Trend Prediction of Unemployment in Kenya using Singular Spectrum Analysis

by Flora Mwikali

Submission date: 08-Aug-2019 06:01PM (UTC+0300)
Submission ID: 1158623362
File name: Flora.pdf (1.58M)
Word count: 11927
Character count: 60237

University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in Applied Mathematics

Trend Prediction of Unemployment in Kenya using Singular Spectrum Analysis

Research Report in Mathematics, Number 11, 2019

Flora Mwikali Katiwa

September 2019

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Applied Mathematics at University of Nairobi and is deposited in the University library to be made available to borrowers under the rules of the library

Master Project in Applied Mathematics **University of Nairobi**
September 2019

Trend Prediction of Unemployment in Kenya using Singular Spectrum Analysis

Research Report in Mathematics, Number 11, 2019

Flora Mwikali Katiwa

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Applied Mathematics at the University of Nairobi and is deposited in the University library to be made available to borrowers under the rules of the library.

Submitted to: The Graduate School, University of Nairobi, Kenya

Abstract

Unemployment is a worldwide problem that affects not only the unemployed but also their family with respect to income, health and mortality. This work combines the theory of Approximation and Numerical linear algebra tools especially Singular Value Decomposition and least squares problems. It shows the limitations of least squares method in trend estimation of time series applied to unemployment rates data in Kenya. This makes Singular Spectrum Analysis the best method in predicting the rate of unemployment in Kenya.

Master Thesis in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

Signature Date

Flora Mwikali Katiwa
Reg No. I56/8494/2017

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

Signature Date

Dr Charles Nyandwi
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: nyandwisr@uonbi.ac.ke

Signature Date

Dr Timothy Kamanu
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: timkuria@gmail.com

Dedication

This project is dedicated to me, my Daddy Josiah Katiwa, my mum Priscah Mutanu, my only sister Leah Kalumba, my brothers Ben Munyoki and Derick Kyatha.

Acknowledgements

First and foremost, I would like to thank God Almighty for giving me the strength, knowledge, ability and opportunity to undertake this research study and to persevere and complete it satisfactorily. Without his blessings, this achievement would not have been possible.

Many thanks to my supervisors Dr. Charles Nyandwi and Dr. Timothy Kamanu who offered continuous advice and encouragement throughout the course of my work.

I must express my very profound gratitude to my beloved parents, my granny and my siblings for their moral and financial support throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Special thanks to Julius Muigai, for providing me with unfailing support and continuous encouragement.

I wish also to acknowledge the University of Nairobi more so school of mathematics for the chance awarded to me to pursue my MSc. studies. I will also very much want to thank Ann Mumo and the director school of mathematics, Prof P.G.O Weke.

Finally, am thankful to my classmates Simon Muoria, Paul Muthama, Lydia Atuya, Lewis Kivuva, Felix Mureithi, Stephen Mathenge for the help and support you gave me during my studies.

Flora Mwikali

Nairobi, 2019.

Contents

List of Figures

Symbols and Acronymns

SVD	Singular Value Decomposition.
SSA	Singular spectrum analysis.
$\mathbb{X}_N$	Ordered collection of numbers.
N	Size of $\mathbb{X}$.
L	Window length.
K	The number of columns in the trajectory matrix X; in the 1-D case, K=N-L+1.
$\mathbf{X}$	The trajectory matrix of dimension L * K associated with $\mathbb{X}$.
λ_i	i^{th}eigenvalue of matrix $\mathbf{X}\mathbf{X}^T$.
U_i	i^{th} eigenvector of matrix $\mathbf{X}\mathbf{X}^T$.
$V_i = \frac{\mathbf{X}^T U_i}{\sqrt{\lambda_i}}$	i^{th} factor vector of the matrix $\mathbf{X}$.
$(\sqrt{\lambda_i}, U_i, V_i)$	i^{th} eigentriples of SVD of the trajectory matrix $\mathbf{X}$.

1 INTRODUCTION

1.1 Singular value decomposition

Matrix decomposition, also known as matrix factorization, involves describing a given matrix using its constituent elements.
All matrices have an SVD, making it more stable than other methods, such as the eigendecomposition. It is often applied in compressing, denoising and data reduction.

1.2 Time series

Time series insures information about Socioeconomic, physical or biological that produced it. The main objective of time series analysis depends on pattern of time series, to determine some of main properties, to understand how time series has behaved in the past and to understand and predict future behavior. There are many well known decomposition methods which are used in time series analysis.

The main SSA method basically contains two stages: reconstruction and decomposition stages and both stages contain two separate steps. The decomposition stage consists: embedding and singular value decomposition steps; then the reconstruction stage consists of grouping and diagonal averaging steps [26].
Structure of the SSA algorithm is as follows. First,a one dimensional time arrangement is changed over into a higher measurement framework called the *trajectory matrix*. The dimension of the trajectory matrix is called the *window length*. SVD is then applied to the trajectory matrix and eigenvalues and eigenvectors are found. The next step is grouping which includes parting the elementary matrices in various groups and summing the matrices in each of the grouping. To get the reconstructed components, we take the diagonal averaging along the diagonals of each group, then by combining them into a one dimensional time series, we obtain the approximated time series [26].
An important advantage of SSA is that after reconstruction of the time series it allows the production of forecast for the reconstructed components which is known as SSA forecasting algorithm. The purpose of this work is to understand SVD, least squares method, study and understand SSA method, SSA forecasting algorithm and make numerical experiments on unemployment data.

1.3 Problem statement and research objectives

Lining up with the exploration inspirations, the issue explanation of this research is that in a nation with nearly 55 million people living in it like Kenya, having the largest citizens

unemployed is a huge problem. According to the latest research 42.01 percent of the population is declared unemployed. Despite of the fact that employment in Kenya has been growing fast for the past decade, still, many Kenyans are jobless.
This study aims at understanding least squares, SVD and the basic algorithm of SSA then later predict on unemployment rate in Kenya from 2018 to 2028 on annually basis.

1.3.1 Main objective

To study and predict the future of unemployment rates in Kenya.

1.3.2 Specific objectives

- To study the basic algorithm of SSA.
- To study numerical linear algebra especially SVD and least squares problems.

1.4 Literature review

1.5 Historical background of SVD

The differential geometers, developed singular value decomposition, where they wished to decide if a genuine bi-linear structure could be made equivalent to another by autonomous symmetrical changes of the two spaces it follows up on. **E. Beltrami** and **C. Jordan** discovered separately, in (1873) and (1874) respectively, that the singular values of the bi-linear forms, spoken to as a matrix, structure a total arrangement of invariants for bi-linear frames under symmetrical substitutions.
In (1889) **Joseph Sylvester** discovered the singular value decomposition of real square matrices. **Sylvester** called singular values of a matrix, the canonical multipliers.
Autonne discovered the singular value decomposition in (1915) who arrived at it via polar decomposition. **Gale Young** et al in (1936), did the first proof of SVD for rectangular and complex matrices, they show it as speculation of Principal axis change for Hermitian matrices. **Schmidt** characterized an analog of singular values for integral operators; it seems he was unaware of the parallel work on singular values of finite matrices in (1907). This theory was further developed by Emile Picard in (1910), who called the numbers σ_k singular values [4]. Practical method for computing the SVD date back to **Kogbetliantz** in (1954), (1955) and **Hestenes** in (1958) resembling closely the Jacobi eigenvalue algorithm, which uses plane rotation. However this were replaced by method of **Gene Golub** and **Willian Kahan** in (1965) which uses Householders reflections. In (1970) Golub and Christian published a variant of Golub/Kahan algorithm still one most used today.

1.6 History of least squares

The least squares method emerged out of the fields of space science and geodesy as researchers and mathematicians looked to offer answers for the difficulties confronting Earths seas during **the age of exploration**.
The advancement of foundation can be assessed to demonstrate when the arrangement with least blunder has been accomplished. **Laplace** attempted to give a numerical formulation of probability density of mistakes of estimation. Laplace utilized symmetric two sided exponential dispersion and utilized whole of total deviation as mistake of estimation. The primary clear strategy for least squares was distributed by **Legendre** in (1805). This an arithmetical method for fitting straight conditions to data.

1.7 History of singular spectrum analysis

Singular spectrum analysis was developed simultaneously and independently in (1986) by **Broomhead**, **King** and **Fraedrich**. Broomhead and King applied singular spectrum analysis to the issues of dynamical systems theory and the singular spectrum approach to the method of delays was suggested to remove some of the limitations and ambiguities experienced with the method of delays. By combining SVD and embedding theorems, they laid the mathematical basis used for SSA.
They also investigated some preliminary artificial time arrangement to outline [8] the advantages of using singular spectrum analysis as a statistical tool for qualitative analysis and for the removal of especially white noise from time series. **Fraedrich** (1986) utilized watched climate and atmosphere factors to give data to depictions of the properties of the attractors of these dynamical frameworks and to acquire a gauge of the most modest number of factors important to clarify the framework elements. Further groundbreaking work in the methodological development of the singular spectrum analysis toolkit and substantial research on the possibilities of the technique, was done by **Robert Vautard** and **Michael Ghil**. Robert and Michael (1989) extended the previous research done by Broomhead and King (1986) and refined certain aspects of the application.
After applying SSA to various paleoclimatic time series, they found the technique to be very flexible and incisive. They concluded that, even though SSA is related to ordinary spectral analysis, it is considerably more robust to the nonstationarities that can be found in climatic records.
Robert et al. (1992) distinguish among three major cases encountered when performing data analysis. The first is where the evolution equations governing the data are known and these equations are relatively insensitive to the initial values of the system. The second type of data analysis occurs when the governing equations are also known, but long-term prediction of the data is impossible due to the sensitivity of the system for the initial values. The last class of data analysis is where the evolution equations for the system are completely unknown and often only noisy measurements of one of the variables in a high-dimensional system are available.

It is especially with respect to this last class of data that Robert et al. (1992) identified the potential of SSA. Even though the work focused on single-channel SSA, they already saw the possibilities of multi-channel SSA to account for the cross-correlation between several variables that were measured simultaneously. The outline of the thesis is per the following:

Chapter 2: chapter 2 discusses the SVD and its overview.

Chapter 3:
Chapter 3 discusses least squares and it's overview. Then the link SVD and least squares and finally its application to Unemployment rate data in Kenya from 1991 to 2018.

Chapter 4:
Chapter 4 discusses the basic algorithm of singular spectrum analysis. Chapter 5 is about the application of SSA to unemployment rate data of Kenya. Finally there is conclusion and recommendation for future work.

2 THE SINGULAR VALUE DECOMPOSITION

The SVD is given by;

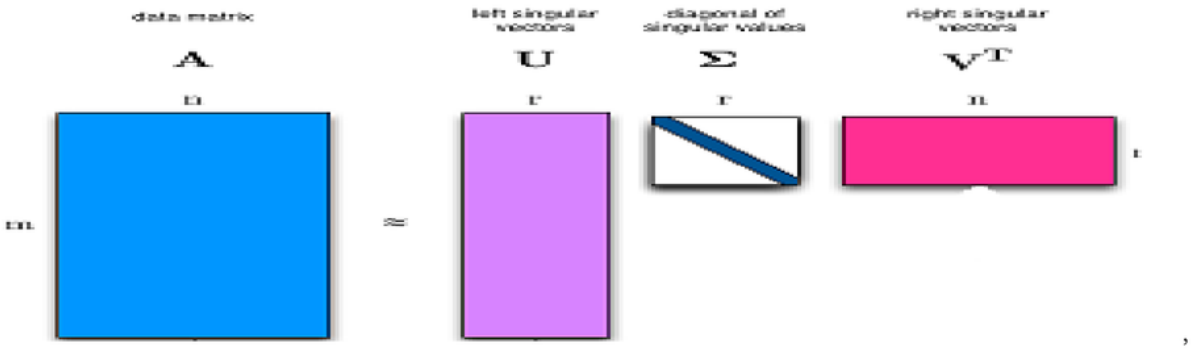


Figure 1. SVD.

where $A \in \mathbb{R}^{m\times n}$, then $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are symmetrical matrices and $\Sigma \in \mathbb{R}^{m\times n}$ is a matrix whose diagonal elements satisfy $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$ and with non diagonal entries being zeros

$$\Sigma = \begin{bmatrix} \sigma_1 & \ldots & 0 & \ldots \\ 0 & \sigma_2 & \ldots & \ldots \\ \ldots & 0 & \sigma_n & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix}.$$

The $\sigma_j s$ are special and known as the singular values of A. The eigenvectors of matrix AA^T being column vectors of U. The eigenvectors of the matrix A^TA being column vectors of V.

Theorem 2.0.1. *(singular value decomposition)*

If matrix $A \in \mathbb{R}^{m\times n}$ then A can be decomposed into singular values.

Proof. Let A^TA be a symmetric matrix. Then all its eigenvalues are symmetrical and real it has an diagonalizing (symmetrical) matrix V, it's eigenvalues are real and positive. Let λ be an eigenvalue of A^TA and x be eigenvector associated with λ, then

$$\|Ax\|^2 = x^TA^TAx$$

$$= \lambda x^T x$$
$$= \lambda ||x||^2.$$

Hence

$$\lambda = \frac{||Ax||^2}{||x||^2} \geq 0.$$

We expect that the columns of V are arranged in the descending order

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_0.$$

The singular values of matrix A are given by

$$\sigma_i = \sqrt{\lambda_i}, i = 1, \ldots, n.$$

□

The rank of matrix A is denoted by r. The matrix $A^T A$ will also have rank r. Since $A^T A$ is symmetric, its rank equals the number of nonzero eigenvalues. Thus

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_0 \quad \textbf{(1)}$$

and

$$\lambda_{r+1} = \lambda_{r+2} = \lambda_n = 0.$$

The same relation holds for the singular values

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$$

and condition numbers, singular values and matrix norms

$$\sigma_{r+1} = \sigma_{r+2} = \sigma_n = 0.$$

Getting the number of nonzero singular values of Σ, gives us the rank of matrix A by the use SVD. Let

$$V_1 = (v_1, ..., v_r),$$
$$V_2 = (v_{r+1}, ..., v_n)$$

and

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_r \end{bmatrix},$$

thus

$$\Sigma \in \mathbb{R}^{r \times r}.$$

2.1 Existence of SVD

Supposed to prove that any $A \in \mathbb{R}^{m\times n}$ with rank $(A) = R$ can be written as $A = U\Sigma V^T$.
We can write:

$$A^T A = \sum_{n=1}^{N} \lambda_n v_n v_n^T, \tag{2}$$

since $A^T A$ is symmetric positive semi-definite, where the $\mathbf{v}_n$ are orthonormal and the λ_n are real and non-negative. Since rank $(\mathbf{A}) = R$, we also have rank $(A^T A)$=R and so $\lambda_1, \ldots, \lambda_R$ are strictly positive .
Set $\mathbf{u}_m = \frac{1}{\sqrt{\lambda_m}} A v_m$, for $m = 1, \ldots, R, U = [u_1, \ldots, u_R]$,
where u_m are orthonormal, as

$$\langle u_m, u_\ell \rangle = \frac{1}{\sqrt{\lambda_m \lambda_\ell}} v_\ell^T A^T v_m = \sqrt{\frac{\lambda_m}{\lambda_\ell}} v_\ell^T v_m = \begin{cases} 1, m = \ell; \\ 0, m \leq \ell \end{cases} . \tag{3}$$

These u_m are also eigenvectors of AA^T, as

$$AA^T u_m = \frac{1}{\sqrt{\lambda_m}} AA^T A v_m = \sqrt{\lambda_m} A v_m = \lambda_m u_m. \tag{4}$$

Now let $u_{R+1}, \ldots, u_m$ are an orthobasis for full space of U^T concating these sets into $u_1, \ldots, u_m$ forms an orthobasis for all of $\mathbb{R}^m$.

2.2 Properties of SVD

Assume

$$A = U\Sigma V^T. \tag{5}$$

The following are the five properties of $U\Sigma V^T$;

1. The particular estimations of a matrix A are equivalent to the square foundations of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ of the matrix $A^T A$ $(\sigma_i = \sqrt{\lambda_i})$.
2. The number of singular values of any matrix A gives the rank of that matrix.
3. The largest singular values is equal to the euclidean norm A $\|A\|_2 = S_i$.
4. The main r sections of the matrix U structure orthonormal basis for the space projected by the columns of matrix A.
5. The primary r segments of the matrix V structure orthonormal basis for the space anticipated by the rows of matrix A.

2.3 Matrix norms and SVD

The SVD gives a method for getting the 2- norm of a matrix say A, since $\| A \|_2 = \sqrt{\sigma_1}$.
Suppose A is invertible, then can say that $\|A\|_2 = \sqrt{\frac{1}{\sigma_1}}$.
The SVD gives a method for getting the Frobenius norm. If K is a matrix then

$$\|A\|_F^2 = trace(A^T A) = trace(AA^T).$$

Lemma 2.3.1. .

Suppose $V \in R^{n\times n}$ symmetrical matrix then U is an $m \times m$ symmetrical matrix and

$$\|UAV\|_F^2 = \|A\|_F^2.$$

`Proof.` proof of lemma 2.3.1

$$\|UA\|_F^2 = trace((UA)^T(UA)$$

$$= trace(A^T U^T)UA$$
$$= trace(A^T IA)$$
$$= A^T A$$
$$= \|A\|_F^2.$$

□

The Frobenius norm is fixed under left augmentation by a symmetrical matrix.

$$\|AV\|_F^2 = trace((AV)(AV)^T)$$

$$= trace(AV)(V^T A^T)$$
$$= traceAA^T$$
$$= \|A\|_F^2.$$

and

$$\|RA\|_F^2 = trace(A^T R^T RA)$$
$$= trace(A^T A)$$
$$= \|A\|_F^2.$$

Here we have used the symmetrical nature of R

$$R^T R = RR^T = 1$$

and the cyclic nature of the trace

$$trace(XYZ) = trace(ZXY),$$

it also satisfies

$$\|A^T A\|_F = \|AA^T\|_F \leqslant \|A\|_F^2$$

$$\|A+B\|_F^2 = \|A\|_F^2 + \|B\|_F^2 + 2 < A, B >_F,$$

where $< A, B >_F$ is the Frobenius inner products. Since $\|A\|_2 = \sqrt{\sigma_1}$ if A is invertible then $\|A\|_2 = \sqrt{\frac{1}{\sigma}}$. The SVD gives a way of computing the Frobenius norm. Given A is a general matrix

$$\|A\|_F^2 = trace(A^T A) = trace(AA^T).$$

Lemma 2.3.2. *If $U \in R^{m\times m}$ and $V \in R^{n\times n}$ are symmetrical matrices then $\|UAV\|_F^2 = \|A\|_F^2$.*

Proof.

$$\begin{aligned}\|UA\|_F^2 &= trace((UA)^T(UA)) \\ &= trace(A^T U^T)(UA) \\ &= trace(A^T IA) \\ &= traceA^T A \\ &= \|A\|_F^2.\end{aligned}$$

The Frobenius norm is fixed under left augmentation by an symmetrical matrix.

$$\|AV\|_F^2 = trace((AV)^T(AV))$$

$$\begin{aligned}&= trace(AV)(V^T A^T) \\ &= trace(A^T A) \\ &= \|A\|_F^2.\end{aligned}$$

The Frobenius norm is fixed under right multiplication by an symmetrical matrix. □

Theorem 2.3.3.

$$\|A\|_F = \left(\sum_{j=1}^{r} \sigma_j^2\right)^{\frac{1}{2}}.$$

Proof. By SVD U and V are symmetrical matrices such that

$$\|A\|_F = \|U\Sigma V^T\|_F$$

$$\|\Sigma\|_F.$$

From the above lemma, the (just) non-zero entries are Σ are the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ so

$$\|A\|_F = \left(\sum_{j=1}^{r} \sigma_j^2\right)^{\frac{1}{2}}.$$

□

2.4 Spectral decomposition

The spectral theorem provides a canonical decomposition, called the spectral decomposition, eigenvalue decomposition, or eigendecomposition of the underlying vector space on which the operator acts.

Theorem 2.4.1. *Any n square matrix A can be characterized by a biorthogonal system of left and right eigenvectors which form the matrices P and Q.*

Proof. $QP^T = PQ^T = I$.
Q^TAP = diag $[\lambda_1, \lambda_2, \ldots, \lambda_n]$,
where λ_i are eigenvalues of A.
A has number of linearly independent eigenvectors $x_1, \ldots, x_n$ with the corresponding eigenvectors

$$P := [x_1, \ldots, x_n],$$

then we have

$$AP = Pdiag[\lambda_1, \lambda_2, \ldots, \lambda_n]. \tag{6}$$

Therefore, taking transpose in both sides and rearranging, we get

$$A^T(P^T)^{-1} = (P^T)^{-1}diag[\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_n].$$

The eigenvectors of A^T can, therefore, be selected as the columns of $(P^T)^{-1}$ which we denote by $Q = (P^T)^{-1}$. Hence, we have a bi-orthogonal system of left and right singular vectors forming respectively the matrices Q and P and satisfying

$$Q^TP = I$$

and in the view of (6),

$$Q^T AP = diag(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

□

Therefore we write the theorem as

$$A = Pdiag(\lambda_1, \lambda_2, \ldots, \lambda_n)Q^T = \lambda_1 PD_1 Q^T + \lambda_2 PD_2 Q^T + \ldots + \lambda_n PD_n Q^T = \lambda_1 Q_1 + \lambda_2 Q_2 + \ldots + \lambda_n Q_n, \tag{7}$$

where D_i = a diagonal matrix with diagonals equal to zero but for i^{th} which is 1,
$Q_i = PD_iQ^T$.

2.5 Low rank approximation

The rank of a matrix is given by singular value decomposition, which is the quantity of nonzero singular values of matrix A. If

$$A_{m\times n} = U_{m\times r}\Sigma_{r\times r}V^T_{r\times n},$$

then it's rank is $r = min(m,n)$,

$$A = [u_1 u_2 \ldots u_r] \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_r^T \end{bmatrix},$$

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T,$$

where $u_1, u_2, \ldots, u_r$ and $v_1, v_2, \ldots, v_r$ are columns of $U_{m\times r}$ and $V_{n\times r}$ respectively. The matrix A can be represented by the sum of external products of vectors. Using SVD we can approximate a matrix A by low rank approximation The rank k approximation (also called the truncated or partial SVD) of A, A_k where $k < r$ is given by zeroing out the r-k trailing singular values of A, that is

$$A_k = U_{m\times k}(\Sigma_k)_{k\times k}V^T_{k\times n} = \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

Here $\Sigma_k = diag(\sigma_1, \sigma_2, \ldots, \sigma_k), U_{m\times k} = u_1, u_2, \ldots, u_k$ and $V^T_{k\times n} = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k^T \end{bmatrix}$. Then it's seen that

$$A_k = U_{m\times k} U^T_{k\times m} A = (\sum_{i=1}^{k} u_i u_i^T) A \tag{8}$$

and

$$A_k = A V_{n\times k} V^T_{k\times n} = A(\sum_{i=1}^{k} v_i v_i^T), \tag{9}$$

i.e A_k is the projection of the A onto the space spread over by the top k singular vectors of A. The following theorem states that the above approximation is the best rank k approximation in both Frobenius and spectral norm.

Theorem 2.5.1. *Let A_k be the rank -k approximation of A achieved by SVD truncation as above. Then A_k is the closest rank-k matrix to A, i.e.*

$$||A - A_k||_F \leqslant ||A - b||_F,$$

where B,s are rank-k matrices. The minimal error is given by the Euclidean norm of the singular values that have been in the process

$$||A - A_k||F = \sqrt{\sigma^2_{k+1} + \ldots + \sigma^2_r},$$

where $||.||_F$ is Frobenius norm. SVD also gives the best low rank approximation in spectral norm:

$$||A - A_k||_2 = \min_{rank(B)=k} ||A - B||_2 = \sigma_{k+1}.$$

2.6 Fundamental subspaces

The four central subspaces of a matrix are given by SVD.

1. The linear subspace is the range of matrix A. $R(A) = y|y = \{Ax \text{ for arbitrary } x\}$, assuming that A has rank r

$$\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_n = 0.$$

2. Then utilizing the external product structure we have

$$y = Ax = \sum_{j=1}^{r} \sigma_j u_j v_j^T x$$

$$= \sum (\sigma_j v_j^T x) u_j$$

$$= \sum_{j=1}^{r} \alpha_j u_j.$$

3. The linear subspace is the null-space of the matrix A. $N(A) = \{x | Ax = 0\}$, since

$$y = Ax$$

$$= \Sigma_{j=1}^{r} \sigma_j u_j v_j^T x,$$

then any vector

$$= \Sigma_{r+1}^{n} \beta_j v_j$$

is the null space.

$$AZ = (\sum_{j=1}^{r} \sigma_j u_j v_j^T)(\sum_{j=r+1}^{n} \beta_j v_j) = 0.$$

2.7 Moore Penrose inverse

A generalization of the inverse of a matrix which is defined for any matrix and its unique. Special properties of the mp-inverse.

1. $A^\dagger$ provides atleast the least squares solution.
   $\Rightarrow$Eliminates the influence of error orthogonal to the range of A.
2. $A^\dagger$ has minimum spectral norm among all left inverse of A.
   $\Rightarrow$ Recovers x but doesn't blow up the noise.

2.7.1 Definition and characterizations

We consider the case of $A \in R^{m \times n}$. Every $A \in R^{m \times n}$ has a pseudoinverse denoted by $A^\dagger \in R^{n \times m}$ is unique.

Theorem 2.7.1. *let $A \in Rm \times n$ then $J = A^\dagger$ if and only if*

(P1) $AJA = A$

(P2) $JAJ = J$

(P3) $(AJ)^T = AJ$

(P4) $(JA)^T = JA$

where $A^\dagger$ *always exists and its unique and matrix* J *purpots to be the pseudoinverse.*

some proofs

1.
$$\begin{aligned} AJ^\dagger A &= U\Sigma V^T V\Sigma^\dagger U^T U\Sigma V^T \\ &= U\Sigma\Sigma^\dagger V^T \\ &= U\Sigma V^T \\ &= A. \end{aligned}$$

2.
$$\begin{aligned} (AA^\dagger A)^T &= (U\Sigma V^T V\Sigma^\dagger U^T)^T \\ &= (U\Sigma\Sigma^\dagger U^T)^T \\ &= U(\Sigma\Sigma^\dagger)^T U^T \\ &= U\Sigma\Sigma^\dagger U^T \\ &= U\Sigma V^T V\Sigma^\dagger U^T \\ &= AA^\dagger. \end{aligned}$$

Limit definition of the pseudoinverse.

Theorem 2.7.2. *Let* $A \in R^{m\times n}$*, then*

$$\begin{aligned} A^\dagger &= \lim_{\delta\to 0}(A^T A+)\delta^2 I)^{-1}A^T \\ &= \lim_{\delta\to 0} A^T(AA^T + \delta^2 I)^{-1}. \end{aligned}$$

$A^\dagger = A^T(AA^T)^{-1}$: if A is onto i.e has linearly independent rows (A is right invertible). $A^\dagger = (A^T A)^{-1}A^T$: if A is one to one i.e has linearly independent columns (A is left invertible). For any scalar α

$$\alpha^\dagger = \begin{cases} \alpha^{-1}, \alpha \neq 0 \\ 0, \alpha = 0 \end{cases} . \tag{10}$$

2.7.2 Properties

Let $A \in R^{m\times n}$ and suppose $U \in R^{m\times m}$, $V \in R^{n\times n}$ are symmetrical. Then

$(UAV)^\dagger = V^T A^\dagger U^T$.

$(A^\dagger)^\dagger = A$.

$(A^T A)^\dagger = (A^\dagger (A^T)^\dagger, (AA^T)^+ = (A^T)^+ A^+$.

$R(A^\dagger) = R(A^T) = R(A^\dagger A) = R(A^T A)$.

$N(A^\dagger) = N(AA^\dagger) = N((AA)^\dagger) = N(AA^T) = N(A^T)$.

If A is normal then $A^K A^\dagger = A^\dagger A^K$, for all $K > 0$ and $(A^K)^\dagger = (A^\dagger)^K$ for all $K > 0, A \in R^{m\times n}$ is normal if $AA^T = A^T A$.

Theorem 2.7.3. *For all $A \in R^{m\times n}$*

1. $A^\dagger = (A^T A)^\dagger A^T = A^T (AA^T)^\dagger$
2. $(A^T)^\dagger = (A^\dagger)^T$.

Proof.

$$
\begin{aligned}
(A^T)^\dagger &= \lim_{\delta\to 0} (AA^T + \delta^2 I)^{-1} \\
&= \lim_{\delta\to 0} [A^T (AA^T + \delta^2 I)^{-1}]^T \\
&= [\lim_{\delta\to 0} A^T (AA^T + \delta^2 I)^{-1}]^T \\
&= (A^\dagger)^T.
\end{aligned}
$$

□

Combining the above two theorems we can compute the Moore-penrose inverse of any matrix (AA^T and $A^T A$) are symmetric SVD of A

$$A = U\Sigma V^T,$$

$$A^\dagger = V\Sigma^\dagger U^T,$$

$$\Sigma^\dagger = \begin{bmatrix} S^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

2.8 Perturbation analysis

Let $B = A + E$ be the perturbated matrix. The perturbation in $A^{\dagger}$ becomes unbounded when the perturbation $||E||_2 \longrightarrow 0$, for any change in the rank. This makes the theory muddled. Unimportant model of this is $A(\varepsilon) = \begin{pmatrix} \sigma & 0 \\ 0 & \varepsilon \end{pmatrix}$, for which rank$(A) = 2$, if $\varepsilon \neq 0$, but rank $(A(0)) = 1$ and $||(A+E)^{\dagger} - A^{\dagger}||_2 = |\varepsilon|^{-1} = \frac{1}{||E||_2}$. The condition

$$rank(A) = rank(B) = rank(P_{R(A)} B P_{R(A^H)}) \tag{11}$$

characterizes the perturbations for which the Moore penrose-inverse is well behaved. If the equation (11) holds then the matrix B is an acute Perturbation of A.

Theorem 2.8.1. *If rank (A+E) = rank (A)= r and* $\eta = ||A^{\dagger}||_2 ||E||_2 < 1$, *then* $||(A+E)^{\dagger}||_2 \leq \frac{1}{1-\eta} ||A^{\dagger}||_2$.

Proof.

$$\frac{1}{||(A+E)^{\dagger}||_2}$$

$$= \sigma_r(A+E) \geq \sigma_r(A) - ||E||_2$$

$$= \frac{1}{||A^{\dagger}||_2 - ||E||_2} > 0.$$

□

2.9 Perturbation theory of SVD

Let A be an $n \times m$ matrix. Then

$$U^T A V = \begin{pmatrix} \Sigma \\ 0 \end{pmatrix},$$

U and V are unitary matrices. Let $\tilde{A} = A + E$ be a perturbation of A and let

$$\tilde{U}^T \tilde{A} \tilde{V} = \begin{pmatrix} \tilde{\Sigma} \\ 0 \end{pmatrix}$$

be singular value decomposition of $\tilde{A}$ [13]. A and $\tilde{A}$ are compared using perturbation bound and perturbation expansion. A perturbation bound gives an upper bound on the

difference between perturbed quantity and its original; say between σ_i and $\tilde{\sigma}_i$ in terms of a norm of E. A perturbation expansion sets to approximate $\tilde{\sigma}_i$ as a function of E. First order perturbation expansion expresses $\tilde{\sigma}_i$ in the form

$$\tilde{\sigma}_i = \sigma_i + \Psi(E)0(||E||^2),$$

where Ψ is a linear function.
By definition,

$$||E||_2 = \lim_{||x||_2=1} max||Ex||_2$$

the spectral norm is written as $||.||_2$ Both the spectral and Frobenius norms are unitary fixed in that $||U^H EU|| = ||E||$ for all the unitary matrices U and V; where the spectral norm of E is the biggest singular value of E and the sum of squares of the singular values of E is the square of the Frobenious norm.

2.9.1 Singular values perturbation bounds

Weyl and Mirsky are behind the basic perturbation bounds for the singular values of a matrix

Theorem 2.9.1.

- No singular values can move than the norm of pertubation

$$|\tilde{\sigma}_i - \sigma_i| \leq ||E||_2.$$

-

$$\sqrt{\Sigma(\tilde{\sigma}_i)) - \sigma_i^2} \leq ||E||_F).$$

2.10 Outer product form and blocked matrices of SVD

In Σ, the outer columns and rows of matrix A can be eliminated if matrix product $A = U\Sigma V^T$ is expressed using sub-matrices as

$$A = [u_1 \dots u_k | u_{k+1} \dots u_m] \left(\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_k & \\ \hline & 0 & & 0 \end{array}\right) \begin{pmatrix} v_1^T \\ \vdots \\ v_k^T \\ - \\ v_{k+1}^T \\ \vdots \\ v_n^T \end{pmatrix}.$$

Multiplying the sub-matrices we have

$$A = \begin{pmatrix} u_1 \dots u_k \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_k^T \end{pmatrix} \oplus \begin{pmatrix} u_{k+1} & \dots & u_m \end{pmatrix} (0) \begin{pmatrix} v_{k+1}^T \\ \vdots \\ v_n^T \end{pmatrix}.$$

The second part vanishes since only the first k of the u_i and v_i make contributions to matrix A.

$$A = (u_1 \dots u_k) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix} \begin{pmatrix} v_1^T \\ \vdots \\ v_k^T \end{pmatrix}.$$

We have 2 rectangular matrices u_i and v_i $(m \times k)$ and $(k \times n)$ respectively. Then the diagonal matrix is square $(k \times k)$ which is alternative formulation of SVD.

Proposition 2.10.1. *Let* $\boldsymbol{k}$ *be the rank of matrix* A *of dimension* $(m \times n)$ *be communicated in the structure* $A = U\Sigma V^T$ *given* $U \in R^{m \times k}$, $\Sigma \in R^{k \times k}$ *diagonal matrix with positive entries and* $V \in R^{n \times k}$ *matrix.*

In matrix product rows are multiplied by columns

$$WZ = \sum_{i=1}^{2} w_i z_i^T,$$

where w_i are columns of W and Z_i^T are rows of Z. Let

$$W = (u_1 \dots u_k) \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix} = (\sigma_1 u_1 \dots \sigma_k u_k)$$

$$Z = \begin{pmatrix} v_1^T \\ \vdots \\ v_k^T \end{pmatrix}.$$

With $A = WZ$ expressed as outer product expansion

$$A = \sum_{i=1}^{k} \sigma_i u_i v_i^T,$$

thus

$$A\vec{x} = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

$$A\vec{x} = \sum_{i=1}^{k} v_i^T \vec{x} \sigma_i u_i,$$

where u_i is a vector formed as a result of linear combination of $A\vec{x}$.

3 LEAST SQUARE PROBLEMS

3.1 Linear least squares problems and their normal equations

Linear least squares problems occurs when comprehending linear systems.
When reducing norms of the residual vector, we may locate an important estimated solutions for overdetermined solutions which has no solutions.
Given a matrix $A \in \mathbb{R}^{m\times n}$, where $n < m$ and $\vec{b} \in \mathbb{R}^m$, we are searching for $\vec{x} \in \mathbb{R}^n$ for which the norm of the residual r is reduced i.e.

$$\| r \| = \| \vec{b} - A\vec{x} \| .$$

When we choose $2-$ norm, we simplify the calculation. The square of the residual vector length is thus reduced

$$\| r \|_2^2 = r_1^2 + r_2^2 + \ldots + r_2^m = \sum_{i=1}^{m} r_i^2.$$

The minimum is accomplished by $\vec{x} \in \mathbb{R}^n$, we note that $E = \{\vec{b} - A\vec{x} \mid x \in \mathbb{R}^n\}$ is closed, a non empty and convex subset of $\mathbb{R}^m$.

E has a unique element with the least norm hence $\exists$ an $\vec{x} \in \mathbb{R}^n$ such that $\| \vec{b} - A\vec{x} \|_2$ is reduced. From the reducing problem

$$\| r \|_2^2 = r_1^2 + r_2^2 + \ldots + r_2^m$$

emerges the name least squares method. The following theorem characterizes the least squares solution.

Theorem 3.1.1. *Least square solution*
Let $L = \{\vec{x} \in \mathbb{R}^n with \| \vec{b} - A\vec{x} \|_2\} \to min$ *be the set of solutions and assume* $r_x = \vec{b} - A\vec{x}$ *be the residual for a particular* $\vec{x}$*. Then*

$$\vec{x} \in L \Longleftrightarrow A^T r_x = 0 \Longleftrightarrow r_x \perp R(A). \qquad (12)$$

where the columns of A are the basis of the subspace $R(A)$

Proof. We start by demonstrating the underlying identicalness.
" $\Leftarrow$ " Let $A^T r_x = 0$ and $z \in \mathbb{R}^n$ be an arbitrary vector, it pursues that

$$r_z = \vec{b} - Az = \vec{b} - A\vec{x} + A(x - z)$$

thus

$$r_z = r_x + A(x - z).$$

Now

$$\| r_z \|_2^2 = \| r_x \|_2^2 + 2(x - z)^T A^T r_x + \| A(x - z) \|_2^2 .$$

But $A^T r_x = 0$ and therefore $\| r_z \|_2 \geq \| r_x \|_2$ since this holds for z then $\vec{x} \in L$.

" $\Rightarrow$ " we demonstrate this by logical inconsistency, let $A^T r_x = z \neq 0$ we consider $u = x + \varepsilon z$ with $\varepsilon > 0$

$$r_u = by - Au = \vec{b} - A\vec{x} - \varepsilon Az = r_x - \varepsilon Az.$$

Now

$$\| r_u \|_2^2 = \| r_x \|_2^2 - 2\varepsilon Z^T A^T r_x + \varepsilon^2 \| AZ \|_2^2 .$$

Because

$$A^T r_x = z$$

we obtain

$$\| r_u \|_2^2 = \| r_x \|_2^2 - 2\varepsilon \| Z \|_2^2 + \varepsilon^2 \| AZ \|_2^2$$

for adequate small ε we can acquire $\| r_u \|_2^2 < \| r_x \|_2^2$. Since in this case $\vec{x}$ can't be in the set of solutions, we reason this is an inconsistency. Consequently the suspicions was not valid i.e. we must have $A^T r_x = 0$, this demonstrates the underlying comparability in condition equation(12).

□

Gauss-markoff theorem expresses the least squares solutions which has an important statistical property. By linear relation let the $\vec{b}$ of perceptions be identified with an obscure parameter $\vec{x}$

$$A\vec{x} = \vec{b} + \varepsilon,$$

where $A \in \mathbb{R}^{m \times n}$ is a known matrix and ε is a vector of random errors.

Theorem 3.1.2. *Gauss markoff Consider the standard linear model*

$$A\vec{x} = \vec{b} + \varepsilon.$$

Then the best linear unbiased estimator of any linear function $C^T \vec{x}$ is the least square solution of $\| A\vec{x} - \vec{b} \|_2^2 \to min$.

The equation (12) can be utilized to decide the least square arrangement structure $A^T r_x = 0$ where, $A^T(\vec{b} - A\vec{x}) = 0$ and the Gauss normal equations can be acquired

$$A^T A\vec{x} = A^T\vec{b}. \tag{13}$$

To approximate $\vec{b}$, we need to get the linear combination of columns of A. The range of A is the spanned space by columns of A. The hyperplane in $\mathbb{R}^m$ is $R(A)$ and $\vec{b}$ doesn't lie in this hyperplane. Thus reducing $\| \vec{b} - A\vec{x} \|_2$ is similar to reducing the residual vector length r and thus be $R(A)$ has to be symmetrical to r. The normal equation (13) concentrate data since $Y = A^T A$ is a small $n \times n$ matrix whereas A is $m \times n$. The matrix Y is symmetric and it's rank $(A) = n$, then it is also positive definite. Cholesky decomposition is the common way of solving the normal equations.

1. From $Y = A^T A$ (we process just the upper triangle since Y is symmetric then $c = A^T\vec{b}$).
2. Since R is an upper triangular matrix, then $Y = R^T R$ (cholesky decomposition)
3. The arrangement by forward substitution $R^T\vec{b} = c$ and back substitution $R\vec{x} = \vec{b}$.

The methods for computing the least squares solution all use orthogonal matrices (i.e matrices Y for which $Y^T Y = I$).
Using Gaussian elimination method to solve linear system of the form $A\vec{x} = \vec{b}$ with equal equations and unknowns, we diminish the framework to triangular framework structure utilizing the way that proportionate frameworks have similar arrangements.
$A\vec{x} = \vec{b} \Longleftrightarrow YA\vec{x} = BY\vec{b}$ if Y is non singular. Confine ourselves to the class of symmetrical matrices $A\vec{x} \approx \vec{b} \Longleftrightarrow \vec{b}A\vec{x} \approx BY\vec{b}$ if Y is symmetrical then the least squares problems remains equivalent. Since $r = \vec{b} - A\vec{x}$ and $Yr = \vec{b}Y$, $YA\vec{x}$ have the same length

$$\| Yr \|_2^2 = (Yr)^T(Yr) = r^T Y^T Yr = r^T r = \| r \|_2^2 . \tag{14}$$

3.2 QR Decomposition.

Theorem 3.2.1. *Let A be m by n with $m \geq n$. Suppose that A has full column rank. Then there exists a unique m by n orthogonal matrix Q $(Q^T Q = I_n)$ and a unique n by n upper triangular matrix R with positive diagonals such that $A = QR$.*

Proof. Algorithm
procedure(Gramschmidt)
The column vectors of matrix A

$$A = [g_1 \mid g_2 \mid \ldots \mid g_n],$$

then

$$u_1 = g_1$$

$$e_1 = \frac{u_1}{\| u_1 \|}$$
$$u_2 = g_2 - (g_2.e_1)e_1$$

$$e_2 = \frac{u_2}{\| u_2 \|}$$
$$u_{k+1} = g_{k+1} - (g_{k+1}.e_1)e_1 \ldots (g_{k+1}.e_k)e_k$$
$$e_{k+1} = \frac{u_{k+1}}{\| u_{k+1} \|}.$$

The resulting QR factorization is;

$$A = [g_1 \mid g_2 \mid \ldots \mid g_n][e_1 \mid e_2 \mid \ldots \mid e_n]$$

$$= \begin{bmatrix} g_1e_1 & g_2e_2 & \ldots & g_ne_1 \\ 0 & g_2e_2 & \ldots & g_ne_2 \\ \vdots & & & \\ 0 & 0 & \ldots & g_ne_n \end{bmatrix}.$$

□

The classical Gram-Schmidt is numerically unstable.

3.3 Condition number and perturbation theory for least squares problems

The condition number of a matrix A is characterized by SVD. If singular values satisfy;

$$\sigma_1 \geq \ldots \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0,$$

then the rank of A is R. where $p = min(m,n)$. Then

$$\kappa(A) = \frac{\sigma_1}{\sigma_r}$$

defines the condition number

Theorem 3.3.1. *Assume that the matrix $A \in \mathbb{R}^{m\times n}$, where $m \geq n$ be the solution of the least squares problem $min\, \vec{x}||A\vec{x} - \vec{b}||_2$. Let δA and $\delta\vec{b}$ be perturbations such that*

$$n = \frac{||\delta A||_2}{\sigma_n} = \kappa\varepsilon_A < 1$$

$$\varepsilon_A = \frac{||\delta A||}{||A||_2}.$$

The perturbated matrix $A + \delta A$ has full rank and the perturbation of the solution $||\delta x||_2 \leq \frac{\kappa}{1-\eta}(\varepsilon_A||x||_2) + \frac{||\delta b||_2}{||A||_2} + \varepsilon_A\kappa\frac{||r||_2}{||A||_2}$,
where $r = \vec{b} - A\vec{x}$ is the residual.

The number κ determines the condition of the Least squares, the equality becomes the linear system of equation if $m = n$, where the residual r is equal to zero.
For the case of over-determined system , the residual is usually not equal to zero where the conditioning depends on κ^2. This maybe significant if the norm of the residual is large.

3.4 SVD and Least squares problems

The tool for analyzing and solving linear least squares problems is SVD. It also plays a key role in algorithms for approximating a given matrix with a matrix of lower rank. This is applied in data compression.
By truncating the SVD expansion of matrix A, we obtain the best approximation of $A \in \mathbb{C}^{m\times n}$ by a matrix of lower rank .

3.4.1 SVD and the Moore Penrose inverse

Theorem 3.4.1. *Consider the least square problem $min_x\, ||A\vec{x} - \vec{b}||_2$, where $A \in \mathbb{C}^{m\times n}$ and rank $(A) \leq$ min(m,n). Assume that*

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix}.$$

U_1 and V_1 are r columns and the diagonal matrix $\Sigma > 0$. Then the solution

$$x = V_1 \Sigma_1^{-1} U_1^H \vec{b}$$

is the unique pseudoinverse.

Proof. Let $\vec{x} = Vz$, using unitary invariance of the spectral norm, then

$$\left\| \vec{b} - A\vec{x} \right\|_2 = \left\| U^H (\vec{b} - AVz) \right\|_2$$

$$\left\| \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \right\| - \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \left\| \begin{bmatrix} c_1 - \Sigma_1 z_1 \\ c_2 \end{bmatrix} \right\|$$

$c_i = U_i^H \vec{b}$, i=1,2.
Minimum value is attained by the residual norm which is equal to $||c_2||_2$ for $z_1 = \Sigma_1^{-1}$ and z_2 are arbitrary. The choice of $z_2 = 0$ reduces $||x||_2 = ||Vz||_2 = ||z||_2$.
The pseudoinverse of matrix A is

$$A^\dagger = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{bmatrix} \Sigma_1^1 & 0 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} U_1^H \\ U_2^H \end{pmatrix} = V_1 \Sigma_1^1 U_1^H \in \mathbb{C}^{m \times n}.$$

The above represents the SVD of $A^\dagger$ and maps $\vec{b}$ to $\vec{x}$. Can also be written as

$$\vec{x} = \sum_{i=1}^{r} \frac{c_i}{\sigma_i} v_i, c_i = u_i^H \vec{b}.$$

□

Theorem 3.4.2. *If $A = \in \mathbb{C}^{m \times r}, B = \in \mathbb{C}^{r \times n}$ and rank (A)=rank(B)=r, hence*

$$(AB)^\dagger = B^H (BB^H)^{-1} (A^H A)^{-1} A^H = B^\dagger A^\dagger.$$

Proof. $A^H A$ and BB^H are square matrices with rank r and therefore nonsingular. By demonstrating that the Penrose conditions are fulfilled gives the evidence. □

An inner inverse is matrix A^{-1} satisfying the initial Moore Penrose- inverse condition $AA^{-1}A = A$. An outer inverse is one fulfilling the second condition.
Let A^{-1} be an inner inverse of A, then solution is $\vec{x} = A^{-1}\vec{b}$, where the system $A\vec{x} = \vec{b}$ is consistent for all $\vec{b}$, where

$$\vec{x} = A^{-1}\vec{b} + (I - A^{-1}A)z, z \in \mathbb{C}^n$$

is the general solution. For any inner inverse of A,

$$(AA^{-1})^2 = AA^{-1}AA^{-1} = AA^{-1}$$

and

$$(A^{-1}A)^2 = A^{-1}AA^{-1}A = A^{-1}A.$$

Thus the idempotent projectors are AA^{-1} and $A^{-1}A$. Let $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$. When $\vec{x}$ satisfies the normal equations $A^H A\vec{x} = A^H\vec{b}$, then $||A\vec{x} - \vec{b}||_2$ is reduced. $(AA^{-1})^H = AA^{-1}$ is satisfied by the generalized inverse A^{-1} Then the orthogonal projector onto R(A) is AA^{-1} where A^{-1} is a least squares inverse. If $\vec{x} = A^{-1}\vec{b}$ satisfies the normal equations and has least square solution, then we have

$$A^H = (AA^{-1}A)^H = A^H AA^{-1}$$

[3]. Conversely, when $\vec{x} = A^{-1}\vec{b}$, $||A\vec{x} - \vec{b}||_2$ is reduced for all $b \in \mathbb{C}^m$ if $A^{-1} \in \mathbb{C}^{n\times m}$.

3.5 SVD and matrix approximation

Theorem 3.5.1. *Assume $A \in \mathbb{C}^{m\times n}$ with singular values $\sigma_1 \geq \ldots \geq \sigma_p \geq 0$, where $p = min(m, n)$. Then, if $\mathbb{S}$ fulfills a linear subspace of $\mathbb{C}^n$, one has that $\sigma_i = maxmin||A\vec{x}||_2 = minmax||A\vec{x}||_2$ [4].*

Theorem 3.5.2. *The ordered singular values σ_i of $A \in \mathbb{C}^{m\times n}$ interlace singular values σ_i of the bordered matrix*

$$\hat{A} = \begin{pmatrix} A \\ u^H \end{pmatrix},$$

as follows

$$\hat{\sigma}_1 \geq \sigma_1 \geq \hat{\sigma}_2 \geq \ldots \geq \hat{\sigma}_m \geq \sigma_m \geq \hat{\sigma}_{m+1}, m < n$$

$$\hat{\sigma}_1 \geq \sigma_1 \geq \hat{\sigma}_2 \geq \ldots \geq \hat{\sigma}_{n-1} \geq \sigma_{n-1} \geq \hat{\sigma}_n, m \geq n.$$

When A is bordered by a column, then a similar result holds.

Proof. For Hermitian matrices, the theorem is as a result of Cauchy's interlacing theorem. Also the eigenvalues of the leading principal minor of order $n-1$ of a Hermitian matrix B interlace those of B. Since

$$\begin{pmatrix} A^H \\ u^H \end{pmatrix} \begin{pmatrix} A & u \end{pmatrix} = \begin{pmatrix} A^H A & A^H u \\ u^H A & u^H u \end{pmatrix}$$

.

$$\begin{pmatrix} A \\ v^H \end{pmatrix} \begin{pmatrix} A^H & v \end{pmatrix} = \begin{pmatrix} AA^H & Av \\ v^H A^H & v^H v \end{pmatrix}.$$

It's seen that the singular values of A are the positive square underlying the foundations of the eigenvalues of $A^H A$ and AA^H [4]. □

Theorem 3.5.3. *Let $A \in \mathbb{C}^{m\times n}$ and $B_k = X_k Y_k^H$, where $X_k, Y_k \in \mathbb{C}^{m\times k}$. Then rank $(B_k) \leq k < \min m, n$ and*

$$\sigma_1(A - B_k) \geq \sigma_{k+1}(A),$$

where i^{th} singular value of its argument is denoted by $\sigma_i(\dot{u})$ [4].

Proof. Let the right singular vectors of A be $v_i, i = 1 : n$. Since k is the rank Y and $k < n$, there is a vector $v = c_1 v_1 + \ldots + c_k v_k + c_{k+1} v_{k+1}$ such that $||v||_2^2 = c_1^2 + \ldots + c_{k+1}^2$ and $Y^H v = 0$. It follows that

$$\sigma_i^2(A - B_k) \geq v^H (A - B_k)^H (A - B_k) v = v^H A^H A v$$

$$|c_1|^2 \sigma_1^2 + \ldots + |c_{k+1}|^2 \sigma_{k+1}^2.$$

□

Theorem 3.5.4. *Let $A \in \mathbb{C}^{m\times n}, m \geq n$ has polar decomposition $A = PH$ with $P \in \mathbb{C}^{m\times n}$ unitary, $P^H P = I_n$ and $H \in \mathbb{C}^{m\times n}$. Hermitian and positive semi-definite. H is positive definite iff rank $(A) = n$ and this decomposition is unique.*

Proof. Let $A = U_1 \Sigma V^H, U_1 \in \mathbb{C}^{m\times n}$, be the "thin" SVD and set

$$P = U_1 V^H, H = V \Sigma V^H.$$

Since $V^H V = I$, then

$$PH = U_1 V^H V \Sigma V^H = U_1 \Sigma V^H = A.$$

□

The SVD of A can help us obtain the polar decomposition as shown by the proof. Having the polar decomposition $A = PH$, SVD can be constructed from spectral decomposition $H = V \Sigma V^H$ i.e $A = (PV)\Sigma V^H$ [4].

3.5.1 Normal equations

For n distinct points $(t_1,\ldots,t_n)$ and data $(b_1,\ldots,b_n) \in \mathbb{R}$, there exists a least squares polynomial $P_m(t) = c_0 + c_1 t + \ldots + c_m t^m$ of degree m which fits the data in the sense that

$$\| \vec{r} \|_2 = \| \vec{b} - A\vec{x} \|_2 \tag{15}$$

is minimum, where $\vec{x}=(c_0, c_1, \ldots, c_m)^T$ is of unknown coefficients and can be determined by use of system of normal equations and use of a Vandermonde matrix. The system of normal equations is :

$$\sum_{i=0}^{m} c_i \left[\sum_{k=1}^{n} t_k^{i+\ell} \right] = \sum_{k=1}^{n} t_k^{\ell} b_k, \ \ell = 0, 1, \ldots, m$$

or, in matrix form

$$\begin{bmatrix} n+1 & \sum t_k & \sum t_k^2 & \ldots & \sum t_k^n \\ \sum t_k & \sum t_k^2 & \sum t_k^3 & \ldots & \sum_k^{n+1} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \sum t_k^n & \sum t_k^{n+1} & \sum t_k^{n+2} & \ldots & \sum t_k^{n+m} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} \sum b_k \\ \sum b_k t_k \\ \sum b_k t_k^2 \\ \vdots \\ \sum b_k t_k^n \end{bmatrix}. \tag{16}$$

3.5.2 Polynomial curve fitting

Suppose we are given m distinct points $(t_1,\ldots,t_m)$ and data $(b_1,\ldots,b_m) \in \mathbb{R}$ at these points. Then $\exists$ a unique polynomial degree n.

$$p(t) = c_0 + c_1 t + \ldots + c_{n-1} t^{n-1} + c_n t^n. \tag{17}$$

For some $n < m$ such a polynomial is a least squares fit to the data if it minimizes the sum of the squares of the deviation from the data

$$\sum_{i=1}^{r} |b_i - pt_i)|^2. \tag{18}$$

This sum of squares is equal to the norm squared of the residual for the rectangular Vandermonde system [11]
Basically to determine the coefficients $c_0, c_1, c_2, \ldots, c_{n-1}, c_n$ of equation (17) such than it interpolates the n points

$$(t_1, b_1), \ldots, (t_n, b_m)$$

is to write a linear system of equations as follows

$$\begin{aligned} P_n(t_1) = b_1 &\Rightarrow c_0 + c_1 t_1 + c_2 t_1^2 + \ldots c_{n-1} t_1^{n-1} + c_n t_1^n \\ P_n(t_2) = b_2 &\Rightarrow c_0 + c_1 t_2 + c_2 t_2^2 + \ldots c_{n-1} t_2^{n-1} + c_n t_2^n \\ \vdots &\Rightarrow \vdots \\ P_n(t_n) = b_m &\Rightarrow c_0 + c_1 t_n + c_2 t_n^2 + \ldots c_{n-1} t_{n-1}^{n-1} + c_n t_n^n \end{aligned}$$

or, in matrix form:

$$\underbrace{\begin{bmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^{n-1} & t_1^n \\ 1 & t_2 & t_2^2 & \ldots & t_2^{n-1} & t_2^n \\ 1 & t_3 & t_3^2 & \ldots & t_3^{n-1} & t_3^n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{n-1} & t_{n-1}^2 & \ldots & t_{n-1}^{n-1} & t_{n-1}^n \\ 1 & t_n & t_n^2 & \ldots & t_n^{n-1} & t_n^n \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \\ c_n \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}}_{\vec{b}}, \tag{19}$$

We have a system of a linear equations to solve but with more equations than unknowns. Hence system is described as $\vec{b} = A\vec{x}$. Since A is a rectangular matrix, the MP- inverse can be performed [4].
Using least squares estimation

$$\vec{x} = (A^T A)^{-1} A^T \vec{b}, \tag{20}$$

we can obtain the coefficients.

3.5.3 Application to unemployment rates data in Kenya

The Unemployment rates data in Kenya from 1991 to 2018.

Table 1. Unemployment rates data in Kenya

T(Years)	1991	1992	1993	1994	1995	1996	1997
U. Rate	33.9110	34.0979	34.3160	34.5469	34.7630	34.9800	35.1139

Table 2. Unemployment rates data in Kenya

T(Years)	1998	1999	2000	2001	2002	2003	2004
U. Rate	35.2580	35.3809	36.3810	37.4580	38.5509	39.6940	40.8409

Table 3. Unemployment rates data in Kenya

T(Years)	2005	2006	2007	2008	2009	2010	2011
U. Rate	41.9819	41.9230	41.7900	41.5680	42.1730	42.0390	41.9189

Normal equations

The below equations (21) and (22) are the polynomials of degree 3 and 5 from equation (15)

$$\begin{bmatrix} 28 & 56.126 & 112.506394 & 225.5263912 \\ 56.126 & 112.506394 & 225.5263912 & 452.0896742 \\ 112.506394 & 225.5263912 & 452.0896742 & 1785.766069 \\ 225.5263912 & 452.0896742 & 1785.766069 & 1816.770941 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1092.739006 \\ 2191.084116 \\ 4393.47875 \\ 8809.776316 \end{bmatrix} \tag{21}$$

$$\begin{bmatrix} 28 & 56.126 & 112.506 & 225.526 & 452.089 & 1785.766 \\ 56.126 & 112.506 & 225.526 & 452.089 & 1785.766 & 1816.770 \\ 112.506 & 225.526 & 452.089 & 1785.766 & 1816.770 & 3642.072 \\ 225.526 & 452.089 & 1785.766 & 1816.770 & 3642.072 & 7301.363 \\ 452.089 & 1785.766 & 1816.77 & 3642.072 & 7301.363 & 14637.483 \\ 1785.766 & 1816.770 & 3642.072 & 7301.363 & 14637.483 & 29345.122 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix} = \begin{bmatrix} 1092.739 \\ 2191.084 \\ 4393.478 \\ 8809.776 \\ 17665.588 \\ 69984.043 \end{bmatrix} \tag{22}$$

The system can be computed by Gauss elimination method or any other direct computation to solve the linear system of equations, since the matrices formed are square matrices. Below is the table for the order of polynomial, the coefficients value and the condition number.

Table 4. Unemployment rates data in Kenya

T(Years)	2012	2013	2014	2015	2016	2017	2018
U. Rate	41.8779	41.9070	41.9580	42.0139	42.0960	42.0960	42.1030

Order of Polynomial	Coeffients	Condition number
3	-716.6696 376.9998 -0.0000 -0.0000	$71.59 * 10^3$
5	39.2927 -0.0076 -0.0069 -0.0050 -0.0033 -0.0020	$276.0352 * 10^4$

The results obtained are not significant because the matrices obtained are ill conditioned and this is one of the limitations. The use of normal equation to solve linear systems is the fastest technique but least accurate.

Vandermonde matrix

A logical extension to the linear regression curve fit is to use a higher- order polynomial such as

$$p(t) = c_0 + c_1 t + \ldots + c_{n-1} t^{n-1}. \tag{23}$$

A third degree or cubic polynomial fit is of the form

$$p(t) = c_0 + c_1 t + c_2 t^2 c + c_3 t^3. \tag{24}$$

The fifth degree polynomial is of the form

$$p(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3 + c_4 t^4 + c_5 t^5. \tag{25}$$

The first degree polynomial is similar to linear. The below equations (26) and (27) are from the polynomials in equations (24) and (25) respectively and similar system follows for any polynomial from equation (23)

$$\begin{bmatrix} 1 & 1.991 & 3.964081 & 7.8925 \\ 1 & 1.992 & 3.9680 & 7.9044 \\ 1 & 1.993 & 3.9720 & 7.9163 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2.017 & 4.0682 & 8.2057 \\ 1 & 2.018 & 4.0723 & 8.2179 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{27} \\ c_{28} \end{bmatrix} = \begin{bmatrix} 33.9110 \\ 34.0979 \\ 34.3160 \\ \vdots \\ 42.0960 \\ 42.1030 \end{bmatrix}. \tag{26}$$

$$\begin{bmatrix} 1 & 1.991 & 3.9640 & 7.8925 & 15.7139 & 31.2864 \\ 1 & 1.992 & 3.9680 & 7.9044 & 15.7455 & 31.3650 \\ 1 & 1.993 & 3.9720 & 7.9163 & 15.7771 & 31.4439 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 2.017 & 4.0682 & 8.2057 & 16.550 & 33.3832 \\ 1 & 2.018 & 4.0723 & 8.2179 & 16.5838 & 33.4661 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{27} \\ c_{28} \end{bmatrix} = \begin{bmatrix} 33.9110 \\ 34.0979 \\ 34.3160 \\ \vdots \\ 42.0960 \\ 42.1030 \end{bmatrix}. \tag{27}$$

We pick the level of polynomial for which the coefficients are figured by equation (20) is a base or when there is no noteworthy abatement in its incentive as the level of polynomial is expanded. If a third order polynomial is chosen, we get the residual from equation (18)

$$p(\vec{x}) = 10^7 * 1.2757 + 10^7 * -19122T + 10^7 * 0.9554T^2 + 10^7 * -0.1591T^3 + \varepsilon. \tag{28}$$

Below is the table for the order of polynomial, the coefficients value and the residual value

Order of Polynomial	Coefficients	Residuals
3	$1.2757 * 10^7$ $-1.9122 * 10^7$ $0.9554 * 10^7$ $-0.1591 * 10^7$	3.7903
4	$1.2068 * 10^7$ $-1.8016 * 10^7$ $0.8927 * 10^7$ $-0.1449 * 10^7$ $-0.0009 * 10^7$	3.7979
5	$2.8297 * 10^9$ $-5.2881 * 10^9$ $3.5138 * 10^9$ $-0.8756 * 10^9$ $0.0001 * 10^9$ $0.0217 * 10^9$	3.2197
10	$-1.9551 * 10^9$ $-0.8561 * 10^9$ $0.5002 * 10^9$ $1.6075 * 10^9$ $0.0000 * 10^9$ $0.6337 * 10^9$ $-1.0226 * 10^9$ $-0.9741 * 10^9$ $1.5135 * 10^9$ $-0.6085 * 10^9$ $0.0808 * 10^9$	2.6002

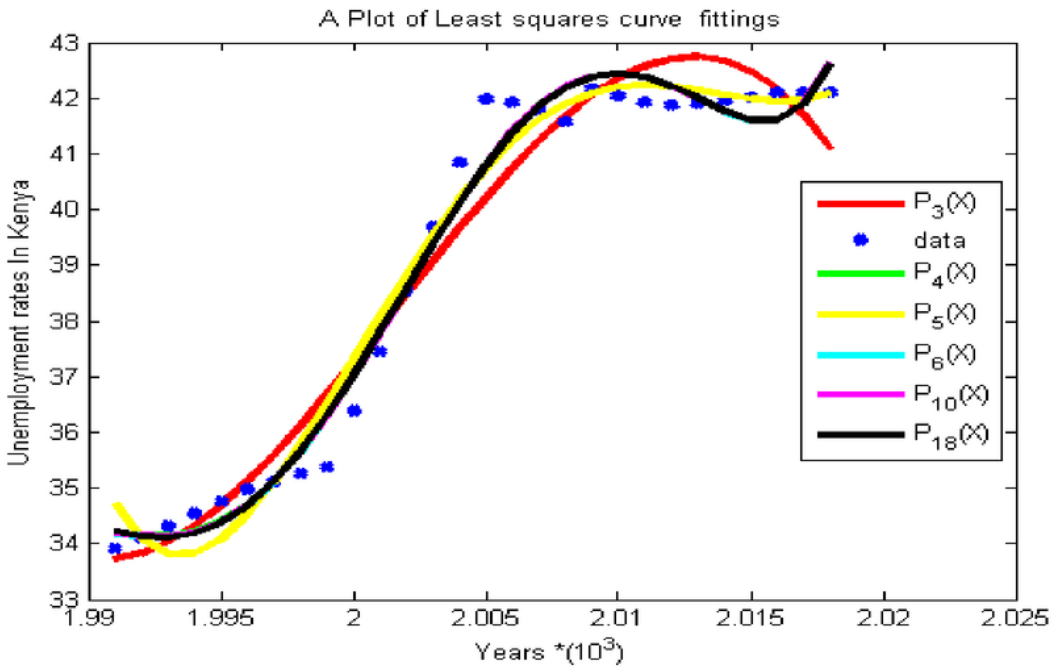


Figure 2. Different degrees polynomial of least squares fit to 28 data points of unemployment rates in Kenya.

From figure 2 it has modelled the expected value of $\vec{b}$ as a $3^{rd}, 4^{th}, 5^{th}$ 10^{th} and 18^{th} degree polynomial yielding the general polynomial. The values of the constants $c_0, c_1, \ldots, c_k$, where $k = 1, \ldots, 10$ are achieved by the use of equation (20). The higher the degree the lower the error. From equation (2) the error is minimal when the polynomial is of degree 18, obtained from equation (18)

4 SINGULAR SPECTRUM ANALYSIS IN THE STUDY OF TIME SERIES

Spectral decomposition or eigenvalue is the origin of *singular spectrum* of a matrix say A. These eigenvalues, λ then make the matrix $||A - \lambda I||$ singular. The spectral decomposition of matrices of multivariate data is also referred to as singular spectrum analysis.

4.1 Time series

Time series is a masterminded progression of estimations of a variable at comparably isolated time intervals. If a single variable is measured then the time series is multivariate otherwise univariate. When time series is measured at finite steps then it is discrete time series and data used in this work is a discrete time series is $X_t; t = 0, 1, 2, \ldots$.
Times series has some components like seasonal and trend, where trend in time series occur when there is a pattern of continous increase, decrease or stagnation over time.

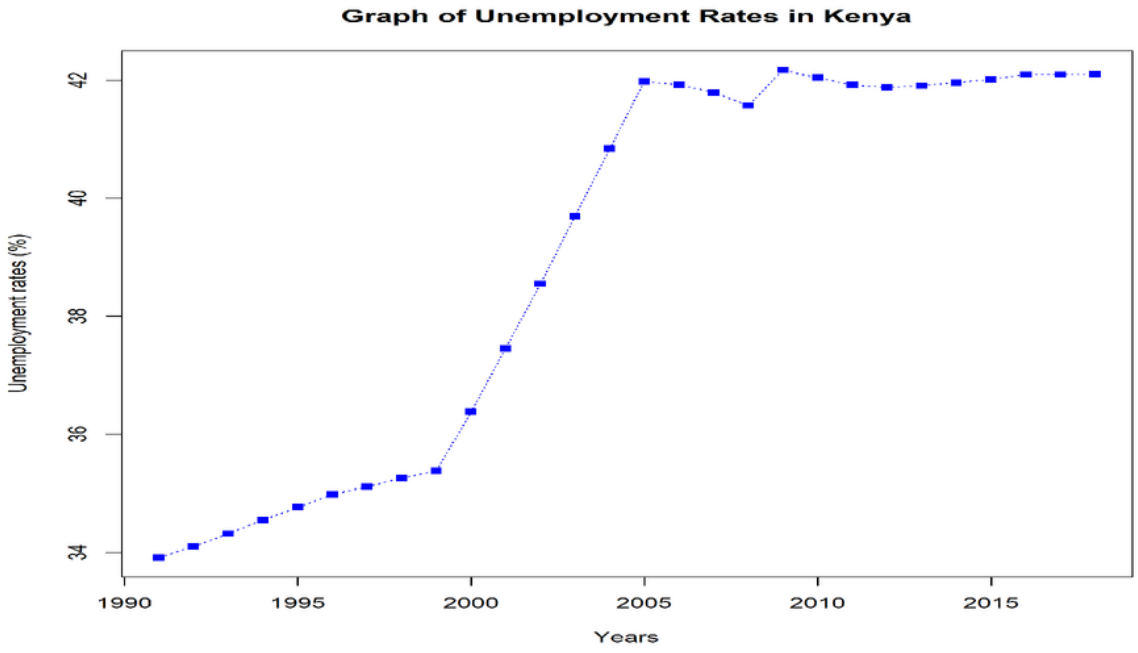


Figure 3. Graph of unemployment rates in Kenya from 1991 to 2018

The utilization of time series is two fold [23]

- Acquire an appreciation of the concealed powers and structure that made the watched data.
- Estimating observing or on the other hand even analysis and feed forward control. Time series is utilized in numerous applications, for example, deals estimating, remaining task at hand projections, evaluation investigation, budgetary examination.

4.2 Basic spectrum analysis

SSA is a non parametric time series method that decomposes, reconstructs and forecasts time series. It basically incorporates tools from time series analysis, multivariate data, dynamical systems and signal processing. SSA seeks to decompose the unique series into a whole of modest number of interpretable segments.
It is based on the singular value decomposition of a specific matrix constructed upon the time series. The basic SSA basically includes: decomposition stage and reconstruction stage.
The decomposition comprises of embedding and SVD. The reconstruction stage comprises of eigentriple grouping and diagonal average.

The structure of SSA algorithm is as follows: Initial, a one dimensional time series is changed over into a higher measurement matrix known as the trajectory matrix. Then dimension of the trajectory matrix is known as the window length.
Second, SVD is then applied to the trajectory matrix and eigenvalues and eigenvectors found. The next step is grouping step which basically involves splitting the elementary matrices into several groups and then summing the matrices in each group. The approximated series of the initial series is given by combining the reconstructed components obtained by taking the mean across the diagonals of all groups and combining them into a single series.
The advantage of SSA is that, after reconstruction of the time series under study it allows to produce forecast for the reconstructed components which is called SSA forecasting algorithm.

4.2.1 Decomposition stage

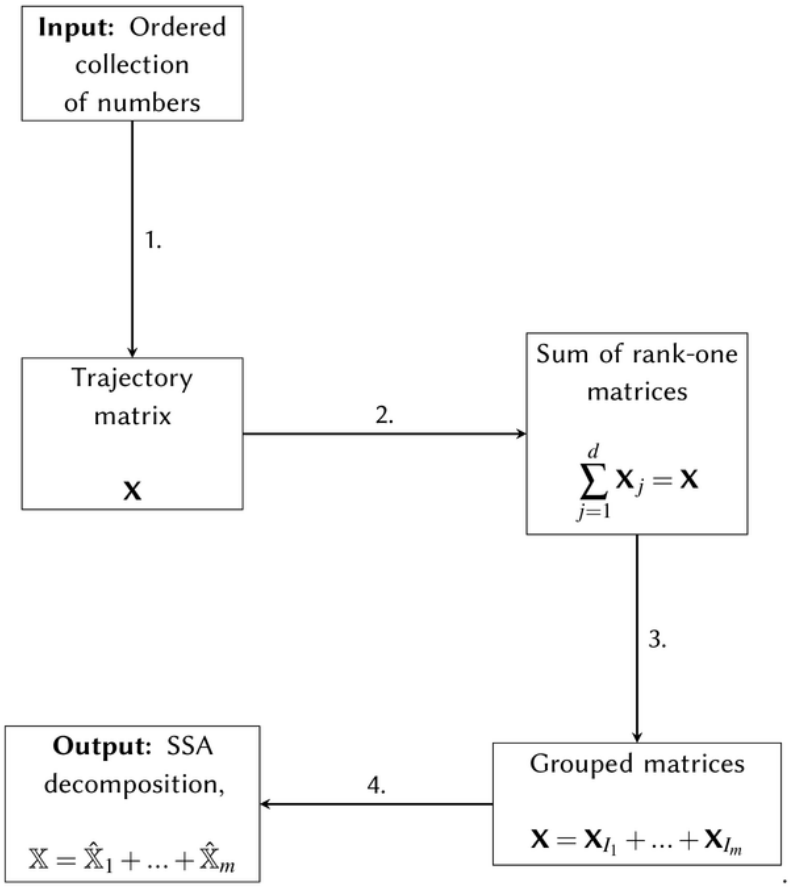
(a) Embedding
The purpose of the *embedding step* is to expand the single time series into a multidimensional matrix, called a trajectory matrix, which can then in turn be decomposed into various components.
This embedding is done by giving the time series a certain lag, usually one, and then

combining the resulting lagged column vectors into a matrix of a specific window length(number of columns). Consider $F = F_t = (f_i, \ldots, f_N)$, the time series of length N, where $N \geq 2$ and F is a non zero series: that is there exists at least one i such that $f_i \neq 0$. Let L be some integer known as the window length, which is $1 < L < N$. Then let $K = N - L + 1$.
Mapping the initial time series into a sequence of lagged vectors of size L performs the embedding by forming $K = N - L + 1$ *lagged vectors.*

$$X_i = (f_i, \ldots, f_{i+l-1})^T.$$

The *trajectory matrix* F can then be constructed by combining the lagged column vectors into a single matrix

$$X = [F_1, F_2, \ldots, F_N].$$



- An arranged accumulation of N numbers is the info.
-

$$X = \hat{X}_1 + \ldots + \hat{X}_m \tag{29}$$

is a decomposition of $\mathbb{X}$ in a sum of identifiable components:

The mathematical definition of the combined trajectory matrix therefore is:

$$X = \begin{bmatrix} f_1 & f_2 & f_3 & \cdots & f_K \\ f_2 & f_3 & f_4 & \cdots & f_{K+1} \\ f_3 & f_4 & f_5 & \cdots & f_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_L & f_{L+1} & f_{L+2} & \cdots & f_N \end{bmatrix}. \tag{30}$$

The above principle of the construction of the trajectory matrix can be illustrated by the following example.

Example 4.2.1. *Let $Y_t = [0,2,3,4,5,6,7,8,9,10]$. The time series can now be given a lag of 1 and be embedded in a matrix with window length of 5, which results to the below trajectory matrix, X:*

$$X = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 5 & 6 & 7 \\ 4 & 5 & 6 & 7 & 8 \\ 5 & 6 & 7 & 8 & 9 \\ 6 & 7 & 8 & 9 & 10 \end{bmatrix}.$$

The optimal size of the embedding window depends on the nature of the time series and it is vital in the analysis of the time series to determine the most favorable window length. This window length should be wide enough to sufficiently capture the global behavior of the system, but it should be kept in mind that the complexity of the analysis increases with the increase in the number of columns in the trajectory matrix. Because the width of embedding window is one of the two most fundamental parameters of SSA. The trajectory matrix has two main essential properties:

(a) The subseries of the initial series is made of both rows and columns are of **X**.

(b) **X** is a Hankel matrix since it has equivalent components on its enemies of diagonals.

(b) Singular Value Decomposition (SVD)
Once the trajectory matrix X has been formed, singular value decomposition (SVD)

can be performed [26].
Let $C = XX^T$ and denote by $\lambda_1, \lambda_2, \ldots, \lambda_L$ the eigenvalues of C in decreasing order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L \geq 0$. Let $U_1, U_2, \ldots, U_L$ be the orthonormal eigenvector's of the matrix C corresponding to those eigenvalues [26].
Let $V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}, (i = 1, 2, \ldots, d)$, where (**d**) is equal to the rank of the matrix X such that $\lambda_i > 0$ in any case, all things considered, series we typically have $d = L^*$, with L^* = min (L, K).
The triple $(\sqrt{\lambda_i}, V_i, U_i)$ is called the i^{th} eigentriple of the SVD [26]

4.2.2 Reconstruction stage

(c) Eigentriple Grouping
This progression relates to partitioning the elementary matrices X_i into several groups and summing the matrices in each group.
Let $I = i_1, \ldots, i_m$ and P_j has several X_i^j and $X_i = \sigma_i V_i^T U_i$. The expansion of $X = X_1 + \ldots + X_d$ leads to the decomposition.

(d) Diagonal Averaging
We change each matrix of the grouped decomposition equation (29) into a series of length N. Assume $Y \in L \times K$ with elements $y_{ij}, 1 \leq i \leq L, 1 \leq j \leq K$. Set $L^* = min(L, K), K^* = max(L, K)$ and $N = L + K - 1$. Assume $y_{ij}^* = y_{ij}$ otherwise averaging transfers the matrix Y to the series $y_1, y_2, \ldots, y_N$ using the formula

$$y_k = \begin{cases} \frac{1}{k} \sum_{i=1}^{k} y_{m,k-m+1}^* for & 1 \leq k < L^*; \\ \frac{1}{L^*} \sum_{i=1}^{L^*} y_{i,k-m+1}^* for & L^* \leq k < K^*; \\ \frac{1}{N-k+1} \sum_{m=k-K^*+1}^{N-K^*+1} y_{m,k-m+1}^* & for K^* \leq k < N. \end{cases} \tag{31}$$

This corresponds to the averaging of the matrix elements over the diagonal $i + j = k + 2$: The choice $k = 1$ gives $y_1 = y_{1,1}$, for $k = 2$ we have $y_2 = (y_{1,2} + y_{2,1})/2$, for $k = 3; y_3 = (y_{1,3} + y_{3,1} + y_{2,2})/3$.
Diagonal averaging
The series $\tilde{x}^k = (\tilde{y}_1^k, \ldots \tilde{y}_N^k)$ is produced by applying equation (31) applied to a resultant matrix X_{lk}, where the reconstructed series is $\tilde{X}^k$ and the initial series $Y = (y_1, y_2, \ldots, y_N)$ is decomposed into the sum of reconstructed series [26]

$$y_n = \sum_{k=1}^{m} \tilde{y}_n^k \qquad (n = 1, 2, \ldots, N). \tag{32}$$

Equation (30) can be explained in a more practical manner by the following example. If a matrix X_{rec} were obtained by summation of SVD to be retained, the original signal would

be extracted by calculating the averages of the respective diagonals, as illustrated

$$X_{rec} = \begin{bmatrix} 4 & 3 & 4 & 5 & 1 \\ 3 & 5 & 6 & 5 & 6 \\ 3 & 9 & 4 & 5 & 8 \\ 9 & 7 & 8 & 9 & 8 \\ 2 & 7 & 5 & 4 & 1 \\ 3 & 4 & 9 & 6 & 4 \\ 9 & 5 & 6 & 8 & 7 \end{bmatrix}.$$

The resulting series will then be: $Z_{rec} = [4, 3, 4, 7.25, 3.8, 4.8, 7, 6.5, 4.3, 6, 7]$. Elementary reconstructed series is the reconstructed series obtained by the elementary grouping. In the following sections this method will be applied to a real time series.

5 APPLICATION OF SSA-THE CASE OF UNEMPLOYMENT DATA

To represent the utilization of SSA method, the time series formed by annually unemployment data in Kenya from 1991 to 2018. This area is of interest because unemployment is one of the current issues in Kenya. This makes unemployment a pressing, disturbing economic and social issue in Kenya.

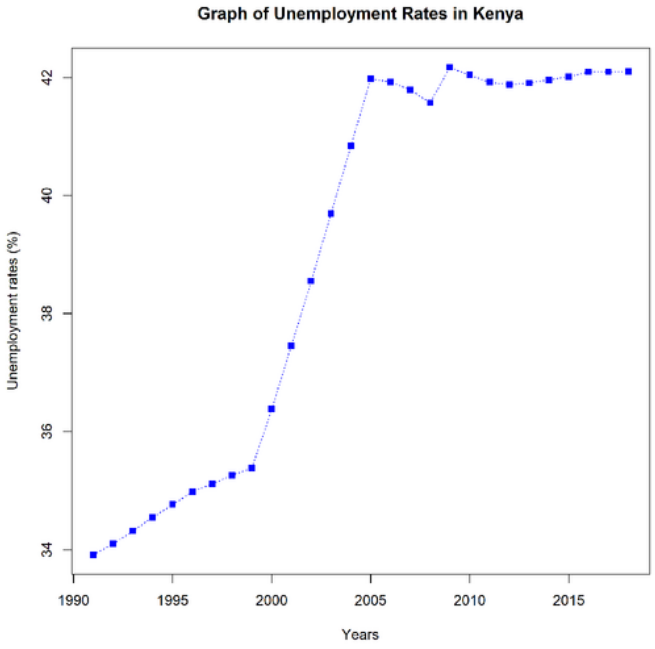


Figure 4. Time series graph for unemployment rate in Kenya between 1991 to 2018.

The time series plot of Unemployment in Kenya, in indicates an upward trend with a slow down after 2004. However, the trend is not regular. This shows that we cannot model the resultant trend using simple linear regression model or using an exponential model. The time series components of these series can be withdrawn using sequential SSA with a given window length of say, 7. With sequential SSA, some components are extracted with the first window length and others with the second window length weighted-correlation matrix This matrix has weighted correlations between the reconstructed time series components. This matrix helps in separability of the components based on the strength of their correlation. Correlated components are grouped into one group.
All the analysis was done using R[28]. The SSA part of the analysis was conducted using

the RSSA package and code in [26].

5.1 Choice of SSA parameters

The choice of parameters depends on the data we have and the analysis to perform. Basically SSA has two parameters, the window length L and the number of components r for reconstruction. Using the information given by time series, the values for L and r could be defined[32].

5.1.1 Selection of the *window length*

In the decomposition stage, the only parameter is the window length **L**. Selection of the best *window length* depends on the problem in hand. To get a better separability of this component, it is recommended to choose the window length proportional to the period since time series may have a periodic component. Basically L should be large enough but not greater than N /2 [26].
The number of components r: The basis for the definition of r is how well the components can be separated. The variance of X, evaluated as $\frac{\lambda_i}{\Gamma}(\Gamma = \sum_{i=1}^{d} \lambda_i)$ is the main criterion and is based on the contribution of the two parameters. Select r out of the components so that the sum of their contributions is at least a predetermined threshold.

The W-correlation matrix helps to decide the window length. The W-correlation matrix is a matrix of elementary reconstructed components. If two elementary reconstructed components are W-orthogonal, it means they are forcefully dissociable[26]. If two elementary reconstructed components are associable, it means they are highly W-correlated.
Then W-correlation matrix for $L = 4$ is shown below.

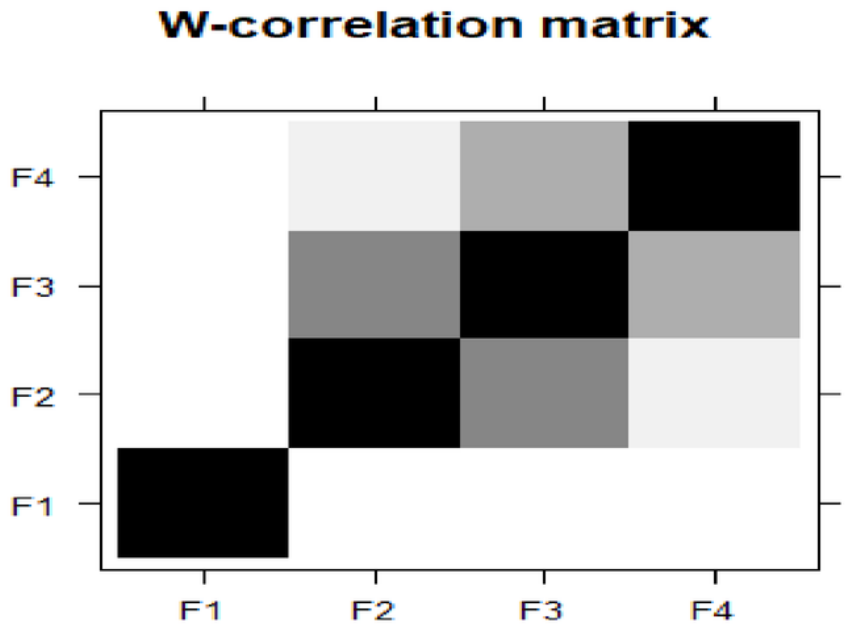
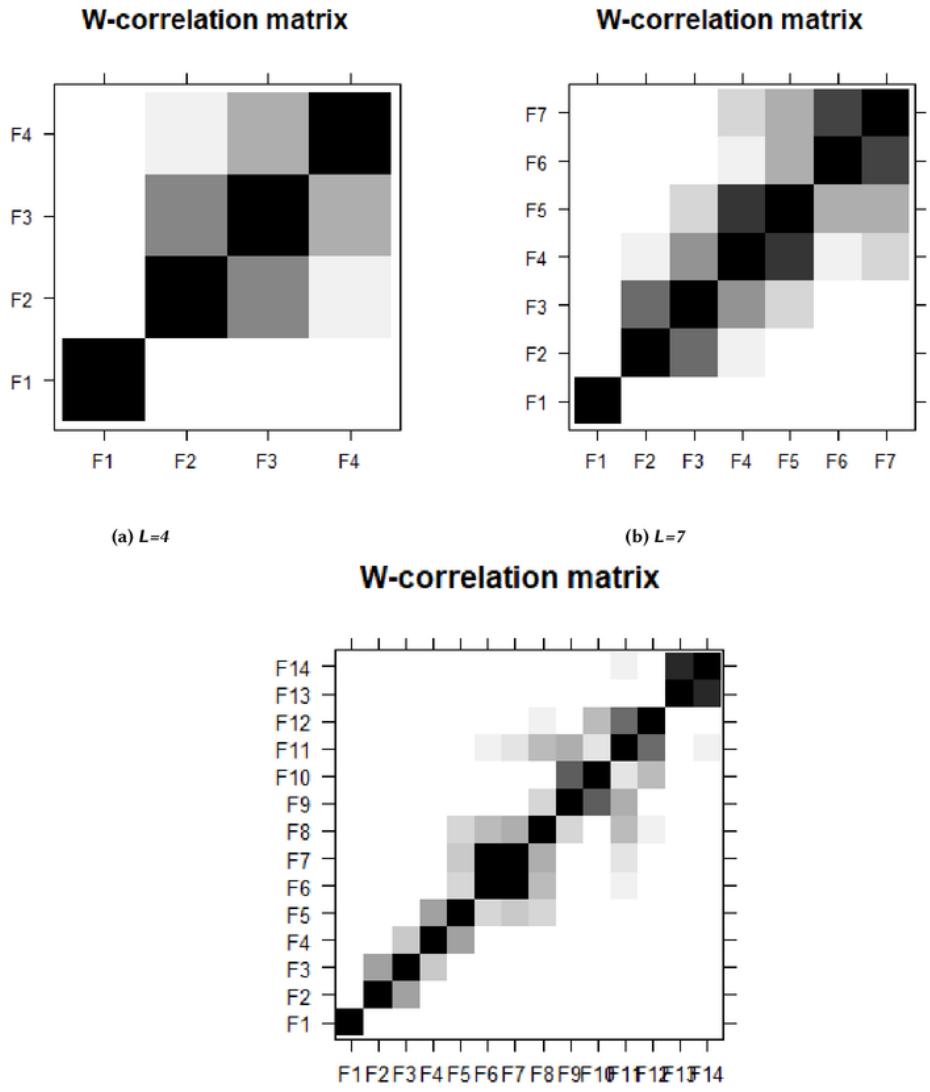


Figure 5. W-correlation matrix of unemployment data with L=4.

The strength of the W-correlation between two components is represented by the shade of each square in figure 5. From figure 5 it can be seen that the first component is W-uncorrelated with the rest of the components. Hence the trend is described by the first eigentriple.

The highly correlated components are the second and third components with the rest and the fourth one is slightly correlated. Similar relationship is observed in the W-correlation matrices in $L = 7, 14$ even when the components change because the matrix X changes when we change the length L.

Figure 6a almost suggests a diagonal shape because the eigentriples tend to be correlated mainly with the neighboring components and not with distant components. As L increases, the components tend to be correlated with more different components even if the correlation is light sometimes. Comparing the plots with different window length L, its seen that as L increases we see the last eigentriples be somehow correlated with more eigentriples. For instance when L = 7, the correlated matrix is almost diagonal shape. When L is increased to 14, the W-correlation matrix almost becomes a funnel shape. Its therefore preferable to work with window length L=7.



(a) *L=4* (b) *L=7*

(c) *L=14*

Figure 6. Comparing different window lengths of unemployment data $L = 4, L = 7 L = 14$.

5.1.2 Scree plot and eigenvectors plot

Scree plot is a basic line section plot of an eigenvalues in lessening request of their sizes i.e $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L$. The key in the scree plot is the "Shark break", which is the second component. Hence, the firs component is the first one. The first turning point is known as the "Shark break". The components before the "Shark break" are significant ones. Hence the first principal component matters most as illustrated by the eigenvectors plot in figure 8.

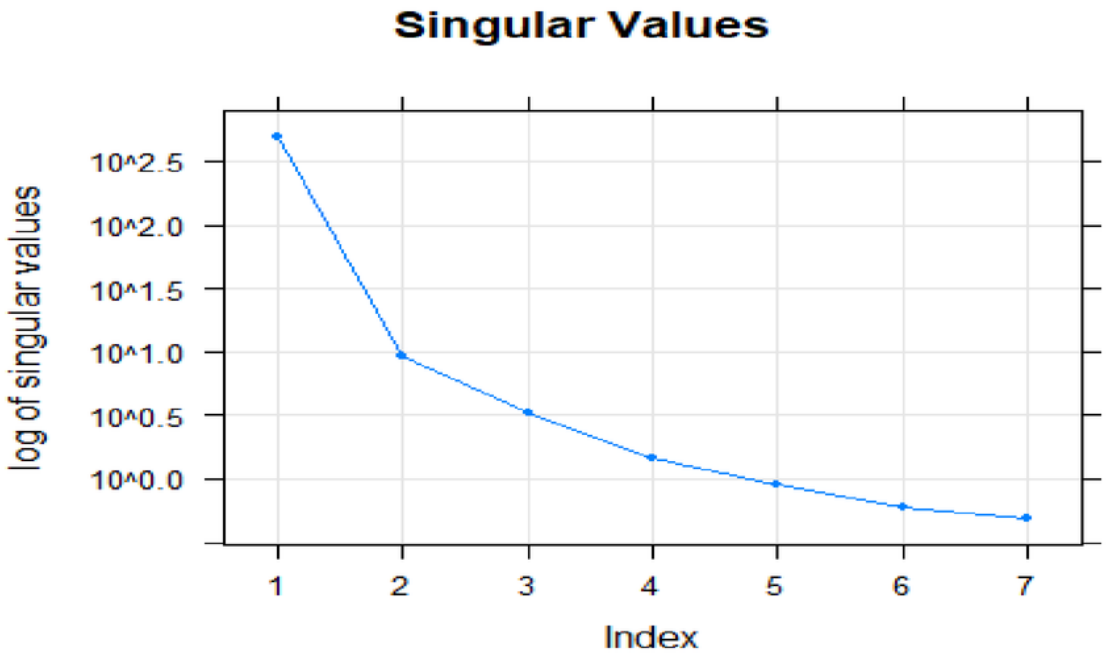


Figure 7. Scree plot of SSA for unemployment data.

Figure 7 shows the Scree plot for the unemployment data.

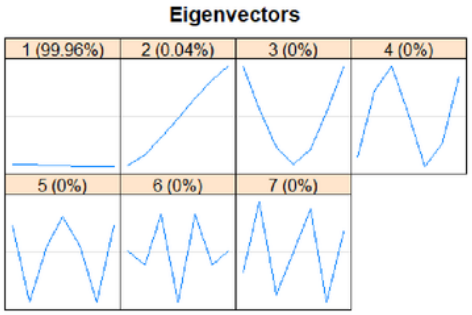


Figure 8. Eigenvectors plot for SSA for unemployment data.

The eigenvectors plot in figure 8 indicates that the leading eigenvector has almost constant coordinates. This kind of behavior of the eigenvectors is interpretable as the trend.

5.1.3 Diagonal averaging

This transforms the series of unemployment into the reconstructed series.

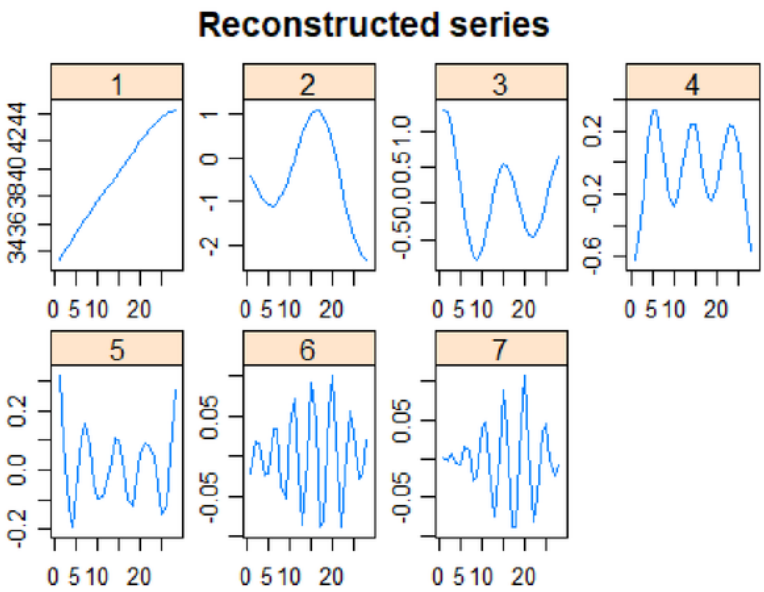


Figure 9. Reconstructed Series.

Figure 9 is the plot for the reconstructed series. The first component corresponds to the trend the rest corresponds to high frequency components which are not related to the trend.

5.2 Extraction of the seasonal part

The residuals are calculated by subtracting the values of the trend from the original time series. The residuals [10] are showed in figure 10.

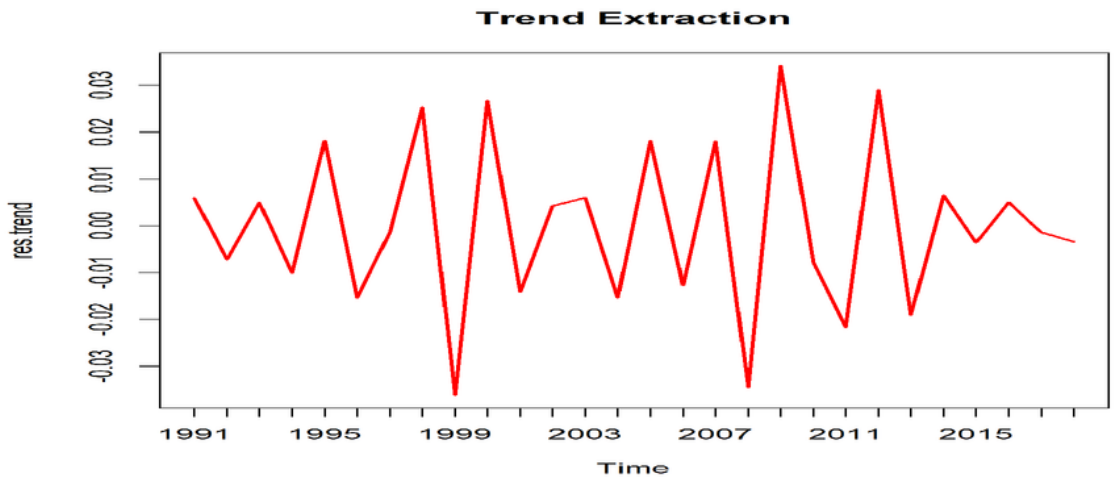


Figure 10. Residuals.

5.3 Forecasting method

In the previous section, we applied the main algorithm of SSA to unemployment data from Kenya. First, we created trajectory matrix and found the singular values and singular vectors. Then constructed principal and reconstructed components. In this section the goal is to forecast the unemployment rate for the next 10 years. the below figure 11 where last 10 points are predicted by SSA forecasting algorithm and figure 12 complete reconstructed components for (N+10).

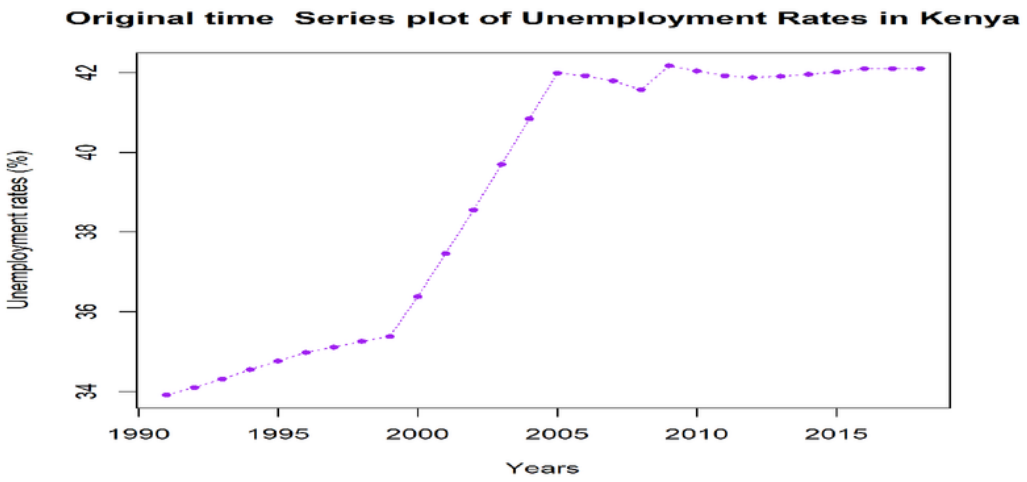


Figure 11. Original time series for unemployment rates in Kenya from 1991: 2018.

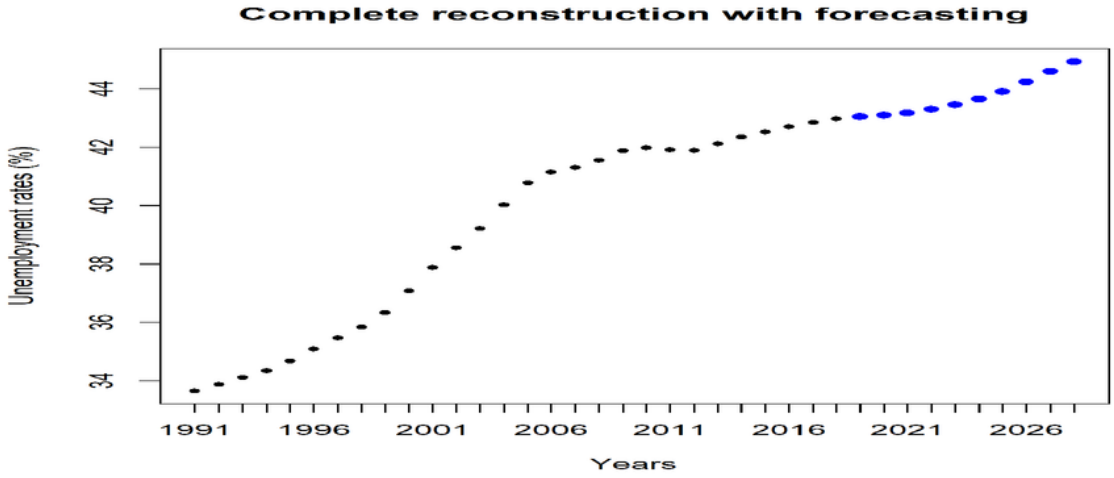


Figure 12. Complete reconstruction with forecasting for unemployment rates in Kenya.

Forecasting isn't done quickly by just looking at a graph 4 since the rate of Unemployment is . There is an increased Unemployment rate from 2018 to 2028. This may depend on several factors e.g courses offered in the education system and how relevant they are to the current economy.

6 CONCLUSION AND RECOMMENDATION

SSA has been considered on of the powerful technique for analysing a variety of time series has appeared over the past 20 years. Despite the fact that it's roots lie in the normal sciences, and the series arisen from such processes, it also be applied in different fields [21].

SSA is a non parametric method of time series analysis that decomposes time series into trend and built reconstructed components which are used for forecasting. It uses linear algebra tools. In this work we analysed and presented theoretical results on SSA applied unemployment data. By carrying out experiment with different window lengths we observed that by increasing the window length the forecast accuracy gets smaller and the forecast is more exact.

We also get different forecasting results by changing the number of components. It's clear that by using the sum of even indices reconstructed components we obtain with negative value which can't be "the best" selection for forecasting algorithm. However using the first two reconstructed components we got more exact forecasting. The comparison forecasting results showed that there is an influence in the selection of SSA parameters to forecasting. It should be mentioned that the application of the SSA forecasting algorithm has given us some expected results but has not yet showed its potential. In future work it is crucial to study forecasting accuracy with even indices reconstructed components. Also can apply the SSA Forecasting algorithm and compare the results with other forecasting methods.

Bibliography

[1] A. Horn. On the eigenvalues of a matrix with prescribed singular values *Proc. Amer. Math. Soc.*, 1954.

[2] A. Horn. On the singular values of a product of completely continous operators.*Proc. Nat. Acad. Sci.*USA, 1951.

[3] A. B. Israel and T. N. E. Greville. *Generalized Inverses: Theory and Apllication. Second Edition.* Springer, 2003.

[4] Ake Bjorck. "Numerical methods in Matrix Computations", Springer Science and Business Media LLC, 2015.

[5] Alter O, Brown P.O. Botstein D. Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci USA 2000..

[6] Aldrich, John, "Eigenvalue, eigenfunction, eigenvector, and related terms", in Jeff Miller (Editor), Earliest Known Uses of Some of the Words of Mathematics, retrieved 2006.

[7] B. Rajendra. *Matrix Analysis*, 1997.

[8] C. Aldrich. "Process system identification strategies on the use of singular spectrum analysis", Minerals Engineering, 2003.

[9] C. H. Park and H.Park. Nonlinear discriminant analysis using kernel functions and the generalized singular value decomposition.*SIAM Journal on Matrix Analysis and Applications*, 2005.

[10] Danilov, D. and Zhigljavsky, A. Principal components of time series Saint Petersburg Press, Saint Petersburg 1997.

[11] Fan, Jianquing. *Local Polynomial Modelling and Its Applications: From Linear to nonlinear regression.* Monograpghs on Statistics and Applied Probability, 1996.

[12] G. W. Stewart and J. G. Sun.*Matrix Pertubation Theory*. Academic Press, New York, 1990.

[13] G. W. Stewart and J. G. Sun. *Matrix Pertubation Theory*. Academic Press, New York, 1990.

[14] G. E. Forsythe, M. A. Malcolm, and C. B. Moler. Computer Methods for Mathematical Com-putations. Prentice-Hall, 1977.

[15] G. Strang. Introduction to Linear Algebra. Wellesley-Cambridge Press, 1993.

[16] Gass, S.I "Singular value decomposition in AHP," 2004.

[17] Golub, G.H., and Van Loan, C.F. Matrix Computations, 2nd ed. (Baltimore: Johns Hopkins University Press) 1989.

[18] Hefferon, Jim, Linear Algebra, Online book, St Michael's College, Colchester, Vermont, USA 2001.

[19] Hassani, H. Singular Spectrum Analysis: Methodology-and Comparison. *Journal of Data Science 5*, 2007.

[20] Héctor Briceñoa, Claudio M. Roccoa, Enrico Zio, Singular Spectrum Analysis For Forecasting Of Electric Load Demand. 2013. Aidic.it. Retrieved from **http://www.aidic.it/cet/13/33/154.pdf**

[21] H. Hassani and D. Thomakos, A review on Singular Spectrum Analysis for Economic and Financial Time Series, Statistics and its Interface, 3 2010.

[22] James W. Demmel. "3 Linear Least Squares Problems", Society for industrial and Applied Mathematics (SIAM), 1997.

[23] Monica Chis , Crina Grosan. " Evolutionary Hierarchical Time Series Clustering", Sixth International Conference on Intelligent Systems Design and Applications, 2006.

[24] M. Ghil, M. R. Allen, M. Dettinger, K. Ide, D.Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P.Yiou, Advanced Spectral Methods For Climatic Time Series, Reviews of Geophysics 40 2002.

[25] M. Erdmann. Lecture notes for 16-811 Mathematical Fundamentals for Robotics. The Robotics Institute, Carnegie Mellon University, 1998.

[26] N. Golyandina, V. Nekrutkin and Zhigljavsky, Analysis of Structure of Time Series, Springer, New York 2013.

[27] N. Golyandina, A. Zhigljavsky, Basic Singular Spectrum Analysis and Forecasting with R, submitted to Computational Statistics and Data Analysis, 2013.

[28] R Core Team. R: A language and environment for statistical computing.R Foundation for Statistical Computing, Vienna, Australia, 2012.
htt.//www.r-project.org/

[29] S. Chandrasekaran and I. C. F. Ipsen. on rank-revealing factorisations. *SIAM Journal on Matrix Analysis and Applications*, 1994.

[30] T. F. Chan. Rank revealing QR factorizations. *Linear Algebra and its Applications*, 1987.

[31] T. Alexandrov, A method of trend extraction using Singular Spectrum analysis, REVSTAT 7 (1) 2009.

[32] V. Poornima and V. Jothiiprkash. "Daily rainfall forecasting for one year in a single run using SingularSpectrum Analysis", Journal of Hydrology, 2018.

[33] William Ford. "The Singular Value Decomposition", Elsevier BV, 2015.

Trend Prediction of Unemployment in Kenya using Singular Spectrum Analysis

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

PAGE 54

PAGE 55

PAGE 56

PAGE 57

PAGE 58

PAGE 59

PAGE 60

PAGE 61

PAGE 62

PAGE 63

PAGE 64

PAGE 65