University of Cape Coast https://ir.ucc.edu.gh/xmlui

UNIVERSITY OF CAPE COAST

PRINCIPAL COMPONENT EXTRACTION WITH NO INFORMATION LOSS

VICTORIA ESI WILSON-SEY

2023

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

©Victoria Esi Wilson-Sey

University of Cape Coast

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

UNIVERSITY OF CAPE COAST

PRINCIPAL COMPONENT EXTRACTION WITH NO INFORMATION LOSS

BY

VICTORIA ESI WILSON-SEY

Thesis submitted to the Department of Statistics of the School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast, in partial fulfilment of the requirements for the award of Master of Philosophy degree in Statistics

FEBRUARY 2023

Digitized by Sam Jonah Library

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature .... Date ....

Name: Victoria Esi Wilson-Sey

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Supervisor's Signature .... Date ....

Name: Prof. Bismark Kwao Nkansah

Co-Supervisor's Signature .... Date ....

Name: Dr. Arimiyaw Zakaria

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

ABSTRACT

In Principal Component (PC) analysis of an $r \times p$ variance-covariance (V-C) matrix, there is always a loss of information when the first few set of $r(< p)$ PCs are retained. This study derives a new reduced set of PCs (NRPCs) that is simply a constant multiple of the first r original PCs (OPCs). Thus, the OPCs are just a normalization of the NRPCs. The normalizing constant represents the common variance explained by each of the components in the set of r NRPCs. Further features of the NRPCs are examined both analytically and practically in Multivariate Multiple Reduced Rank Regression (MMRRR) modelling. It is found that for the NRPCs extracted from regular (unweighted) V-C matrix, the analytical relationship between the NRPCs and the OPCs are preserved in MMRRR modelling. However, if OPCs are based on weighted V-C matrix, then the analytical relationship between the two types of PCs does not hold practically in MMRRR modelling. The results of the study shows that in order to determine the real spread of PC scores for further analysis, the use of the NRPCs would be more useful.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

KEY WORDS

Classical Principal Component

Generalized Principal Component

Orthogonality

Parsimonious

Variance-Covariance

Weights

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

ACKNOWLEDGEMENTS

I thank God Almighty, for his grace has brought me this far.

Good and kind people are also to be acknowledged all the time, especially those who support you to achieve something in life. I wish to express my sincerest gratitude to my supervisor, Prof. Bismark K. Nkansah, for his immense guidance, direction and patience. I believe I would not have finished without him. Again, I wish to say thank you to my co-supervisor, Dr. Arimiyaw Zakaria, for his help as well.

I say a big thank you to Mr. Solomon Kweku Afful, Dr. Agnes Adom-Konadu and Mrs. Suzzy Hurson for providing the necessary supports and encouragements, and to my sons and sister, Cephas, Michael and Joana, for their perfect support, understanding and patience particularly during my numerous absence from home while studying.

Lastly, a very cordial words of thanks to my course mates, especially Gyamu-Atta Pius who stood by me through out and was very instrumental in the writing of this thesis.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

DEDICATION

To my children and Mr. Solomon Kweku Afful

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

TABLE OF CONTENTS

Digitized by Sam Jonah Library

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

LIST OF TABLES

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

LIST OF FIGURES

Page

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

LIST OF ABBREVIATIONS

CMD	Coefficient of Multiple Determination
PC	Principal Component
MLR	Multiple Linear Regression
MMRRR	Multivariate Multiple Reduced Rank Regression
NRPCs	New Reduced Principal Components
OPCs	Original Principal Components
PCR	Principal Component Regression
RRR	Reduced Rank Regression
SVD	Singular Value Decomposition
V-C	Variance-Covariance

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

CHAPTER ONE

INTRODUCTION

Dimensionality reduction is the process of transforming multidimensional data into a significant representation of one with decreased dimensionality (Van der Maaten, 2007). Ideally, the decreased representation has a dimensionality that is equal to the intrinsic dimensionality of the data. The intrinsic dimensionality of the data is the minimum number of parameters needed to account for the observed properties of the data. Dimensionality reduction is central in many areas, since it makes classification, visualization, and compression of high-dimensional data much easier. A fundamental technique for dimensionality reduction is the Principal Component Analysis (PCA). This technique basically obtains an orthogonal transformation of a multivariate data. Such a tranformation may further be used by concentrating on the features of the data that are captured along the first few transformed variables or components that retains the geometry of the original data as much as possible. Such a procedure is widely referred to as dimensionality reduction. In spite of widespread application of dimensionality reduction, it comes along with its own curse of loss of information. This study attempts to examine the various perspectives that dimensionality reduction has been exploited and to make a novel attempt at addressing the issue of loss of information that is associated with dimensionality reduction.

Background to the Study

The concept of Principal Components (PCs) is well-documented in notable texts on multivariate statistical analysis (Anderson, 2003; Johnson & Wichern, 2002). The purpose of extracting PCs from dimensional multivariate data whether by the classical or generalized approach, is to achieve two main objectives; to transform the original data into one for which the variables are uncorre-

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

lated such that the new components account for decreasing amounts of variation in the data, and to use the first few components to account for variation in the data. Invariably, the use of PC is motivated by the second objective. This is why, for example, a method of Truncated Singular Value Decomposition (TSVD) (Xu, 1998) is widely used. However, if the level of multicollinearity in the data is low as a result of low dependence among the variables, all new components which are equal in number as the original set of variables will account for some amount of variation. This means that by using the first few components, there is a tradeoff between parsimonious use of variables and a substantial loss of information contained in the last few variables that are discarded. The information loss as a result of the tradeoff could be high depending on the number of components retained and the information they capture. The loss of information could affect applications of the method in many ways.

PCs are computed from either the covariance matrix or the correlation matrix. If PCA are extracted from the correlation matrix, then what is effectively being done in such analyses is to standardize the variables, and then finding linear mathematical relation of these standardized variables which in turn maximize variation. The importance of standardizing the variables is to give all variables equal weight, whereas the original variables may have huge differences in their variances. In the latter situation, the variables with high variances will have a significant influence on the first few PCS, which is often undesirable, although sometimes it can be precisely what is desired. Another reason for using the correlations is that the variables may be measured in different units. In that case, the relative sizes of the variances and covariances depend critically, and arbitrarily, on the units used to measure the various different variables. Jollife (2002) recounts that standardization of the variables is a good strategy for overcoming this arbitrariness.

The linear combinations of the original variables give the principal com-

Digitized by Sam Jonah Library

ponents. This is expressed mathematically as

$$y_i = \sum_{j=1}^{p} w_{ij} x_j; \qquad i = 1, 2, \cdots, p. \tag{1.1}$$

where y_i is the ith principal component, x_j is the jth variable, and w_{ij} is the weight of the jth variable for the ith principal component. Most importantly, it is essential to provide meanings to the linear combinations that are formed after the transformation. One popular way of achieving this ideal is to use the loadings of the PCs, which are mathematically represented as

$$l_{ij} = \frac{w_{ij}}{s_j} \sqrt{\lambda_i} \tag{1.2}$$

In Equation (1.2), l_{ij} is the loading of the jth variable for the ith principal component, λ_i is the variance or eigenvalue of the ith principal component and s_j is the standard deviation of the jth variable. The lower the loading of a variable, the less influential it is in the formation of the PCs and vice versa. The loadings can therefore be used to decide which variables significantly influence the construction of the PCs. Depending on the significance of the loadings, appropriate meaning or label can be given to each of the p PCs. However, it may not be of practical use to interpret all p new components. A rule of thumb is that an influential variable in the formation of a PC should have a loading whose absolute value is greater than or equal to 0.5.

For PC to be used as a dimensionality reduction technique, one is usually concerned about the relative importance of the first few PCs that are to be retained within the context of the study. Another point of concern is the amount of information that may be lost as a result.

Digitized by Sam Jonah Library

Singular Value Decomposition

Singular value decomposition (SVD) is that which represents any matrix $\mathbf{X}$ of column rank r (where $n \geq p$) with dimension $n \times p$ as a multiplication of three matrices, $A, D,$ and B such that

$$\mathbf{X} = ADB' \tag{1.3}$$

where matrix A and B' are of dimensions $n \times r$ and $r \times p$ respectively, and D is an $r \times r$ diagonal matrix. A and B are orthonormal.

According to Baker (2005), three perspectives that are compatible with one another can be used to examine SVD . It can be viewed as a technique for converting a set of correlated variables into a set of uncorrelated ones that more clearly reveal the numerous relationships between the original data items. At the same time, SVD is a technique for figuring out and ranking the dimensions along which data points display the dominant variation. The third way of viewing SVD, is that once the dominant variation has been located, we can use fewer dimensions to determine the best approximation of the original data points. Thus, SVD can be seen as a method for data reduction. As an illustration of these ideas, we consider a 2-dimensional data points in Figures 1 and 2. In Figure 1, the regression line through the points displays the most accurate representation of the original data using a one-dimensional object (a line). This is because it is the line with the shortest distance between each original point and the line.

A condensed representation of the original data that captures as much of the original variance as feasible would be obtained if we took the perpendicular line from each point to the regression line and used the intersection of those lines as the approximation of the original datapoint. This second regression line,

Digitized by Sam Jonah Library

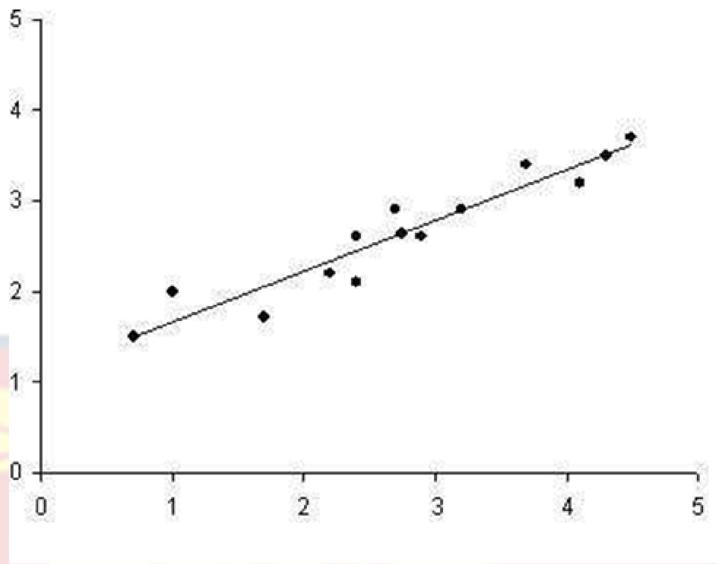

Figure 1: Best-fit regression line regression line reduces data from two dimensions into one

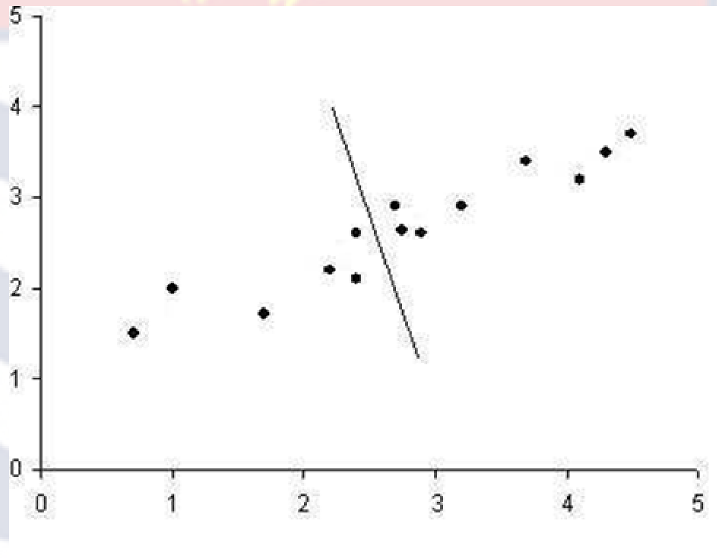

Figure 2: Regression line along second dimension that captures less variation in the original data

perpendicular to the first, is shown in Figure 2.

This line effectively brings out the variability along the original data set's second dimension. It can however be seen that unlike that in 1, this line does a deficient work of estimating the original data. This is because it matches a dimension presenting less variability. It is feasible to create a set of uncorrelated data points using these regression lines that will reveal subgroupings in the original data that might not be obvious at first glance.

Digitized by Sam Jonah Library

Application of PCA in Regression

A notable application of Principal Component Analysis is in regression modelling (Liu et al., 2001). The procedure in this application is what is usually referred to as Reduced Rank Regression (RRR) modelling. In order to introduce the RRR, the general concept of regression is briefly highlighted.

Linear Regression

Simple regression analyses describe the relationship that exists between two variables by fitting a straight line, normally called the line of best fit, through the set of data points. The simple linear regressions model is represented as

$$Y = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \tag{1.4}$$

where Y is the value of the response (dependent) variable, X is the value of the predictor (independent) variable, β_0 and β_1 are the regression coefficients, and ε is the random error term. ε is assumed to be standard normal. Two common methods used to estimate β_0 and β_1 are the eyeball fitting method and the method of least squares.

An extension of the simple linear regression is the multiple linear regression. In MLR, Y is the dependent variable with more than just one X variable. Suppose there are p independent variables, the MLR model is

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} \tag{1.5}$$

where $\beta_0, \beta_1, \beta_2, \cdots, \beta_p$ are the regression coefficients that must be estimated from sample data.

MLR demonstrates that a set of independent variables explains a significant amount of the variation in a response variable through a significant test

Digitized by Sam Jonah Library

known as the coefficient of determination (R^2). The R^2 is the percent of variance in the dependent variable, explained collectively by all of the independent variables. Mathematically,

$$R^2 = 1 - \frac{SSe}{SST} \tag{1.6}$$

where SSe is the error sum of squares and SST is the total sum of squares. In addition, the relative predictive importance of the independent variables can be ascertained by comparing the beta weights. To explore the curvilinear effects, quadratic terms of the independent variables can be added as independent variables. Cross-product terms can also be added as independent variables to explore interaction effects. Several R^2s can be compared to determine if removing or adding an independent variable to the model helps significantly. The regression coefficient $\beta_j, j = 0, 1, 2, \cdots, p$ is the average amount the dependent changes in magnitude when the independent increases one unit and the other predictors are held constant. The smaller the magnitude of β_j, the less the dependent variable changes for each unit change in the independent, and vice-versa. The $\beta_j, j = 0, 1, 2, \cdots, p$ is the unstandardized simple regression coefficient for the case of one independent. When there are two or more independents, each coefficient is a partial regression coefficient, though it is commonly called regression coefficient. When they are standardized, they are usually known as weights.

Reduced Rank Regression

Reduced Rank Regression (RRR) is a multivariate linear regression method where several response variables are related to the same set of explanatory variables, and the estimated matrix of regression coefficients is of reduced rank. This means that the number (r) of estimated regression coefficients is now less than the original number (p) of original predictor variables. When several response variables are studied simultaneously, we are in the sphere of multivariate

Digitized by Sam Jonah Library

regression. The usual description of the multivariate regression model that relates the set of m multiple responses to a set of reduced r predictor variables, assumes implicitly that the $m \times r$ regression coefficient matrix is of full rank. The literature does not appear to have focused on the effect of the reduced number of predictors on the subsequent regression model. It should be of interest to examine the effect of the procedure that led to the reduction in the number of predictors on the subsequent modelling process.

Extraction of Principal Components

The classical principal component (CPC) is extracted based on the various covariance matrix (S) of the data. The use of the CPC has been found to be fraught with challenges especially when the data contains extreme values. In order to obtain the desired PCs that reflect the real dominant dimensionality of the data, various studies (Da Costa et al., 2011; Hong et al., 2018) have adopted the notion of a weighted principal component analysis (WPCA). The WPCA has been approached from two perspectives. One approach may extract PCAs based on a weighted variance covariance matrix that pre-assigns relative weights to the observations based on some similarity measure such as the Mahalanobis distance. An example of such a matrix is the one that gives rise to the generalized principal component analysis (GPCA). It will be apparent in the literature that the use of the PCA has focused on reducing the effect of anomalies in the variance covariance structure on the results of the true dimensionality of the data. It appears that the loss of information associated with component extraction has not been the interest for research.

Statement of the Problem

Dimensionality reduction techniques are all known to be associated with information loss. The amount of information loss no matter the size, could have

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

important implication depending on the sensitivity of the area to which the techniques are applied. In spite of the known consequences, one has no choice so far than to truncate or reduce dimensions only with the motivation of obtaining a parsimonious representation of the data. It will be ideal to remove the tradeoff between a parsimonious choice and loss of information. PCA is popular for multivariate orthogonal data transformation that has been widely used for dimensionality reduction. As pointed out in the background, PCA has seen widespread application in various fields, and research is also ongoing that seeks to improve upon the original PCs to achieve various results. It is noticeable that attempts in this direction have sought to obtain weighted PCs in order to overcome the challenge introduced by a defective variance covariance matrix, a matrix that is not positive definite or has some other structural issues that makes it inadequate for certain statistical analysis. It is clear however that there is no attempt in the literature to achieve the ideal of removing the information loss associated with the procedure. As a result, there is the need for an attempt at producing reduced dimensions in a high dimensional dataset such that the information in the original dataset could be preserved along the reduced dimensions.

Objectives of the Study

In this study therefore, the intention is to derive a variant of reduced set of PCs that retains all information. This will be done by incorporating into the desired few components, the required graduated amounts of information that each of the few components must retain so that they collectively explain the total information.

The study's specific objectives are to:

1. propose a new component extraction approach that does not lead to information loss in the original data

2. examine the characteristics of the new components in relation to the orig-

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

inal PC extracted based on different variance-covariance structures

3. examine specific features of the new components applied in reduced rank regression modelling

Significance of the Study

Apart from the benefit of retaining all information in a few reduced components, there are other specific use for this proposal. In this study, an application will be made to MLR analysis. The approach will serve as a useful alternative to Best Subset Regression. It will be possible therefore to make a geometric representation of MLR by reducing it to simple linear so that the relationship among the variables could be presented in the plane. This application could be extracted to MMLR such that analysis of data obtained on two sets of variables could be reduced to the case of Simple Linear Regression (SLR). By this, it will be possible to determine the exact correlation co-efficient between one set of variables and another set, both of which constitute a sing multivariate multiple data. This expected result would be a further summary of results of Canonical Correlation Analysis (CCA).

Scope and Delimitation of the Study

The study is centered on reducing the dimensions of multivariate data in a manner that there will be no loss of information. Applications to this novel component extraction will be made to RRR. Although there are many dimensionality reduction techniques, the study will focus on PCA.

Limitation

As an application to RRR, coefficient of determination (R^2) in the original data may not be the same as that of the transformed data in the third chapter. This may influence the interpretation of the findings from our research.

Digitized by Sam Jonah Library

Organization of the Study

The study is divided into five chapters: Chapter one contains the introductory part of the study, which highlights the background to the study, statement of the problem, objectives, significance, limitations and delimitations.
The Chapter two entails literature review of other related literature, which show methods adopted by previous researchers.
The third chapter discusses the methodology. The fourth chapter also shows the analysis and discussion. The fifth and final chapter covers the summary, conclusion and recommendations, with suggestions for further research, references and appendices.

Chapter Summary

In this chapter, the problem of the classical principal component has been introduced and the motivation for the study has been outlined. The pertinent problem identified is that in the classical PCA, the amount of information loss after truncation, no matter the size, can have an important implication depending on the sensitivity of the area to which the technique is applied. It makes clear the intention of the study to derive a variant of reduced set of PCs that retains all information in a given dataset. Of significance, the chapter outlines how the approach will be applied to Reduced Rank Regression modelling, and how it will serve as a useful alternative to best subset regression.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter shows review of works done by previous researchers. The review focuses on the Classical PCA, Weighted PCA and Principal Component Regression (PCR). It is expected that review brings to light the importance of the various PC techniques and applications considered in this chapter. In the last section of this chapter, a brief overview of the illustrative datasets that will aid in the implementation of results is presented.

Principal Component Analysis

As presented in the first chapter of this work, Principal Component Analysis (PCA) creates new variables from linear combinations of the original ones such that there is no correlation among the new variables. The total number of potential new variables is the same as the total number of initial variables. The advantages PCA brings has seen many researchers using it in their works.

Usman et al. (2012) investigated the rate of crime in Sokoto State, Nigeria, using PCA. They analyzed how many principal components should be kept out of seven variables on crime that were received from the Criminal Investigation Department of the Sokoto State Police Headquarters. The original variables used for the study were Murder, Assault, Robbery, Theft, Store breaking, Grievous Harm and Wounding (GHW), and False. According to the findings of the statistical study, three components accounted for up to 89.40% of the overall variability of the data set. The first PC described crime as those done against people and those done against properties. Assault and GHW were discovered to be the most frequent and serious crimes done against people in Sokoto State, whereas store breaking was discovered to be the most serious crime performed

Digitized by Sam Jonah Library

against property. Subsequently, the second PC classified the crime into two categories with respect to the rate of occurrence. Assault, GHW and Store breaking were found to occur frequently, while Theft, Murder and Robbery were observed to occur less often.

Aboagye and Mensah (2016) analysed students academic performances in Mathematics and Statistics courses in the Department of Mathematics and Statistics in the University of Cape Coast. Results for level 300 students in the department for the 2013/2014 academic year were obtained. Ten courses made up this list, six of which were in statistics and four of which were in mathematics. Each of the ten disciplines was used as a variable to be analyzed together with a number of observations, which were the students' grades in the various courses. The researcher revealed his motivation that PCA was employed in his study because the principal components it produced might be utilized as indicators of how well the pupils were performing. Three key elements were chosen from the analysis to serve as guidelines or indices for grouping students' performance. To categorize pupils' overall performance as good, average, below average, or excellent, the first principal component was used. Again, based on the results of the study, the second PC was used to group students according to their semester performances. It was discovered that the third principal component could be used to categorize student performance based on subject. By this, it would be easy to determine where a student is mathematically inclined or statistically inclined. Therefore, by using the three retained PCs, observations revealed that the majority of pupils performed consistently but only averagely in both courses and both semesters. Few demonstrated particular prowess in either of the two subjects. This study supported the idea that only a small number of students could major in just one subject in the department and do so well.

Digitized by Sam Jonah Library

Weighted Principal Component Analysis

This section concentrates on works done by researchers on Weighted Principal Component Analysis.

Delchambre (2014) presented a direct PCA based on diagonalizing the weighted variance–covariance matrix through the power iteration and Rayleigh quotient iteration spectral decomposition approaches. By reviewing the conventional PCA , the paper considered a situation where we have n variables (n_{var}), from a data matrix $\mathbf{X}$, each having n observations n_{obs}, from which we want to retrieve n PCs (n_{comp}). It reviewed the often used matrices along with their corresponding dimensions: $\mathbf{W}$, the weight of each variable within each observation, $\mathbf{P}$ the orthogonal matrix of principal components, ($n_{\text{var}} \times n_{\text{var}}$), $\mathbf{P}_{\mathbf{i}}^{\text{col}}$ being the ith principal component, C the principal coefficient matrix ($n_{\text{var}} \times n_{\text{var}}$), and σ^2, the symmetric matrix of variance-covariance ($n_{\text{var}} \times n_{\text{var}}$) associated with $\mathbf{X}$. The author added that regarding the classical PCA, the aim will be to obtain as many principal components as the number of variables; that is, $n_{\text{comp}} = n_{\text{var}}$, then PCA will aim at finding a decomposition

$$\mathbf{X} = \mathbf{PC} \tag{2.1}$$

such that

$$\mathbf{D} = \mathbf{P}'\sigma^{\mathbf{2}}\mathbf{P} = \mathbf{P}'\mathbf{X}\mathbf{X}' \tag{2.2}$$

is diagonal and for which

$$\mathbf{D}_{\mathbf{ii}} \geq \mathbf{D}_{\mathbf{jj}}; \qquad \text{for all } \mathbf{i} < \mathbf{j} \tag{2.3}$$

Inferring from Equation (2.2), it was noted that $\mathbf{P}$ will be orthogonal. It recorded that $\mathbf{P}$ can be seen as a change of basis which will aid in maximizing the variance within $\mathbf{D}$ and thus minimizing the off-diagonal elements that corresponds

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

to the covariance.

It was revealed that unlike the classical PCA, the goal WPCA will be to explain the whole data set variance according to a given number of principal components.

In the work of Da Costa et al. (2009) on a weighted principal component analysis and its application to gene expression data, they preceded with some new developments in PCA. In the second part, they introduced a new method that can be used for variable selection. Their concentration was on situations where the values of the various variables did not all have the same importance and where the data could have been infected with noise and contained outliers. It reviewed that in the usual PCA, the ratios of the original variables' coefficients in linear combinations are built from the eigenvectors of the covariance matrix or the Pearson's correlation matrix. However, the Pearson's correlation coefficient is very sensitive to the presence of outliers and noise. In order to cope with outliers and noise, the original data used for the study was ranked, and a correlation was obtained in a Spearman ranked form as

Digitized by Sam Jonah Library

It was noted that the distance between two ranks in Spearman's coefficient given by

$$D_i^2 = (r_i - q_i)^2 \tag{2.4}$$

does not consider the rank importance. An alternative distance given as

$$\begin{aligned} WD_i^2 =& (r_i - q_i)^2 \Big((n - r_i + 1) + (n - q_i + 1) \Big) \\ =& D_i^2 (2n + 2 - r_i - q_i) \end{aligned} \tag{2.5}$$

rank correlation coefficient was proposed. In Equation (2.5), the D_i^2 represents the distance between r_i and q_i while the second term of the product is a linear weighting function which describes the importance of both r_i and q_i. The weighted rank correlation was obtained as

$$R_w = 1 - \frac{6 \sum_{i=1}^{n} (r_i - q_i)^2 (2n + 2 - r_i - q_i)}{n^4 + n^3 - n^2 - n} \tag{2.6}$$

Nevertheless, it was reviewed that this correlation cannot be used when there are tied values. Subsequently, in order to apply higher weights to the higher absolute expression values inside each variable, a new weighted rank correlation coefficient given as

$$M + N \sum_{i=1}^{n} (r_i - q_i)^2 (2n + 2 - r_i - q_i)^2 \tag{2.7}$$

was obtained. In Equation (2.7) the constants M and N take values between -1 and $+1$. The PCs obtained from Equation (2.7) were referred as weighted PCs. In the second part of their work, they proposed a new PCA-based algorithm used to iteratively select the most important genes in a micro-array data set. They showed that the algorithm produced better results when their WPCA was used instead of the usual PCA.

Digitized by Sam Jonah Library

Hong et al. (2018) revealed that modern data are increasingly both high-dimensional and heteroscedastic. In his paper on ”Optimally weighted PCA for high-dimensional heteroscedastic data”, he considered the challenge of estimating underlying principal components from high-dimensional data with noise that is heteroscedastic across samples. He recounts that such heteroscedasticity naturally arises, for example, when combining data from diverse sources or sensors. By that, an appropriate approach to help account for this heteroscedasticity is to give noisier blocks of samples less weight in PCA by using the leading eigenvectors of a weighted sample covariance matrix. The difficulty of choosing weights to best recover the underlying components was considered. It was revealed that generally, one cannot know these optimal weights since they depend on the underlying components that are to be estimated. However, it was shown that under some natural statistical assumptions, the optimal weights converge to a simple function of the signal and noise variances for high-dimensional data. Surprisingly, the optimal weights were not the inverse noise variance weights commonly used in practice. The theoretical results were demonstrated through numerical simulations and comparisons with existing weighting schemes.

In the article published by Niu and Qui (2010), The multi-feature fusion weighted principal component analysis (WPCA) and upgraded support vector algorithms (SVMs) were used to create a novel face emotion detection technique. They employed WPCA with multi-features to extract the facial expression feature and the SVMs to classify human facial expression. The weights were determined using the distribution of action units in the different facial areas. They provided empirical proof that, the proposed approach using the WPCA had a higher recognition rate for all the basic expressions using the classical PCA.

Digitized by Sam Jonah Library

Principal Component Regression

Principal Components Regression (PCR) is one of the frequently used statistical techniques. (Artigue & Smith, 2019). A PC transformation of the original independent variables is used to form a set of eigenvectors which are perpendicular to one another, and the variation in the original data is represented by their corresponding eigenvalues. PCR aims at reducing a large number of independent variables in a regression model down to a small number of principal components. The PCs chosen for the multiple regression model are then based on the amount of variation explained by each PC. Most often than not, these PCs are truncated for the purpose of the regression.

Illustrative Dataset

As used in Chapter Four, the illustrative dataset on the performance of sales personnel of a company (labelled as Dataset I in the study) is contained in several texts (Johnson & Wichern, 2014; Anderson, 2003; Mardia, Kent & Bibby, 1979). In all of these the data is highlighted to be a typical data for illustrating the technique of factor analysis. The data is also one of several datasets that have been used in studies of problems associated with factor analysis by Benyi (2018) and on the Kaiser-Meier-Olkin's measure of sampling adequacy (or simply KMO) (Nkansah, 2018). These studies reveal interesting features of the data that show that although it is suitable for factor analysis, no reasonable factor solution is found for it. Using a procedure for dimensionality detection, it is found, however, that only one dimension adequately underlies the data. The findings show that although the data is theoretically suitable for factor extraction, it practically has only one-factor solution, which is also not suitable. Additionally, a factor solution beyond one shows problems of the factor solutions, which cover contrasting factors and one-indicator factor solutions, Adu (2022) gives a further and more advanced explanation to this data problem explored

Digitized by Sam Jonah Library

by Benyi (2018) by using an automated threshold approach in the determination of dimensionality of multivariate data. This approach confirms the factor-suitability of the data and reveals that the true measure of factor-suitability of the data is even slightly higher than what is given by the traditional KMO measure. The lack of practical factor-suitability is however found to be attributable to the difficulty to generate a unique homogeneous set with the highest KMO and the highest number of indicators.

Dataset 2 is on student performance on nine subjects has also been studied by Benyi (2018) to identify dimensions in the data. Two homogeneous sets are identified in this data indicating a dimensionality of two. A cut-off value as low as 0.2 is used by subjective choice to support identification of homogeneous groupings in the data. The cut-off value chosen for that study is used to buttress the point that the choice is dependent on the data structure and that a good choice is required to identify appropriate dimensionality of the data.

The data labelled as Dataset 2 in the study is an important data for illustrating principal components analysis and is usually referred to as the USAFood data. It covers prices on five food items from twenty three (23) cities of the USA (Sharma, 1996).

Digitized by Sam Jonah Library

Chapter Summary

The chapter looked at several works worked on by other researchers using the classical PCA, WPCA, and PCR. It is presented that the advantages PCA brings has seen many researchers employ in their works. The chapter has reviewed that the literature on WPCA aim at describing the variation in a dataset based on a given number of PCs. It has also reviewed that the method also permits one to obtain a given number of PCs which are perpendicular, among the most important ones for the case of challenges with weighted or missing data. Comparatively, the literature shows that WPCA produces better results than the classical PCA. The chapter also highlighted the illustrative datasets that will aid in the analyses.

Digitized by Sam Jonah Library

CHAPTER THREE

METHODOLOGY

Introduction

This chapter focuses on the various methodologies used in this study. In the early sections, idea and methods behind PCA, the eigenstructure of a covariance matrix, and the singular value decomposition will be reviewed. The final sections will look at the V-C matrix for PC extraction, and linear transformation of random vectors which forms the bases for our new transformation.

Principal Component Analysis

The Principal Component Analysis (PCA) technique is used to create new variables that are linear combinations of the original variables. The amount of original variables plus as many new variables as possible can be formed, and the new variables are uncorrelated with one another.
PCA objectives can be broadly categorized as geometric or analytic.
PCA's geometric goal is to identify a new set of orthogonal axes such that:

1. The coordinates of the observations with respect to each of the axes give values for the new variables. The values of the new variable are known as principle component scores, and the new variables are known as principal components.

2. The PCs are linear combinations of the original variables.

3. The initial PC explains the most variation in the data.

4. The greatest variation that the first PC did not take into account is taken into account by the second new variable.

5. The third new variable accounts for the subsequent maximum variance that the first two have not accounted for.

Digitized by Sam Jonah Library

6. The variation not taken into account by the first $p-1$ variables is subsequently explained by the pth new variable.

7. There is no correlation among the p new variables.

Now, if a significant portion of the overall variation in the data can be explained by a small number of principal components or new variables—preferably considerably fewer—then the researcher can employ these few principal components.

In the analytical sense, supposing that there are p variables, we are interested in forming the following p linear combinations:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_r \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} & \cdots & \lambda_{1p} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} & \cdots & \lambda_{2p} \\ \lambda_{31} & \lambda_{32} & \lambda_{23} & \cdots & \lambda_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{r1} & \lambda_{r2} & \lambda_{r3} & \cdots & \lambda_{rp} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \lambda_{p3} & \cdots & \lambda_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_r \\ \vdots \\ x_p \end{bmatrix} \tag{3.1}$$

where the PCs are $Y1, Y2, Y3, \cdots, Y_r, Yp$, and λ_{ij} is the weight of the jth variable for the ith principal component. The weights are estimated using λ_{ij} so that:

1. The first principal component, Y_1 explains the most variation in the data. The second principal component Y_2, explains the second-most variation that has not been explained by the first principal component, and so on.

2. $\lambda_{i1}^2 + \lambda_{i2}^2 + \lambda_{i3}^2 + \cdots + \lambda_{ir}^2 + \cdots + \lambda_{ip}^2 = 1,\ i = 1, 2, \cdots, p$

3. $\lambda_{j1} + \lambda_{j2} + \lambda_{j3} + \cdots + \lambda_{jr} + \cdots + \lambda_{jp} = 0,$ for all $i \neq j$

The requirement in (2), which is somewhat arbitrary, is that the squares of the weights add up to one. Because it is possible to increase the variance of a

Digitized by Sam Jonah Library

linear combination by changing the scale of the weights, this condition is used to set the scale of the additional variables. The new axes are guaranteed to be orthogonal to one another by the condition provided by (3). How to get the weights of (2) so that the aforementioned requirements are met is a question in calculus which the next section will seek to address.

The properties specified in (2) and (3) above may be given alternatively by matrix-vector form as

$$\begin{aligned} \mathrm{D}(Y) =& \mathrm{D}(P'X) \\ =& P'\mathrm{D}(X)P \\ =& P'\Sigma_X P \\ =& P'P\Lambda P'P \\ =& I\Lambda I \\ =& \Lambda \end{aligned}$$

What this means is that:

1. All PCs together account for all variation in the original data.
2. The PCs are independent of each other.
3. Variations explained by PCs are in decreasing order.

Eigenstructrue of Covariance Matrix

Let $X = (X_1, X_2, \cdots, X_p)'$ with variance-covariance matrix given as Σ. Let $\Lambda' = \lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_r, \cdots, \lambda_p$ be needed in forming the original variables, and $\mathbf{Y} = \mathbf{\Lambda}'\mathbf{X}$ be the current variable, a combination of initial variables. The variation in the new variables is given by $\mathrm{E}(YY')$ which is the same as $\mathrm{E}(\lambda' XX' \lambda)$ or $\lambda'\Sigma\lambda$, the variation in the original variables. The problem is now reduced to determining the vector of weights λ' such that the variation in the

Digitized by Sam Jonah Library

new variable, $\lambda\Sigma\lambda$, is optimum over the class of linear combinations that can be constructed subject to the constraint $\lambda'\lambda = 1$. The maximization problem can be solved as follows:

$$\vartheta = \lambda'\Sigma\Lambda - \chi(\lambda'\lambda - 1) \quad (3.2)$$

where χ is the Lagrange multiplier. The $p-$ component vector of the partial derivative is given by

$$\frac{\partial\vartheta}{\partial\lambda} = 2\Sigma\nu - 2\chi\nu \quad (3.3)$$

By setting the above vector of partial derivatives to zero, we obtain

$$(\Sigma - \chi\mathbf{I})\lambda = \mathbf{0} \quad (3.4)$$

For the above system of homogeneous equations to have a non-trivial solution, the determinant of $\Sigma - \chi\mathbf{I}$ should be zero. That is

$$|\Sigma - \chi\mathbf{I}| = \mathbf{0} \quad (3.5)$$

Equation (3.5) is a function in χ of degree p, and thus, has p roots. Let $\chi_1 \geq \chi_2 \geq \cdots \geq \chi_p$ be the p roots. That is, Equation (3.5) gives p solutions for χ and each of the solutions is known as the eigenvalue. By solving Equations (3.6) and (3.7) simultaneously, each solution of χ results in a set of weights given by the $p-$ component vector λ.

$$(\Sigma - \chi\mathbf{I})\lambda = \mathbf{0} \quad (3.6)$$

$$\lambda'\lambda = 1 \quad (3.7)$$

Digitized by Sam Jonah Library

Therefore, the principal eigenvector, λ_1 which corresponds to the principal eigenvalue, χ_1, is found by solving Equations (3.8) and (3.9).

$$(\Sigma - \chi_1 \mathbf{I})\boldsymbol{\lambda_1} = \mathbf{0} \tag{3.8}$$

$$\lambda_1' \lambda_1 = 1 \tag{3.9}$$

Premultiplying Equation (3.8) by λ_1' gives

$$\lambda_1'(\Sigma - \chi_1 \mathbf{I})\boldsymbol{\lambda_1} = 0$$

$$\lambda' \Sigma \lambda_1 = \chi_1 \lambda_1' \lambda_1'$$

$$\lambda' \Sigma \lambda_1 = \chi_1$$

The left hand side is the variation of the first PC, Y_1, and is the same as the eigenvalue, χ_1. The principal PC, therefore, is given by the eigenvector, λ_1 which corresponds to the largest eigenvalue, χ_1.

Let λ_2 be the second $p-$ component vector of weights to form another linear combination. The following linear combination can be found such that the variance of $\lambda_2' X$ is the greatest subject to the constraints $\lambda_1' \lambda_2 = 0$ and $\lambda_2' \lambda_2 = 1$. It can be shown that λ_2' is the eigenvector of χ_2, the second largest eigenvalue of Σ. Similarly, it can be proven that the remaining PCs, $\lambda_3', \lambda_4', \cdots, \lambda_p$ are the eigenvectors corresponding to the eigenvalues, $\chi_3, \chi_4, \cdots, \chi_p$ of the covariance matrix, Σ respectively. The eigenvalues reflect the variation of the new variables, while the eigenvectors provide the vectors of weights.

Singular Value Decomposition

Singular value decomposition (SVD) is that which denotes any $n \times p$ matrix (for cases where $n \geq p$) as a multiplication of three matrices, $A, D,$ and B

Digitized by Sam Jonah Library

such that

$$\Sigma = ADB' \tag{3.10}$$

where X is an $n \times p$ matrix of column rank r, A is an $n \times r$ matrix, D is an $r \times r$ diagonal matrix, and B' is an $r \times p$ matrix. Matrix A is orthonormal to matrix B. Thus,

$$A'A = I$$

and

$$B'B = I$$

Singular Value Decomposition of the Data Matrix

If X can be represented as a $n \times p$ data matrix, then it has the assumption that its rank is p. B is a square symmetric matrix its columns give the eigenvectors D is a diagonal matrix such that the diagonal values gives the square root of the eigenvalues corresponding to the original data matrix.

Let the matrix of principal components scores be denoted by η with dimension $n \times p$. Then

$$\begin{aligned}\eta &= XQ \\ &= (PDQ')Q \\ &= PDQ'Q \\ &= PD\end{aligned}$$

Digitized by Sam Jonah Library

The variance-covariance matrix, Σ_Y of the PCs is given by

$$\begin{aligned}\Sigma_Y = \mathrm{E}(\eta'\eta) =& \mathrm{E}\Big[(PD)'(PD)\Big] \\ =& \mathrm{E}(D'P'PD) \\ =& \mathrm{E}(D^2) \\ =& \frac{1}{n-1}D^2\end{aligned}$$

Since D is a diagonal matrix, there is no correlation among the new variables. The SVD of X also gives the principal component analysis solution. The major component scores are provided by PD, while the matrix Q provides the weights for generating the new variables. The variations in the PCs are given by $\frac{D^2}{n-1}$

Spectral Decomposition of a Matrix

The spectral decomposition of a matrix is another name for the singular value decomposition of a square matrix that is symmetric. This decomposition assumes that the covariance of any symmetric matrix X can be written as a multiplication of two matrices, P and Λ, such that

$$\Sigma = P\Lambda P' \tag{3.11}$$

In Equation (3.11), P is a symmetric orthogonal matrix with dimension $p \times p$ which contains the eigenvectors of the data matrix, and Λ, contains the eigenvalues of the X matrix. Again,

$$P'P = PP' = I$$

Digitized by Sam Jonah Library

Spectral Decomposition of the Covariance Matrix

As aforementioned, Σ is a square symmetric matrix and can be written in a spectral decomposition form as

$$\Sigma = P\Lambda P' \tag{3.12}$$

where Λ is a diagonal matrix whose elements ae the eigenvalues $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_p$ of the symmetric matrix, Σ. P is a $p \times p$ orthogonal matrix whose jth column is the eigenvector corresponding to the jth eigenvalue.

The principal component scores are denoted by the matrix $\eta = XP$ and the variance-covariance matrix of the principal components scores is given by

$$\begin{aligned}\Sigma_Y = \text{E}(\eta'\eta) =& \text{E}\Big[(XP)'(XP)\Big]\\ =& \text{E}(P'X'XP)\\ =& P'\Sigma P\end{aligned}$$

Making substitution for the covariance matrix Σ, we get

$$\begin{aligned}\Sigma_Y =& P'P\Lambda P'P\\ =& \Lambda\end{aligned}$$

as $P'P = \mathbf{I}$. Thus, the PCs, $Y_1, Y_2, \cdots, Y_p$ are uncorrelated with variances equal to $\lambda_1, \lambda_2, \cdots, \lambda_p$ respectively. In addition, the trace of Σ is given by

$$\text{tr}(\Sigma) = \sum_{j-1}^{p} \sigma_{jj}^2$$

Digitized by Sam Jonah Library

where σ_{jj}^2 is the variation of the jth variable. The trace of Σ can also by represented as

$$\begin{aligned}\text{tr}(\Sigma) =& \text{tr}(P\Lambda P')\\ =& \text{tr}(P'P\Lambda)\\ =& \text{tr}(\Lambda)\\ =& \text{tr}(\Sigma_Y)\end{aligned}$$

which is the same as the sum of the eigenvalues of the covariance matrix, Σ of the original data. The findings from the previous calculations demonstrate that the total variation in the original variables and that in the new variable are equal.

In conclusion, PCA reduces to finding the eigenvalues and eigenvectors of the covariance matrix, or finding the SVD of the original data matrix, X, or obtaining the spectral decomposition of the covariance matrix.

Variance-Covariance Matrix for Principal Component Extraction

PCA is one of the methods that are usually used for outlier detection. However, it is not specifically designed for outlier detection as it focuses on maximal dispersion (Jollife, 2002). The PCA is carried out on the usual variance-covariance matrix of the random vector $X = (X_1, X_2, \cdots, X_p)$ given

$$\begin{aligned}\text{D}(X) =& \frac{1}{n-1}(X-\bar{X})(X-\bar{X})'\\ \text{or}&\\ \text{D}(X) =& \frac{1}{n-1}\sum_{j=1}^{n}(X_j-\bar{X})(X_j-\bar{X})'\end{aligned} \tag{3.13}$$

A variant of the PCA is the Generalized PCA (GPCA) which is actually designed to reveal dimensions that display outliers. Unlike the PCA, GPCA searches for the eigenvectors of the matrix product $\mathbf{SS^{*-1}}$ that are associated with the q

Digitized by Sam Jonah Library

largest eigenvalues, where S is the usual variance-covariance matrix, that is, $S = \mathrm{D}(X)$, and S^* is defined as

$$S^* = \frac{\sum_{j=1}^{n} K\Big(||\mathbf{x}_j - \mathbf{x}^*||^2_{s^{-1}}\Big)(\mathbf{x_j} - \mathbf{x}^*)(\mathbf{x_j} - \mathbf{x}^*)'}{\sum_{j=1}^{n} K\Big(||\mathbf{x}_j - \mathbf{x}^*||^2_{s^{-1}}\Big)} \tag{3.14}$$

where x^* is a vector of means. The measure $||X||^2_m$ is defined as

$$||X||^2_m = X'MX$$

and K is a decreasing function given as

$$K(u) = e^{-hu}$$

and $h = 0.1$ as recommended by Caussinus and Ruiz (1990). The expression in Equation(3.14) suggest that observations that are distant from the center of the data are given less weight than those that are close to the center. The method of GPCA is thus based on the spectral decomposition of a scatter estimator relative to another scatter estimator.

In this study, it will be important to compare the performance of our PCA extraction based on components exracted from S and the one extracted from SS^{*-1}, an example of a weighted variance-covariance matrix.

Linear Transformation of Random Vectors

Let $Y = (Y_1, Y_2, \cdots, Y_m)'$ and $X = (X_1, X_2, \cdots, X_p)'$ be a random vector with mean vector μ and variance-covariance Σ. If $Y = AX$ where $A = (a_{ij})$ is an $m \times p$ matrix of constants. Then,

1. $\mathrm{E}(Y) = A\mu$
2. $\mathrm{D}(Y) = A\Sigma A'$

Digitized by Sam Jonah Library

Chapter Summary

This chapter was organized as follows: Principal Component Analysis, Eigenstructure of covariance matrix, Singular Value Decomposition of Data matrix, Spectral Decomposition of a Matrix, Spectral Decomposition of a matrix, and the Variance-Covariance matrix for Principal Component Extraction. The methodology has presented extensive review of the procedure for conducting principal component analysis. All the new transformed PCs were extracted from the Variance-Covariance matrix and the correlation matrix of the original datasets used in this work. All the datasets were multivariate.

Digitized by Sam Jonah Library

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

This chapter looks at the derivation of a novel principal component with no loss of information from a multivariate dataset. The chapter follows up with an application of this novel component extraction to three distinct multivariate datasets. The intention is to compare the total variation explained by the reduced set of PCs and that in the corresponding set of the original PC. In the final section of this chapter, Reduced Rank Regression (RRR) models will be built for reduced PCs extracted from an ordinary variance-covariance matrix and weighted variance-covariance matrix. These models will then be compared with the full model based on the original variables (MLR). Discussion of results will be presented in the end.

Proposition for Extracting New Principal Components

Instead of truncating to retain the desired PCs with attendant loss of information, we can force all the other PCs into the first few PCs, such that there will be no loss of information. This calls for further transformation. In the illustrative data, we may retain only r PCs out of the original p variables.
We obtain a transformation of X in the form

$$Y = P'X \tag{4.1}$$

where the matrix P whose columns are the eigenvectors of the variance-covariance matrix of X, Σ is obtained by the spectral decomposition in Equation (3.10) given by

$$\Sigma = P\Lambda P' \tag{4.2}$$

Digitized by Sam Jonah Library

From Equation (4.1), it can be inferred that if

$$\mathrm{f}(\Sigma) = P\mathrm{f}(\Lambda)P'$$

Then

$$\Sigma^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P'$$

Now we know that

$$\begin{aligned}\Sigma =& \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} \\ \Rightarrow \Sigma =& P\Lambda^{\frac{1}{2}}P'P\Lambda^{\frac{1}{2}}P' \\ =& P\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}P'\end{aligned}$$

Now let $S = P\Lambda^{\frac{1}{2}}$

$$\Rightarrow \Sigma = SS' \tag{4.3}$$

where

$$S = P\Lambda^{\frac{1}{2}} \tag{4.4}$$

Noting clearly that $S \neq \Sigma^{\frac{1}{2}}$

Since $\Sigma^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P'$,

$$\Sigma^{\frac{1}{2}} = SP' \tag{4.5}$$

Suppose the data is transformed by

$$Y = S^{-1}X \tag{4.6}$$

Digitized by Sam Jonah Library

We examine the properties of Y by finding Σ_Y, the variance-covariance matrix of Y. That is,

$$\begin{aligned} \text{D}(Y) =& \text{D}(S^{-1}X) \\ =& S^{-1}D(X)S^{-1'} \\ =& S^{-1}\Sigma_X S^{-1'} \end{aligned}$$

Substituting for Σ_X from Equation (4.3),

$$\begin{aligned} \text{D}(Y) =& S^{-1}(SS')S^{-1'} \\ =& S^{-1}SS'S^{-1'} \\ =& I \end{aligned}$$

This means that the transformation gives components of Y that are independent with unit variances.

The transformation given by Equation (4.6) is in terms of the original data X. Suppose now that a similar transformation is carried out on the centered data $(X - \mu)$. Then,

$$Y = S^{-1}(X - \mu)$$

Examining the expectation $E(Y)$ and the variance-covariance matrix $\text{D}(Y)$ gives

$$\begin{aligned} \text{E}(Y) =& S^{-1}[\text{E}(X) - \mu] = 0 \qquad \text{and} \\ \text{D}(Y) =& S^{-1}\text{D}(X - \mu)S^{-1} \\ =& S'\text{D}(X)S^{-1} \\ =& I \end{aligned}$$

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

The transformation given by S^{-1} thus gives the same result for both the original and centered data.

Further transformations based on the PCs

In this section, a few transformations are first explored to highlight the motivation for the intended extraction method proposed in the study. The overriding principle is to obtain new component that accounts for all information in the data. In the process, it is also important to examine how each of the new components contributes to the overall information.

We suppose in the meantime that the transformed components explain equal (but not necessarily unity) variances. That is,

$$\text{var}(Y_i) = \frac{1}{p}Tr(\Lambda), \qquad i = 1, 2, \cdots, p$$

This means that the corresponding transformation may be given by

$$Y = \text{D}_Y S^{-1} X \tag{4.7}$$

where

$$\text{D}_Y = \text{diag}\left(\frac{1}{p}Tr(\Lambda), \cdots, \frac{1}{p}Tr(\Lambda)\right)$$

This means that the V-C matrix of Y is given as

$$\begin{aligned}
\text{D}(Y) &= \text{D}(\text{D}_Y S^{-1} X) \\
&= \text{D}_Y S^{-1} \Sigma_X S^{-1\prime} \text{D}_Y' \\
&= \text{D}_Y S^{-1} S S' S^{-1\prime} \text{D}_Y' \\
&= \text{D}_Y I \text{D}_Y' \\
&= \text{D}_Y^2 \\
&= \text{diag}\left[\left(\frac{1}{p}Tr(\Lambda)\right)^2, \cdots, \left(\frac{1}{p}Tr(\Lambda)\right)^2\right]
\end{aligned}$$

Digitized by Sam Jonah Library

Thus, Equation (4.7) gives a transformation that yields components with equal amount of variance explained.

In order to obtain the exact variation explained by each component, the transformation above should rather be specified as

$$Y = \left(D_Y^{\frac{1}{2}} S^{-1}\right) X \tag{4.8}$$

The V-C matrix of the revised Y now becomes

$$\begin{aligned}
\text{D}(Y) =& D_Y^{\frac{1}{2}} S^{-1} \text{D}(X) S^{-1} D_Y^{\frac{1}{2}} \\
\text{D}(Y) =& D_Y^{\frac{1}{2}} S^{-1} \Sigma_X S^{-1} D_Y^{\frac{1}{2}} \\
\text{D}(Y) =& D_Y^{\frac{1}{2}} S^{-1} (SS') S^{-1} D_Y^{\frac{1}{2}} \\
=& D_Y^{\frac{1}{2}} D_Y^{\frac{1}{2}} \\
=& D_Y
\end{aligned}$$

Equation (4.7) gives another transformation that yields independent components with equal(one each) variance explained.

It can be noted that

1. Components are independent
2. Components may be expressed to have zero means
3. Variance explained are the same (and equal to 1 if extraction is based on the correlation matrix rather than the V-C matrix) for all components

The transformation demonstrated above shows that it is possible to learn a particular transformation that would yield some desired characteristics. In the next section, we proceed with this strategy to derive the proposed methodology for extracting new PCs.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Proposed Methodology

Case I

Suppose without laws of generality that we want only one component to contain all the information. Then define the matrix of transformation

$$
\begin{aligned}
Q_1 =& P\text{diag}(\lambda_1 + \lambda_2 + \cdots + \lambda_p, 0, 0, \cdots, 0)^{\frac{1}{2}} \\
=& P\text{diag}\Big(\text{tr}(\Lambda), 0, 0, \cdots, 0\Big)^{\frac{1}{2}} \\
=& P\Lambda_1^{\frac{1}{2}}
\end{aligned}
$$

where

$$\Lambda_1^{\frac{1}{2}} = \text{diag}\Big(tr(\Lambda^{\frac{1}{2}}, 0, ..., 0)\Big) \tag{4.9}$$

Suppose all the original variables are transformed onto only one PC, then this PC is given by

$$Y^{(1)} = \Lambda_1^{\frac{1}{2}} S$$

where $S = P\Lambda^{\frac{1}{2}}$ as in Equation (4.4).

From Equation (4.9), it can be observed that all the information will be explained by only the first reduced PC.

Case II

Suppose all the original variables are transformed onto only two PCs, so that all the information in the data is explained by the two reduced PCs. Then the transformation that yields these reduced PCs is given by

$$
\begin{aligned}
Q_2 =& P\text{diag}\left[\lambda_1\Big(1 + \frac{\sum_{j=3}^{p}\lambda_j}{\sum_{j=1}^{2}\lambda_j}\Big), \lambda_2\Big(1 + \frac{\sum_{j=3}^{p}\lambda_j}{\sum_{j=1}^{2}\lambda_j}\Big), 0, \cdots, 0\right]^{\frac{1}{2}} \\
=& P\Lambda_2^{\frac{1}{2}}
\end{aligned}
$$

Digitized by Sam Jonah Library

where

$$\Lambda_2^{\frac{1}{2}} = \text{diag}\left[\lambda_1\Big(1 + \frac{\sum_{j=3}^{p}\lambda_j}{\sum_{j=1}^{2}\lambda_j}\Big), \lambda_2\Big(1 + \frac{\sum_{j=3}^{p}\lambda_j}{\sum_{j=1}^{2}\lambda_j}\Big), 0, \cdots, 0\right] \tag{4.10}$$

Suppose all the original variables are transformed onto only the two PCs, then the two new PCs are given by

$$Y^{(2)} = \Lambda_2^{\frac{1}{2}} S$$

The remaining information in the other original PCs will be forced onto the two retained PCs in accordance with the proportion of variance explained. This means that the total variation accounted for by each of the two reduced PCs will be given as follows:

$$\begin{aligned}
\text{Var}(Y_1^{(2)}) =& \lambda_1 + \frac{\lambda_1}{\lambda_1 + \lambda_2}\sum_{j=3}^{p}\lambda_j \\
=& \lambda_1 + \frac{\lambda_1}{\lambda_1 + \lambda_2}\text{SS2} \\
=& \lambda_1\Big(1 + \frac{SS2}{SS1}\Big)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\text{Var}(Y_2^{(2)}) =& \lambda_2 + \frac{\lambda_2}{\lambda_1 + \lambda_2}\sum_{j=3}^{p}\lambda_j \\
=& \lambda_2\Big(1 + \frac{SS2}{SS1}\Big)
\end{aligned}$$

Case III

Suppose all the original variables are transformed onto only r reduced PCs ($r \leq p$), so that all the information in the data is explained by the r reduced PCs.

Digitized by Sam Jonah Library

Then the transformation that yields these reduced PCs is given by

$$\begin{aligned} Q_r =& P\text{diag}\Bigg[\lambda_1\Big(1+\frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\Big), \lambda_2\Big(1+\frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\Big), \cdots, \\ & \lambda_r\Big(1+\frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\Big), 0, \cdots, 0\Bigg]^{\frac{1}{2}} \\ =& P\Lambda_r^{\frac{1}{2}} \end{aligned}$$

where

$$\begin{aligned} \Lambda_r =& \text{diag}\Bigg[\lambda_1\Big(1+\frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\Big), \lambda_2\Big(1+\frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\Big), \cdots, \\ & \lambda_r\Big(1+\frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\Big), 0, \cdots, 0\Bigg] \end{aligned} \tag{4.11}$$

Denote

$$\Psi_r = \lambda_r\left(1+\frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\right)$$

which is the measure of variation explained by $Y_r^{(r)}$, the last of the r reduced set of new PCs.

The generalization of the technique of extracting new components is that if $X = (X_1, X_2, \cdots, X_p)$ is transformed onto only the r PCs, then the r new PCs are given by

$$Y^{(r)} = \Lambda_r^{\frac{1}{2}} S^{-1} \tag{4.12}$$

Equation (4.12) is the transformation that gives the proposed extraction of reduced PCs with no loss of information. In the next section, we will examine the features of the new components. First, it will be verified that the variation accounted for by the set of r reduced components accounts for the entire information (variation) in the original data.

Digitized by Sam Jonah Library

Examination of Features of the Extracted New Components

Consider the projected data on the new component defined in Equation (4.12) given by

$$\mathbf{W}^{(r)} = \Lambda_r^{\frac{1}{2}} S^{-1} X' \tag{4.13}$$

or $\mathbf{W}^{(r)} = TX'$, where $T = \Lambda_r^{\frac{1}{2}} S^{-1}$.

The transformed data matrix $\mathbf{W}^{(\mathbf{r})}$ may thus be partitioned as

$$\mathbf{W}^{(r)} = \left[(TX')'_r \;\middle|\; \mathbf{0} \right] \tag{4.14}$$

In equation (4.14) $(TX')'$ is the reduced transformed data of dimension $(n \times r)$ and the remaining $(p - r)$ columns are set to zero.

Following this data structure, it can be shown therefore that the total variation in $\mathbf{W}^{(\mathbf{r})}$ is the same as that in X. That is,

$$\text{tr}\Big(\mathbf{W}^{(r)}\Big) = \text{tr}(\Sigma_X) \tag{4.15}$$

Proof of Equation (4.15)

From Equation (4.13),

$$\begin{aligned}
\mathbf{W}^{(r)} &= \Lambda_r^{\frac{1}{2}} S^{-1} X' \\
&= \Lambda_r^{\frac{1}{2}} (P\Lambda^{\frac{1}{2}})^{-1} X' \\
&= \Lambda_r^{\frac{1}{2}} \Lambda^{-\frac{1}{2}} P' X' \\
\text{D}\Big(\mathbf{W}^{(\mathbf{r})}\Big) &= \text{D}\Big(\Lambda_r^{\frac{1}{2}} \Lambda^{-\frac{1}{2}} P' X'\Big) \\
&= \Lambda_r^{\frac{1}{2}} \Lambda^{-\frac{1}{2}} P' \text{D}(X) P \Lambda^{-\frac{1}{2}} \Lambda_r^{\frac{1}{2}}
\end{aligned}$$

Digitized by Sam Jonah Library

Noting that $\mathrm{D}(X) = \Sigma_X = P\Lambda P'$, and $\Lambda_r^{\frac{1}{2}}$ is as defined in Equation (4.11). Making substitutions and simplifying gives

$$\begin{aligned}
\mathrm{D}\left(\mathbf{W}^{(r)}\right) &= \Lambda_r^{\frac{1}{2}}\Lambda^{-\frac{1}{2}}P'(P\Lambda P')P\Lambda^{-\frac{1}{2}}\Lambda_r^{\frac{1}{2}} \\
&= \Lambda_r^{\frac{1}{2}}\Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}}\Lambda_r^{\frac{1}{2}} \\
&= \Lambda_r
\end{aligned}$$

Now taking the trace of

$$\begin{aligned}
\mathrm{tr}(\Lambda_r) &= \sum_{k=1}^{p} \lambda_j \left(1 + \frac{\sum_{j=r+1}^{p} \lambda_j}{\sum_{j=1}^{r} \lambda_j}\right) \\
&= \sum_{k=1}^{r} \lambda_j \left(1 + \frac{\sum_{j=r+1}^{p} \lambda_j}{\sum_{j=1}^{r} \lambda_j}\right)
\end{aligned}$$

Since $\lambda_k = 0$, for $k > r$

$$\begin{aligned}
\mathrm{tr}(\Lambda_r) &= \left(1 + \frac{\sum_{j=r+1}^{p} \lambda_j}{\sum_{j=1}^{r} \lambda_j}\right) \sum_{j=1}^{r} \lambda_j \\
&= \sum_{j=1}^{r} \lambda_j + \sum_{j=r+1}^{p} \lambda_j \\
&= \sum_{j=1}^{p} \lambda_j \\
&= \mathrm{tr}(\Lambda)
\end{aligned}$$

It can be seen that

$$\begin{aligned}
\mathrm{tr}\left(\Sigma_X\right) &= \mathrm{tr}\left(P\Lambda P'\right) \\
&= P\ \mathrm{tr}(\Lambda)P' \\
&= PP'\mathrm{tr}(\Lambda) \\
&= \mathrm{tr}(\Lambda)
\end{aligned}$$

Digitized by Sam Jonah Library

Hence, Equation (4.15) is proofed.

This means that the amount of information explained by the reduced transformed r new variables $Y^{(r)}$ is the same as that in the data on the original set of variables $(\mathbf{X})$.

Relationship between the new components and the original PCs

From Equation (4.12)

$$Y^{(r)} = \Lambda_r^{\frac{1}{2}} S^{-1}$$

where $S = P\Lambda^{\frac{1}{2}}$ and $\Lambda = \text{diag}\Big(\lambda_1, \lambda_2, \cdots, \lambda_p\Big)$.

It is not too clear from this equation the explicit relationship between $Y^{(r)}$ and P, that is, the new components and the original components. In order to specify clearly the exact relationship between the two sets of components, we carry out systematic simplification of the matrix products involved. Rewriting Equation (4.12) gives

$$Y^{(r)} = \Lambda_r^{\frac{1}{2}} \Lambda^{-\frac{1}{2}} P'$$

From Equation (4.11)

$$\begin{aligned}\Lambda_r =& \text{diag}\left[\lambda_1\Big(1 + \frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\Big), \lambda_2\Big(1 + \frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\Big), \cdots,\right. \\ &\left.\lambda_r\Big(1 + \frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\Big), 0, \cdots, 0\right]\end{aligned}$$

It has been denoted that

$$\Psi_r = \lambda_r\left(1 + \frac{\sum_{j=r+1}^{p}\lambda_j}{\sum_{j=1}^{r}\lambda_j}\right)$$

Digitized by Sam Jonah Library

Now for any particular set of r reduced components, the expression

$$1 + \frac{\sum_{j=r+1}^{p} \lambda_j}{\sum_{j=1}^{r} \lambda_j} = c$$

which is a constant for all components in the set.

Thus,

$$\Lambda_r^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1 c} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2 c} & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{\lambda_r c} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}_{(p \times p)}$$

The transformation is then expanded as

$$Y^{(r)} = \begin{bmatrix} \sqrt{\lambda_1 c} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2 c} & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{\lambda_r c} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \times \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sqrt{\lambda_r}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots & \frac{1}{\sqrt{\lambda_p}} \end{bmatrix} \times$$

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1r} & \cdots & p_{1p} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2r} & \cdots & p_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & p_{r3} & \cdots & p_{rr} & \cdots & p_{rp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \\ p_{p1} & p_{p2} & p_{p3} & \cdots & p_{pr} & \cdots & p_{pp} \end{bmatrix}$$

Digitized by Sam Jonah Library

where $P_j = (p_{j1} \quad p_{j2} \quad \cdots \quad p_{jr} \quad \cdots \quad p_{jp})', \quad j = 1, \cdots, p$ is the jth original principal component. Further simplification gives

$$Y^{(r)} = \begin{bmatrix} \sqrt{c} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{c} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{c} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \times \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1r} & \cdots & p_{1p} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2r} & \cdots & p_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & p_{r3} & \cdots & p_{rr} & \cdots & p_{rp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \\ p_{p1} & p_{p2} & p_{p3} & \cdots & p_{pr} & \cdots & p_{pp} \end{bmatrix}$$

In partitioned matrices form, if $\mathbf{I}_{r\times r}$ is an identity matrix, the above product may be written as

$$Y^{(r)} = \left[\begin{array}{c|c} \sqrt{c}\mathbf{I}_{r\times r} & \mathbf{0}_{r\times(p-r)} \\ \hline \mathbf{0}_{(p-r)\times r} & \mathbf{0}_{(p-r)\times(p-r)} \end{array}\right] \begin{bmatrix} P_1' \\ P_2' \\ \vdots \\ P_r' \\ \hdashline P_{r+1}' \\ \vdots \\ P_p' \end{bmatrix}$$

$$= \left[\begin{array}{c:c} \sqrt{c}P_{r\times r}^{(r)'} & \mathbf{0}_{r\times(p-r)} \\ \hdashline \mathbf{0}_{(p-r)\times r} & \mathbf{0}_{(p-r)\times(p-r)} \end{array}\right]$$

Therefore, the relationship between the new components and the original is simply stated as

$$Y^{(r)} = \sqrt{c}P^{(r)} \tag{4.16}$$

Digitized by Sam Jonah Library

where $P_{(r)} = (P_1, P_2, \cdots, P_r)$ is the set of first r original components.
It can therefore be seen that the normalization of the new components gives the r original components.

Deductions from Results on New Components

It is observed that if $Y^{(r)}$ is the new $r-$dimensional new PC obtained from the original PCs $P^{(r)}$, then

$$Y^{(r)} = \sqrt{c}P^{(r)}$$

Let

$$\mathbf{W^{(r)}} = \Lambda_r^{\frac{1}{2}} S^{-1} X'$$

the projection of the data onto the components.
The V-C matrix of $\mathbf{W^{(r)}}$ is given as

$$\begin{aligned} \mathrm{D}\Big((W^{(r)}\Big) =& Y^{(r)'}\mathrm{D}(X)Y^{(r)} \\ =& (\sqrt{c})^2 P^{(r)'}\Sigma_X P^{(r)'} \\ =& c\Sigma_Z \end{aligned}$$

where $\Sigma_Z = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r)$ is the V-C matrix of the projected data onto the first $r-$original PC, $P^{(r)}$ and $Z = P^{(r)}X'$.

Reduced Rank Regression Modelling

In this section, we futher examine the features of the new reduced PCs (NRPCs) in Multivaritate Multiple Reduced Rank Regression (MMRRR) modelling The multiple coefficient of determination of regression of Y on several predictors
Let $Y^{(r)}$ be the regression model of the transformed data $W^{(r)}$ of the original

Digitized by Sam Jonah Library

data on $X = (X_1, X_2, \cdots, X_p)'$. Then

$$Y^{(r)} = t'W^{(r)},$$

$\mathbf{W^{(r)}} = \left(\mathbf{1} \,\middle|\, W^{(r)} \right)$ where $\mathbf{1} = \text{ones}(n, 1)$, a vector of n ones and $t = \left(t_0 \,\middle|\, t_{(r)} \right)'$ is the regression coefficients consisting of the intercepts t_0 and the set of r coefficients on the NRPCs.

Noting that the estimated model of a set of response variable $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_m)'$ is thus given as

$$Y = b'X, \qquad \mathbf{b} = (X'X)^{-1}(X'Y)$$

The corresponding Least Squares Error (LSE) model in terms of $\mathbf{W}^{(r)}$ is given by

$$\mathbf{t} = \left(\mathbf{W}^{(r)\prime}\mathbf{W}^{(r)}\right)^{-1}\left(\mathbf{W}^{(r)\prime}Y\right) \tag{4.17}$$

We seek to determine the exact analytical relationship between the components of $\mathbf{b}$ and $\mathbf{t}$. Expanding and simplifying Equation (4.17) gives

$$\begin{aligned}
t &= \left[\left(\mathbf{1} \,\middle|\, W^{(r)} \right)'\left(\mathbf{1} \,\middle|\, W^{(r)} \right)\right]^{-1}\left(\mathbf{1} \,\middle|\, W^{(r)} \right)'Y \\
&= \left(\begin{array}{c|c} n & \mathbf{1}'W^{(r)} \\ \hline W^{(r)\prime}\mathbf{1} & W^{(r)\prime}W^{(r)} \end{array}\right)^{-1}\left(\mathbf{1} \,\middle|\, W^{(r)} \right)'Y \\
&= \frac{F}{\text{Det}\left(W^{(r)\prime}W^{(r)}\right)}\left(\begin{array}{c|c} W^{(r)\prime}W^{(r)} & -\mathbf{1}'W^{(r)} \\ \hline -W^{(r)\prime}1 & n \end{array}\right)\left[\begin{array}{c} \mathbf{1}'Y \\ W^{(r)\prime}Y \end{array}\right]
\end{aligned}$$

Digitized by Sam Jonah Library

where

$$\begin{aligned}\text{Det}\left(\mathbf{W}^{(\mathbf{r})'}\mathbf{W}^{(\mathbf{r})}\right) &= \text{D}\left(\mathbf{W}^{(\mathbf{r})}\right)\\ &= n\mathbf{W}^{(\mathbf{r})'}\mathbf{W}^{(\mathbf{r})} - \left(\mathbf{W}^{(\mathbf{r})'}\mathbf{1}\right)\left(\mathbf{1}'\mathbf{W}^{(\mathbf{r})}\right)\\ &= n\mathbf{W}^{(\mathbf{r})'}\mathbf{W}^{(\mathbf{r})} - \mathbf{W}^{(\mathbf{r})'}\mathbf{1}\mathbf{1}'\mathbf{W}^{(\mathbf{r})}\\ &= \Sigma_W\end{aligned}$$

and $\Sigma_W = c\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r) = c\Sigma_Z$, since

$$\begin{aligned}\text{D}(\mathbf{W}) &= n\mathbf{W}^{(\mathbf{r})'}\mathbf{W}^{(\mathbf{r})} - \left(\mathbf{W}^{(\mathbf{r})'}\mathbf{1}\right)\left(\mathbf{1}'\mathbf{W}^{(\mathbf{r})}\right)\\ &= ncP_{(r)}(X'X)P_{(r)'} - \left(\sqrt{c}P_{(r)}X'\mathbf{1}\right)\left(\sqrt{c}\mathbf{1}'XP_{(r)'}\right)\\ &= ncP_{(r)}(X'X)P'_{(r)} - kP_{(r)}X'\mathbf{1}\mathbf{1}'XP'_{(r)}\\ &= c\left[nP_{(r)}(X'X)P_{(r)} - P_{(r)}X'\mathbf{1}\mathbf{1}'XP_{(r)'}\right]\\ &= cP_{(r)}\left[n(X'X) - X'\mathbf{1}\mathbf{1}'X\right]P_{(r)'}\\ &= cP_{(r)}\text{D}(X)P_{(r)'}\\ &= c\Sigma_Z\end{aligned}$$

Thus,

$$\begin{bmatrix} t_0 \\ \hline t_{(r)} \end{bmatrix} = \begin{bmatrix} \left[W^{(r)'}W^{(r)}\right]\left(\mathbf{1}'Y\right) - \left(\mathbf{1}'W^{(r)}\right)\left(W^{(r)'}Y\right) \\ \hline \left[-W^{(r)'}\mathbf{1}\right]\left(1'Y\right) + nW^{(r)'}Y \end{bmatrix}\Sigma_W^{-1}$$

Now noting that

$$\begin{aligned}W^{(r)'}W^{(r)} &= \left(\sqrt{c}P_{oc}^{(r)}X'\right)\left(\sqrt{c}P_{oc}^{(r)}X'\right)'\\ &= cP^{(r)}\left(X'X\right)P^{(r)'}\end{aligned}$$

Digitized by Sam Jonah Library

Therefore, the intercept t_0 is further simplified as follows:

$$t_0 = \left[\left\{ cP_{(r)}(X'X)P_{(r)'} \right\} \left\{ \mathbf{1}'Y \right\} - \left\{ \sqrt{c}\mathbf{1}'XP_{(r)'} \right\} \left[\left\{ \sqrt{c}P_{(r)}X' \right\} Y \right] \right] \Sigma_W^{-1}$$
$$= \left[c\left\{ P_{(r)}(X'X)P_{(r)'} \right\} (\mathbf{1}'Y) - c(\mathbf{1}'XP_{(r)'})(P_{(r)}X'Y) \right] \Sigma_W^{-1}$$
$$= \left[\left\{ P_{(r)}(X'X)P_{(r)'} \right\} (\mathbf{1}'Y) - (\mathbf{1}'XP_{(r)'})(P_{(r)}X'Y) \right] \Sigma_W^{-1}$$

Introducing $(\frac{1}{n}\mathbf{1}'Y)\mathrm{D}(W)$ and canceling it off gives

$$t_0 = \left[\frac{1}{n}(\mathbf{1}'Y)\mathbf{D}(W) - \frac{1}{n}(\mathbf{1}'Y)\mathbf{D}(W) + \left\{ P_{(r)}(X'X)P_{(r)'} \right\} (\mathbf{1}'Y) - \right.$$
$$\left. (\mathbf{1}'XP_{(r)'})(P_{(r)}X'Y) \right] \Sigma_W^{-1}$$
$$= \left[\frac{1}{n}(\mathbf{1}'Y)\mathbf{D}(W) - \frac{1}{n}(\mathbf{1}'Y)\left\{ ncP_{(r)}(X'X)P'_{(r)} - cP_{(r)}X'\mathbf{1}\mathbf{1}'(XP'_{(r)}) \right\} + \right.$$
$$\left. c\left[\left\{ P_{(r)}(X'X) \right\} P_{(r)}(\mathbf{1}'Y) - (\mathbf{1}'XP'_{(r)})(P_{(r)}X'Y) \right] \right] \Sigma_W^{-1}$$
$$= \left[\frac{1}{n}(\mathbf{1}'Y)\mathbf{D}(W) - \frac{1}{n}\mathbf{1}'(XP'_{(r)})\left\{ ncP_{(r)}X'Y - cP_{(r)}X'\mathbf{1}\mathbf{1}'Y \right\} \right] \Sigma_W^{-1}$$
$$= \left[\frac{1}{n}\mathbf{1}'Y\mathbf{D}(W) - \frac{c}{n}\mathbf{1}'(XP'_{(r)})\left\{ n(P_{(r)}X')Y - P_{(r)}X'\mathbf{1}\mathbf{1}'Y \right\} \right] \Sigma_W^{-1}$$
$$= \frac{1}{n}\mathbf{1}'Y - \frac{1}{n}\mathbf{1}'(XP'_{(r)})\left\{ n(P_{(r)}X')Y - P_{(r)}X'\mathbf{1}\mathbf{1}'Y \right\} \Sigma_W^{-1}$$
$$= \bar{Y} - \frac{1}{n}\mathbf{1}'(XP'_{(r)})\Sigma_{ZY}\Sigma_{ZZ(p)}^{-1}$$
$$= \bar{Y} - \Sigma_{ZY}\Sigma_{ZZ}\bar{Z}$$

Digitized by Sam Jonah Library

Taking the coefficient vector $t_{(r)}$ gives

$$\begin{aligned}
t_{(r)} &= \left[nW^{(r)'}Y - (W^{(r)'}\mathbf{1})(\mathbf{1}'Y) \right] \Sigma_W^{-1} \\
&= \left[n\sqrt{c}(P_{(r)}X')Y - \sqrt{c}(P_{(r)}X')\mathbf{1}(\mathbf{1}'Y) \right] \Sigma_W^{-1} \\
&= \sqrt{c}\left[n(P_{(r)}X')Y - (P_{(r)}X')\mathbf{1}(\mathbf{1}'Y) \right] \frac{1}{c}\Sigma_Z^{-1} \\
&= \frac{\sqrt{c}}{c}\Sigma_{ZY}\Sigma_{ZZ}^{-1} \\
&= \frac{1}{\sqrt{c}}\Sigma_{ZY}\Sigma_{ZZ}^{-1}
\end{aligned}$$

Thus, the regression coefficients based on the original PC multiplied by the reciprocal of $\sqrt{c}$ gives the regression coefficients based on the new PCs except for the intercept which remain the same for both types of components.

Now, the coefficient of determination (R^2) of the model for $Y_i; i = 1, 2 \cdots, n$ in terms of $X = (X_1, X_2, \cdots, X_p)'$ is given as

$$r^2 = \frac{SSR}{SST}$$

where the Sum of Squares Regression (SSR) is given as

$$\begin{aligned}
SSR &= b'X'Y - \frac{1}{n}\left(\sum Y\right)^2 \\
&= b'X'Y - \frac{1}{n}\left(\mathbf{1}'Y\right)^2
\end{aligned}$$

and the Sum of Squares Total (SST) is given as

$$SST = Y'Y - \frac{1}{n}\left(\mathbf{1}'Y\right)^2$$

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Now denoting by $SSR^{(r)}$, the Sum of Square Regression explained in the model $\mathbf{t}$ for $\mathbf{W}^{(r)}$. Then

$$\begin{aligned}
SSR^{(r)} =& t'W^{(r)'}Y - \frac{1}{n}\left(\mathbf{1}'Y\right)^2 \\
=& \begin{pmatrix} \frac{1}{n}\left(\mathbf{1}'Y\right) - \frac{1}{n}\Sigma_{YZ}\Sigma_{ZZ}^{-1}\left(P_{(r)}X'\right)\mathbf{1} \\ \frac{1}{\sqrt{c}}\Sigma_{YZ}\Sigma_{ZZ}^{-1} \end{pmatrix}' \begin{pmatrix} \mathbf{1}' \\ \hdashline W^{(r)'} \end{pmatrix} Y - \frac{1}{n}\left(\mathbf{1}'Y\right)^2 \\
=& \left[\frac{1}{n}(\mathbf{1}'\cdot Y) - \frac{1}{n}\Sigma_{YZ}\Sigma_{ZZ}^{-1}(P^{(r)}X')\mathbf{1}\right](\mathbf{1}'Y) + \\
& \frac{1}{\sqrt{c}}\Sigma_{YZ}\Sigma_{ZZ}^{-1}\left(\sqrt{c}(P^{(r)}X')\right)Y - \frac{1}{n}(\mathbf{1}\cdot Y)^2 \\
=& \Sigma_{YZ}\Sigma_{ZZ}^{-1}(P^{(r)}X')\cdot Y - \Sigma_{YZ}\Sigma_{ZZ}^{-1}(P^{(r)}X')\mathbf{1}\cdot\mathbf{1}'Y \\
=& \Sigma_{YZ}\Sigma_{ZZ}^{-1}\left[(P^{(r)}X')Y - \frac{1}{n}(P^{(r)}X')\mathbf{1}\mathbf{1}'Y\right] \\
=& \frac{1}{n}\Sigma_{YZ}\Sigma_{ZZ}^{-1}\left[n(P_{(r)}X')\cdot Y - (P_{(r)}X')\mathbf{1}\mathbf{1}'\cdot Y\right] \\
=& \frac{1}{n}\Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZY}
\end{aligned}$$

The result shows that $SSR^{(r)}$ is in terms of V-C matrices involving the original components PR. It means that the performance of the NRPCs in regression is the same as that of the reduced set of original PCs.

Description of Illustrative Datasets

The study makes use of relevant datasets to illustrate the characteristics of the new components derived in the study. Three datasets are found suitable for the illustrations, and are labelled as Dataset 1, 2 and 3. All three datasets are secondary and are contained in various texts on multivariate statistics such as Johnson and Wichern (2014), Anderson (2003) and Sharma (1996). Dataset I which is frequently referred to in this thesis as 'sales performance' data or 'sales' data is one that covers seven variables $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_7)$ with fifty responses. The variables are

$\mathbf{X}_1$— Sales Growth

Digitized by Sam Jonah Library

$\mathbf{X}_2$— Sales Profitability

$\mathbf{X}_3$— New Account Sales

$\mathbf{X}_4$— Creativity

$\mathbf{X}_5$— Mechanical Reasoning

$\mathbf{X}_6$— Abstract Reasoning

$\mathbf{X}_7$— Mathematics

The data consists of performance scores in assessment tests of sales personnel of a firm on the seven variables which measure the quality of the sales staff, and are believed to reveal performance in sales. The measures on the first three variables $\mathbf{X}_1, \mathbf{X}_2,$ and $\mathbf{X}_3$ are converted to a scale, with 100 indicating "average" performance. Measurements are also taken for each of the 50 individuals on each of four tests on the remaining variables $\mathbf{X}_4, \mathbf{X}_5, \mathbf{X}_6,$ and $\mathbf{X}_7$ which appears to measure creativity, mechanical reasoning, abstract reasoning, and mathematical ability, respectively.

As a result of the composition of this data, it is found convenient to partition it into two as $\mathbf{X}' = (\mathbf{X}^{(1)}|\mathbf{X}^{(2)})$ where $\mathbf{X}^{(1)} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ and $\mathbf{X}^{(2)} = (\mathbf{X}_4, \mathbf{X}_5, \mathbf{X}_6, \mathbf{X}_7)$. This partition structure is also adopted on studies in Canonical Correlation Analysis (Apanyin, 2021). In this study, the first subset may be regarded as a set of response variables, whilst the second subset constitutes the set of predictor variables. It is convenient therefore to write $\mathbf{Y} = \mathbf{X}^{(1)}$ and $\mathbf{X} = \mathbf{X}^{(2)}$. This dataset is particularly useful for illustrating the performance of the NRPC in multivariate multiple reduced rank regression (MMRRR) modelling. Only the second subset $\mathbf{X}^{(2)}$ is used for illustrating the relationship between new components NRPC, and the corresponding original principal components (OPCs). The partition of this dataset is guided to obtain a meaningful linear relationship between the two subsets, a condition that is required to make the technique of MMRRR meaningful. The selected dataset shows that any component of the first set of variables may be influenced by components of the second set of variables. As outlined in the literature review, this

Digitized by Sam Jonah Library

particular data has been used extensively in multivariate studies. It is, however, being applied for the first time in this study from the perspective of MMRRR modelling. The data labelled as Dataset 2 is usually referred to as the USAFood data. It covers prices on five food items collected from twenty three (23) cities of the United States of America (Sharma, 1996). The food items are Bread, Burger, Milk, Oranges and Tomatoes.

Implementation of Results

Tables 1 to 3 give the results of the weighted PCs for various reduced dimensions for various datasets of dimension p based on ordinary component extraction. That is, the transformations are obtained by weighting the principal components extracted from the ordinary variance-covariance matrix. For each set of r reduced component dimension, the corresponding weight (c) is provided. In each case, the last set of $r = p$ weighted components become the same as the original principal component. It can be verified that in each case, dividing the new component by the weight c gives the corresponding ordinary PC. It can be noticed that any set of r reduced dimensions explains 100% of information in the data. In particular, the single reduced dimension (for $r = 1$) explains 100% information. This means that the total information in the original p dimensions is now contained in a single transformed dimension.

Table 4 also gives the transformed component for various reduced dimensions for Dataset 2 based on a weighted variance-covariance matrix that yields generalized PC. It can be seen that the features of the weighted principal component are the same as those in Tables 1 to 3. That is, there is no loss of information even when the weighting is carried out on components extracted from weighted variance-covariance matrix.

Digitized by Sam Jonah Library

Table 1: **Weighted PCs for various Reduced Dimensions for Dataset 1 (SalesData) based on Ordinary PC Extraction**

Var	$r = 1; \sqrt{c} = 1.0913$	$r = 2; \sqrt{c} = 1.0289$	
1	0.1901	0.1792	0.8761
2	0.2150	0.2028	0.4736
3	0.1232	0.1162	-0.0725
4	1.0456	0.9859	-0.2481
% Exp	100.00	88.7198	11.0778
Cum	100.00	88.7198	100.0000

Table 1 continued

Var	$r = 3; \sqrt{c} = 1.0100$		
1	0.1759	0.8600	-0.4592
2	0.1990	0.4649	0.8324
3	0.1140	-0.0712	0.3173
4	0.9677	-0.2436	-0.1251
% Exp	85.6558	10.6952	3.6491
Cum	85.6558	96.3509	100.0000

Table 1 continued

Var	$r = 4; \sqrt{c} = 1.0100$			
1	0.1742	0.8515	-0.4546	0.1948
2	0.1971	0.4603	0.8242	-0.2646
3	0.1129	-0.0705	0.3141	0.9400
4	0.9582	-0.2412	-0.1239	-0.0918
% Exp	83.9677	10.4844	3.5771	1.9708
Cum	83.9677	94.4521	98.0292	100.00

Digitized by Sam Jonah Library

Table 2: **Weighted PCs for various Reduced Dimensions for Dataset 2 (US Food) based on Ordinary PC Extraction**

Var	$r = 1; \sqrt{c} = 1.3037$	$r = 2; \sqrt{c} = 1.0945$	
1	0.0371	0.0312	-0.1809
2	0.2609	0.2190	-0.6919
3	0.0543	0.0456	-0.4839
4	1.2240	1.0276	0.3441
5	0.3593	0.3016	-0.57781
% Exp	100.00	70.481	29.519
Cum	100.00	70.481	100.0000

Table 2 continued

Var	$r = 3; \sqrt{c} = 1.0336$		
1	0.0294	-0.1709	0.0221
2	0.2069	-0.65359	0.2628
3	0.0431	-0.4570	-0.9187
4	0.9705	0.3249	-0.1254
5	0.2849	-0.5457	0.3731
% Exp	62.861	26.328	10.811
Cum	62.861	89.189	100.0000

Table 2 continued

Var	$r = 4; \sqrt{c} = 1.0041$			
1	0.0286	-0.1660	0.0214	-0.1905
2	0.2009	-0.6348	0.2552	-0.6613
3	0.0418	-0.4440	-0.8924	0.1081
4	0.9427	0.3156	-0.1218	-0.0693
5	0.2767	-0.5301	0.3625	0.7198
% Exp	59.318	24.844	10.201	5.6367
Cum	59.318	84.162	94.363	100.00

Table 2 continued

Var	$r = 5; \sqrt{c} = 1.0000$				
1	0.0285	-0.1653	0.0214	0.1897	0.9672
2	0.2001	-0.6322	0.2542	-0.6586	-0.2488
3	0.0417	-0.4422	-0.8887	0.1077	-0.0361
4	0.9389	0.3144	-0.1214	-0.0690	0.0152
5	0.2756	-0.5279	0.3610	0.7168	0.0343
% Exp	58.835	24.642	10.118	5.5909	0.8138
Cum	58.835	83.477	93.595	99.186	100.00

Digitized by Sam Jonah Library

Table 3: **Weighted PCs for various Reduced Dimensions for Dataset 3 (Subscores) based on Ordinary PC Extraction**

Var	$r = 1; \sqrt{c} = 1.2618$	$r = 2; \sqrt{c} = 1.13405$	
1	-0.8003	-0.7193	0.0070
2	-0.7493	-0.6734	0.0028
3	-0.2070	-0.1861	-0.0190
4	-0.5892	-0.5295	0.0256
5	0.0161	0.0145	1.1336
% Exp	100.00	80.769	19.231
Cum	100.00	80.769	100.0000

Table 3 continued

Var	$r = 3; \sqrt{c} = 1.0674$		
1	-0.6770	0.0066	0.3709
2	-0.6339	0.0027	0.3481
3	-0.1751	-0.0179	-0.0230
4	-0.4984	0.0231	-0.9380
5	0.0136	1.0670	0.0168
% Exp	71.554	17.037	11.409
Cum	71.554	88.5913	100.0000

Table 3 continued

Var	$r = 4; \sqrt{c} = 1.0254$			
1	-0.6504	0.0063	0.3563	0.7066
2	-0.6089	0.0026	0.3344	-0.7125
3	-0.1682	-0.0172	-0.0221	-0.2102
4	-0.4788	0.0222	-0.9011	0.0201
5	0.0131	1.0250	0.0161	-0.0065
% Exp	66.040	15.724	10.530	7.706
Cum	66.040	81.7643	92.294	100.00

Table 3 continued

Var	$r = 5; \sqrt{c} = 1.0000$				
1	-0.6342	0.0062	0.3475	0.6890	0.0465
2	-0.5938	0.0025	0.3261	-0.6948	-0.2414
3	-0.1641	-0.0168	-0.0215	-0.2050	0.9645
4	-0.4669	0.0217	-0.8788	0.0196	-0.0945
5	0.0128	0.9996	0.0157	-0.0064	0.0186
% Exp	62.805	14.954	10.014	7.3287	4.8990
Cum	62.805	77.759	87.772	95.101	100.00

Digitized by Sam Jonah Library

Table 4: **Weighted PCs for various Reduced Dimensions for Dataset 2 (USFood) based on Generalized PC Extraction**

Var	$r = 1; \sqrt{c} = 2.0501$	$r = 2; \sqrt{c} = 1.4951$	
1	-3.3530	-2.4454	3.0154
2	-0.5512	-0.4020	-2.0117
3	-1.5702	-1.1452	-0.5788
4	-0.5188	0.3783	0.3730
5	-0.5477	-0.3995	1.3185
% Exp	100.00	53.1889	46.8111
Cum	100.00	53.1889	100.0000

Table 4 continued

Var	$r = 3; \sqrt{c} = 1.2456$		
1	-2.0373	2.5122	2.3756
2	-0.3349	-1.6760	-1.5931
3	-0.9541	-0.4822	0.8124
4	-0.3152	0.3107	-0.6464
5	-0.3328	1.0985	0.9835
% Exp	36.9175	32.4908	30.5917
Cum	36.9175	69.4087	100.0000

Table 4 continued

Var	$r = 4; \sqrt{c} = 1.0975$			
1	-1.7950	2.2134	2.0931	2.1780
2	-0.2951	-1.4767	-1.4037	-0.0640
3	-0.8406	-0.4249	0.7158	-0.3688
4	-0.2777	0.2738	-0.5695	-0.0133
5	-0.2932	0.9678	0.8665	-0.3357
% Exp	28.6591	25.2226	23.7483	22.3701
Cum	28.6591	53.8817	77.6299	100.00

Table 4 continued

Var	$r = 5; \sqrt{c} = 1.0000$				
1	-1.6356	2.0168	1.9071	1.9845	-0.7335
2	-0.2689	-1.3455	-1.2790	-0.0583	0.4773
3	-0.7659	-0.3872	0.6522	-0.3360	-0.6344
4	-0.2530	0.2495	-0.5189	-0.0121	-0.3061
5	-0.2672	0.8819	0.7896	-0.3059	0.6949
% Exp	23.7937	20.9406	19.7166	18.5724	16.2766
Cum	23.7937	44.7344	64.4510	83.0234	100.00

Digitized by Sam Jonah Library

Figures 3 to 5 further illustrate graphically the component extraction with no loss of information. The full information explained by the r reduced dimensions is seen in the larger spread of the box plot of projections onto the components extracted from the three datasets described in Section 4.1. Figure 3 for example, gives the boxplot based on dataset 2. In this figure, NP1 is the plot for the new single reduced dimension and PC1 is the first original component. The full information explained in NP1 is seen in the longer plot compared to a much smaller information (58.8%). The second and third plots are the information contained in the two 2-reduced dimension components. The structure is the same for the other two figures. Another observation in the plot shows that when there is extreme observation in the plot along the original PC, the plot for the corresponding new dimension also shows the extreme observation.

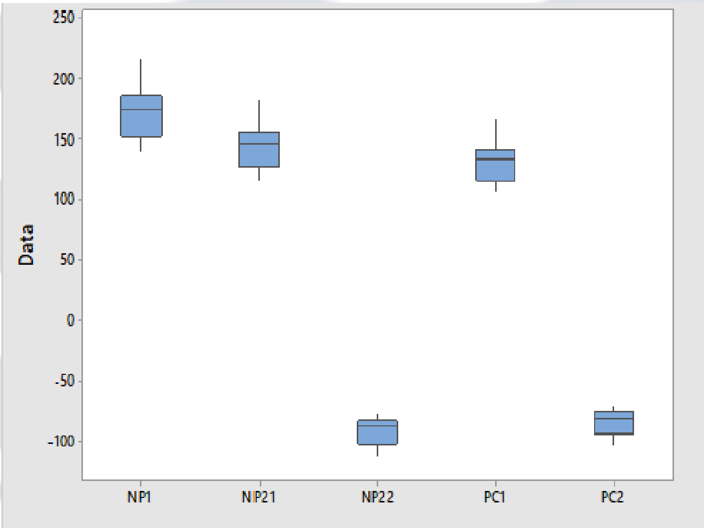

Figure 3: Plot showing variation in projected data on reduced 1 and 2 new and original PC dimensions for Dataset 2 based on classical PC extraction

Note: *NP21 implies first new PC of the two retained PCs.*

NP22 implies second PC of the two retained PCs.

Digitized by Sam Jonah Library

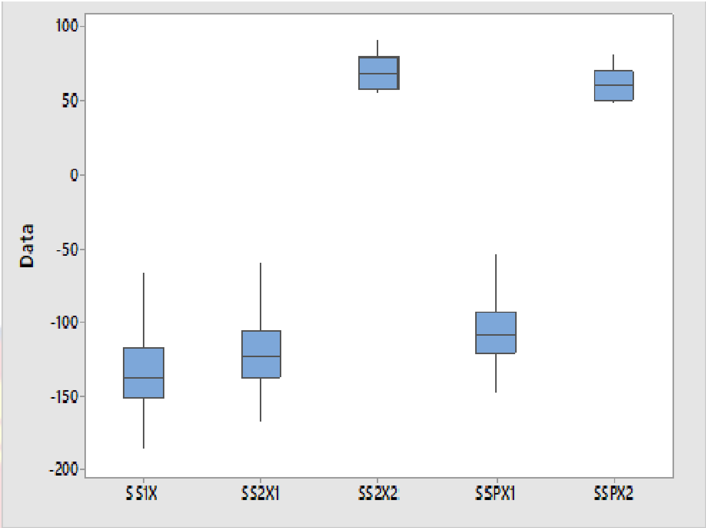

Figure 4: Plot showing variation in projected data on reduced 1 and 2 new and original PC dimensions for Dataset 3

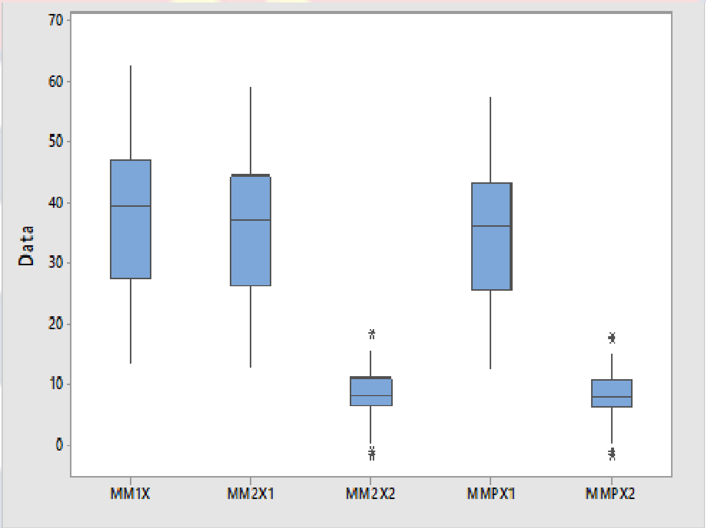

Figure 5: Plot showing variation in projected data on reduced 1 and 2 new and original PC dimensions for Dataset 1 based on classical PC extraction

Figure 6 is the graphical representation of the information explained along the r reduced dimensions compared to the corresponding original components extracted from weighted variance covariance matrix that generates generalized components (GPC). An interesting observation is that projection on the second of the two 2-reduced dimension component (WGUS22) (i.e., the third plot) shows extreme observations whiles projection on the corresponding second original GPC (USP2) does not show extreme observations. This is in contrast to a

Digitized by Sam Jonah Library

feature of the projection when PC is extracted from the ordinary variance covariance matrix demonstrated in Figures 3 to 5. It can be seen that there is a much greater spread in the single reduced component (WGUS11) than the corresponding original component USP1 that reflects the wide difference in the percentage information explained (100%) by the single weighted PC and the information (23.8%) in the first ordinary GPC (USP1).

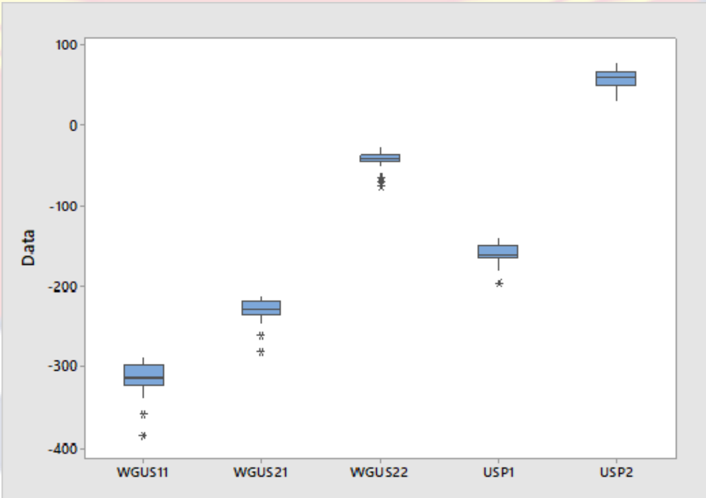

Figure 6: Plot showing variation in projected data on reduced 1 and 2 weighted and original PC dimensions for Dataset 2 based on Generalized PC

Implementation of Reduced Regression Modelling

In Tables 5 and 6, various reduced regression models for various response variables in Dataset 1 are provided for various reduced dimensions based on PCs extracted form ordinary variance-covariance matrix. As explained from the formulations in Section 4.1, it is not expected that the full information contained in the r reduced dimensions would translated into a full coefficient of multiple determination when the set of r reduced components are used in regression modelling of a response variable in the dataset. As expected, the tables show that the R^2 accounted for by the model in terms of the set of new weighted r reduced components is the same as that for the model in terms of the corresponding set of original components. As shown in the derivation of the technique, in each

Digitized by Sam Jonah Library

case of r reduced dimensions, the intercepts are the same for each pair of models in terms of the set of new reduced components and the other in terms of the original components. It can also be verified that dividing the non-intercept coefficients of the models in terms of the original PCs gives the corresponding set of non-intercept coefficients for the model in terms of the weighted PCs.

A number of other observations are worth-noting in the model tables. It can be seen that for all models in terms of the original PCs, the introduction of additional PCs does not change the coefficients of the PCs initially in the model. This is not the case for the models in terms of the new weighted PCs. Thus, for reduced rank regression (RRR) modelling in terms of the weighted PC, the resulting coefficients are not affected by the orthonormality of the original PCs. Again, as expected, the pair of models when $r = p$ are the same for the two types of components.

The last column of the tables give the original multiple linear regression (MLR) models in terms of the original variables. It can be seen that its coefficients are different from the full model in terms of the components. However, the full coefficient of multiple determination (CMD) value (96.52% in Table 5) for the original MLR is the same as those of the full models in terms of the components.

Table 7 gives various reduced regression models for Variable 2 in Dataset 1 for various reduced dimensions based on generalized PC (using the weighted variance-covariance matrix). Unlike the models in terms of PCs extracted from regular variance-covariance matrix, the table shows that the R^2 accounted for by the model in terms of the set of new weighted r reduced components is not the same as that for the model in terms of the corresponding set of original GPC. In fact, in this case, the R^2 accounted for by the weighted components could be extremely small. Another deviation is that for each case of r reduced dimensions, the intercepts are not the same for each pair of models in terms of the set of weighted components and the other in terms of the original GPC. It can also be

Digitized by Sam Jonah Library

verified that dividing the non-intercept coefficients of the models in terms of the original GPC does not yield the corresponding set of non-intercept coefficients for the models in terms of the weighted PC.

The results therefore show that the theoretical relation that has been derived to exist between the coefficients of the models in terms of the original GPC and the corresponding set of coefficients for the models in terms of weighted PC may not apply in the case where components are not extracted from regular variance-covariance matrix.

Digitized by Sam Jonah Library

Table 5: **Reduced Regression Modelling for Response Variable 2 for various Reduced Dimensions for Dataset 1 based on Classical PC**

	Type of model			
	$r = 1; c = 1.0913$		$r = 1; c = 1.0913$	
Coefficient	WPC1	PC	WPC	PC
t_0	76.0040	76.0040	73.1287	73.1287
t_1	0.8147	0.8891	0.8641	0.8891
t_2			0.3423	0.3522
t_3				
t_4				
R-sq	0.9217	0.9217	0.9397	0.9397

Table 5 continued

	Type of model				
	$r = 3; c = 1.0100$		$r = 4; c = 1.0000$		
Coefficient	WPC1	PC	WPC	PC	MLR
t_0	70.7593	75.3579	75.3579	75.3579	75.3579
t_1	0.8803	0.8891	0.8891	0.8891	0.1225
t_2	0.3487	0.3522	0.3522	0.3522	0.8673
t_3	0.3772	0.3809	0.3809	0.3809	-0.5723
t_4			-0.8166	-0.8166	0.7947
R-sq	0.9469	0.9469	0.9652	0.9652	0.9652

Digitized by Sam Jonah Library

Table 6: **Reduced Regression Modelling for response variable 1 for various Reduced Dimensions for Dataset 1 based on Regular PC**

	Type of model			
	$r = 1; c = 1.0913$		$r = 1; c = 1.0289$	
Coefficient	WPC1	PC	WPC	PC
t_0	76.9095	76.9095	74.6041	74.6041
t_1	0.5834	0.6367	0.6188	0.6367
t_2			0.2744	0.2824
t_3				
t_4				
R-sq	0.8999	0.8999	0.9220	0.9220

Table 6 continued

	Type of model				
	$r = 3; c = 1.0100$		$r = 4; c = 1.0000$		
Coefficient	WPC1	PC	WPC	PC	MLR
t_0	72.8674	72.8674	68.9209	68.9209	68.9209
t_1	0.6304	0.6367	0.6367	0.6367	0.3609
t_2	0.2796	0.2824	0.2824	0.2824	0.3001
t_3	0.2764	0.2792	0.2792	0.2792	0.7985
t_4			0.7008	0.7008	0.4431
R-sq	0.9294	0.9294	0.9550	0.9550	0.9550

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 7: **Reduced Regression Modelling for Response Variable 2 for various Reduced Dimensions for Dataset 1 based on GPC**

	Type of model			
	$r = 1; c = 1.0913$		$r = 1; c = 1.0289$	
Coefficient	WPC1	PC	WPC	PC
t_0	102.5688	78.7516	98.1181	69.8695
t_1	-0.1798	-0.9707	-0.2512	-0.9868
t_2			0.2306	0.7262
t_3				
t_4				
R-sq	0.0193	0.8478	0.0296	0.9163

Table 7 continued

	Type of model				
	$r = 3; c = 1.0100$		$r = 4; c = 1.0000$		
Coefficient	WPC1	PC	WPC	PC	MLR
t_0	79.5762	76.5175	75.3579	75.3579	75.3579
t_1	-0.3050	1.2262	-0.3419	1.8341	0.1225
t_2	0.2800	-0.2256	0.3139	-0.1098	0.8673
t_3	0.8348	1.9707	0.9359	1.9309	-0.5723
t_4			-0.7398	-0.7851	0.7947
R-sq	0.8913	0.9518	0.9652	0.9652	0.9652

Digitized by Sam Jonah Library

Chapter Summary

The chapter has learnt and derived a new set of reduced PCs from an original set of PCs extracted from a given V-C matrix. It has subsequently examined the features of the new reduced PCs (NRPCs). It is shown analytically that the NRPCs is a constant multiple of the corresponding set of reduced original PCs. Thus, the original PCs are found to be a normalization of the NRPCs. The relationship between the V-C matrices of both the NRPCs and the original PCs is also determined. The derived NRPCs have been implemented using suitable datasets in the literature. The features of the NRPCs are further examined in application to Multivariate Multiple Reduced Rank Regression (MMRRR) modelling. Thus, the NRPCs have been examined both analytically and practically. The application is carried out for component extraction using regular (unweighted V-C) matrix as well as weighted V-C matrix. It has been found that the analytical relationship between the NRPCs and the original PCs hold in MMRRR modelling when PC extraction is based on regular V-C matrix. However, the analytical relationship between the two types of components is not found to hold when component extraction is based on weighted V-C matrix.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Overview

This chapter presents an overview of the entire work and conclusions based on the discussion of results in Chapter Four. Based on that, recommendations will be given.

Summary

In Chapter One, the problem of the classical principal component was introduced and the motivation for the study was outlined. The pertinent problem identified was that in the classical PCA, the amount of information loss after truncation, no matter the size, can have an important implication depending on the sensitivity of the area to which the technique is applied. It made clear the intention of the study to derive a variant of reduced set of PCs that retains all information in a given dataset. Of significance, the chapter outlined how the approach would be applied to Reduced Rank Regression modelling, and how it would serve as a useful alternative to best subset regression.

Chapter two considered various works done by other researchers using the classical PCA, WPCA, and PCR. It is presented that the advantages PCA bring have seen many researchers employ in their works. The chapter also reviewed that the literature on WPCA aim at explaining the overall variation in a dataset according to a given number of PCs. It has also reviewed that the method also allows one to recover a given number of perpendicular PCs among the most important ones for the case of problems with weighted or missing data. Comparatively, the literature shows that WPCA produces better results than the classical PCA.

The third chapter was organized as follows: Principal Component Analy-

Digitized by Sam Jonah Library

sis, Eigenstructure of covariance matrix, Singular Value Decomposition of Data matrix, Spectral Decomposition of a Matrix, Spectral Decomposition of a matrix, and the Variance-Covariance matrix for Principal Component Extraction. The methodology presented extensive review of the procedure for conducting principal component analysis. All the new transformed PCs were extracted from the Variance-Covariance matrix and the correlation matrix of the original datasets used in this work. All the datasets were multivariate.

The chapter four learnt and derived a new set of reduced PCs from an original set of PCs extracted from a given V-C matrix. It subsequently examined the features of the new reduced PCs (NRPCs). It showed analytically that the NRPCs is a constant multiple of the corresponding set of reduced original PCs. Thus, the original PCs was found to be a normalization of the NRPCs. The relationship between the V-C matrices of both the NRPCs and the original PCs were also determined. The derived NRPCs were implemented using suitable datasets in the literature. The features of the NRPCs were further examined in application to Multivariate Multiple Reduced Rank Regression (MMRRR) modelling. Thus, the NRPCs have been examined both analytically and practically. The application was carried out for component extraction using regular (unweighted V-C) matrix as well as weighted V-C matrix. It was found that the analytical relationship between the NRPCs and the original PCs hold in MMRRR modelling when PC extraction is based on regular V-C matrix. However, the analytical relationship between the two types of components was not found to hold when component extraction is based on weighted V-C matrix.

Digitized by Sam Jonah Library

Conclusions

It is a feature of all dimensionality reduction techniques to have an appreciable amount of loss of information when the first few PCs are retained. The loss of information due to dimensionality reduction means that the spread or variation in the component scores on the retained PCs does not reflect the actual variation in the original data. As a result of this phenomenon, extracting reduced components could have quite crucial implications in sensitive areas of human endeavor. It therefore becomes necessary to consider reduced PC extraction that eliminates the loss of information.

The attempt made in this study to address the identified problem learns from the basic principle that a V-C matrix may be factorized as a product of a matrix and the transpose of that matrix, and this matrix is further derived from the spectral decomposition of the V-C matrix. A new reduced PC (NRPC) extraction simply makes use of a product of the factorization of the V-C matrix. Other background methodology that is employed is the function of a V-C matrix that is simplified by the spectral decomposition. Too main functions of the V-C matrix employed in the process involves the square root function and the trace function

With this study, we have been able to show that it is possible to bring down the dimensions of a dataset with no loss of information in the original data. This study derives a new reduced set of PCs (NRPCs) that is simply a constant multiple of the first r original PCs (OPCs). This means that the OPCs are just a normalization of the NRPCs. The normalizing constant represents the common variation explained by each of the components in the set of r NRPCs. Irrespective of the number r of NRPCs, all the information in the original data is explained by the NRPCs. This means that even a single dimensional NRPC accounts for 100% information. Further features of the NRPCs are examined both analytically and practically in Multivariate Multiple Reduced Rank Regression

Digitized by Sam Jonah Library

(MMRRR) modelling for extraction based on both regular (unweighted) and weighted V-C matrices of a dataset. It is found that for the NRPCs extracted from unweighted V-C matrix, the analytical relationship between the NRPCs and the OPCs are preserved in MMRRR modelling. However, if OPCs are based on weighted V-C matrix, then the analytical relationship between the two types of PCs does not hold practically in MMRRR modelling. In this case, the RRR modelling is found to perform very poorly except when almost all new components are used in the model. The results of the study shows that in order to determine the real spread of PC scores for further analysis, the use of the NRPCs would be more useful.

Recommendations

The result of the study has shown that it is possible to obtain a reduced set of unstandardized PCs that account for all information in the data. It should therefore be a preferred procedure for extracting PCs rather than the usual PCA in the literature.

The study has demonstrated that the derived features of the new PCA are exhibited by one that is based on the ordinary PCA that are extracted from the usual V-C matrix. However, some of these features are not exhibited by the new PCA based on weighted V-C. A natural extension of this work would be to examine the characteristics of PCs based on V-C matrix that do not allow for those derived features when transformed by the procedures studies in this work. This information could provide a guide for a more generalized procedure for extracting PCs with no information loss.

Another useful area for consideration from this study is an application of the new PCs that would also lead to zero information loss. For example in this study, the application of the extracted $r-$reduced dimensions in Reduced Rank Regression Modelling does not explain an amount of variation that is equal to

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

the Coefficient of Multiple Determination (CMD) of the original Multiple Linear Regression Modelling in terms of the original variables. It would therefore be interesting to obtain reduced dimensions that could also lead to zero information loss applied in regression in the original data.

Digitized by Sam Jonah Library

REFERENCES

Aboagye, E. A., & Mensah, C. (2016). Principal Component Analysis of Students' Academic Performance in Mathematics and Statistics. *American Based Research Journal, 5.*

Adu, A. A. (2022). *Automated Threshold Approach for Dimensionality Detection in Multivariate Data,* PhD Thesis, University of Cape Coast.

Anderson, T. W. (2003). Introduction to multivariate statistical analysis. New York: Wiley.

Apanyin, F. (2021). *Application of canonical correlation in multivariate multiple linear regression analysis,* MPhil Thesis, University of Cape Coast.

Artigue, H., & Smith, G. (2019). The principal problem with principal components regression. *Cogent Mathematics & Statistics, 6*(1), 1622190.

Baker, K. (2005). Singular value decomposition tutorial. *The Ohio State University, 24.*

Da Costa, J. F. P., Alonso, H., & Roque, L. (2009). A weighted principal component analysis and its application to gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 8*(1), 246-252.

Delchambre, L. (2014). Weighted principal component analysis: a weighted covariance eigen decomposition approach. *Monthly Notices of the Royal Astronomical Society, 446*(4), 3545-3555.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Hong, D., Fessler, Johnson. A., & Balzano, L. (2018). Optimally weighted PCA for high-dimensional heteroscedastic data.*arXiv preprint arXiv:1810.12862*.

Johnson, R. A., & Wichern, D. W. (2002).*Applied multivariate statistical analysis.* New Jersey: Prentice Hall.

Jollife, I. T (2002). Principal Component Analysis, Second Edition. New York: Springer.

Nkansah, B. K. (2018). On the Kaiser-Meier Olkin's Measure of Sampling adequacy. *Journal of Mathematics Theory and Modelling, 8*(7), 52-76.

Niu, Z., & Qiu, X. (2010). Facial expression recognition based on weighted principal component analysis and support vector machines. *International Conference on Advanced Computer Theory and Engineering (ICACTE), 3*(1), 174.

Tavoli, R., Kozegar, E., Shojafar, M., Soleimani, H., & Pooranian, Z. (2013). Weighted PCA for improving Document Image Retrieval System based on keyword spotting accuracy. *2013 36th international conference on telecommunications and signal processing (TSP)* (pp. 773-777).

Xu, P. (1998). Truncated SVD methods for discrete linear ill-posed problems. Geophysical Journal International, 135(2), 505-514.

Van der Maaten, L. J. P. (2007). An introduction to dimensionality reduction using MATLAB. Report, 1201(07-07), 62.

Digitized by Sam Jonah Library