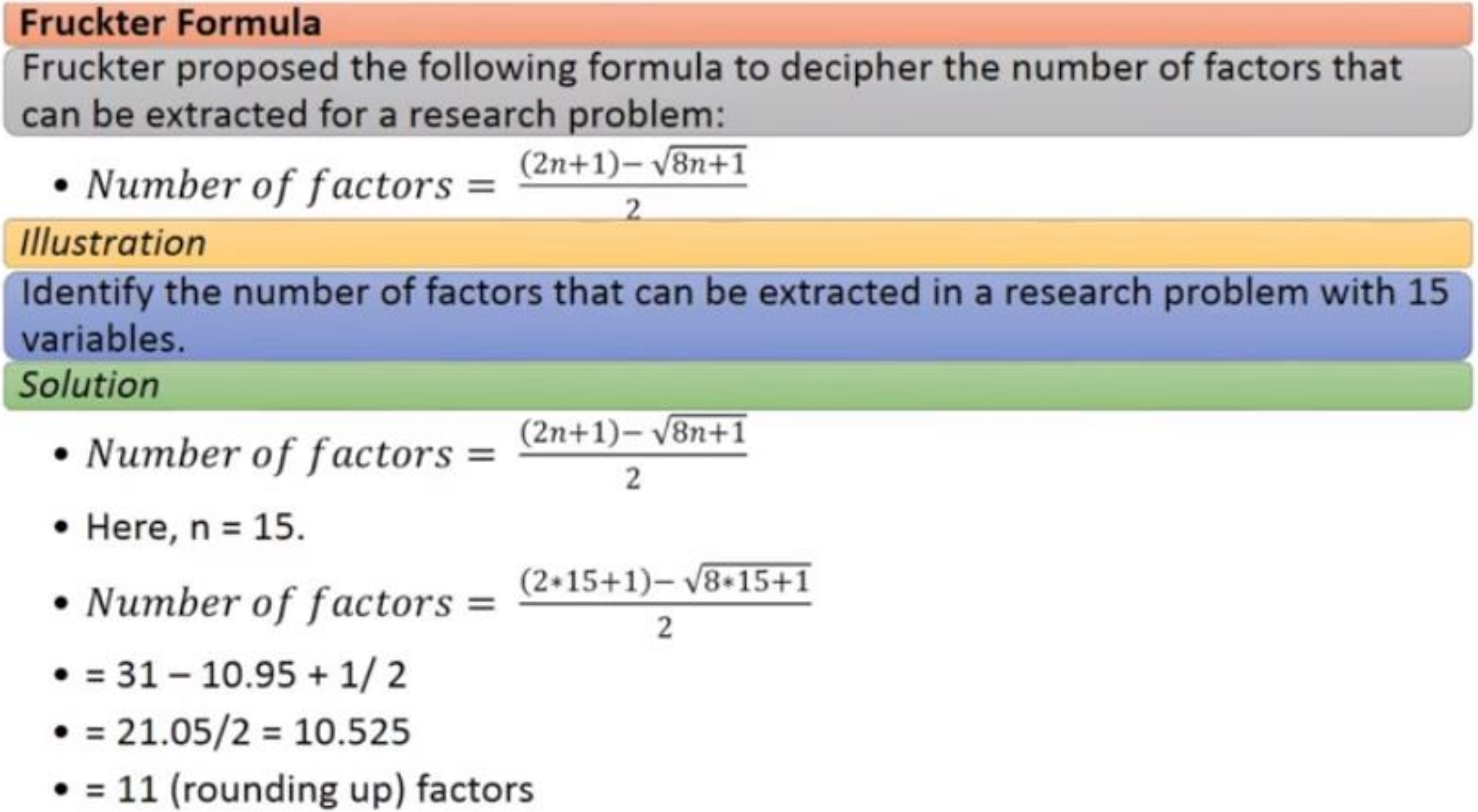
	Factor 1	Factor 2
1	0.58	0.3
2	0.64	0.2
3	0.61	-0.23
4	0.83	-0.13
5	0.71	-0.24

FA – How many factors can be extracted?

- For a 10x10 correlation matrix, maximum number of factors that can be extracted is maximum 10
- But, it can be decided based on the following 3 methods

1. Fruckter formula
2. Eigen value index - calculate the eigen values and extract the factors only if the eigen value is 1 or more than 1
3. Residual correlation matrix - stop the factor extraction if most of the values in the residual correlation matrix is zero or approaching zero

Fruckter formula



Fruckter Formula

Fruckter proposed the following formula to decipher the number of factors that can be extracted for a research problem:

- $Number\ of\ factors = \frac{(2n+1)-\sqrt{8n+1}}{2}$

Illustration

Identify the number of factors that can be extracted in a research problem with 15 variables.

Solution

- $Number\ of\ factors = \frac{(2n+1)-\sqrt{8n+1}}{2}$
- Here, n = 15.
- $Number\ of\ factors = \frac{(2*15+1)-\sqrt{8*15+1}}{2}$
- = 31 – 10.95 + 1/ 2
- = 21.05/2 = 10.525
- = 11 (rounding up) factors

Eigen value index

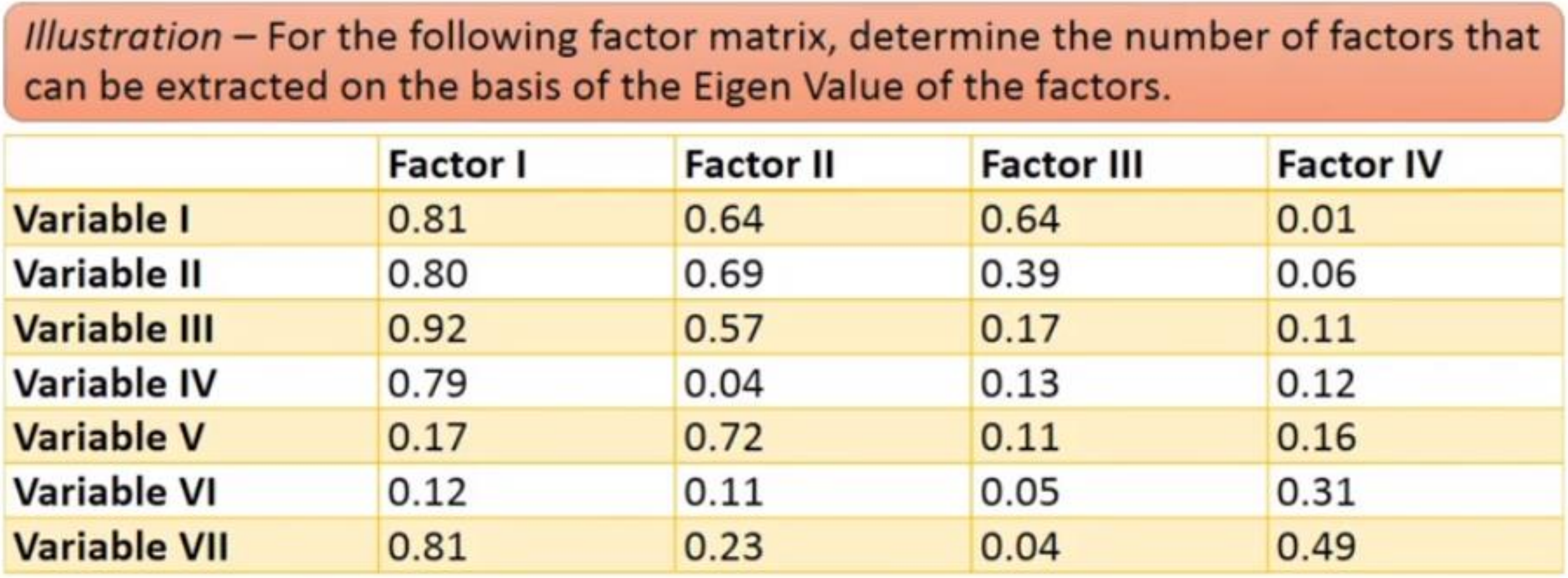
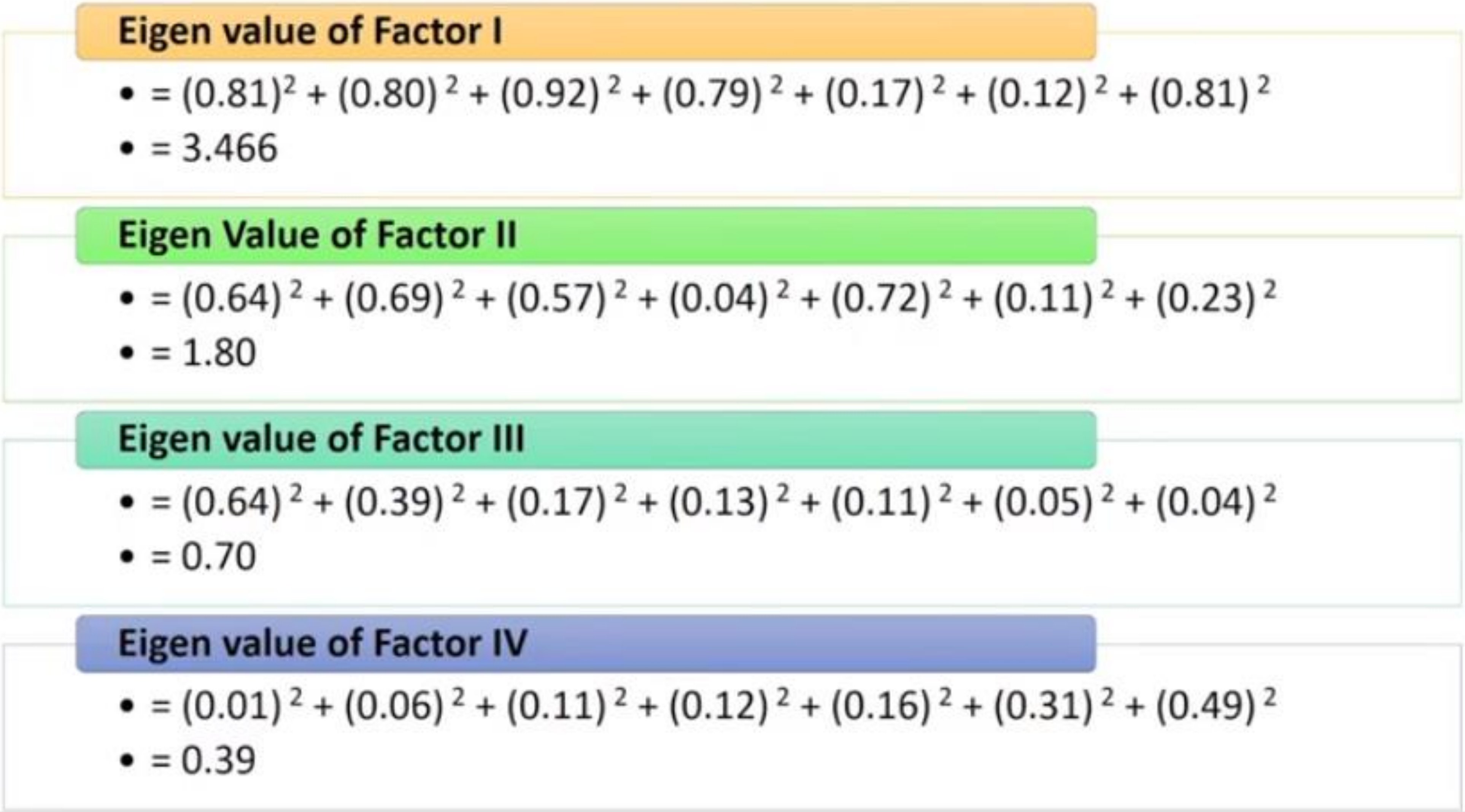


Illustration – For the following factor matrix, determine the number of factors that can be extracted on the basis of the Eigen Value of the factors.

	Factor I	Factor II	Factor III	Factor IV
Variable I	0.81	0.64	0.64	0.01
Variable II	0.80	0.69	0.39	0.06
Variable III	0.92	0.57	0.17	0.11
Variable IV	0.79	0.04	0.13	0.12
Variable V	0.17	0.72	0.11	0.16
Variable VI	0.12	0.11	0.05	0.31
Variable VII	0.81	0.23	0.04	0.49

Eigen value index



Eigen value of Factor I

- $= (0.81)^2 + (0.80)^2 + (0.92)^2 + (0.79)^2 + (0.17)^2 + (0.12)^2 + (0.81)^2$
- $= 3.466$

Eigen Value of Factor II

- $= (0.64)^2 + (0.69)^2 + (0.57)^2 + (0.04)^2 + (0.72)^2 + (0.11)^2 + (0.23)^2$
- $= 1.80$

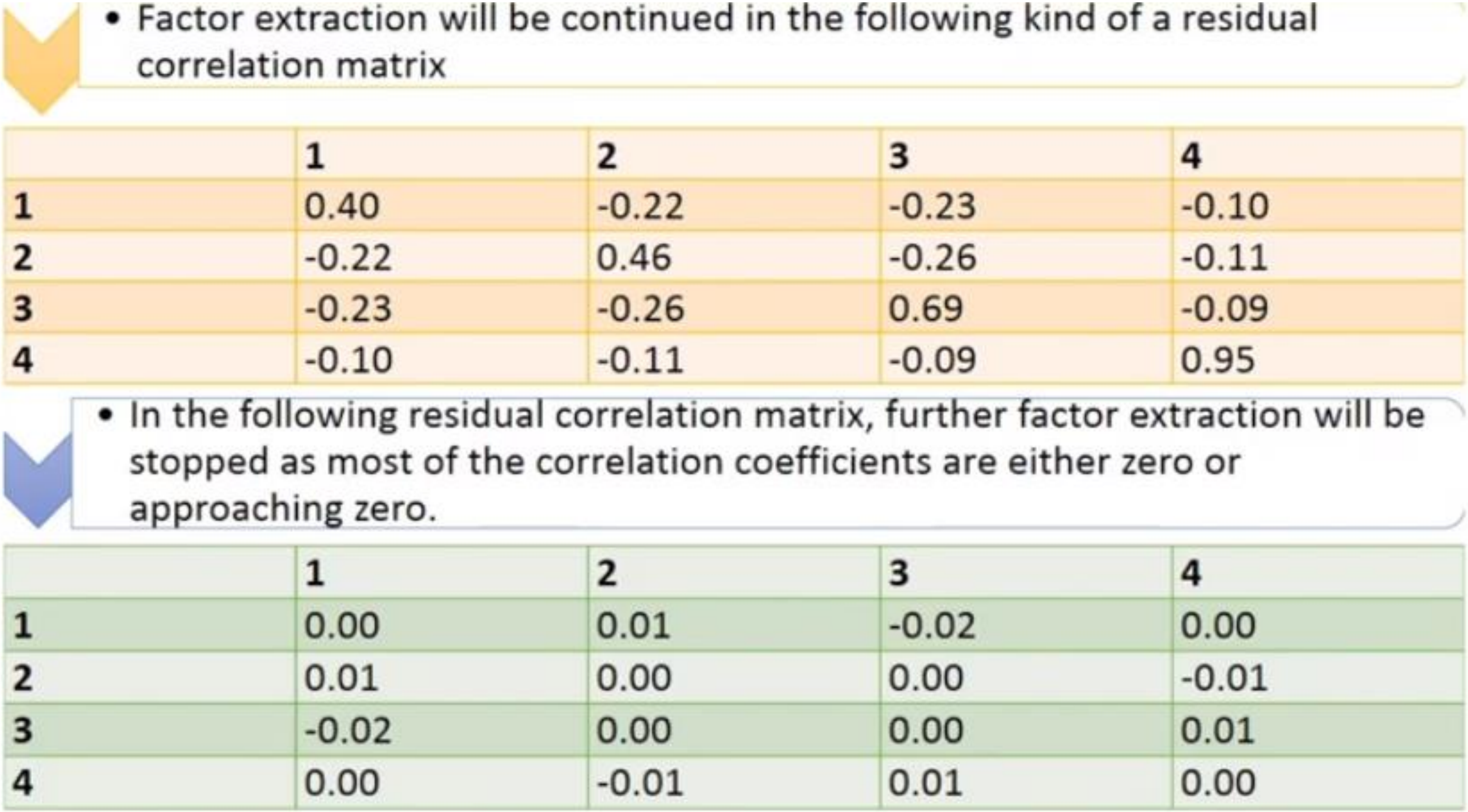
Eigen value of Factor III

- $= (0.64)^2 + (0.39)^2 + (0.17)^2 + (0.13)^2 + (0.11)^2 + (0.05)^2 + (0.04)^2$
- $= 0.70$

Eigen value of Factor IV

- $= (0.01)^2 + (0.06)^2 + (0.11)^2 + (0.12)^2 + (0.16)^2 + (0.31)^2 + (0.49)^2$
- $= 0.39$

Residual correlation matrix



- Factor extraction will be continued in the following kind of a residual correlation matrix

	1	2	3	4
1	0.40	-0.22	-0.23	-0.10
2	-0.22	0.46	-0.26	-0.11
3	-0.23	-0.26	0.69	-0.09
4	-0.10	-0.11	-0.09	0.95

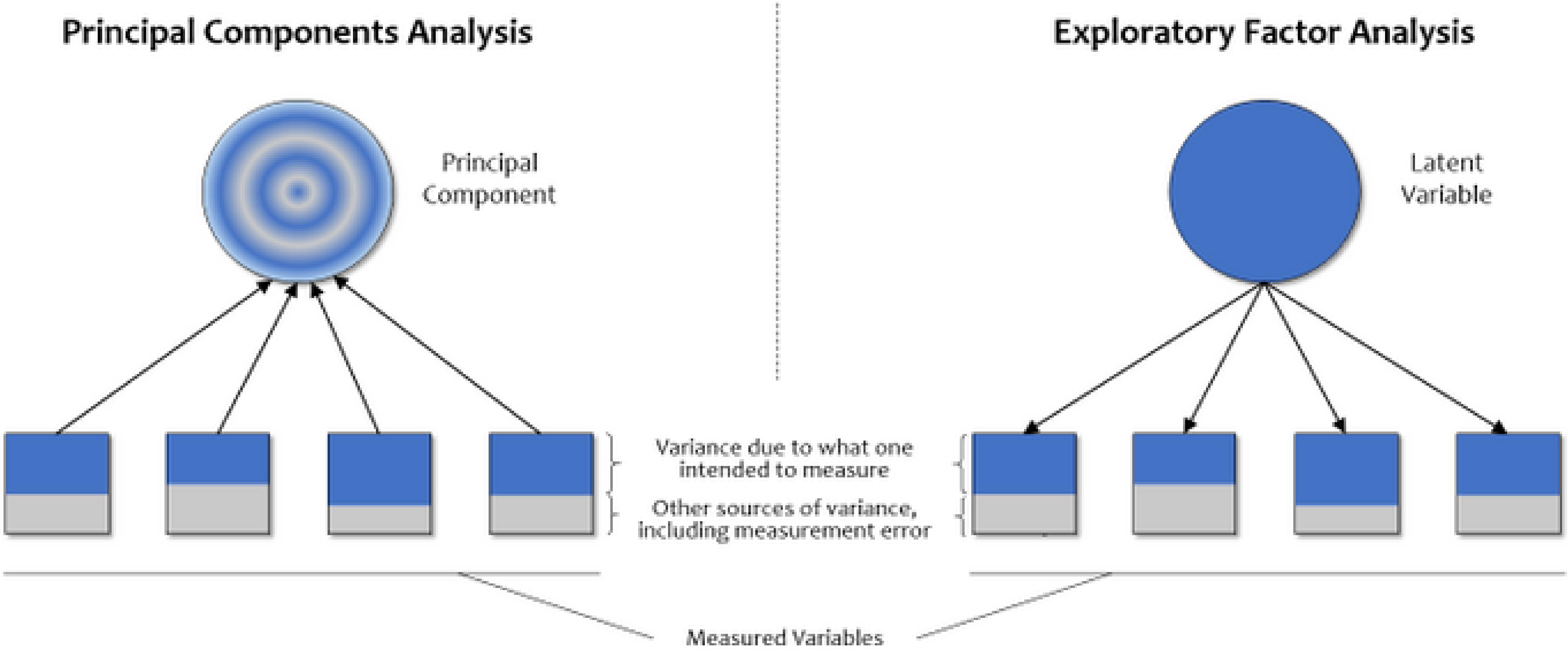
- In the following residual correlation matrix, further factor extraction will be stopped as most of the correlation coefficients are either zero or approaching zero.

	1	2	3	4
1	0.00	0.01	-0.02	0.00
2	0.01	0.00	0.00	-0.01
3	-0.02	0.00	0.00	0.01
4	0.00	-0.01	0.01	0.00

FA - Summary

- In FA, we are trying to find the latent variables that explains the pattern of observed variables
- Latent factors are assumed to exist but cannot be measured directly
- When not sure how many latent variables exist, we can us Exploratory FA (mostly preferred)
- When we can guess the number of latent variables and wanted to check with the model then we can go with Confirmatory FA

PCA vs FA


Principal Components Analysis
Exploratory Factor Analysis
Principal Component
Latent Variable
Variance due to what one intended to measure
Other sources of variance, including measurement error
Measured Variables

PCA vs FA

- PCA is useful for reducing the number of variables while retaining the most amount of information in the data, whereas EFA is useful for measuring unobserved (latent), error-free variables.
- When variables don’t have anything in common, EFA will not find a well-defined underlying factor, but PCA will find a well-defined principal component that explains the maximal amount of variance in the data.
- PCA is a technique for reducing the dimensionality of one’s data, whereas EFA is a technique for identifying and measuring variables that cannot be measured directly (i.e., latent variables or factors).
- When the goal is to measure an error-free latent variable but PCA is used, the component loadings will most likely be higher than they would’ve been if EFA was used. This would mislead analysts into thinking they have a well-defined, error-free factor when in fact they have a well-defined component that’s an amalgam of all the sources of variance in the data.
- When the goal is to get a small subset of variables that retain the most amount of variability in the data but EFA is used, the factor loadings will likely be lower than they would’ve been if PCA was used. This would mislead analysts into thinking they kept the maximal amount of variance in the data when in fact they kept the variance that’s in common across the measured variables.

Exercise

- Consider the covariance matrix given in table 1 and find the factors corresponding to it.

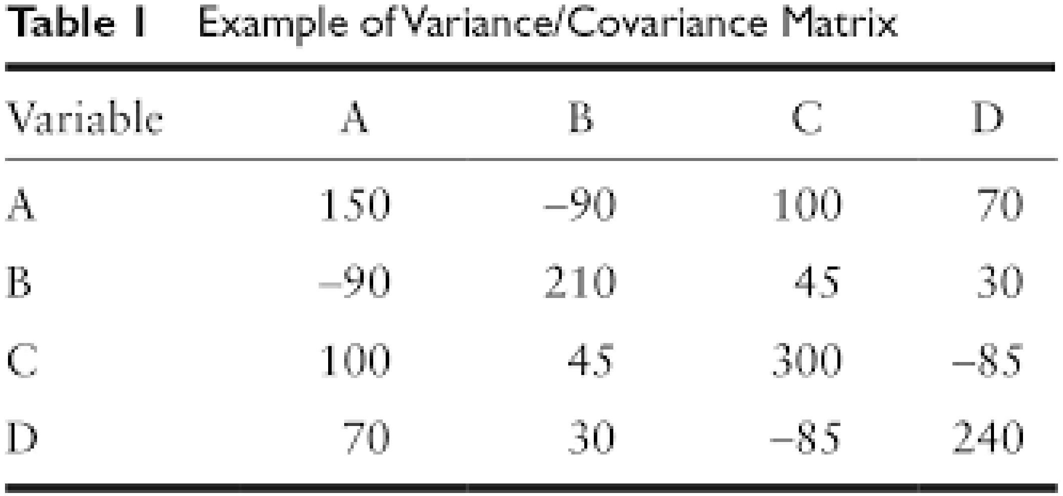


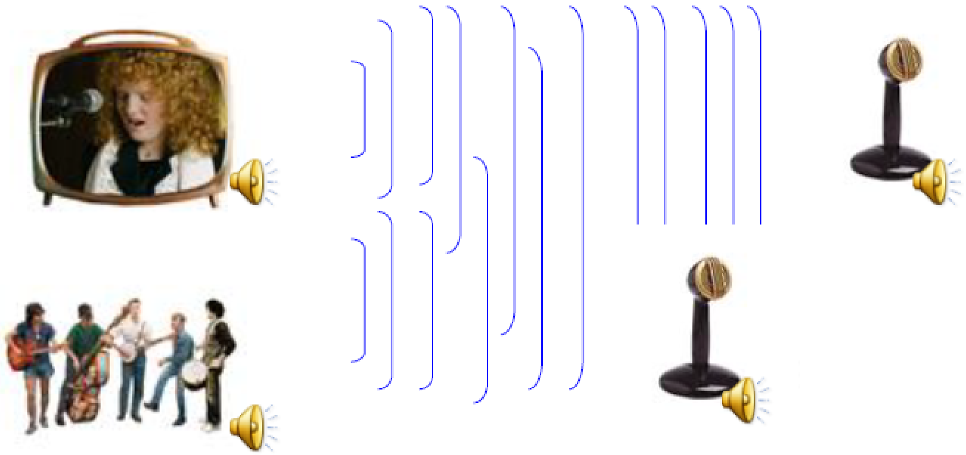
Table 1 Example of Variance/Covariance Matrix

Variable	A	B	C	D
A	150	–90	100	70
B	–90	210	45	30
C	100	45	300	–85
D	70	30	–85	240

Independent Component Analysis (ICA)

Motivation - Cocktail-Party Problem

- Simple scenario:
  - Two people speaking simultaneously in a room.
  - Speeches are recorded by two microphones in separate locations.



Motivation - Cocktail-Party Problem

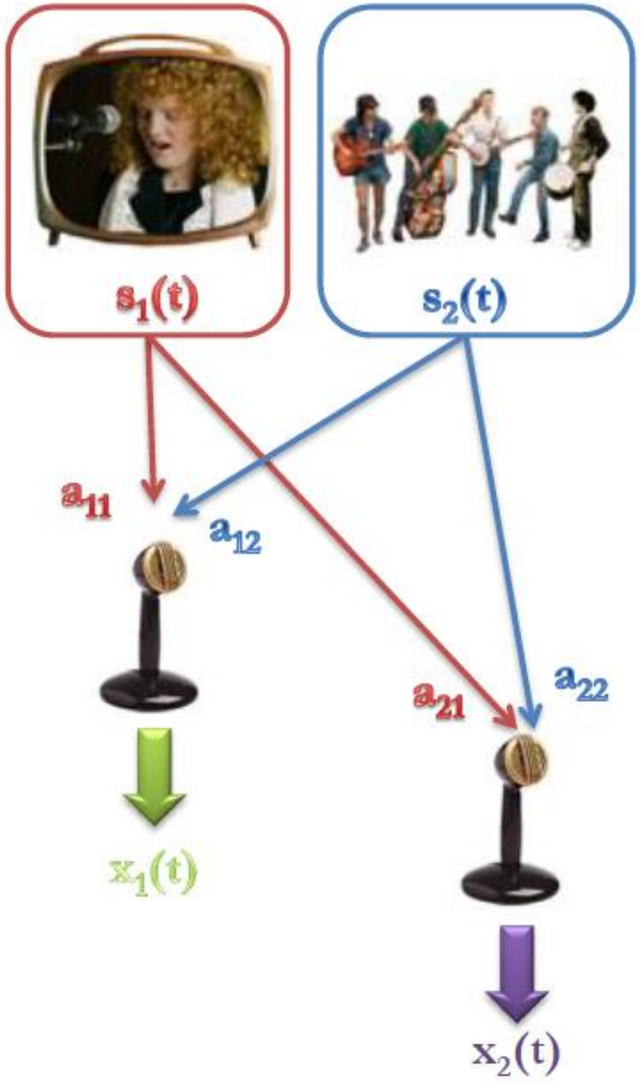
- Let $\mathbf{s_1(t)}$, $\mathbf{s_2(t)}$ be the speech signals emitted by the two speakers.
- Recorded time signals, by the two microphones, are denoted by $\mathbf{x_1(t)}$, $\mathbf{x_2(t)}$.
- The recorded time signals can be expressed as a linear equation:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t)$$

where parameters in matrix **A** depend on distances of the microphones to the speaker, along with other microphone properties

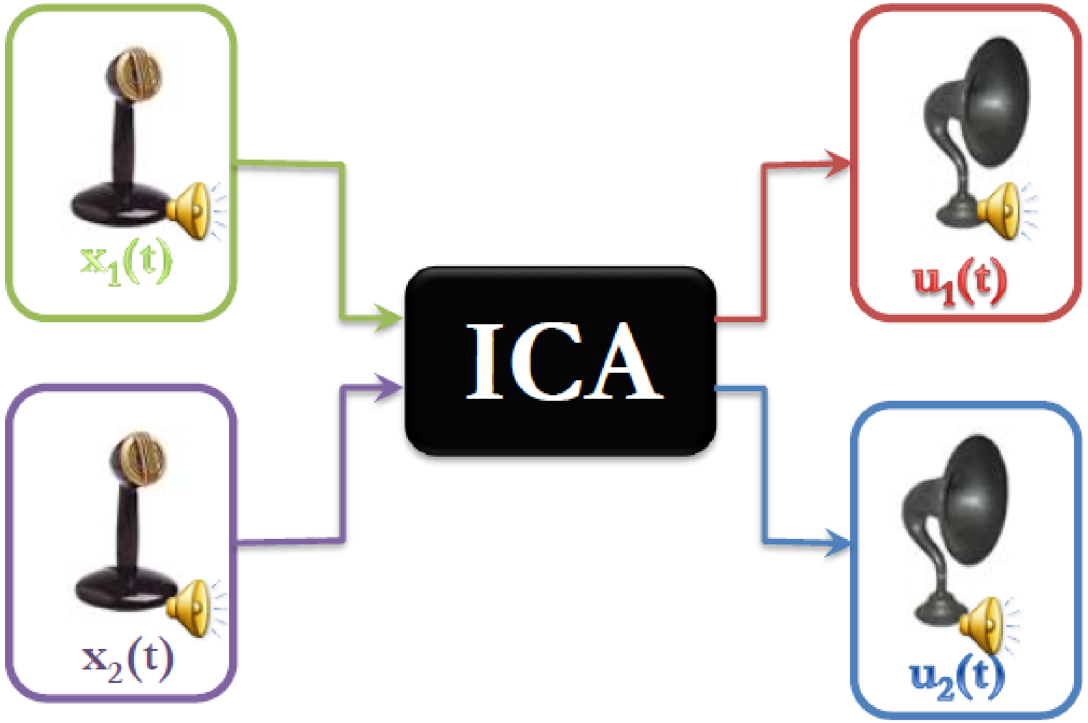
- Assume $\mathbf{s_1(t)}$ and $\mathbf{s_2(t)}$ are *statistically independent.*



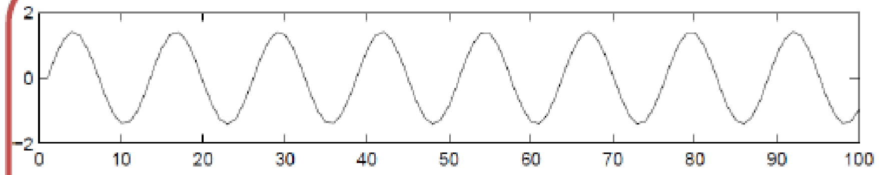
Motivation - Cocktail-Party Problem

Goal:

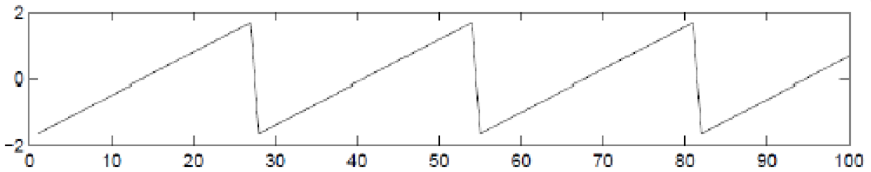
- Recover the unmixed speech signals, best estimate $\mathbf{u}_i(\mathbf{t})$, without knowing $\mathbf{A}$ or $\mathbf{s}_i(\mathbf{t})$.



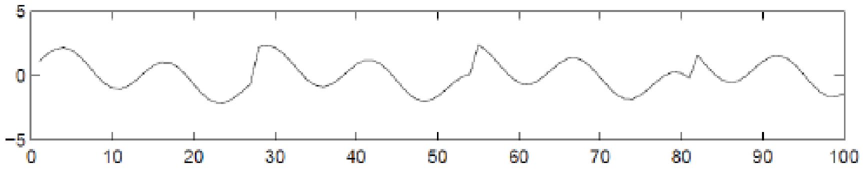
Motivation - Cocktail-Party Problem



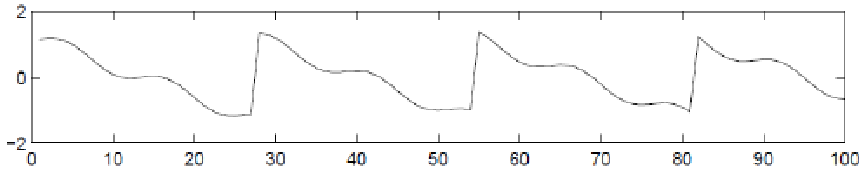
$s_1(t)$: original signal



$s_2(t)$: original signal

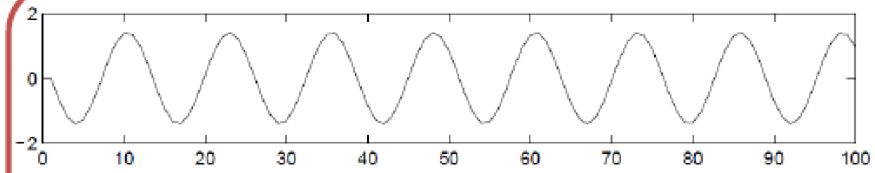


$x_1(t)$: observed mixture

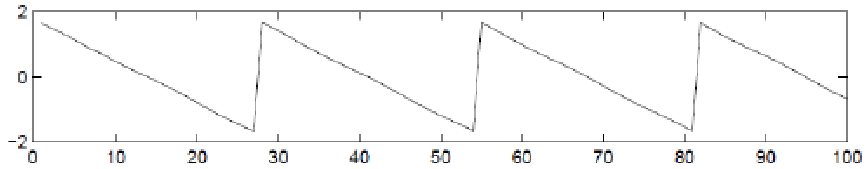


$x_2(t)$: observed mixture

ICA



$u_1(t)$: recovered estimate



$u_2(t)$: recovered estimate

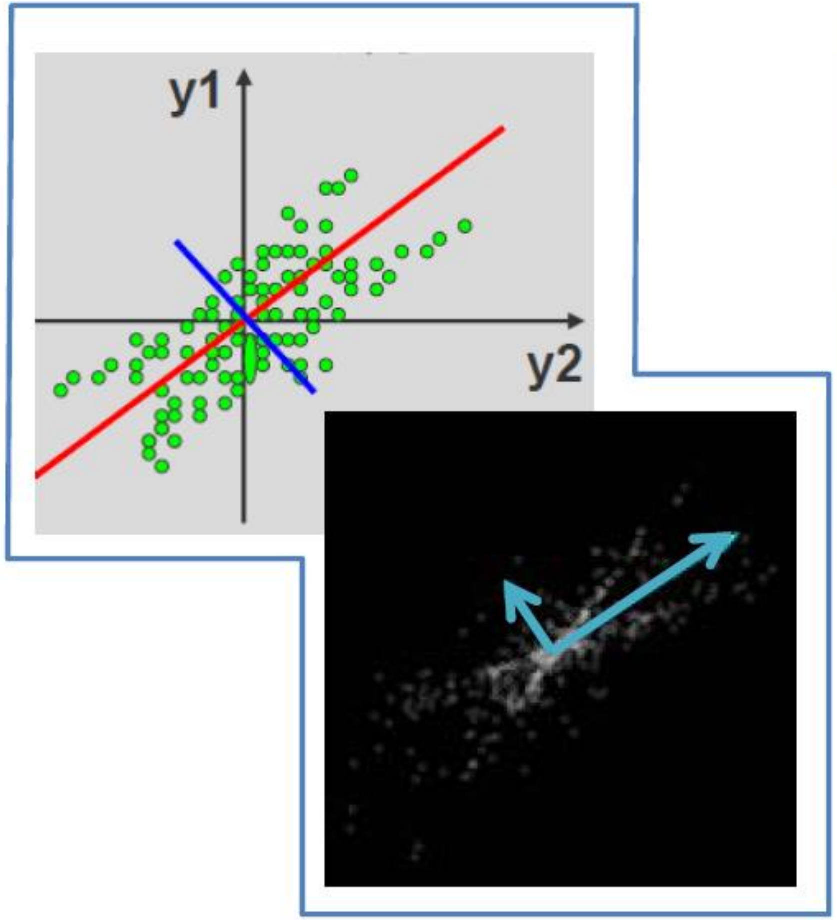
The original signals were very accurately estimated, up to multiplicative signs

ICA versus PCA

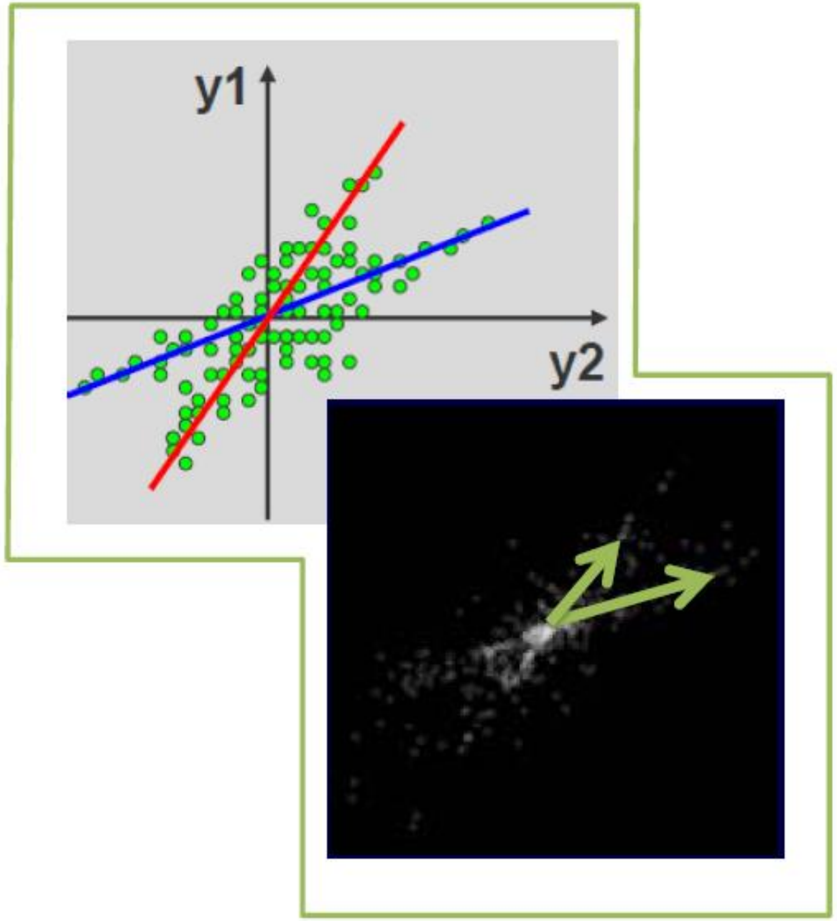
- Similarity
  - Feature extraction
  - Dimension reduction
- Difference
  - PCA uses up to *second order moments* of the data to produce uncorrelated components.
  - ICA strives to generate components as independent as possible through minimizing both the second-order and higher-order dependencies in the given data.

First order moment – mean
Second order moment – variance
Third order moment – skewness
Higher order moments – skewness and kurtosis

ICA versus PCA



PCA finds directions of maximal variance (using second order statistics)



ICA finds directions which maximize independence (using higher order statistics)

ICA - Definition

- Assume that we have *n* mixtures $\mathbf{x}_1$, …, $\mathbf{x}_n$ of *n* independent components:

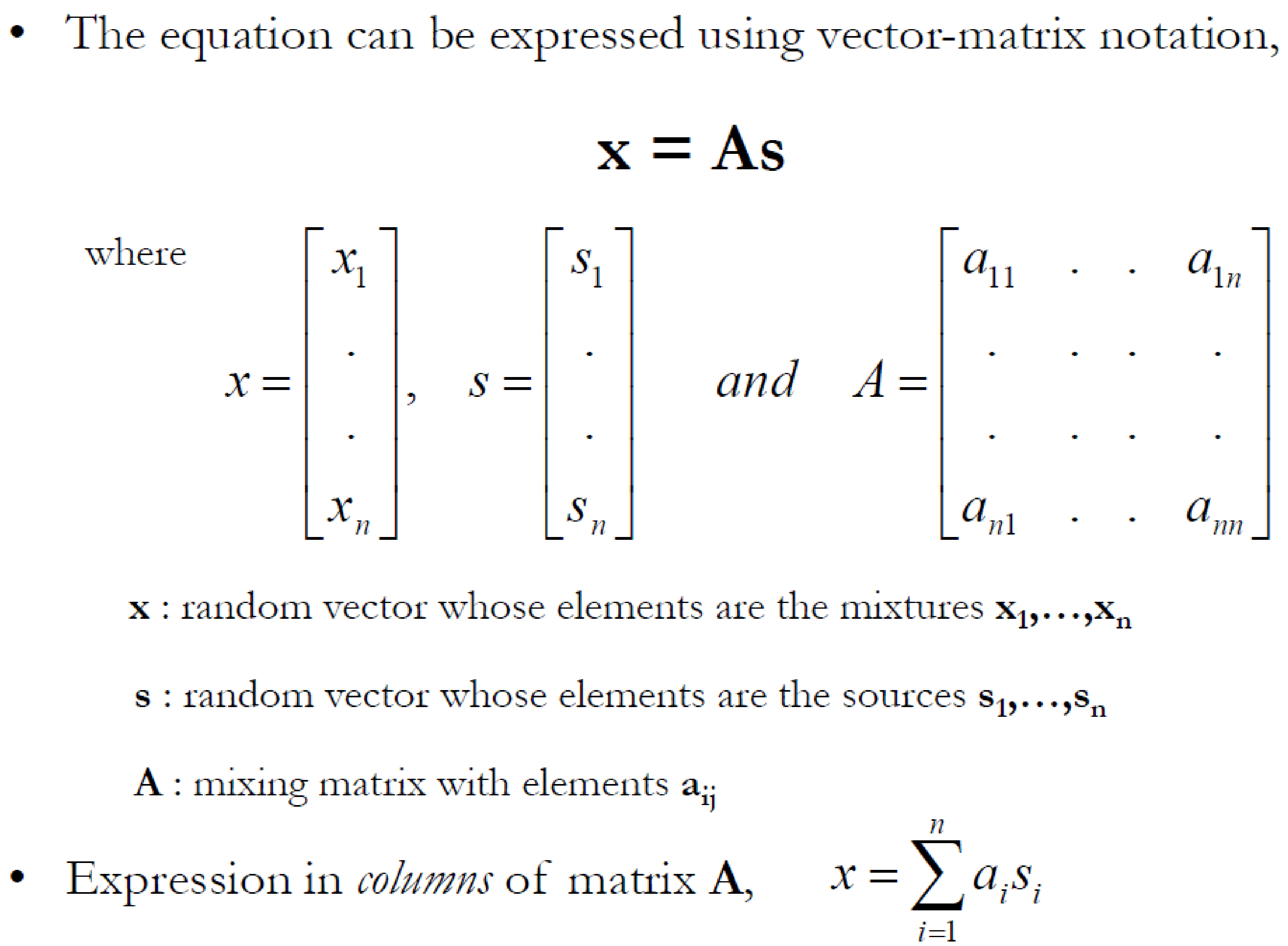
$$\mathbf{x}_j = \mathbf{a}_{j1}\mathbf{s}_1 + \mathbf{a}_{j2}\mathbf{s}_2 + \ldots + \mathbf{a}_{jn}\mathbf{s}_n \quad \text{for all j}$$

The time index t has dropped in ICA model, since we assume that each mixture and individual components are random variables instead of a proper time signal. Thus the observed values $x_j(t)$, e.g. the microphone signals in the cocktail party problem, are then a sample/realization of this random variable.

- Without loss of generality, we can assume that both the mixture variables and the independent components have zero mean.

If this is not true, then the observable variables x_j can always be centered by subtracting the sample mean, which makes the model zero-mean.

ICA



- The equation can be expressed using vector-matrix notation,

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

where

$$x = \begin{bmatrix} x_1 \\ . \\ . \\ x_n \end{bmatrix}, \quad s = \begin{bmatrix} s_1 \\ . \\ . \\ s_n \end{bmatrix} \quad and \quad A = \begin{bmatrix} a_{11} & . & . & a_{1n} \\ . & . & . & . \\ . & . & . & . \\ a_{n1} & . & . & a_{nn} \end{bmatrix}$$

x : random vector whose elements are the mixtures $\mathbf{x_1,\ldots,x_n}$

s : random vector whose elements are the sources $\mathbf{s_1,\ldots,s_n}$

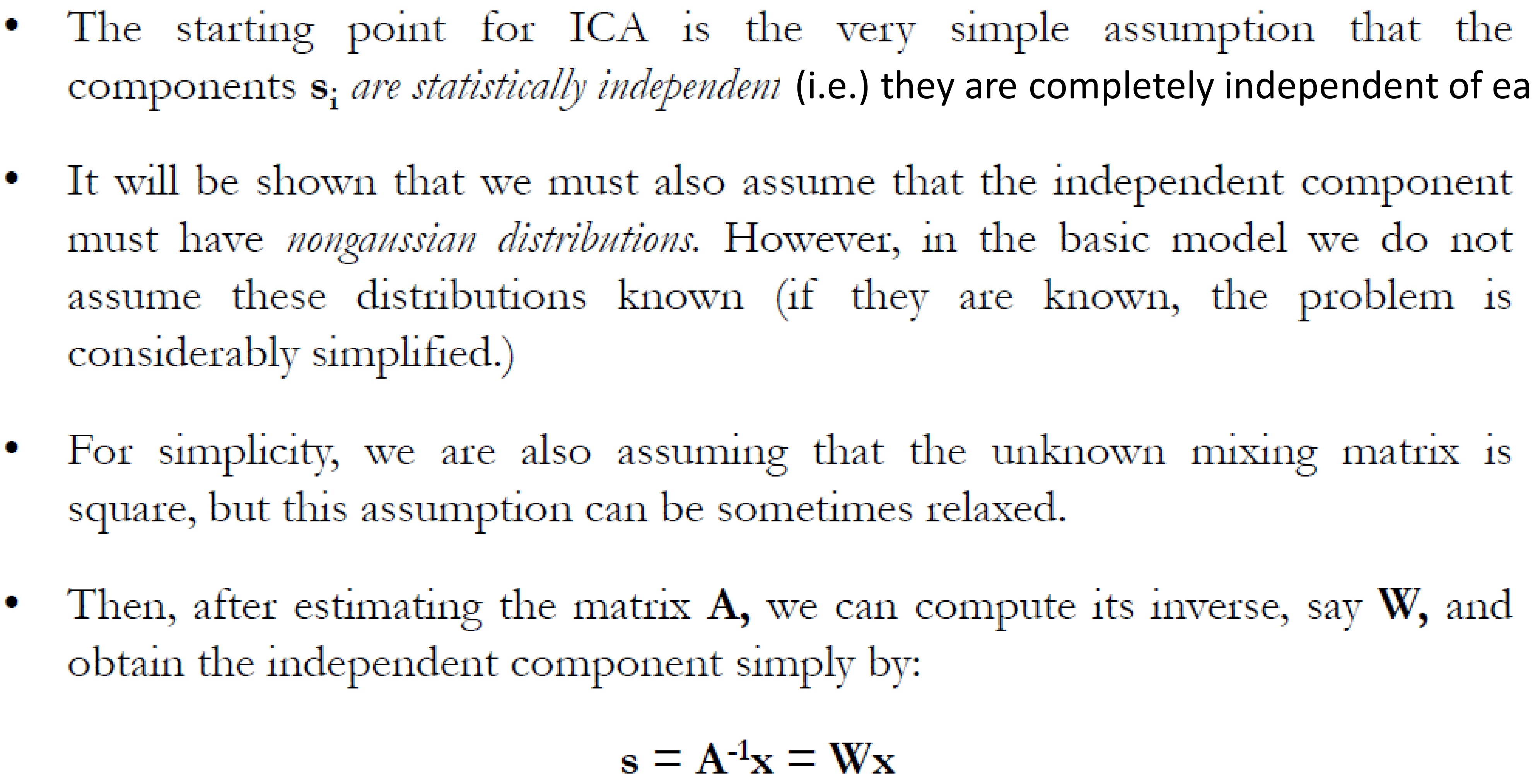
A : mixing matrix with elements $\mathbf{a_{ij}}$

- Expression in *columns* of matrix **A**, $x = \sum_{i=1}^{n} a_i s_i$

ICA

- This statistical model is called *independent component analysis*, or **ICA** model.
- ICA model is a *generative* model, since it describes how the recorded data are generated by mixing the individual components.

ICA - Assumptions



- The starting point for ICA is the very simple assumption that the components $\mathbf{s}_i$ *are statistically independent* (i.e.) they are completely independent of each other

- It will be shown that we must also assume that the independent component must have *nongaussian distributions.* However, in the basic model we do not assume these distributions known (if they are known, the problem is considerably simplified.)

- For simplicity, we are also assuming that the unknown mixing matrix is square, but this assumption can be sometimes relaxed.

- Then, after estimating the matrix **A,** we can compute its inverse, say **W,** and obtain the independent component simply by:

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{x} = \mathbf{W}\mathbf{x}$$

BSS - Blind Source Separation

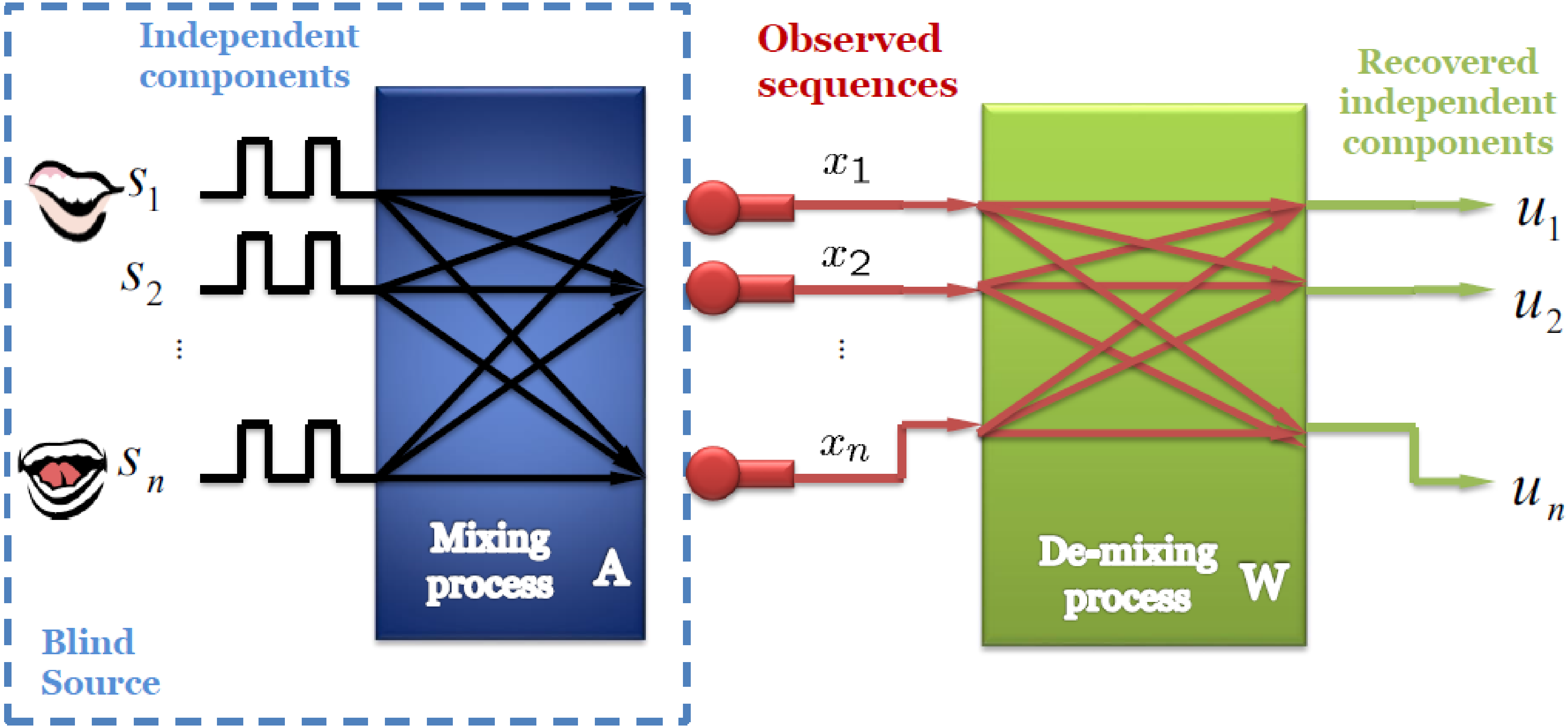
- ICA is very closely related to the method called *blind source separation (BSS) or blind signal separation.*

- *A* "**source**" means here an original signal, i.e. independent component, like the speaker in a cocktail party problem.

- "**Blind**" means that we know very little, if anything, on the mixing matrix **A**, and make little assumptions on the source signals.

- ICA is one method, perhaps the most widely used, for performing blind source separation.

BSS - Blind source separation



Problem Formulation

The goal of ICA is to find a linear mapping **W** such that the unmixed sequences **u**,

$$\mathbf{u}(t) = \mathbf{W}\,\mathbf{x}(t) = \mathbf{W}\,\mathbf{A}\,\mathbf{s}(t)$$

are maximally *statistically independent*.

ICA Algorithm

1. Center **x** by subtracting the mean
2. Whiten **x**
3. Choose a random initial value for the de-mixing matrix **w**
4. Calculate the new value for **w**
5. Normalize **w**
6. Check whether algorithm has converged and if it hasn’t, return to step 4
7. Take the dot product of **w** and **x** to get the independent source signals

$$S = Wx$$

Whitening

- Before applying the ICA algorithm, we must first “*whiten*” our signal. To *“whiten”* a given signal means that we transform it in such a way that potential correlations between its components are removed (covariance equal to 0) and the variance of each component is equal to 1. Another way of looking at it is that the covariance matrix of the whitened signal will be equal to identity matrix.

$$I_1 = [1],\ I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ \cdots,\ I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Whitening

- The actual way we set about whitening a signal involves the **eigen-value decomposition of its covariance matrix**.

$$\tilde{x} = ED^{-1/2}E^T x$$

- where **D** is a diagonal matrix of eigenvalues (every lambda is an eigenvalue of the covariance matrix) and **E** is an orthogonal matrix of eigenvectors

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots \\ 0 & \lambda_2 & 0 & \dots \\ 0 & 0 & \lambda_3 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

De-mixing

- Once we’ve finished preprocessing the signal, for each component, we update the values of the de-mixing matrix **w** until the algorithm has converged or the maximum number of iterations has been reached.
- Convergence is considered attained when the dot product of **w** and its transpose is roughly equal to 1.

De-mixing

$$for\, 1\, to\, the\, number\, of\, components\, :$$

$$repeat\,\, until\,\, w_p^T w_{p+1} \approx 1 :$$

$$w_p = \frac{1}{n}\sum_i^n X g(W^T X) - \frac{1}{n}\sum_i^n g'(W^{TX})W$$

$$w_p = w_p - \sum_{j=1}^{p-1}(w_p^T w_j)w_j$$

$$w_p = \frac{w_p}{||w_p||}$$

$$W = [w1, w2...]$$

where

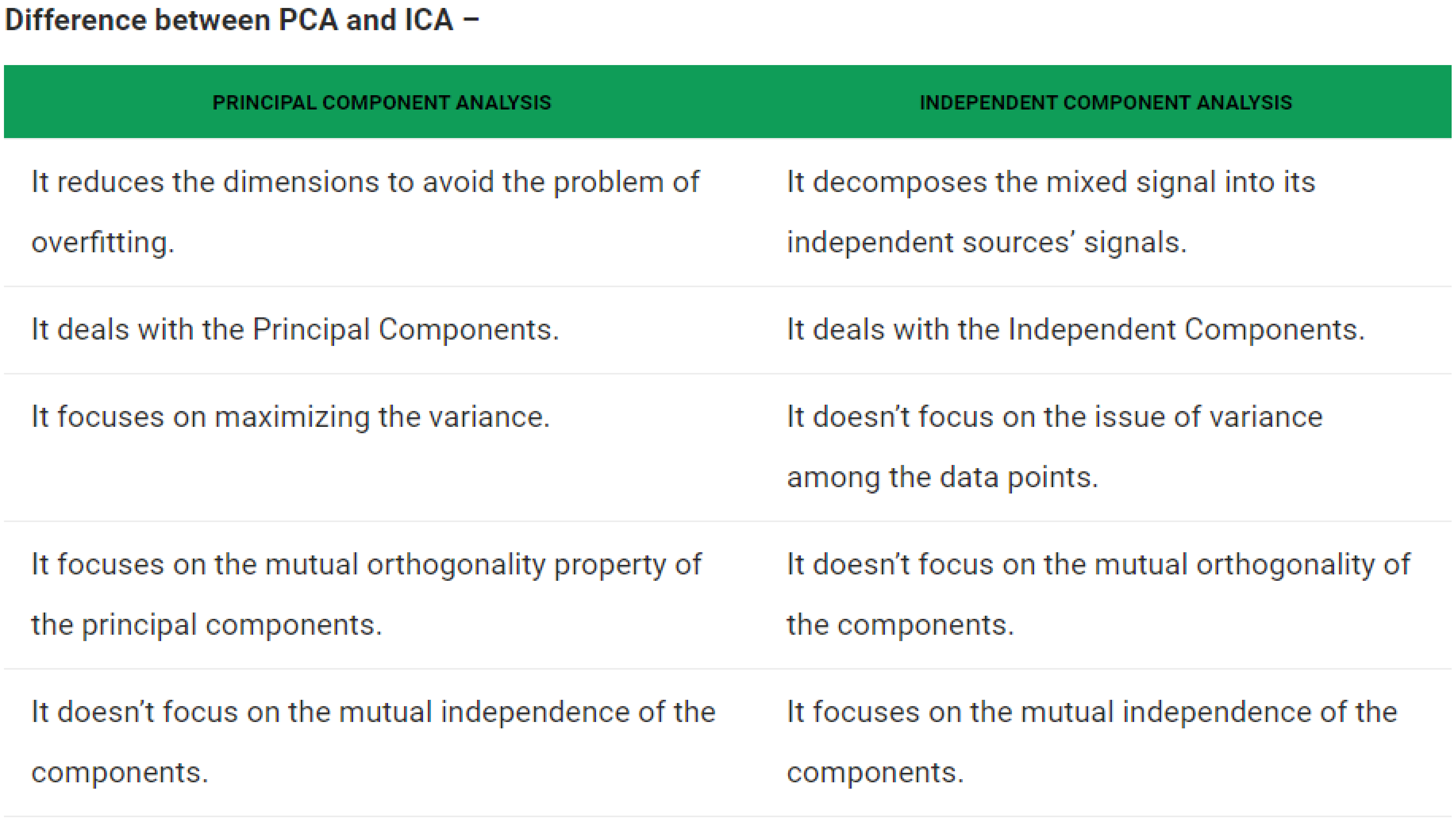
$$\mathbf{g}(u) = \tanh(u)$$

$$\mathbf{g}'(u) = 1 - \tanh^2(u)$$

Restrictions on ICA

- The independent components generated by the ICA are assumed to be statistically independent of each other.
- The independent components generated by the ICA must have non-gaussian distribution.
- The number of independent components generated by the ICA is equal to the number of observed mixtures.

PCA and ICA - Difference



Difference between PCA and ICA –

PRINCIPAL COMPONENT ANALYSIS	INDEPENDENT COMPONENT ANALYSIS
It reduces the dimensions to avoid the problem of overfitting.	It decomposes the mixed signal into its independent sources' signals.
It deals with the Principal Components.	It deals with the Independent Components.
It focuses on maximizing the variance.	It doesn't focus on the issue of variance among the data points.
It focuses on the mutual orthogonality property of the principal components.	It doesn't focus on the mutual orthogonality of the components.
It doesn't focus on the mutual independence of the components.	It focuses on the mutual independence of the components.