Principal Component Analysis and Its Derivation From Singular Value Decomposition

Orumie, Ukamaka Cynthia[1] & Ogbonna Onyinyechi[1]

[1] Department of Mathematics/Statistics, University of Port Harcourt, Nigeria

Correspondence: Department of Mathematics/Statistics, University of Port Harcourt, Nigeria

Abstract

Generally, today data analysts and researchers are often faced with a daunting task of reducing high dimensional datasets as large volume of data can be easily generated given the explosive activities of the internet. The most widely used tools for data reduction is the principal component analysis. Merely in some cases, the singular value decomposition method is applied. The study examined the application and theoretical framework of these methods in terms of its linear algebra foundation. The study discovered that the SVD method is a more robust and general method for a change of basis and low rank approximations. But.in terms of application, the PCA method is easy to interpret as illustrated in the work.

Keywords: principal component, singular value decomposition, eigen values and eigen vectors, simulation, linear algebra

1. Introduction

Over time, new mathematical methods tend to spring up owing to complex empirical problems. Just like the saying 'necessity is the mother of all interventions', when confronted with a mathematical impasse, the human mind has the capability to rediscover new and better ways of solving problems. A brainchild of such mind-probing process is the Singular Value Decomposition (SVD) method often abbreviated as SVD. The SVD method in its rights has expanded the frontiers and uses of linear algebra as it underpins several methodologies like Principal Component Analysis, Orthogonal Function Analysis, Eigen Decomposition, Matrix Decomposition, Cholesky Decomposition, Hessenberg Decomposition etc. In this study, the focus of interest of such methods is the Principal Component Analysis (PCA) and its derivation from SVD.

According to Jollife (2002), the SVD methodology is older likened to PCA as it culminated through the works of five mathematicians: Eugenio Beltrami, Camille Jordan, James Joseph Sylvester, Erhard Schmidt and Herman Weyl. These mathematicians did not just discover the SVD methodology; they also laid the theoretical foundation. Beltrami and Jordan are the progenitors of the decomposition methods. Beltrami gave a proof of the outcome for real, invertible matrices with distinct singular values in 1873. Subsequently, Jordan refined the concept and eliminated the unnecessary restrictions imposed by Beltrami. Sylvester, apparently unfamiliar with Beltrami and Jordan's work, rediscovered the result in 1889 and suggested its importance.

PCA is a dimension reducing tool employed to reduce a large data set of variables to a few set of variables without much loss of information. As a mathematical procedure, it converts many correlated variables into a number of uncorrelated variables termed principal components.

PCA is deeply connected to SVD in terms of dimensionality reduction purposes and change of basis. However, SVD is a highly robust method due to its ability to decompose any matrix $\mathbf{A}$ with rank r into $\boldsymbol{U\Sigma V}^{\mathbf{T}}$, with orthogonal matrices $\boldsymbol{U}$ and $\mathbf{V}$ and diagonal $\boldsymbol{\Sigma}$. The non-zero value along the diagonal of $\boldsymbol{\Sigma}$, termed singular values$(\sigma_{1,\dots,}\sigma_r)$ are positive, and $\sigma_n \geq \sigma_{n+1}$. In this same vein, PCA applies the Eigen Decomposition method on square symmetric matrices by searching for new abstract components (eigen-vectors) which clarify most of the information variation in a new organized system. Although compared with EVD used in PCA, SVD is a more precise, robust and reliable method with no need to compute the input correlation/covariance matrix according to Will (1999). Classical PCA is not useful when the number of variables is bigger than the number of observation. In microarray data analysis, Lim, (2013) encountered this situation commonly when genes are accepted as the variables. In such instances, it is pertinent to use PCA using SVD (Deshunk and Purohit , 2007). For applying PCA using SVD, the DNA microarray data was in use for the small round blue cell tumors (SRBCT) of childhood by Khan *et al* (2001).

Nonetheless, in terms of linear algebra, the common underlying idea that coordinates these two methods is the change of basis. Succinctly, the idea behind changing the basis of a vector connotes a new presentation of the vector in a different coordinate system that best depicts the underlying structure of the vector components.

In PCA, the original data set is transformed so that the eigen vectors are the basis vectors, then the new coordinates of the data points are formed with respect to this new basis. This is the change of basis transformation. It is possible because the covariance matrix of the data set is real symmetric. Hence, it's eigen vector matrix is an orthogonal matrix due to an eigen decomposition.

However, there is a more generic transformation that occurs in SVD, which is even possible for non-square matrices. This method is quite powerful for several reasons and one of those is its ability to produce orthonormal bases for the four fundamental bases. Due to this robust nature, SVD is considered computationally efficient and numerically stable as such that it has been used to solve least square problems and other related problems such as pseudo-inverse, etc. (Lyche, 2018).

There are four basic methods to solving the least square problems. These are:

Normal equations, QR decomposition, and SVD.

The first method provides very quick and easy least square solution though these solutions are not that accurate. The second method is more accurate than the first method but requires double as much time. Amongst all, SVD is most widely used for its capability to deliver quick, accurate and stable results even for unconventional cases like the overdetermined system (Least squares problems, 2013).

Between the SVD and PCA methods, there is a large misconception being thrown around with regards to how these methods are being applied to practical problems and its theoretical interconnectedness. This study aims at providing a theoretical framework on the relationship that exist between these methods. More so, the study demonstrates the interpretative power of PCA over SVD by carrying out an exploratory data analysis on a financial data.

Understanding the underlying theoretical foundation which these methods utilize in resolving complex structures has become somewhat invaluable, if not critical. By unraveling the mathematical mysteries behind these methods, it becomes quite easy to navigate the intricacies of dealing with multidimensional data, and also increasing precision of data analysis.

2. Methodology and Data Collection

The research adopted a financial time series data consisting of the Domestic Debt of State Governments for the thirty-six (36) states of Nigeria and its capital (Abuja). As a secondary data, it was sourced from Debt Management office (DMO) via Nigeria Statistical Bulletin. A 6-year time period (2011-2016) was considered as the variables (features) to be studied against the states (observations). This obtains a 36×6 matrix denoted as **X**, where there is 36 number of observations and 6 features to be understudied.

2.1 Principal Component Analysis

Mathematically, the PCA method basically obtains the most important basis used to re-express an ambiguous data into its principal components. That is, it involves a linear transformation of a high dimensional dataset into simpler groups that uncover hidden structures of a garbled data.

At this point, the questions that would readily come to mind are; how would these new bases be obtained and why are they important? In a bid to answer these questions, an arbitrary dataset is assumed and an attempt would be made to implement the underlying rigors of the PCA methodology.

Given an $m \times n$ matrix X where the samples are the *n* columns (e.g. observations) and the *m* rows are variables, the objective is to transform the matrix X linearly to another matrix Y of similar dimension $m \times n$. Hence, for some $m \times m$ matrix P,

$$Y = PX \qquad (1)$$

The matrix P is a transformation matrix which is expected to change the basis from X to Y. Geometrically, P is a rotation and a stretch which moves X to Y.

The rows of P are thought to be the vectors row $p_1, p_2 \ldots, p_m$ and columns of X to be the vectors of the column $x_1, x_2 \ldots, x_n$, then (**2.0**) can be construed this way,

$$PX = (Px_1, Px_2, \ldots, Px_n) = \begin{pmatrix} p_1 \cdot x_1 & p_1 \cdot x_2 \ldots & p_1 \cdot x_n \\ p_2 \cdot x_1 & p_2 \cdot x_2 \ldots & p_2 \cdot x_n \\ \vdots & \vdots & \vdots \\ p_m \cdot x_1 & p_m \cdot x_2 \ldots & p_m \cdot x_n \end{pmatrix} = Y \qquad (2)$$

Such that $p_j. x_j \in \mathbb{R}^m$.

In essence, $p_j. x_j$ is just the Euclidean inner (dot) standard product. This means that the main data is being projected onto columns of P. Hence, the new foundation for column representation of X is given as the rows of P, $(p_1, p_2 \dots, p_m)$.

By assuming linearity, the problem reduces to finding the appropriate change of basis. The vector row $(p_1, p_2 \dots, p_m)$ in the transformation becomes the X principal component.

More underlying questions at this stance would be;

- What is the most profound way to re-express X. That is, how can independence between the new basis in the principal components be defined?

- What is a good basis choice for P?

Principal component analysis defines independence by considering the variance of the data in the original basis. Its purpose is to search for new directions that maximize the variance which defines the new bases. Recall that variance of X with mean μ is given as,

$$\sigma^2{}_x = E[(X - \mu)^2] \tag{3}$$

For instance, if given a vector of p discrete measurements, that is,

$$f = (f_1, f_2, .., f_p) \, with \, mean \, \mu_f$$

If the mean is subtracted from individual measurements, then we get a translated measurement set. Thus, the relation for the measurement variance measurements is given thus,

$$\sigma^2{}_f = \frac{1}{n} f f^T \tag{4}$$

The normalization constant $\frac{1}{n}$ becomes $\frac{1}{n-1}$ as that will provide a biased estimation of the variance particularly for small sample sizes. Hence, equation (1.3) becomes

$$s^2{}_f = \frac{1}{n-1} f f^T \tag{5}$$

Additionally, if given another vector $l = (l_1, l_2, .., l_p)$ again with absent mean, the idea can be generalized to derive the covariance of f and l - covariance is considered a degree of how much two variables alternate simultaneously while variance is essentially a special covariance case where we find two identical variables. The covariance of f and l is given as,

$$s^2{}_{fl} = \frac{1}{n-1} f l^T \tag{6}$$

For an $m \times n$ matrix X where the data matrix is considered with regards to m row vectors, individual length n, that is

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,1} & \cdots & x_{m,n} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \in \mathbb{R}^{m \times n} \tag{7}$$

$where \, x^T{}_j \in \mathbb{R}^{m \times n}$

Since we own vector for individual variables, each variable contains one specific variable samples. Then the ensuing matrix product can be thought as

$$C_X = \frac{1}{n-1} X X^T = \frac{1}{n-1} \begin{pmatrix} x_1 x_1{}^T & x_1 x_2{}^T & \cdots & x_1 x_m{}^T \\ x_2 x_1{}^T & x_2 x_2{}^T & \cdots & x_2 x_m{}^T \\ \vdots & \vdots & \ddots & \vdots \\ x_m x_1{}^T & x_2 x_m{}^T & \cdots & x_m x_m{}^T \end{pmatrix} \tag{8}$$

A closer view at the matrix elements, it is obvious that the possible covariance pairs amongst m variables have been computed. This matrix is hence termed the covariance matrix or the dispersion matrix, denoted by Σ_x.

Going back to equation (**1**)

$$Y = PX$$

Where X is the foremost data matrix,

P is the transformation matrix &

Y is the new matrix

A decision has to be agreed on the best possible method to re-express Σ_y to Σ_x and what characteristics to look out for in Σ_y.

The covariance measures how well the correlation between two variables. A key assumptions made about PCA method was independence. This means, the desired Σ_x matrix has to be un-correlated as possible and that is to say that these variables covariances in Σ_y should be near zero. Large variances are rather an interesting feature because they tell you so much about the structure. Consequent upon that, it becomes imperative to maximize inherent information (measured by the variance) by maximizing the variance and reducing the covariance between variables.

Reducing the Σ_y matrix to a minimum covariance (zero) and maximizing the variances to as much information each component can hold for, this ensues in a matrix which is diagonal. In essence, the matrix of transformation **P** will be chosen to result in the diagonalization of Σ_y.

The foregoing's implication is that the new basis $p_1, p_2 \ldots, p_m$ should be orthogonal. From linear algebra, the row vectors of an orthogonal matrix are orthonormal vectors. That is each row has a unit length and the vectors are equally perpendicular.

By this assumption, PCA seeks for a normalized direction in m-directional space along which the variance is maximized. Also, by the orthogonality condition, it circumscribes this search to directions that are perpendicular to pre-determined directions; this makes intuitive sense since geometrically the axes of Cartesian systems are usually perpendicular. This continues until m directions are selected. These **P**'s are the principal components.

The orthonormality condition lessens the challenge to a stage of being solvable by linear algebra decomposition tools. As we know, direct and effective solutions can be obtained easily by certain decompositions in linear algebra.

For the dispersion matrix Σ_y and by re-expressing **Y** with regards to **P** and **X**, we obtain,

$$\Sigma_y = \frac{1}{n-1} YY^T = \frac{1}{n-1}(PX)(PX)^T$$

$$= \frac{1}{n-1}(PX)(X^T P^T) = \frac{1}{n-1} P(XX^T)P^T$$

$$\Sigma_y = \frac{1}{n-1} PSP^T \; where \; S = XX^T \tag{9}$$

Obviously, S is going to be symmetric since $(XX^T)^T = (X^T)^T(X)^T = XX^T$

At this point, a common theorem in algebra will be introduced.

Theorem 1.0

If a matrix is orthogonally diagonalizable, then it is symmetric.

Proof

We have $Q^T A Q = D$

Multiplying Q on the left and Q^T on the right gives $A = QDQ^T \; (since \; Q^T = Q^{-1})$

Then $A^T = (QDQ^T)^T = (Q^T)^T D^T Q^T = QDQ^T = A$

So A is symmetric. Given that the converse is also true, it can be adopted to support equation (2.8).

Therefore, since the symmetric matrix $S = XX^T$ is diagonalizable by a matrix which is orthonormal of its eigen vectors, such that,

$$S = EDE^T$$

Where **D** is a matrix that is diagonal contains the eigen-values of **S**, **E** is a matrix that is orthonormal whose columns are orthogonal eigen-vectors of $\boldsymbol{S}$.

The rank r of $\boldsymbol{S}$ is the number of orthonormal eigen-vectors that it contains. If r is less than m then the constraint of orthogonality would have to be maintained by generating (m-r) orthonormal vectors to fill in the remaining spaces.

The foregoing galvanizes our choice of the conversion matrix P holding that the P rows become the eigen-vectors of $\boldsymbol{S}$, that is $= \boldsymbol{E^T}$. If substituted into equation **(9)**, $\boldsymbol{\Sigma_y}$ becomes

$$\boldsymbol{\Sigma_y} = \frac{1}{n-1}\boldsymbol{PSP^T} \qquad = \frac{1}{n-1}\boldsymbol{E^T(EDE^T)E}$$

Since a matrix that is orthonormal $\boldsymbol{E^T = E^{-1}}$, hence,

$$\boldsymbol{\Sigma_y} = \frac{1}{n-1}\,D \qquad\qquad\qquad\qquad (10)$$

In conclusion, the objective of diagonalising the covariance matrix $\boldsymbol{\Sigma_y}$ has been achieved. The principal components (the rows of P) are the eigen-vectors of the covariance matrix, $\boldsymbol{XX^T}$ and the rows are in order of importance.

2.2 Singular Value Decomposition

The SVD is intimately related to the familiar theory of diagonalising a symmetric matrix. Recall from theorem 1.0, if $\boldsymbol{A}$ is a symmetric real $n \times n$ matrix, there is an orthogonal matrix $\boldsymbol{V}$ and a diagonal $\boldsymbol{D}$ such that $\boldsymbol{A = VDV^T}$. Here the columns of $\boldsymbol{V}$ are eigenvectors for $\boldsymbol{A}$ and form an orthonormal basis for $\boldsymbol{\mathcal{R}^n}$; the diagonal entries of $\boldsymbol{D}$ are the eigenvalues of $\boldsymbol{A}$. To emphasize the connection with the SVD, we will refer to $\boldsymbol{VDV^T}$ as the eigenvalue decomposition, or EVD, for $\boldsymbol{A}$.

For the SVD we begin with an arbitrary real $m \times n$ matrix $\boldsymbol{A}$. As we shall see, there are orthogonal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ and a diagonal matrix, this time denoted $\boldsymbol{\Sigma}$, such that $\boldsymbol{A = U\Sigma V^T}$. In this case, $\boldsymbol{U}$ is $m \times m$ and $\boldsymbol{V}$ is $n \times n$, so that $\boldsymbol{\Sigma}$ is rectangular with the same dimensions as $\boldsymbol{A}$. The diagonal entries of $\boldsymbol{\Sigma}$, that is the $\boldsymbol{\Sigma_{ii} = \sigma_i}$, can be arranged to be nonnegative and in order of decreasing magnitude. The positive ones are called the singular values of $\boldsymbol{A}$. The columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ are called left and right singular vectors, for $\boldsymbol{A}$.

The analogy between the EVD for a symmetric matrix and SVD for an arbitrary matrix can be extended a little by thinking of matrices as linear transformations. For a symmetric matrix $\boldsymbol{A}$, the transformation takes $\boldsymbol{\mathcal{R}^n}$ to itself and the columns of $\boldsymbol{V}$ define an especially nice basis. When vectors are expressed relative to this basis, we see that the transformation simply dilates some components and contracts others, according to the magnitudes of the eigenvalues (with a reflection through the origin tossed in for negative eigenvalues). Moreover, the basis is orthonormal, which is the best kind of basis to have.

Now let's look at the SVD for an $m \times n$ matrix $\boldsymbol{A}$. Here the transformation takes $\boldsymbol{\mathcal{R}^n}$ to a different space, $\boldsymbol{\mathcal{R}^m}$, so it is reasonable to ask for a natural basis for each of domain and range. The columns of $\boldsymbol{V}$ and $\boldsymbol{U}$ provide these bases. When they are used to represent vectors in the domain and range of the transformation, the nature of the transformation again becomes transparent: it simply dilates some components and contracts others, according to the magnitudes of the singular values, and possibly discards components or appends zeros as needed to account for a change in dimension. From this perspective, the SVD tells us how to choose orthonormal bases so that the transformation is represented by a matrix with the simplest possible form, that is, diagonal.

2.3 The Derivation of PCA From SVD

Consider the $n \times m$ matrix, **A**, where we have a singular decomposing value, $\mathbf{A} = \mathbf{U\Sigma V^T}$. There exists a theorem from linear algebra which stipulates that the non-zero singular values of **A** are the nonzero eigenvalues square roots of $\mathbf{AA^T}$ or $\mathbf{A^TA}$. The first assertion for the $\mathbf{A^TA}$ case is proven in the succeeding way:

$$\boldsymbol{A^TA = (U\Sigma V^T)^T(U\Sigma V^T) = (V\Sigma^T U^T)(U\Sigma V^T) = V(\Sigma^T\Sigma)V^T} \qquad (11)$$

We note that $\mathbf{A^TA}$ is same as $\mathbf{\Sigma^T\Sigma}$, and hence it has similar eigenvalues. Since $\mathbf{\Sigma^T\Sigma}$ is a $(m \times m)$ square, diagonal matrix, the eigenvalues are hence the entries done diagonally, which are the singular values squares. Note that the non-zero eigenvalues of individual covariance matrices, $\mathbf{AA^T}$ and $\mathbf{A^TA}$ are very identical.

It is imperative to note that an eigenvalue matrix decomposition, $\mathbf{A^TA}$ has been carried out. Certainly, since $\mathbf{A^TA}$ is symmetric which is an orthogonal diagonalisation and hence the eigenvectors of $\mathbf{A^TA}$ are the **V** columns. This is pertinent in making the practical linking between the PCA and SVD of the **X** matrix, which comes up next.

Recalling the initial $m \times n$ data matrix, **X**, let us describe a new $n \times m$ matrix, **Z**:

$$Z = \frac{1}{\sqrt{n-1}} X^{\square} \tag{12}$$

Recalling that since the rows of **X** on m contained the samples of n data, we removed the average of the row from individual entries to ensure no mean existed across the rows. Hence, the new **Z** matrix, has columns without mean. Consider establishing the $m \times m$ matrix, $\mathbf{Z^T Z}$:

$$Z^T Z = \left(\frac{1}{\sqrt{n-1}} X^T\right)^T \left(\frac{1}{\sqrt{n-1}} X^T\right)$$

$$= \frac{1}{n-1} X X^T \tag{14}$$

$$i.e.\ \mathbf{Z^T Z = \Sigma_x}$$

We observe that describing **Z** in this pattern ensures that $\mathbf{Z^T Z}$ is similar to the **X**, $\mathbf{\Sigma_x}$. covariance matrix. From the conversation in the last section, the main components of **X** (subject to be identified) are the $\mathbf{\Sigma_x}$ eigenvectors. Thus, if we carry out an individual value matrix decomposition of $\mathbf{Z^T Z}$, the main components will be the orthogonal matrix, V columns.

The final step is to link the SVD of $\mathbf{Z^T Z}$ back to the change of basis signified by equation **(1)**:   **Y = PX**

We wish to project the original data onto the directions described by the principal components. Since we have the relation $\mathbf{V = P^T}$, this is simply:

$$\mathbf{Y = V^T X}$$

Having the relative $\mathbf{V = P^T}$, this is basically:

$$\mathbf{Y = V^T X}$$

If recovering the original data is what we wish to do, we basically compute (employing orthogonality of **V**):

$$\mathbf{X = VY}$$

3. Analysis, Results and Discussion

Principal Component Analysis

Table 1.0. Data representing Domestic Debt of State Governments for the thirty-six (36) states of Nigeria and its capital (Abuja) as shown in appendix A. The table is analysed using the minitab computer package to obtain the eigen analysis as shown in **Table 1.1** below.

Table 1.1. An eigen analysis showing the eigenvalues and the proportion of variation explained by each component

Eigen analysis of the Covariance Matrix

Eigenvalue	16097419729	1565058147	540198305	301840604	134415535	88759329
Proportion	0.860	0.084	0.029	0.016	0.007	0.005
Cumulative	0.860	0.943	0.972	0.988	0.995	1.000

The proportion of each eigenvalue is given in the second row and it shows how much information is being captured by the eigenvalues. For example, the first eigenvalue has a relatively high proportion of **0.860** which is about **86%** of variability explained while the following eigenvalue contributes **8.4%** and etc. These proportions are obtained by dividing each eigenvalue by the sum total of all eigen values.

The cumulative proportion which occupies the third column is obtained by adding the successive proportions of variation explained to obtain the running total. The cumulative proportion provides an easy way of selecting principal components based on the percentage of variation explained by these components. It suffices to choose the first and the second components for this study since they both explain **94.3%** of information in the dataset.

Table 1.2. A table of eigenvectors and corresponding variables

Eigenvectors

Variable	PC1	PC2	PC3	PC4	PC5	PC6
2011	0.268	**0.421**	0.374	-0.166	**0.476**	**-0.597**
2012	0.371	**0.622**	0.308	0.141	**-0.440**	**0.408**
2013	0.391	0.189	**-0.749**	-0.066	-0.309	-0.389
2014	**0.433**	-0.073	-0.192	**-0.558**	**0.432**	**0.522**
2015	**0.459**	**-0.578**	**0.406**	-0.244	**-0.428**	-0.220
2016	**0.490**	-0.246	-0.051	**0.760**	0.339	0.072

The principal components are the linear combinations of the original variables that account for the variance in the data. The maximum number of components extracted always equals the number of variables. In essence, the coefficients indicate the relative weight of each variable to each component.

From Table 1.2, PC1 has relatively high values for 2014, 2015 and 2016 (**0.433, 0.459** and **0.490** respectively). High correlation values are in bold numbers in the table above. These values imply presence of correlations between these variables and the first principal component. That is, an increase in the first principal component will result to an increase in these variables. On the other hand, the second principal component has negative values for 2015(**-0.578**) which would mean a negative relationship between that year and the second principal component.

Finally, a high correlation between 2012 (**0.622**) and the second principal component indicating an increase in the second principal component will result to an increase in debt values in 2012 but a decrease in 2015.

Table 1.3. A table of scores for each observation with respect to PC1 and PC2

States	Abia	Adamawa	Akwa-Ibo	Anambra	Bauchi	Bayelsa	Benue	Borno	Cross-Rive	Delta	Ebonyi
PC1 Score	74579.45	85755.63	279524.1	13043.96	91320.57	309073.5	80403.42	53919.28	265624.8	452328.7	59102.87
PC2 score	-12813.1	-15454	-20781.4	8905.222	-29985.4	119114.6	-13110.2	-1725.59	10514.06	-150572	10260.96

States	Edo	Ekiti	Enugu	Gombe	Imo	Jigawa*	Kaduna	Kano	Katsina*	Kebbi	Kogi
PC1 Score	113056	108728.9	64806.67	85076.19	109073.1	21997.94	82723.01	105661.3	17030.07	50159.6	76184.86
PC2 score	23507.68	-14710.9	-17611	-17823.4	-43049.4	-15348.6	-14880.4	-50723.2	-10472.9	-38295.2	-17579.3

States	Kwara	Lagos	Nassarawa	Niger	Ogun***	Ondo	Osun	Oyo	Plateau	Rivers**	Sokoto
PC1 Score	69506.55	605693.9	77806.15	56503.17	150267.8	85179.1	186743.4	97207.88	167171.4	274748.4	24324.84
PC2 score	3913.91	39732.72	-28381.7	881.1676	-15530.4	19029.3	-88442.1	-43939.7	-54783.3	-9255.76	-7808.22

States	Taraba	Yobe	Zamfara	FCT
PC1 Score	54390.33	11613.4	74839.52	285839.8
PC2 score	-6013.1	-2121.35	-21467.2	6051.316

Scores are obtained for each observation by a linear combination of the coefficients of each principal component from the data. That is, eigenvectors, with coefficients corresponding to each variable, are used to calculate the principal component scores for each observation. For example, the score for the first observation (Abia state) on PC1 can be given as,

PC1= 0.268(2011) + 0.371(2012) + 0.391 (2013) + 0.433(2014) + 0.459(2015) + 0.490(2016)

= 0.268(24202.24) + 0.371(8663.79) + 0.391(31736.72) + 0.433(25126) + 0.459(33530.53) + 0.490(53525.31)

= 74607.03

From **Table 1.3**, the PC1 score value for Abia state can be approximately equated to the calculated score (i.e. **74579.5~74607.03**). The values are not exactly equal due to the aggregations of approximations made by the software, and large values associated with the data. Nonetheless, these disparities do not affect the interpretations.

Table 1.4. Summary of Scores from PC1 and PC2

	Scores	States
Max Value for PC1	605694	Lagos
Max Value for PC2	119115/-150572	Bayelsa/Delta
Min Value for PC1	11613.4	Yobe
Min Value for PC2	881	Niger

Scores can be loosely defined as a function of the performance of each state in a particular principal component. From **Table 1.4**, by examining the magnitude of these values, Lagos state had the highest score (605694). This is plausible considering its robust economic activities with the capacity to attract high debts just like United States of America. While on PC2, Bayelsa and Delta states had high magnitudes but opposite sign which means a negative relationship in their values.

The minimum scores for PC1 and PC2 were associated with Niger and Yobe states. However, general high and low values are obtained from the first component since it accounts for the most variation in the data. So even if the least score (in terms of magnitude and regardless of sign) is associated with Niger state (881), Yobe state, with a score of 11613.4 on the first principal component, is the state with the least outstanding debt.
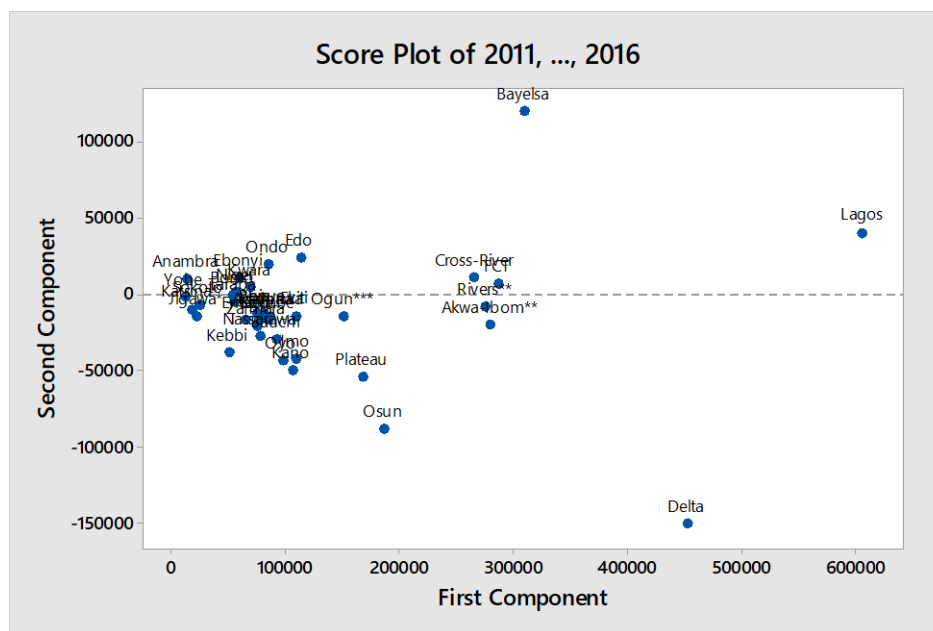

Figure 1.0. A score plot for the individual observations

To corroborate the understanding of the score table, the score plot pictorially buttresses the results of the scores by associating observations located at the extreme of both axes to high and low outstanding debt values of each state.

That is, the outstanding debt of states located closer to these variables on the PC2 axis of the plot score will be very high. For example, Bayelsa and Delta states can be viewed at the second component axis extreme while Niger state is the nearest to its origin though not so apparent owing to the overlapping of points.

Alternatively, the first component axis has Lagos state at its extreme, that is, Lagos state has the biggest outstanding debt value on this component. This can be generalized for the total dataset just same as the interpretation above. However, Yobe state being nearer to zero can be commonly attributed to the state with the least outstanding debt value.

4. Conclusion

One of the major conclusions that can be drawn from this work is that the PCA method is an offshoot of the SVD method. Both methods can be used to achieve the same purpose though the latter has a more robust application. PCA is applied by obtaining the eigenvectors and eigenvalues of the covariance matrix whereas SVD is applied on the matrix of observations by decomposing the data matrix into three factors that contain the eigenvectors and eigenvalues in the PCA method. The study demonstrated the power of the PCA tool by carrying out an exploratory data analysis on a financial data. Even when the correlation matrix is quite difficult to analyze for such a multidimensional dataset, the PCA method was able to decompose the data into simpler components that could be easily interpreted. More so, deep inherent structures were uncovered in terms of relationships that exist between and within data structure.

References

Deshmunkh, S. R., & Purohit, S. G. (2007). Microarray Data: Statistical Analysis Using R. New York: Alpha Science International Ltd.

Jolliffe, I. T. (2002). *Principal Component Analysis*. (2nd. Ed) New York: Springer.

Khan, J. K., Wei, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., … & Meltzer, P. S. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Publishing Group, 7*, 673-679

Least Squares problems. (2013). Retrieved from https://www.google.com/search?client=firefox-b-

Lim, D. H. (2013). Principal Component Analysis using Singular value Decomposition of Microarray Data. *International Journal of Mathematical, Computational, physical, Electrical and Computer Engineering*, 7(9).

Lyche, T. (2018).Singular value decomposition and least squares problems. Retrieved from https://www.uio.no/studier/emner/matnat/ifi/nedlagte-emner/INF

Will, T. (1999). Introduction to the singular value decomposition. Retrieved from https://hadrienj.github.io/posts/Deep-Learning-Book-Series-2.8-Singular-Value-Decomposition/

Appendix A

DOMESTIC DEBT OF STATE GOVERNMENTS

State	2011	2012	2013	2014	2015	2016
Abia	24,202.24	8,663.79	31,736.72	25,126.07	33,530.53	53,525.31
Adamaw	25,954.20	24,284.06	15,976.52	26,443.26	47,201.62	62,157.54
Akwa-Ib	41,253.91	108,889.39	125,037.04	81,756.01	147,575.74	155,431.51
Anambr	6,403.32	14,299.99	3,025.80	2,876.18	3,575.77	3,993.89
Bauchi	18,345.73	18,807.27	16,825.51	27,999.81	57,652.77	69,988.36
Bayelsa	162,822.65	222,401.77	69,513.13	91,681.86	103,374.23	140,177.08
Benue	16,631.14	24,402.44	24,987.87	17,772.06	39,944.21	63,526.71
Borno	1,684.56	24,423.20	23,943.15	22,302.79	22,338.73	30,929.43
Cross-Ri	90,750.05	90,872.91	116,061.63	107,342.90	115,522.25	128,142.09
Delta	90,843.57	83,684.01	102,100.20	211,953.21	320,605.71	241,231.44
Ebonyi	40,239.94	28,895.75	13,236.09	6,954.98	34,168.94	28,057.14
Edo	39,044.30	62,274.74	48,190.15	40,050.00	46,289.08	45,091.95
Ekiti	23,667.51	39,587.70	22,376.37	30,460.63	52,564.98	85,049.68
Enugu	10,887.17	17,354.19	12,061.40	22,625.69	37,550.23	48,417.54
Gombe	7,170.42	30,243.54	27,992.84	29,591.44	53,454.40	48,312.23
Imo	25,419.40	16,700.73	12,633.53	28,946.45	71,743.51	93,267.76
Jigawa*	1,590.54	2,081.43	1,612.29	1,569.94	22,194.83	19,005.55
Kaduna	34,771.71	22,855.93	9,831.84	16,683.75	49,847.91	63,276.47
Kano	5,867.29	5,867.29	32,207.01	31,423.63	65,007.33	93,715.18
Katsina*	2,059.88	918.93	269.65	586.70	11,495.03	21,449.61
Kebbi	7,291.05	2,716.01	853.68	17,271.45	63,793.34	20,650.99
Kogi	34,122.12	14,979.19	7,109.87	10,304.74	42,034.63	71,381.26
Kwara	25,254.47	29,776.56	22,416.65	22,147.54	31,966.82	38,136.72
Lagos	157,536.16	230,432.88	278,867.07	268,065.02	218,538.87	311,755.80
Nassara	5,336.06	7,096.14	28,848.54	34,525.70	40,557.05	59,033.75
Niger	16,975.51	17,802.50	24,731.75	23,454.54	21,501.79	31,984.09
Ogun***	30,143.97	45,726.56	58,382.00	70,193.52	75,921.43	75,921.43
Ondo	48,369.86	36,518.09	30,883.18	19,267.66	26,647.79	53,159.72
Osun	5,463.64	38,600.00	41,400.00	37,820.83	144,699.56	147,069.97
Oyo	4,808.39	11,726.21	19,106.05	12,912.64	47,437.01	115,886.55
Plateau	20,908.12	24,117.32	52,416.33	78,415.07	96,204.85	110,340.67
Rivers**	83,978.39	81,459.19	129,549.65	91,757.57	134,966.60	142,424.09
Sokoto	4,902.05	2,997.31	5,739.57	7,650.12	11,658.21	22,450.25
Taraba	17,974.66	16,701.02	13,883.98	14,395.30	27,646.23	38,868.70
Yobe	2,088.40	3,991.22	1,122.64	1,638.44	3,867.46	13,581.30
Zamfara	12,968.38	15,508.11	28,217.65	11,072.04	46,280.69	58,321.02
FCT	85,563.89	123,992.77	84,324.10	110,139.17	133,900.29	152,804.61

Source: Debt Management Office via Statistical Bulletin

Copyrights