Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

An Exploration of Principal Component Analysis using Module Theory

Istiqomah Suharto[1, *], Sisilia Sylviani[2]

[1, 2] *Universitas Padjadjaran, Jl. Soekarno Km 21 Jatinangor, Kab. Sumedang, West Java 45363, Indonesia*

**Corresponding author: istiqomah22002@mail.unpad.ac.id*

Copyright: ©2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License-(CC-BY-SA) (https://creativecommons.org/licenses/by-sa/4.0/)

Article Info

Article history:
Received: August 16, 2025
Revised: September 30, 2025
Accepted: October 1, 2025
Available online: October 22, 2025

DOI:
https://doi.org/10.71385/idjams.v2i2.25

Abstract

With big data becoming more and more prominent in the world, high-dimensional data is a relevant issue due to the challenges that it imposes to meaningful analysis. The exponential growth in space leads to data becoming sparse thus making it difficult to analyze underlying patterns/relationships. This is where dimension-reducing techniques come in, the most popular one being Principal Component Analysis. Hence, it is important to analyze exactly what makes PCA work and to study it and generalize it in order to leave room for more variants for differing fields and applications to grow. This paper examines PCA from a linear algebra perspective, particularly using module theory. We prove that PCA is an $\mathbb{R}$ −module homomorphism, and, when all principal components are kept, an $\mathbb{R}$ −module automorphism, meaning that it preserves structure and is invertible. We then look into what happens algebraically when only a subset of principal components are kept. The $\mathbb{R}$ −module homomorphism is then only an $\mathbb{R}$ −module epimorphism, not an isomorphism, still structure preserving but not invertible. Through these findings, we find that there are three essential, algebraic properties of Principal Component Analysis, namely (1) the transformation must be linear, (2) it must project the data onto a new orthonormal basis, and (3) it must diagonalize the covariance (or correlation) matrix of the centered dataset. With these properties, we get an algebraic definition of PCA: an $\mathbb{R}$ −module automorphism that diagonalizes the covariance structure of the original dataset via an orthogonal change of basis.

Keywords: Module Theory; Principal Component Analysis; R-module Automorphism; Data Science; Statistics; Dimension-reducing Techniques

I. Introduction

With big data becoming more and more prominent in the world, high-dimensional data is a relevant issue, especially in machine learning and analytics. Paik and Park [1] talk about “The curse of dimensionality” in their 2023 paper ‘Circular Coordinates for Density-Robust Analysis’ as a concept that shows the challenges that high-dimensional data presents to meaningful analysis and interpretation. The “Curse of Dimensionality” was first introduced by Richard Bellman in his 1959 paper “A Mathematical theory of adaptive control processes” [2] to state that as the number of dimensions in a dataset increases, the volume of the space grows exponentially. This exponential growth in space leads to the data points becoming farther and farther apart (sparse) thus making it difficult to detect underlying patterns and/or relationships. This means that high-dimensional data hinders effective model training since it leads to overfitting and poor generalization capabilities [1]. It is then of interest to reduce the dimensions on these high-dimensional datasets using dimension-reduction techniques in order to better analyze and visualize high-dimensional data [1].

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

There are various dimension-reducing techniques, but the most widely used is Principal Component Analysis [3]. Ian Jolliffe describes Principal Component Analysis (PCA) as an analysis technique for high-dimensional data that summarizes vital information using principal components. Essentially, PCA reduces datasets with multiple intercorrelated variables by transforming them into new variables called Principal components which are uncorrelated. From trying to understand your data with, say, 100 variables, PCA can summarize it in less than 5, successfully reducing 100 dimensions to only 5 dimensions while still preserving important information.

Principal Component Analysis is used in multiple and varying fields but has its roots in the field of data science, which is why existing literature generally describes PCA in statistical or geometric terms. For example, Abdi & Williams (2010) explains PCA as a method of multivariate analysis that displays patterns of the observations in the data table as points in maps [4]. However, they do not explain PCA's role as an algebraic structure. Little work has framed PCA within module theory or explained if the map of PCA is invertible or surjective. Hence, there is work to be done in finding an algebraic definition of PCA and understanding how exactly PCA fits into the language of modules and homomorphisms. This paper seeks to analyze it from an algebraic perspective using module theory. By casting PCA in module-theoretic language, we formalize and generalize its definition. We will look at the conditions of the PCA, provide an algebraic characterization of PCA to reveal the minimal algebraic framework (linearity, orthonormal basis, covariance diagonalization) on which PCA functions.

This paper seeks to explore the mathematical foundation of PCA, re-defining it in algebraic terms and generalizing it, to offer a deeper understanding that compliments and goes beyond the current text.

II. Literature Review

Principal Component Analysis was first introduced by Karl Pearson [5]. Pearson explored how to find the line/plane that best fits a set of data in multidimensional space, using the least squares method to determine the equation of the line/plane. This line/plane then becomes the basis for the principal component.

Harold Hotelling took this idea and expanded upon it in 1933 [6]. His crucial advancement was introducing the calculation of eigenvalues and eigenvectors from the data's covariance matrix. The eigenvectors revealed the direction of the principal components while the eigenvalues quantified their relative importance. This generalization transformed PCA. It was no longer limited to reducing data into just 2D or 3D (as Pearson's method did), Hotelling's approach could summarize multidimensional data more flexibly, reveal all principal components (not just the first few), and precisely rank their significance through eigenvalues.

PCA now finds applications in many fields such as finance, criminology, and pharmaceuticals to name a few. Lin et al. explored the financial world using Principal Component Analysis [9]. They used PCA to reveal structures and interactions of the US and Chinese stock market during crises between the years 1992 and 2012 by applying the technique to cross-correlation matrices derived from multiscale multifractal detrended cross correlation analysis (MSMF-DXA). They found 2 principal components that explained the most variance in market behavior and led them to the following findings. Firstly, Markets in the same country or region move more similarly than those that aren't (as shown by US markets and Chinese markets clustered tightly in PCA plots). Secondly, financial crises led global markets to become more similar, regardless of region (as shown by even tighter clustering between US and Chinese markets on the PCA plot during crises). Lastly, Lin et al. found that while the US market dominated in normal times, China overtook them during crises, showing China's influence is not only strong but stronger especially during turmoil [9]. In the health/pharmaceutical field, Yoo & Shahlaei (2018) apply PCA to Quantitative Structure-Activity Relationship, or QSAR, to analyze a dataset of CCR5 antagonists (drug candidates targeting HIV) [10]. From inputting a matrix of 10 molecular descriptors such as atomic weight and polarizability for 42 CCR5 antagonist compounds, they yielded 3 PCs that accounted for about 85% of variance, namely molecular size, electronic properties, and hydrogen-bond acceptors and halogen atoms, effectively summarizing the once intercorrelated complex dataset [10]. Misztal, in 2019, acknowledged PCA's potential in criminology, using it to analyze

Poland's crime rates across 16 different regions from 2000 to 2017 [11]. Misztal separated spatial vs. temporal PCA (to look at regional vs yearly effects). In spatial PCA, he averaged crime rates across years to compare regions, and in temporal PCA, it was the other way around where he averaged crime rates across regions to compare years. These methods found that there was high crime in urbanized regions, mostly property crimes, and in rural regions the lowest crime due to minimal property/violent crimes. Temporal trends also highlighted the impact of implementing policies, since there was a reduction in property crimes in 2013, right when there was a law reclassifying misdemeanor [11]. These results are proof that PCA is beneficial for efficient policymaking as it can guide optimal resource allocation.

In 2016, however, Jolliffe and Cadima state that PCA may not be suitable for all data types or research questions [7]. There are a few things that one should consider before choosing to use PCA. The first is if your data is continuously numeric or not. Jolliffe and Cadima state that PCA requires numerical inputs since categorical data can pose challenges for traditional/classic PCA [7]. There are techniques to convert categories into numbers, but alternative methods such as Multiple Correspondence Analysis (MCA) may be more appropriate for your dataset if it does not consist of continuously numerical data [8]. Researchers must also ask themselves if their data is intercorrelated or not. As stated above, the point of PCA is to reduce high dimensional datasets consisting of many intercorrelated variables and transform them to a new set of uncorrelated PCs (Principal Components). Datasets with weakly correlated variables yield trivial PCA results. If this is the case for your research, the alternative method of Factor Analysis can be considered instead. Additionally, datasets must be large enough (must need dimensional reduction) to begin with. If the number of variables in a dataset is less than 5, you may want to consider another method since there is no necessity to summarize this data. Dimensionality reducing techniques were developed to deal with high-dimensional data where excessive variables blur interpretations [7].

Table 1. Table of summaries of a few notable previous works on Principal Component Analysis

Author(s)	Year	Findings
Pearson [5]	1901	Introduced the very basic concept of PCA by finding the line/plane that best fits a set of data in multidimensional space
Hotelling [6]	1933	Expanded on the basic concept of PCA introduced by Pearson by adding eigenvalues/eigenvectors from the covariance matrix that explained variance.
Lin et al. [9]	2014	Applied PCA as a statistical method in Finance to cross-corerlate matrices of US/China stock markets.
Yoo & Shahlaei [10]	2018	Applied PCA as a statistical method in Pharmaceuticals to reduce 10 moleculor descriptors into 3 PCs which explained 85% variance
Misztal [11]	2019	Applied PCA as a statistical method in Criminology, particulary to Polish crime data, to find that temporal PCA showed policy impacts.
Jolliffe & Cadima [7]	2016	Reviewed PCA's limitations: requires continuous numeric data, strong intercorrelation, and high dimensions (multiple variables).

From the text above, it can be properly concluded that previous works on PCA have developed this statistical method conceptually (Pearson, Hotelling) and applied it across differing fields, using it as a statistical/black-box tool. This paper aims to provide an algebraic reframing of PCA via module theory. We will explore the minimal defining properties of PCA, clarify its morphism type (which previous work has yet to address), and provide a more axiomatic definition of PCA future researchers can generalize and compare with other statistical methods.

A. Principal Component Analysis Methodology

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

There are many variants of PCA. In this paper, I will only be talking and analyzing traditional/statistical sample PCA using the covariance matrix and traditional/statistical sample PCA using the correlation matrix. This variant of PCA is the variant that all the other variants have stemmed from. My hope is that by analyzing this variant, readers can understand the basic gist of PCA and in turn understand other variants of PCA.

1) Principal Component Analysis Using Sample Covariance Matrix

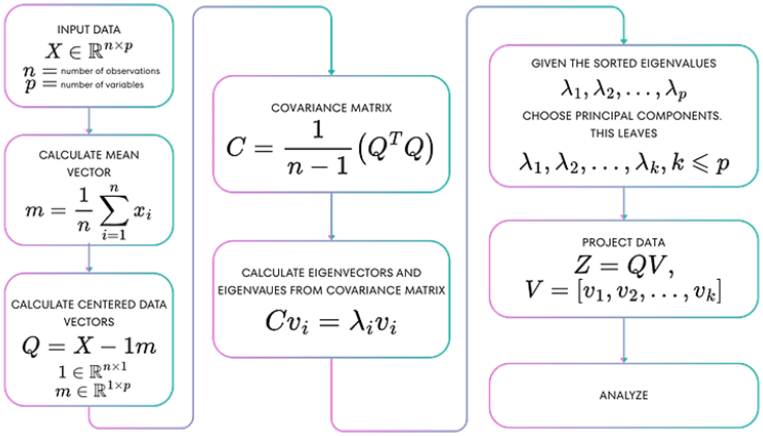


Figure 1. Flowchart of the Process of PCA via Covariance Matrix

The general flowchart for Principal Component Analysis Using Sample Covariance Matrix is shown in figure 1. The first step is to input data and process it by centering. Then, we construct a covariance matrix representation of the data. After this, eigenvalues and eigenvectors are calculated from the covariance matrix. The eigenvectors indicate the direction of maximum remaining variance in the data and forms a new basis. The eigenvalues, on the other hand, indicate exactly how much variance each component explains in the dataset. The larger the eigenvalue, the more "important" or dominant that component is in explaining the variation in data. The last step is to then project this new summary of the dataset using Z (the PCA scores matrix). Z is essentially the original data, rewritten using the principal components. Each column of Z represents one principal component. In other words, Z is what you used PCA for; it is the original data, summarized.

2) Principal Component Analysis Using Sample Correlation Matrix

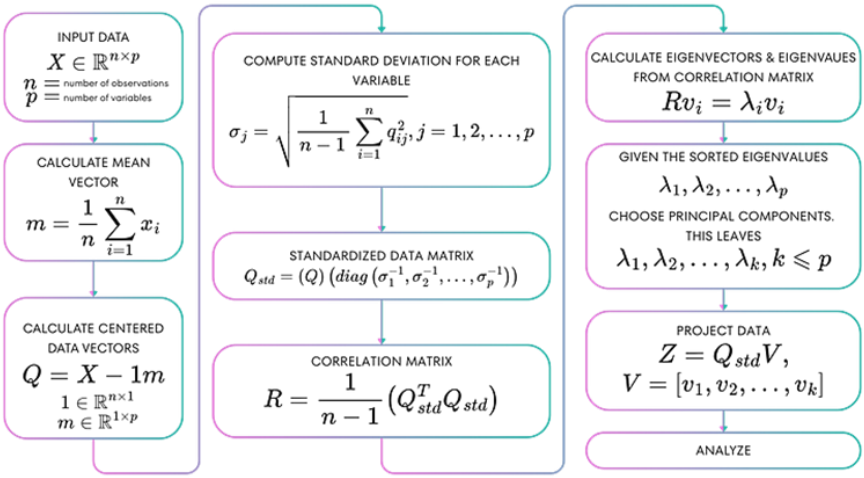


Figure 2. Flowchart of the Process of PCA via Correlation Matrix

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

The process of Principal Component Analysis using the sample correlation matrix follows the same general procedure, except we first center, and standardize our original dataset. Then, using the same formula as PCA via Covariance matrix, we calculate correlation matrix representation of the data, in which each element is a Pearson correlation between variables (-1 to 1). From this point onward, the PCA procedure (eigen decomposition, component selection, etc.) process proceeds identically to the covariance matrix approach.

3) Covariance V.S. Correlation

PCA using a covariance matrix maximizes variance captured by the principal components. However, a problem can arise. If your variables are in different units, interpreting the results can be challenging. More critically, changing measurement units (e.g., kilograms to pounds) will alter the PCA results. The solution to this problem is solved when using the correlation matrix. In this variant, all variables have been standardized to have zero mean and unit variance. Covariance-based PCA is appropriate when all variables have comparable scales, but even then components may become dominated by variables with larger variances, even when these variables aren't necessarily more meaningful. Let's take an example of variables with the same exact units, say USD, if one variable ranges from 0 to 1 USD and another from 0–1,000,000, covariance-based PCA will weight them by their raw variances, even if this disproportionate scaling doesn't reflect their analytical importance. Such analyses are not very informative as they primarily reflect scale differences rather than meaningful underlying patterns. While covariance-based PCA has specific applications, correlation-based PCA is generally preferred. As Jolliffe notes, correlation-based PCA is more common in practice [4]. To reiterate, covariance-based PCA is appropriate when all variables are measured in similar units and their variance reflects their relative importance. If, however, all variables are measured in differing units and their variances don't reflect importance, correlation-based PCA is more appropriate [4].

4) Post-Processing

As mentioned, the result of putting an intercorrelated, high-dimensional dataset through PCA is a set of PCs which are uncorrelated. The question then becomes how many PCs do we choose to summarize/represent our data? And how can we interpret these results?

In the flowcharts above, given p eigenvalues, choose k principal components accounting for most variance in the data. But, what does "most" mean? For each variant of PCA, the process of choosing principal components is different. In this article, I have talked about Traditional/Classic Sample PCA using Covariance and Correlation matrix, so we will be going over the two ways to choose principal components for these 2 specific variants. Again, understanding the process of PCA of these two fundamental variants can give insight into how the process is with other variants. The first and most simple procedure is to set a threshold and stop when the first k components make up for a larger percentage of the overall variation. This threshold can typically range from 70–90% depending on the context of the dataset. The other procedure is based on the values of the eigenvalues. The most popular of this kind of procedure is Kaiser's rule, in which you only keep PCs with eigenvalues that have a value more than 1. Kaiser's logic here is that you should only keep PCs that explain more variance than a single original variable.

III. Preliminary

Definition of Vector Space [14]

Let V be an arbitrary nonempty set of objects on which two operations are defined: addition, and multiplication by numbers called scalars. Addition means a rule for associating with each pair of objects u and v in V an object $u + v$, called the sum of u and v. Scalar multiplication means a rule for associating with each scalar k and each object u in V an object ku. If the following axioms are satisfied by all objects $u, v, w \in V$ and all scalars k_1 and k_2, then we call V a vector space and we call the objects in V vectors.

1. $u, v \in V \rightarrow u + v \in V$

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

2. $u + v = v + u$
3. $u + (v + w) = (u + v) + w$
4. $\exists 0_v \in V \ni 0_v + u = u + 0_v = u$
5. $\forall u \in V, \exists(-u) \in V \ni u + (-u) = (-u) + u = 0_v$
6. $\forall u \in V$, scalar $k_1, k_1 u \in V$
7. $k(u + v) = ku + kv$
8. $(k + m)u = ku + mu$
9. $k(mu) = (km)u$
10. $1u = u$.

Definition of Orthonormal Set [14]

A set of two or more vectors in a real inner product space is said to be orthogonal if all pairs of distinct vectors in the set are orthogonal. An orthogonal set in which each vector has norm 1 is said to be orthonormal.

In other words, a set of vectors $V = \{v_1, v_2, \dots, v_n\}$ is said to be an orthogonal set if $\forall v_i, v_j \in V, i \neq j$

$$v_i \cdot v_j = 0$$

Further, the orthogonal set is said to be orthonormal if $\forall v_i \in V, \|v_i\| = 1$

Definition of Basis of a Matrix [15]

A set of vectors $B = \{b_1, b, \dots, b_n\}$ in a vector space V is called a basis for V, if

1. $b_1, b, \dots, b_n$ are linearly independent
2. Span$\{B\} = V$

Definition of Orthonormal Basis [14]

In an inner product space, a basis consisting of orthonormal vectors is called an orthonormal basis. A set of vectors $B = \{b_1, b, \dots, b_n\}$ in a vector space V is called an orthonormal basis for V, if

1. $b_1, b, \dots, b_n$ are linearly independent
2. Span$\{B\} = V$
3. $\forall b_i, b_j \in V, i \neq j, b_i \cdot b_j = 0$

Definition of Eigenvalues and Eigenvectors [14]

If A is an $n \times n$ matrix, then a nonzero vector x in R^n is called an eigenvector of A if Ax is a scalar multiple of x; that is,

$$Ax = \lambda x$$

For some scalar λ, which is the eigenvalue of A. x is then the eigenvector corresponding to λ.

Definition of Linear Transformation [15]

Let V_1 and V_2 be vector spaces over the field F. $T: V_1 \to V_2$ is a linear transformation if the following are satisfied:

1. $T(x + y) = T(x) + T(y)$
2. $T(kx) = kT(x), \forall x, y \in V_1, k \in F$

Let $dim(V_1) = n, dim(V_2) = m$. The linear transformation is represented by A an $m \times n$ matrix, that $T(x) = T_A(x) = A(x)$.

Theorem 1 [2]

The inverse of an orthogonal matrix is its transpose

Let A be an $m \times n$ orthogonal matrix where a_i is the i^{th} column vector. The ij^{th} element of $A^T A$ is

$$(A^T A)_{ij} = a_i^T a_j = \begin{cases} 1, if\ i = j \\ 0\ otherwise \end{cases}$$

Therefore, because $A^T A = I$, it follows that $A^{-1} = A^T$

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

Theorem 2 [12]
For any matrix A, A^T, and AA^T are symmetric

$$(AA^T)^T = A^{TT}A^T = AA^T$$
$$(A^TA)^T = A^TA^{TT} = A^TA$$

Theorem 3 [12]
A matrix is symmetric if and only if it is orthogonally diagonalizable
[⇐] If A is orthogonally diagonalizable, then A is a symmetric matrix.

By hypothesis, orthogonally diagonalizable means $\exists E \ni A = EDE^T$, where D is a diagonal matrix and E is a matrix which diagonalizes A. Let us prove that $A^T = A$

$$A^T = (EDE^T)^T = E^{TT}D^TE^T = EDE^T = A$$

Evidently, if A is orthogonally diagonalizable, it must also be symmetric.

Theorem 4 [12]
A symmetric matrix is diagonalized by a matrix of its orthonormal eigenvectors.

Theorem 5: Rank Nullity Theorem [15]
Suppose that $T = T_A: F^n \to F^m$ is a linear transformation where A is an $m \times n$ matrix. Then

$$rank(T) + null(T) = dim(F^n)$$

Where

$$null(T) = dim\big(Ker(T)\big)$$

And

$$rank(T) = dim\big(Im(T)\big)$$

Definition of an R-module [13]
Let R be a unitary ring and let M be an additive abelian group. With the left action

$$*: R \times M \to M$$

Defined by

$$(r, m) \to * (r, m) = r * m$$

M is an R −module if the following are satisfied

1. $r_1 * (m_1 + m_2) = r_1 * m_1 + r_1 * m_2$
2. $(r_1 + r_2) * m_1 = r_1 * m_1 + r_2 * m_1$
3. $(r_1 \cdot r_2) * m_1 = r_1 * (r_2 * m_1)$
4. $1_R * m_1 = m_1$

$$\forall r_1, r_2 \in R, m_1, m_2 \in M$$

Definition of R-Module Homomorphism [13]
Let M_1 and M_2 be R −modules over the ring R. $f: M_1 \to M_2$ is an ***R* −module homomorphism, or an *R* −morphism**, if the following are satisfied:

1. $f(x + y) = f(x) + f(y)$
2. $f(kx) = kf(x), \forall x, y \in M_1, k \in R$

An R −morphism f is called an ***R* −monomorphism** if it is injective (or one-to-one). In other words, if $\forall x, y \in M_1, f(x) = f(y)$ then $x = y$.
An R −morphism f is called an ***R* −epimorphism** if it is surjective (or onto). In other words, if $\forall\, z \in M_2, \exists x \in M_1 \ni f(x) = z$.
An R −morphism f is called an ***R* −isomorphism** if it is bijective. In other words, if it is injective and surjective (or one-to-one and onto).
An R −morphism f is called an ***R* −automorphism** if it is bijective and $M_1 = M_2$(it maps to itself).

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

IV. Results and Discussion

A. PCA as a Linear Transformation

In 2014, Shlens described Principal Component Analysis as a change of basis, uncovering the algebraic intuitions behind this dimension-reductionality technique [12]. Shlens's paper, however, discusses a variant of PCA that is more common in signal processing and machine learning, which differs slightly from the traditional statistical version discussed in this paper. Both variants are mathematically equivalent, which means they both have similar mathematical processes and a similar definition of PCA as a linear transformation. Our findings only differ in matrix orientation—and later on, when discussing PCA via correlation matrix as a linear transformation, further distinctions will appear.

This section first outlines Shlens's version of PCA as a linear transformation (used in machine learning) before presenting this paper's variant (the one common in traditional statistics) of PCA as a linear transformation. This is important since the approach in this paper builds directly on Shlens's explanation.

Shlens's explanation of PCA as a linear transformation goes as follows:
Let $X = \mathbb{R}^{m\times n}$ be the original data matrix. Where n =number of observations, and m = number of variables.
Let $Q = \mathbb{R}^{m\times n}$ then be the centered data matrix, which is obtained by subtracting the mean of each row from that row.

Shlens states that PCA is simply re-expressing the original data with a new coordinate system such that the first axis explains the most variance in the data (this is principal component 1), the second axis explains the second most variance in the data (this is principal component 2), and so on. This makes it so the new basis of the original data better captures the patterns and structure of the data. Most importantly, in this new coordinate system, the principal components are uncorrelated. Translating this idea into linear algebra, the axes of the new basis should also be uncorrelated, meaning they form an orthonormal basis (each new axis is perpendicular to the other and has a magnitude of 1).

Let the new orthonormal basis be defined as $P \in \mathbb{R}^{m\times m}$.
PCA as a linear transformation represented by a matrix is then

$$Y = K^T Q \tag{1}$$

where $Y \in \mathbb{R}^{m\times n}$ is the transformed data (the original data expressed in terms of the new basis). The question now becomes "How do we define K?". Let us pay attention to the covariance matrix in this variant, as it tells us how much each variable varies with another variable.
The covariance matrix of the centered data matrix is as below.

$$C_Q = \frac{1}{n}(QQ^T) \in \mathbb{R}^{m\times m} \tag{2}$$

What we want to do is make the new basis have variables that are uncorrelated. In linear algebra terms, we want the covariance matrix of the transformed data to be diagonal.
The covariance matrix of the transformed data matrix $Y \in \mathbb{R}^{m\times n}$ is defined below

$$C_Y = \frac{1}{n}(YY^T) \tag{3}$$

Let us compute the covariance of the transformed data matrix, or Y. Remember that $Y = KQ$, which gives us

$$C_Y = \frac{1}{n}\left((K^T Q)(K^T Q)^T\right) = \frac{1}{n}(K^T QQ^T K) = K^T C_X K \tag{4}$$

We now want to choose K so that

$$C_Y = K^T C_X K \tag{5}$$

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

is a diagonal matrix. This is so the new variables are uncorrelated.
See that

$$(C_Q)^T = \left(\frac{1}{n}(QQ^T)\right)^T = \frac{1}{n}(QQ^T)^T = \frac{1}{n}(Q^T)^T Q^T = \frac{1}{n}(QQ^T) = C_Q \tag{6}$$

So C_Q is symmetric. Because of this (by theorem 3), C_Q can be diagonalized using its eigenvectors.
Let $D \in \mathbb{R}^{m\times m}$ be the diagonal matrix consisting of the eigenvalues of C_Q. Choose $K \in \mathbb{R}^{m\times m}$ as the matrix of eigenvectors of C_Q. Then,

$$C_Q = KDK^T \tag{7}$$

or, equivalently,

$$D = K^T C_Q K \tag{8}$$

Now,

$$C_Y = K^T C_Q K = D \tag{9}$$

Hence, it is proven that transforming the data using the eigenvector matrix K results in a diagonal covariance matrix D, where the variables (principal components) are uncorrelated.
We now have definition of PCA as a linear transformation

$$f: \mathbb{R}^m \to \mathbb{R}^m, f(q) = K \tag{10}$$

where q is each column vector of the centered data matrix Q.
Proof.
We know that $\mathbb{R}^m$ is a vector space over the field $\mathbb{R}$. To prove that f is a linear transformation, it must be proven that $\forall x, y \in \mathbb{R}^m, k \in \mathbb{R}$

1. $f(x+y) = f(x) + f(y)$
2. $f(\alpha x) = \alpha f(x)$

Take $x, y \in \mathbb{R}^m$,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, x_i, y_i \in \mathbb{R}, i = 1,2,\dots,m$$

$$K = \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1m} \\ k_{21} & k_{22} & \dots & k_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ k_{m1} & k_{m2} & \dots & k_{mm} \end{bmatrix} = \begin{bmatrix} k_1^T \\ k_2^T \\ \vdots \\ k_m^T \end{bmatrix}$$

$$f(x+y) = K(x+y) = K\left(\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_p + y_p \end{bmatrix}\right) = \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1p} \\ k_{21} & k_{22} & \dots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \dots & k_{kp} \end{bmatrix} \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_p + y_p \end{bmatrix} =$$

$$= \begin{bmatrix} k_{11}(x_1 + y_1) + k_{12}(x_2 + y_2) + \cdots + k_{1p}(x_p + y_p) \\ k_{21}(x_1 + y_1) + k_{22}(x_2 + y_2) + \cdots + k_{2p}(x_p + y_p) \\ \vdots \\ k_{k1}(x_1 + y_1) + k_{k2}(x_2 + y_2) + \cdots + k_{kp}(x_p + y_p) \end{bmatrix}$$

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

$$\begin{bmatrix} (k_{11}x_1 + k_{11}y_1) + (k_{12}x_2 + k_{12}y_2) + \cdots + (k_{1p}x_p + k_{1p}y_p) \\ (k_{21}x_1 + k_{21}y_1) + (k_{22}x_2 + k_{22}y_2) + \cdots + (k_{2p}x_p + k_{2p}y_p) \\ \vdots \\ (k_{k1}x_1 + k_{k1}y_1) + (k_{k2}x_2 + k_{k2}y_2) + \cdots + (k_{kp}x_p + k_{kp}y_p) \end{bmatrix}$$

$$= \begin{bmatrix} k_{11}x_1 + k_{12}x_2 + \ldots + k_{1p}x_p \\ k_{21}x_1 + k_{22}x_2 + \ldots + k_{2p}x_p \\ \vdots \\ k_{k1}x_1 + k_{k2}x_2 + \ldots + k_{kp}x_p \end{bmatrix} + \begin{bmatrix} k_{11}y_1 + k_{12}y_2 + \ldots + k_{1p}y_p \\ k_{21}y_1 + k_{22}y_2 + \ldots + k_{2p}y_p \\ \vdots \\ k_{k1}y_1 + k_{k2}y_2 + \ldots + k_{kp}y_p \end{bmatrix}$$

$$= \begin{bmatrix} k_{11} & k_{12} & \ldots & k_{1p} \\ k_{21} & k_{22} & \ldots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \ldots & k_{kp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} + \begin{bmatrix} k_{11} & k_{12} & \ldots & k_{1p} \\ k_{21} & k_{22} & \ldots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \ldots & k_{kp} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

$$= K\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}\right) + K\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}\right) = K(x) + K(y) = f(x) + f(y)$$

We also have that

$$f(\alpha x) = K(\alpha x) = K\left(\alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}\right) = K\left(\begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_p \end{bmatrix}\right) = \begin{bmatrix} k_{11} & k_{12} & \ldots & k_{1p} \\ k_{21} & k_{22} & \ldots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \ldots & k_{kp} \end{bmatrix} \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_p \end{bmatrix}$$

$$= \begin{bmatrix} k_{11}\alpha x_1 + k_{12}\alpha x_2 + \ldots + k_{1p}\alpha x_p \\ k_{21}\alpha x_1 + k_{22}\alpha x_2 + \ldots + k_{2p}\alpha x_p \\ \vdots \\ k_{k1}\alpha x_1 + k_{k2}\alpha x_2 + \ldots + k_{kp}\alpha x_p \end{bmatrix} = \begin{bmatrix} \alpha(k_{11}x_1 + k_{12}x_2 + \cdots + k_{1p}x_p) \\ \alpha(k_{21}x_1 + k_{22}x_2 + \cdots + k_{2p}x_p) \\ \vdots \\ \alpha(k_{k1}x_1 + k_{k2}x_2 + \cdots + k_{kp}x_p) \end{bmatrix}$$

$$= \alpha \begin{bmatrix} k_{11}x_1 + k_{12}x_2 + \ldots + k_{1p}x_p \\ k_{21}x_1 + k_{22}x_2 + \ldots + k_{2p}x_p \\ \vdots \\ k_{k1}x_1 + k_{k2}x_2 + \ldots + k_{kp}x_p \end{bmatrix} = \alpha \left(\begin{bmatrix} k_{11} & k_{12} & \ldots & k_{1p} \\ k_{21} & k_{22} & \ldots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \ldots & k_{kp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \right) = \alpha \left(K\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \right) \right)$$

$$= k\big(K(x)\big) = k\big(f(x)\big)$$

Therefore, it is proven that $f: \mathbb{R}^m \to \mathbb{R}^m, f(x) = Px$ is a linear transformation.

1) *Principal Component Analysis via Covariance Matrix* as a Linear transformation

It has been stated above that the variant of PCA discussed in Shlens's 2014 work differs from that which is discussed in this paper. Now that we have discussed Shlens's process and definition of PCA as a linear transformation, we can now explain this for our variant of principal component analysis. It should be noted that the variables below will be the same as the variables in the flowchart given in the previous chapter for consistency purposes.

Let $X = \mathbb{R}^{n \times p}$ be the original dataset. Where n =number of observations, and $p =$ number of variables.

The mean vector is $m = \frac{1}{n}\sum_{i=1}^{n} x_i \in \mathbb{R}^p$. Then, the centered data matrix is $Q = (X - 1m^T) \in \mathbb{R}^{n \times p}$.

The covariance matrix of the centered data matrix is as below.

$$C_Q = \frac{1}{n-1}(Q^T Q) \in \mathbb{R}^{p \times p} \tag{11}$$

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

Since $Q \in \mathbb{R}^{n\times p}$, we use $(Q^TQ) \in \mathbb{R}^{p\times p}$ to obtain a covariance matrix that describes the relationships **between variables**, rather than between observations. This is standard in our variant of PCA which is typically used for statistical purposes. The definition of the covariance matrix used by Shlens (2014) differs since the variant of PCA is also different.
Since C_Q is a real symmetric matrix, it is orthogonally diagonalizable. Hence,

$$C_Q = VDV^T \tag{12}$$

Where $V \in \mathbb{R}^{p\times p}$ is an orthonormal matrix, whose columns are the eigenvectors of C_Q. V^T is then V^{-1} since the eigenvectors are orthonormal. $D \in \mathbb{R}^{p\times p}$ is a diagonal matrix of eigenvalues.
PCA as a linear transformation represented by a matrix is then

$$Z = QV \tag{13}$$

Here, $Z = \mathbb{R}^{n\times p}$ the matrix where each column vector is a projection of the data onto a principal component. The flowchart in the previous chapter also states this, so the concept is matching up.
To confirm this transformation achieves decorrelation, we compute the covariance matrix of the transformed data matrix $Z \in \mathbb{R}^{n\times p}$, which is defined below

$$C_Z = \frac{1}{n-1}(Z^TZ) \tag{14}$$

Remember that $Z = QV$, which gives us

$$C_Z = \frac{1}{n-1}\big((QV)^T(QV)\big) = \frac{1}{n-1}(V^TQ^TQV) = V^TC_QV \tag{15}$$

Again, since C_Q is a real symmetric matrix, $C_Q = VDV^T$ and

$$C_Z = V^TC_QV = V^T(VDV^T)V = D \tag{16}$$

This proves that the covariance matrix of the transformed data Z is diagonal, and each diagonal element is the variance along the corresponding principal component. Furthermore, this proves that PCA transforms the data into a new orthogonal basis where the features are uncorrelated and sorted by variance.
Hence, we have the following **definition of traditional/statistical PCA via Covariance Matrix as a linear transformation:**
Let $X \in \mathbb{R}^{n\times p}$ bet the original dataset. Where n =number of observations, and p = number of variables.
Let the centered data matrix $Q \in \mathbb{R}^{n\times p}$, where each row $q_i = x_i - m, m = \frac{1}{n}\sum_{i=1}^{n} x_i \in \mathbb{R}^p$

$$f: \mathbb{R}^p \to \mathbb{R}^p, f(q) = V^Tq \tag{17}$$

where q is each row vector of the centered data matrix Q.
The proof that f is a linear transformation here is similar to the proof above.

2) *Principal Component Analysis via Correlation Matrix* as a Linear transformation

The definition of PCA via Correlation Matrix as a linear transformation follows the definition of PCA via Covariance Matrix. Only this time the input data matrix is the standardized data matrix Q_{STD}, a matrix where each variable has been rescaled to have variance 1.
Hence, we have the following **definition of traditional/statistical PCA via Correlation Matrix as a linear transformation:**
Let $X \in \mathbb{R}^{n\times p}$ bet the original dataset. Where n =number of observations, and p = number of variables.
Let the standardized data matrix $Q_{STD} \in \mathbb{R}^{n\times p}$

$$f: \mathbb{R}^p \to \mathbb{R}^p, f(q_{STD}) = V^Tq_{STD} \tag{18}$$

where q_{STD} is each row vector of the standardized data matrix Q_{STD}.

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

The proof that f is a linear transformation here is once again similar to the one above.

Since the two definitions are similar, and the only difference is the input, this paper will refer to both definitions as the definition of PCA as a linear transformation

$$f: \mathbb{R}^p \to \mathbb{R}^p, f(q) = V^T q \tag{18}$$

where $q \in \mathbb{R}^p$ is a row vector (from the centered or standardized data matrix), and $V \in \mathbb{R}^{p \times p}$ is a matrix of orthonormal eigenvectors of the covariance or correlation matrix.

B. PCA as an R-morphism

We will first prove that $\mathbb{R}^p$ is an $\mathbb{R}$ -module.
We have the left action

$$*: \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^p$$

Defined by

$$(r, m) \to * (r, m) = r * m$$

Take $v_1, v_2 \in \mathbb{R}^p, \alpha, \beta \in \mathbb{R}$

$$v_1 = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}, v_2 = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}, a_i, b_i \in \mathbb{R}, i = 1,2, \dots, p$$

$$(1)\ \alpha * (v_1 + v_2) = \alpha * \left(\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} \right) = \alpha * \left(\begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_p + b_p \end{bmatrix} \right) = \alpha \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_p + b_p \end{bmatrix} = \begin{bmatrix} \alpha(a_1 + b_1) \\ \alpha(a_2 + b_2) \\ \vdots \\ \alpha(a_p + b_p) \end{bmatrix}$$

$$= \begin{bmatrix} \alpha a_1 + \alpha b_1 \\ \alpha a_2 + \alpha b_2 \\ \vdots \\ \alpha a_p + \alpha b_p \end{bmatrix} = \begin{bmatrix} \alpha a_1 \\ \alpha a_2 \\ \vdots \\ \alpha a_p \end{bmatrix} + \begin{bmatrix} \alpha b_1 \\ \alpha b_2 \\ \vdots \\ \alpha b_p \end{bmatrix} = \alpha \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} + \alpha \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = \alpha * \left(\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \right) + \alpha * \left(\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} \right)$$

$$= \alpha * v_1 + \alpha * v_2$$

$$(2)\ (\alpha + \beta) * v_1 = (\alpha + \beta) * \left(\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \right) = (\alpha + \beta) \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} (\alpha + \beta)a_1 \\ (\alpha + \beta)a_2 \\ \vdots \\ (\alpha + \beta)a_p \end{bmatrix} = \begin{bmatrix} \alpha a_1 + \beta a_1 \\ \alpha a_2 + \beta a_2 \\ \vdots \\ \alpha a_p + \beta a_p \end{bmatrix}$$

$$= \begin{bmatrix} \alpha a_1 \\ \alpha a_2 \\ \vdots \\ \alpha a_p \end{bmatrix} + \begin{bmatrix} \beta a_1 \\ \beta a_2 \\ \vdots \\ \beta a_p \end{bmatrix} = \alpha \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} + \beta \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \alpha * \left(\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \right) + \beta * \left(\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \right)$$

$$= \alpha * v_1 + \beta * v_1$$

$$(3)\ (\alpha \cdot \beta) * v_1 = (\alpha \cdot \beta) * \left(\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \right) = (\alpha \cdot \beta) \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} (\alpha \cdot \beta)a_1 \\ (\alpha \cdot \beta)a_2 \\ \vdots \\ (\alpha \cdot \beta)a_p \end{bmatrix} = \begin{bmatrix} \alpha(\beta \cdot a_1) \\ \alpha(\beta \cdot a_2) \\ \vdots \\ \alpha(\beta \cdot a_p) \end{bmatrix}$$

$$= \alpha \begin{bmatrix} (\beta \cdot a_1) \\ (\beta \cdot a_2) \\ \vdots \\ (\beta \cdot a_p) \end{bmatrix} = \alpha * \left(\begin{bmatrix} (\beta \cdot a_1) \\ (\beta \cdot a_2) \\ \vdots \\ (\beta \cdot a_p) \end{bmatrix} \right) = \alpha * \left(\beta \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \right) = \alpha * \left(\beta * \left(\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \right) \right)$$

$$= \alpha * (\beta * v_1)$$

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

(4) Since $1_{\mathbb{R}} = 1$

$$1_{\mathbb{R}} * v_1 = 1 * \left(\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}\right) = 1\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} 1a_1 \\ 1a_2 \\ \vdots \\ 1a_p \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = v_1$$

The above processes are possible using the properties of $\mathbb{R}^p$ as an additive abelian group, and $\mathbb{R}$ as a unitary ring.
From (1)-(4), we have proven $\mathbb{R}^p$ over $\mathbb{R}$ is an $\mathbb{R}$-module.
Now, referring to the definition of the R-module homomorphism, a function $f: \mathbb{R}^p \to \mathbb{R}^p$ is called an $\mathbb{R}$-module homomorphism if it preserves addition and scalar multiplication.

1. $f(x + y) = f(x) + f(y)$
2. $f(kx) = kf(x), \forall x, y \in \mathbb{R}^p, k \in \mathbb{R}$

From the proof of PCA as a linear transformation, we can conclude that PCA (via correlation matrix and covariance matrix) is also an $\mathbb{R}$-module homomorphism.
We know from the definition of the R-module automorphism that the function $f: \mathbb{R}^p \to \mathbb{R}^p$ must be bijective (both injective and surjective) to be called an $\mathbb{R}$-module automorphism. In other words, f must be structure-preserving and have an inverse, f^{-1}.
Recall our definition of the PCA function in (18).
Where $q \in \mathbb{R}^p$ is a row vector (from the centered or standardized data matrix), and $V \in \mathbb{R}^{p \times p}$ is a matrix of orthonormal eigenvectors of the covariance or correlation matrix. Since V^T is the matrix representation of our function f, to prove that f^{-1} exists, we first prove that V^T is invertible.
Recall that V is orthogonal, meaning that it satisfies $V^T = V^{-1}$. Which means that $V^T V = I = VV^T$
Hence, the matrix V^T is invertible, with the inverse of V^T being V.
Therefore, if $f(q) = V^T q$. Then the inverse of f is $f^{-1}(q) = Vq$. This is easily proven since

$$f^{-1}\big(f(q)\big) = f^{-1}(V^T q) = V(V^T q) = (VV^T)q = I_P q = q \tag{19}$$

$$f\big(f^{-1}(q)\big) = f(Vq) = V^T(Vq) = (V^T V)q = I_P q = q$$

It is then proven that PCA is an $\mathbb{R}$-module automorphism, with the important assumption that all principal components are kept.
When PCA is used for dimensionality reduction, however, the amount of principal components are less than or equal to the amount of variables of the original dataset. Here, V (the matrix representation of the linear transformation is no longer in $\mathbb{R}^{p \times p}$, but in $\mathbb{R}^{p \times k}$, where k is the amount of principal components kept). The transformation then becomes

$$f': \mathbb{R}^p \to \mathbb{R}^k, f(q) = {V_k}^T q, V_k \in \mathbb{R}^{p \times k} \tag{20}$$

We will refer to this function, f', as the definition of dimension-reducing PCA.
We will now analyze this function to see if it is an $\mathbb{R}$-module homomorphism, monomorphism, epimorphism, or isomorphism. We can see clearly that f' is definitely not an $\mathbb{R}$-module automorphism since $\mathbb{R}^p \neq \mathbb{R}^k$ due to the fact that $k < p$.
We will first prove f' is an $\mathbb{R}$-module homomorphism.
Let M_1 and M_2 be R −modules over the ring R. $f: M_1 \to M_2$ is an ***R* −module homomorphism, or an *R* −morphism**, if the following are satisfied:

1. $f(x + y) = f(x) + f(y)$
2. $f(kx) = kf(x), \forall x, y \in M_1, k \in R$

We know that $\mathbb{R}^p, \mathbb{R}^k$ are $\mathbb{R}$ −modules from the proof earlier in this chapter (to prove $\mathbb{R}^k$ is an $\mathbb{R}$ −module, the proof is the same; simply replace $p\ with\ k$).
Take $x, y \in \mathbb{R}^p$,

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}, x_i, y_i \in \mathbb{R}, i = 1,2, \dots, p$$

$$V_k^T = \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1p} \\ k_{21} & k_{22} & \dots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \dots & k_{kp} \end{bmatrix}$$

$$f(x+y) = V_k^T(x+y) = V_k^T\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}\right) = V_k^T\left(\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_p + y_p \end{bmatrix}\right)$$

$$= \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1p} \\ k_{21} & k_{22} & \dots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \dots & k_{kp} \end{bmatrix}\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_p + y_p \end{bmatrix}$$

$$= \begin{bmatrix} k_{11}(x_1 + y_1) + k_{12}(x_2 + y_2) + \cdots + k_{1p}(x_p + y_p) \\ k_{21}(x_1 + y_1) + k_{22}(x_2 + y_2) + \cdots + k_{2p}(x_p + y_p) \\ \vdots \\ k_{k1}(x_1 + y_1) + k_{k2}(x_2 + y_2) + \cdots + k_{kp}(x_p + y_p) \end{bmatrix}$$

$$= \begin{bmatrix} (k_{11}x_1 + k_{11}y_1) + (k_{12}x_2 + k_{12}y_2) + \cdots + (k_{1p}x_p + k_{1p}y_p) \\ (k_{21}x_1 + k_{21}y_1) + (k_{22}x_2 + k_{22}y_2) + \cdots + (k_{2p}x_p + k_{2p}y_p) \\ \vdots \\ (k_{k1}x_1 + k_{k1}y_1) + (k_{k2}x_2 + k_{k2}y_2) + \cdots + (k_{kp}x_p + k_{kp}y_p) \end{bmatrix}$$

$$= \begin{bmatrix} k_{11}x_1 + k_{12}x_2 + \dots + k_{1p}x_p \\ k_{21}x_1 + k_{22}x_2 + \dots + k_{2p}x_p \\ \vdots \\ k_{k1}x_1 + k_{k2}x_2 + \dots + k_{kp}x_p \end{bmatrix} + \begin{bmatrix} k_{11}y_1 + k_{12}y_2 + \dots + k_{1p}y_p \\ k_{21}y_1 + k_{22}y_2 + \dots + k_{2p}y_p \\ \vdots \\ k_{k1}y_1 + k_{k2}y_2 + \dots + k_{kp}y_p \end{bmatrix}$$

$$= \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1p} \\ k_{21} & k_{22} & \dots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \dots & k_{kp} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} + \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1p} \\ k_{21} & k_{22} & \dots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \dots & k_{kp} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

$$= V_k^T\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}\right) + V_k^T\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}\right) = V_k^T(x) + V_k^T(y) = f(x) + f(y)$$

We also have that

$$f(kx) = V_k^T(kx) = V_k^T\left(k\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}\right) = V_k^T\left(\begin{bmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_p \end{bmatrix}\right) = \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1p} \\ k_{21} & k_{22} & \dots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \dots & k_{kp} \end{bmatrix}\begin{bmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_p \end{bmatrix}$$

$$= \begin{bmatrix} k_{11}kx_1 + k_{12}kx_2 + \dots + k_{1p}kx_p \\ k_{21}kx_1 + k_{22}kx_2 + \dots + k_{2p}kx_p \\ \vdots \\ k_{k1}kx_1 + k_{k2}kx_2 + \dots + k_{kp}kx_p \end{bmatrix} = \begin{bmatrix} k(k_{11}x_1 + k_{12}x_2 + \cdots + k_{1p}x_p) \\ k(k_{21}x_1 + k_{22}x_2 + \cdots + k_{2p}x_p) \\ \vdots \\ k(k_{k1}x_1 + k_{k2}x_2 + \cdots + k_{kp}x_p) \end{bmatrix}$$

Indonesian Journal of Applied Mathematics and Statistics (IdJAMS)
ISSN 3064-4062
Volume 2 Issue 2, Desember 2025, pp. 81-96
https://idjams.com/index.php/jms/index

$$= k\begin{bmatrix} k_{11}x_1 + k_{12}x_2 + \ldots + k_{1p}x_p \\ k_{21}x_1 + k_{22}x_2 + \ldots + k_{2p}x_p \\ \vdots \\ k_{k1}x_1 + k_{k2}x_2 + \ldots + k_{kp}x_p \end{bmatrix} = k\left(\begin{bmatrix} k_{11} & k_{12} & \ldots & k_{1p} \\ k_{21} & k_{22} & \ldots & k_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ k_{k1} & k_{k2} & \ldots & k_{kp} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}\right) = k\left(V_k^T\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}\right)\right)$$

$$= k\left(V_k^T(x)\right) = k\big(f(x)\big)$$

Therefore, it is proven that $f'\colon \mathbb{R}^p \to \mathbb{R}^k$ is an $\mathbb{R}$-module homomorphism.
For f' to be an $\mathbb{R}$-module monomorphism, it must be injective. This means that $\forall x, y \in \mathbb{R}^p, f'(x) = f'(y)$, then $x = y$. So, for f' to be injective it must be that $Ker(f') = 0$ and in other words, $dim\big(Ker(f')\big) = 0$. According to Rank-Nullity Theorem, $dim(\mathbb{R}^p) = rank\big(V_k^T\big) + null\big(V_k^T\big)$.
Since $dim\big(Ker(f')\big) = null\big(V_k^T\big)$, $null\big(V_k^T\big)$ should be 0 for f' to be injective.
Since V_k has orthonormal columns, $rank\big(V_k^T\big) = k$. From the Rank-Nullity theorem, we have that $p = k + null\big(V_k^T\big)$so $null\big(V_k^T\big) = p - k$. Since $k < p, p - k > 0$. Hence, $null\big(V_k^T\big) > 0$. Therefore, f' is not an $\mathbb{R}$-module monomorphism because it is not injective.
For f' to be an $\mathbb{R}$-module epimorphism, it must be surjective. This means that $\forall z \in \mathbb{R}^k, \exists x \ni f'(x) = V_k^T(x) = z$. So, for f' to be surjective it must be that $rank(f') = rank\big(V_k^T\big) = dim\big(\mathbb{R}^k\big)$. We know that $dim\big(\mathbb{R}^k\big) = k$ and from above we know that $rank\big(V_k^T\big) = k$. Hence, f' is an $\mathbb{R}$-module epimorphism because it is surjective.

V. Conclusion

PCA is very important and finds multiple applications in numerous fields. Because of this, there are multiple variants of PCA such as the traditional PCA, discussed in this paper, using the covariance or correlation matrix, or the PCA used in machine learning, discussed in Shlens's paper [12], and more. This is a good thing, it means the technique of PCA is adaptable. To be able to apply this powerful dimensionality reduction technique and/or make our own variant, however, it is important to study the original and ask, "what makes PCA PCA?" and "What is the thread that connects all these different variants together?" Through analyzing PCA through the lens of module theory, the answer becomes clear. Despite surface differences, all linear variants of PCA share a common algebraic foundation.

First, PCA was shown to be a linear transformation. Next, this transformation was shown to be an $\mathbb{R}$-module homomorphism since it preserves both addition and scalar multiplication. Furthermore, PCA has also been proven to be an $\mathbb{R}$-module automorphism (when the full orthonormal basis is taken) which means that it preserves linearity and structure of the $\mathbb{R}$-module and proven to be invertible. Finally, PCA as a dimension-reductionality technique (when only a subset of the principal components was taken into consideration at the end, accounting for $< 100\%$ variance in data) was still an $\mathbb{R}$-module homomorphism, but it was no longer invertible. It was, however, still surjective. While PCA as a dimension-reductionality technique is no longer an $\mathbb{R}$-module automorphism or isomorphism, it remains an $\mathbb{R}$-module epimorphism, which means the projection of data into a lower-dimensional subspace using PCA is still structure-preserving.

After our analysis, we can properly conclude that any PCA variant (no matter the field or implementation) is valid so long as it satisfies a few linear-algebraic properties. The first is that is that the transformation must be linear. The second algebraic property is that it maps to a new orthonormal basis of the original centered dataset. And finally, the transformation should diagonalize the covariance (or correlation) matrix (this means the new PCs are decorrelated and have been ranked by variance). These three properties are what make PCA. Therefore, no matter the data orientation, matrix used (correlation or covariance), or application, the PCA variant does and will preserve these algebraic conditions. Hence, we now have a new definition of PCA in module theory: an R-module automorphism that diagonalizes the covariance structure of the original dataset via an orthogonal change of basis.

This paper re-explains PCA with algebraic terms, rather than purely statistical to lay the groundwork for deeper theoretical explorations. Future research can analyze nonlinear and Kernel PCA, identifying the algebraic axioms for each of the statistical methods. Additionally, future papers can also provide a closer generalization of PCA using a different ring other than $\mathbb{R}$.

Acknowledgement

I would like to thank my one-eyed cat (Kerajaan Mata Satu) and the class of Pure Mathematics Unpad 2022 for the motivation.

References

[1] Paik and J. Park, "Circular Coordinates for Density-robust Analysis," *arXiv preprint*, arXiv:2301.12742, 2023. DOI: 10.48550/arxiv.2301.12742
[2] R. Bellman and R. Kalaba, "A Mathematical Theory of Adaptive Control Processes," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 45, no. 8, pp. 1288–1290, 1959. DOI: 10.1073/pnas.45.8.1288
[3] I. Jolliffe, *Principal Component Analysis*. Chichester, U.K.: Wiley, 2005. DOI: 10.1002/0470013192.bsa501
[4] H. Abdi and L. J. Williams, "Principal Component Analysis," *WIREs Comput. Stat.*, vol. 2, pp. 433 – 459, 2010. DOI: 10.1002/wics.101, 2010
[5] K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space," *Phil. Mag.*, vol. 2, no. 11, pp. 559–572, 1901.
[6] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933.
[7] I. Jolliffe and J. Cadima, "Principal Component Analysis: A Review and Recent Developments," *Phil. Trans. R. Soc. A*, vol. 374, no. 2065, p. 20150202, 2016. DOI: 10.1098/rsta.2015.0202
[8] F. Josse and F. Husson, "*Exploratory Multivariate Analysis by Example using R*." Boca Raton, FL, USA: CRC Press, 2016.
[9] A. Lin, P. Shang, and H. Zhou, "Cross-correlations and Structures of Stock Markets Based on Multiscale MF-DXA and PCA," *Nonlinear Dynamics*, vol. 78, pp. 485–494, 2014. DOI: 10.1007/s11071-014-1455-5
[10] C. Yoo and M. Shahlaei, "The Applications of PCA in QSAR Studies: A Case Study on CCR5 antagonists," *Chem. Biol. Drug Des.*, vol. 91, no. 1, pp. 137–152, 2018. DOI: 10.1111/cbdd.13064
[11] M. Misztal, "On the Potential for Using Selected PCA-based Methods to Analyze the Crime Rate in Poland," *Econometrics*, vol. 23, no. 2, pp. 15–32, 2019. DOI: 10.15611/eada.2019.2.02
[12] J. Shlens, "A Tutorial on Principal Component Analysis," *arXiv preprint*, arXiv: 1404.1100, 2014.
[13] T. S. Blyth, "*Module Theory: An Approach to Linear Algebra*." University of St Andrews, 2018. [Online]. Available: https://research-repository.st-andrews.ac.uk/handle/10023/12643
[14] H. Anton and C. Rorres, "*Elementary Linear Algebra: Applications Version*." Hoboken, NJ, USA: Wiley, 2014.
[15] B. Jacob, "*Linear Algebra*." Boston, MA, USA: W. H. Freeman, 1990