2018

Statistics and Biomechanics: An Interdisciplinary Evaluation of the Mathematical, Practical, and Athletic Applications of Principal Component Analysis

Sydney Grace Davis
Gardner-Webb University

Statistics and Biomechanics: An Interdisciplinary Evaluation of the Mathematical, Practical, and Athletic Applications of Principal Component Analysis

A Joint Honors/Mathematics Thesis
Presented to
The University Honors Program
And
The Department of Mathematical Sciences
Gardner-Webb University
6 April 2018

by

Sydney Grace Davis

Accepted by the Honors Faculty

_____________________________	_____________________________
Dr. Olga Poliakova, Thesis Advisor	Dr. Tom Jones; Associate Dean; Univ. Honors
_____________________________	_____________________________
Dr. Candice Rome, Honors Committee	Dr. Lorene Pagcaliwagan, Honors Committee
_____________________________	_____________________________
Dr. Don Olive, Honors Committee	Dr. Anna Sieges Beal, Honors Committee
_____________________________	_____________________________
Dr. Tammy Hoyle, Mathematics Dpt.	Dr. Jason Willis, Mathematics Dpt.

Table of Contents

I. Introduction (Nature and Scope)

Coaches and athletes around the world are in constant pursuit of improving their athletic performance. For some, a routine amount of weight lifting, cardiovascular exercise, agilities and flexibility training with gradual advancement may be enough to see growth. However, many are not satisfied and turn to in-depth analyses of their techniques in order to measure their progress. After an intense workout or competition, coaches spend time breaking down athletic performances based on major movements. By compartmentalizing these activities, they can identify which motions are efficient and which ones hinder fluid motion. From this evaluation and discernment, athletes may then make training adjustments to enhance their performances.

While these observation-based analyses have provided athletes with sufficient feedback, they lack the data necessary for optimal improvement. Thus, for athletes to reach their full potentials, more precision is required. Using mathematical practices to analyze an individual's biomechanics permits evaluation of every detail of an athlete's motion. Not only does this account for motions missed but it also eliminates the subjectivity of the human eye. The addition of this technicality generates all-encompassing, cohesive data that can be practically applied.

In the same way, statisticians are in constant pursuit of an ideal method for data analysis. Using statistics as their platform, mathematicians conduct studies in a variety of fields, organizing and evaluating data from which they may draw conclusions. Whether in pure math or practical applications, various statistical models have proven effective in quantifying results. Still yet, research continues for the perfect statistical model.

Principal component analysis is a century-old method of data analysis that has been implemented over the past several decades. Using statistics as its basis, principal component analysis interprets observations and organizes the data into sets based on repetition and pattern. This organization into sets establishes the **principal components**, or main aspects, of the given data. By consolidating data into fewer variables, principal component analysis allows the researcher to assess information in a much simpler way.

While the concept may appear basic in nature to the naked eye, principal component analysis is dynamic and extends across the entire spectrum of mathematics. While rooted in statistics, it may be more specifically implemented and analyzed through the lens of linear algebra. In fact, several linear algebra concepts are fundamental to one's understanding and the development of principal component analysis. A few relevant linear algebra topics include matrix transformation, orthogonality, eigenvalues and eigenvectors.

In addition to emphasizing the discovery learning process charged by mathematical research, this thesis focuses on principal component analysis as an advanced tool for data analysis in a variety of settings. Each chapter within this paper is structured to portray the history, theory, conceptual development, and mathematical applications of principal component analysis. Additionally, this thesis explores the relationship between this analysis method and the field of biomechanics. More specifically, this relationship is characterized for both generalized athletic movement and competitive running. The various aspects of this concept are designed to build the reader's understanding of principal component analysis while addressing the following research question:

How do the multivariate analysis methods of principal component analysis (PCA) make an impact on our understanding of athletic activity and biomechanical motion?

II. Overview of Principal Component Analysis

Defined very generally, principal component analysis is "a mathematical procedure that transforms a [large] number of possibly correlated variables into a smaller number of uncorrelated variables called PC" (Suryanarayana and Mistry 20). The acronym "PC" is a short-hand way of referring to the identified principal components of a data set. More specifically, principal component analysis is a "quintessential data-crunching procedure" that is used to represent a data set for more efficient interpretation (Meyers et al. 404). It is purposed with both "dimensionality reduction [and] data compression" (*Introduction to Principal Components and Factor Analysis*).

The principal components established through principal component analysis are its characteristic trademark. As an exploratory form of analysis, principal component analysis evaluates all data within a set based on selected variables. In doing so, it determines trends and relationships to develop principal components. Contrary to popular misconception, each component is a "weighted linear combination of the variables being analyzed based on all of the cases in the data file (Meyers et al. 414). In this respect, each principal component is a reflection of the entire collection. The first principal component developed for a set "accounts for as much of the variability in the data as possible" (*Introduction to Principal Components and Factor Analysis*).

Within statistics, principal component analysis may be very closely compared to other statistical models. One model of considerable similarity is factor analysis (FA). In many respects, principal component analysis closely parallels the purpose, structure and function of factor analysis. For this reason, the two are often viewed as interchangeable by individuals who are unfamiliar with their inner workings. While both analyze sets of data with the goal of representing it with reduced elements, each of these two statistical methods approach it in characteristically different ways.

The most striking differences between principal component analysis and factor analysis is the vocabulary used to reference them and their method for approach of data reduction. Although the term “factor analysis” has recently shifted to refer to the general field of both principal component analysis and factor analysis, principal component analysis should not be considered a subset of the other method. According to Meyers et al., factor analysis may be more specifically differentiated as a statistical method for data reduction that “shifts [the] perspective around” (Meyers et al. 422). While principal component analysis focuses on the identification of principal components based on those in the analysis, factor analysis flows from a latent factor back towards the measured indicators.

This may be further illustrated through a psychological example of phrase association with depression (Meyers et al. 422). Using the mechanism of factor analysis, the phrases one commonly associates with depression may be viewed as indicators of the disorder. In this sense, phrases such as “I can’t do this anymore” or “I’m worthless” are the results of someone’s condition. They provide smaller measures for determining the severity of their preceding condition. Principal component analysis, however, approaches this quite differently. Instead of working “backwards,” principal component analysis views a wide variety of variables like these phrases as the causes of depression. The flow of information here is more or less in the "forward direction," allowing depression to be represented and evaluated differently.

By each principal component referencing the entire collection, the purpose of principal component analysis is to provide insight about the total variance of the variables being considered. The total variance is quantified as the total number of variables in the data set being analyzed. With this goal in mind, principal component analysis undergoes one primary procedure for execution. Similar to factor analysis, this procedure takes two distinctive phases: an extraction phase and a rotation phase (Meyers et al. 413).

As one could probably hypothesize, the extraction phase is that in which the principal components are identified. Identified individually, principal components are pulled from the overall data set to represent part of the original variance in the system (Meyers et al. 414). Principal components differ in that they seek to "[explain] the variance in a particular orthogonal dimension" (*Introduction to Principal Components and Factor Analysis*). Because principal component analysis is determined for multidimensional systems, each component may be conceptualized as a line of best fit for part of the variance. Subsequent components to the first must intersect the first component in some way but represent a separate partition of the overall variance (Meyers et al. 416). Each subsequent principal component established represents the next largest portion of the variance in the data.

Within the system, correlation between variables is represented by the distances between them (Meyers et al. 419). Variables that are strongly correlated in a data set are closer together while weak correlations are depicted by increased distance. This, along with the computed weights of principal components, are considered when interpreting the data.

Once all variables have been mathematically extracted and a correlation matrix has been constructed, the extraction phase is complete.

Then, the rotation phase commences. The purpose of the rotation phase is to represent mathematical findings before interpreting the results (Warner 772). The objective of this rotation is to assimilate data into a **simple structure** where most components have high factor loadings for their respective variables. This simple structure is achieved by the "pivoting of the first *n* number of extracted factors around their point of intersection" (Meyers et al. 426). Rotation maintains the integrity of each component by keeping it in the same orientation with respect to the others. Since each principal component retains its respective portion of the variance, a mathematician can effectively "[redistribute] the variance across the factors to facilitate interpretation" (Meyers et al. 426).

Once the principal components and correlation matrix have been sufficiently rotated, interpretation of the principal component analysis results may occur. Because each principal component is representative of a different piece of the variance pointing towards the total variance, each contributes toward the holistic result. When considering the results of principal component analysis, it is crucial to keep other related statistical terminology and concepts in mind. These terms include, but are not limited to, orthogonality and component loading. Orthogonality is the orientation of two items (in this case, principal components) at a 90° angle with one another. Component loading is "correlation coefficients between the variables (rows) and factors (columns)" used when representing principal component analysis in tabular form (*Introduction to Principal Components and Factor Analysis*).

III. <u>Historical Development</u>

Although the singular procedural nature of principal component analysis suggests one root of origin, it is the conglomeration of many statistical efforts that has made it possible. Because of the many aspects needed to make it an effective method, it has not been officially attributed to one individual. However, the statistician that most commonly receives the most credit for the development of principal component analysis is Harold Hotelling. Known for his work with economics and T^2 statistical distributions, Hotelling is acknowledged as being the statistician that "developed PCA more fully in a 1933 paper in the *Journal of Educational Psychology*" (Millsap 105). From Hotelling's work, emphasis is placed on the generation of orthogonal linear combinations that represent variables with maximum variance (De Leeuw 2).

However, the early works of other statisticians have been argued as vital foundations to the developments achieved by Harold Hotelling. One that has been nearly equally acknowledged is Charles Spearman. In 1904, Spearman established techniques characterizing a general version of factor analysis (Meyers et al. 405). As mentioned briefly in the overview of principal component analysis, factor analysis closely resembles the structure and format of the later principal component analysis. In his published work, Spearman "proposed a two-factor theory of intelligence" that "gave way to the extractions of several factors" (Meyers et al. 405).

Still yet, Karl Pearson's published work prior to both Hotelling and Spearman likely provided statistical insight necessary for the success of both subsequent mathematicians. Pearson's basic statistical work at the beginning of the twentieth century provided a platform for raw discussion of the need for a technique like these two factor analyses that were later developed (De Leeuw 2).

Aside from these three primary contributors, many smaller but significant efforts of other mathematicians exist and are worth mentioning for their influence on the development of principal component analysis. Louis Leon Thurstone, a respected mathematician, was working on the further development of factor analysis around the same time frame that Harold Hotelling was developing principal component analysis (Meyers et al. 405). His work on standard deviation and factor analysis likely heavily influenced the simultaneous work of his colleague.

Another early contributor may be Francis Galton. Dating back to the nineteenth century, Galton worked heavily in the field of classical analytic geometry. One of his greatest accomplishments was the development of "principal axes [that] are connected for the first time with the 'correlation ellipsoid'" (De Leeuw 1-2). This ellipsoid in particular was used later on by Pearson to "[cast] the problem in terms of finding low-dimensional subspaces" (De Leeuw 1-2). While it is unknown whether or not Galton was aware of the potential for principal axes, they set the stage for the later development of principal components in other geometric shapes.

IV. Mathematical Methods through the Lens of Personal Development

Upon review of the mathematical methods utilized to undergo this research, perhaps the most significant developments were of personal concern. As I progressed throughout this year-long research, I transitioned from topic to topic. Some topics proved fruitful in my pursuit of applying mathematics to athletics. Others did not. With each connection I discovered, my critical thinking and mathematical reasoning matured. While these appear to be individual attributes, they were crucial to the development of my conceptual understanding of principal component analysis.

Over the course of these two semesters, I have had the opportunity to explore an extension of both my linear algebra and statistics courses through their merging into one holistic idea: principal component analysis. While my path is now clear, my journey did not begin this way. Originally, my research began in linear algebra. As an athlete, my goal with Honors Thesis and Mathematics Research was to apply a linear algebra concept in a practical and personal way by connecting it to the sport I love: track and field. When considering which topic to explore, my mind went to quaternions. There, I foresaw a relationship between quaternion rotations and the dynamic movements associated with throws. I also hypothesized potential links to figure skating.

To understand this concept, I began working through prerequisite material that is traditionally covered in Linear Algebra II. The majority of this work consisted of mastering the content procedurally, being able to represent and evaluate the necessary computations. As I began this process, I also started searching for scientific literature that supported athletic applications. Unfortunately, I quickly realized that these two topics were not as closely related in verified research as I had speculated. While disheartened by this, I continued to work through the Linear Algebra sections. As I progressed into chapter six of Elementary Linear Algebra: Applications Version, I found the Gram-Scmidt Process, QR-Decomposition and approximation theories. These more advanced topics proposed significant challenges at first because of their differences in conceptual processing than the material I had previously studied. However, through reading the theory in this textbook as well as others, I finally began to grasp the concepts. To solidify my understanding, I also underwent the process of proving various linear algebra theorems and conjectures.

In the midst of this process, Dr. Poliakova and I had weekly conversations to discuss the importance of this new content. Because I still longed to apply my research to the biomechanics of track and field, we continued to consider how each of these concepts might contribute to one's understanding of different athletic motions. After much deliberation and conversations with my mother and Gardner-Webb University Head Track and Field Coach Brian Baker, we came across principal component analysis, or PCA. While this topic requires a substantial amount more of background study in order to practice or comprehend, it appeared to also show a plethora of applications.

While it does require more prerequisite material, most of the linear algebra I had spent the first two and a half months studying provided a solid foundation for this extension of my pursuit. At first, I was hesitant about the mathematical branch change from linear algebra to statistics. Out of all the courses I have taken throughout my college career, statistics had presented the largest challenge for me. Additionally, after deciding to select courses other than the second anatomy and physiology, my knowledge of kinematic analysis was limited. However, with the addition of a few extra sections in linear algebra, I was able to start generally exploring principal component analysis.

Very quickly, I realized that the linear algebra and anatomy subjects on which I focused were integral parts as well. After developing a broad view of principal component analysis and its purpose, I transitioned into various chapter readings that explained the last few sections of linear algebra in a way that made the connections between the two fields much more recognizable. I spent the last few weeks of my research semester diving into academic literature that characterized this topic. Although this was frustrating to feel like I was only just beginning my primary topic's research, I gleaned a great deal of knowledge on how the mathematics was applied as well as its importance to bridge the gap between experts in the various fields. I was anxious to start practicing principal component analysis through reading and more studies, evaluating simulations and implementing SPSS technology.

Throughout the second semester, I have continued to read and explore the intricate details of principle component analysis. My primary research has expanded from computational practices and readings to the development of its history and an analysis of its athletic applications. Although the process has been slow due to personal and familial challenges faced this semester, it has been rich. My knowledge of principal component analysis has expanded from procedural and computational fluency of linear algebra topics to a more extensive conceptual understanding through application.

Supported by multiple statistical observations and differentiated representations of data, I have developed a more cohesive image of principal component analysis's wide range of usages. Conceptual understanding was solidified by proofs of various linear algebra theorems and conjectures. Through the continued application of Maple for linear algebra topics and SPSS for principal component analysis, this growth has expanded its roots into the soils of other branches of mathematics. This, therefore, makes it inclusive to much of the overall field of mathematics.

V. Review of Literature

The interpretations and conclusions drawn within this thesis pertaining to the importance of principal component analysis for athletic biomechanics are not without evidential precedent. Prior to this year-long research, countless statisticians, kinematic physiologists and researchers have contributed to the investigation, incorporation and substantiation of principal component analysis. Although rooted in mathematics, experts in various fields have utilized principal component analysis to organize and qualify data representative of varying platforms. Therefore, before demonstrating its importance in understanding the biomechanics of running in a later section of this thesis, this review of literature strives to exemplify its integrity in a variety of other kinematic contexts.

Within the realm of biomechanics, principal component analysis has been used to investigate the performance and technique associated with multiple athletic movements. In the study "The application of principal component analysis to quantify technique in sports," scientists utilized principal component analysis to demonstrate its effectiveness in assisting coaches and athletes with better understanding the intricacies of their sports' respective techniques (Federolf et al. 491). While coaches often observe a progression as "the whole-body movement of an athlete," Federolf et al. used principal component analysis to break down skier posture and technique into smaller, individual movements (492). By doing this, they were able to quantitatively compare the techniques of six different alpine skiers at a much more microscopic level.

In data collection, they were able to categorize ski runs into principal movements based on joint flexion-extension, postural body inclination, distance between skis and several other factors. For all participants, the first principal movement, denoted PM_1, represents "a change of posture that enabled frontal plane body inclination through outer leg extension and inner leg flexion" (Federolf et al. 494). It was also observed that most subjects had similar principal movements but performed them in different chronological sequences. Whereas an outside observer would see all of these factors as one motion, the precision of principal component analysis shed light on these minute differences that affected individual performances. Although sports scientists still face the challenge of "finding an appropriate quantitative methodology that incorporates the holistic perspective of human observers," this study supports the stance that principal component analysis can be a "strategy to handle and extract useful information from [kinematic] data sets in a wide variety of sports" (Federolf et al. 491, 498).

Principal component analysis has also been used to quantify variance and pattern when applied to variables within prolonged intermittent exercise. While prolonged intermittent exercise sports such as tennis and soccer have been associated with "fatigue-related decreases in physical performance in high-intensity running," the retention of comparable vertical-jump heights over time has raised many biomechanical questions (Schmitz et al. 319). To investigate this biomechanical peculiarity, scientists used principal component analysis to quantify the results of sixty collegiate athletes' participation in intermittent exercise for as long as they could maintain their maximum jump height. As a result of this study entitled "Lower-Extremity Biomechanics and Maintenance of Vertical-Jump Height During Prolonged Intermittent Exercise," Schmitz et al. concluded that there were five principal components that contributed to this retention (324).

By the identification of these five components, it may be concluded that their maintenance is necessary for maximal jump height to be continually achieved. While many variables likely contribute, these principal components represent the total variance of factors involved. Moving forward, these participants and their coaches have a basis for measuring jump efficiency while participating in their respective intermittent exercises. Through the incorporation of this study, it may be concluded that principal component analysis is an effective means of interpreting unexpected or peculiar results.

Another athletic field in which principal component analysis has been effectively applied is resistance training. In a study conducted by Sato et al., scientists observed the kinematic positioning of twenty-five participants (n=25) in barbell back squats. Attaching five different receptors to different points along their legs from their hips downward, scientists quantified ten different measurements indicative of individual squat techniques. With the assistance of SPSS technology, they identified two principal components: one above the hip and one below the hip (Sato et al. 4). With this data, they concluded that this data reduction would allow coaches to further focus their efforts on developing the techniques of these two components. This proves principal component analysis's usefulness by minimizing time and energy that would have been otherwise exerted on addressing other elements less essential to an athlete's barbell back squat technique.

Aside from prominently athletic movements, principal component analysis has also been effectively used to quantify other biomechanical efforts. A study entitled "The Force Synergy of Human Digits in Static and Dynamic Cylindrical Grasps" sought to implement principal component analysis to demonstrate the primary movements that overarch the complexities of the human hand (1). Anatomically, the human hand's many muscles are interwoven. Thus, many of these muscles contribute to movements of several digits simultaneously and at different times. In the past, this has led to difficulties in quantitative analysis. However, Kuo et al.'s efforts in this study work are aimed to construct principal components to exemplify the inner workings of a human's cylindrical grasps.

Ultimately, what they found was that principal component analysis was able to establish principal components even amongst the overlaps. The resulting first principal component, PC1, was a result of both the index and middle fingers (Kuo et al. 5). However, the thumb and ring fingers were also recognized as contributors to the first component. This yielded approximately 70% of the variance for vectors in the anti-gravity direction and 97% of the variance in the direction toward the glass simulator (Kuo et al. 4). These results demonstrate the success of principal component analysis in extracting principal components from even the most complex data sets. Therefore, further research using principal component analysis to analyze the human hand holds the potential for enhanced understanding of its anatomical and physiological make-up.

Principal component analysis has also been demonstrated as a tool in data interpretation for biomechanics in clinical settings. In a tutorial study entitled "PCA in studying coordination and variable," principal component analysis was depicted for its proficiency in data reduction for "both kinematic and electromyographic data sets" (Daffertshofer et al. 417). This was executed through its implantation into various simulations that assessed the principal components on walking form. Results indicate that principal component analysis not only contributed in vast data reduction for more efficient interpretation of these whole-body movements, but also provided a "data-driven filter" through which to observe the overall biomechanics (Daffertshofer et al. 424).

Although it is often used within biomechanics as a preventative measure used to retain certain components, clinical biomechanical analysis may also be extended to better understand the different characteristics associated with preexisting conditions and disorders. In a similar manner, principal component analysis can also be used to identify principal components that are indicative to those who are suffering. By identifying these components, this experimental group may be effectively compared to a control group in order to interpret the underlying issues and plan future steps of action. One study that executes this usage of principal component analysis is one entitled "Biomechanical features of gait waveform data associated with knee osteoarthritis: An application of principal component analysis."

In this study, participants with knee osteoarthritis were paralleled to a control group to identify differences in gait biomechanics. With the assistance of technology, scientists collected three-dimensional data on all participants as they completed five walking trials at their own pace (Deluzio et al. 87). Overall, the results yielded eight principal components. Upon comparison of the scores for the control and experimental groups, Deluzio et al. found that four of these indicated characteristic differences between the two groups (91). These features include "the amplification of thee flexion movement, the range of motion of the flexion angle, the magnitude of the flexion moment during early stance, and the magnitude of the adduction moment during stance" (Deluzio et al. 86). Therefore, they were able to conclude that the study of these four principal components would better help healthcare professionals when diagnosing osteoarthritis. This result would not have been possible without the reductive and extractive properties of principal component analysis.

Demonstrated by each of these diverse kinematic applications, principal component analysis clearly has a place within the realm of biomechanics. It bridges the gap in analysis for movements from athletic technique and resistance to training across the spectrum to common movements such as cylindrical grasp and osteoarthritic joints. Still, principal component analysis may be further justified in everyday areas of life outside of sports and biomechanics. Over time, principal component analysis has become a revered tool in several fields, including neuroscience, genetics, demography, risk analysis, computer graphics and quality control (Suryanarayana and Mistry 21).

As a global example of this quality control along with the application of agricultural concepts, one study conducted by colleges in China sought to quantify the efficiency of a water irrigation system using principal component analysis (Jia et al. 1). Because China's water irrigation efficiency is currently less than 50%, their goal was to identify principal components for correction and future improvement. To execute this analysis, scientists identified twenty-two potential variables that affect the water irrigation system's efficiency and ran them through both multiple stepwise regression (MSR) and principal component analysis. With the organizational reduction provided by principal component analysis, they found that their data had five principal components (Jia et al. 8).

Upon reflection of the large difficulties surrounding China's irrigation system, this reduction to five components from the original twenty-two factors allows for the narrowing of politicians' and scientists' focus on addressing their efficiency issues. It is much more tangible to address five issues than it is to be overwhelmed by twenty-two. Through this organization, addressing the principal components would allow the underlying issues of all factors to be addressed. This very much resembles a "kill two birds with one stone" concept.

In addition to providing practical guidance for problem-solving, principal component analysis was also used in this study to fill in the gaps left by multiple stepwise regression. Jia et al. acknowledged a limitation of MSR as being unable to reduce the data that is to be interpreted. Additionally, because multiple stepwise regression addresses linearly correlated variables, principal component analysis provided a process for analyzing the data while maintaining the integrity and validity of the factor analysis results (Jia et al. 1).

Each of the studies mentioned above depicts the many positive attributes of principal component analysis as well as its ability to represent large sets of data with tangible, quantitative efficiency. However, it is equally important to acknowledge that other statistical methods may be favored for data depending on the field. With this being said, principal component analysis is a simple statistical tool that may be used effectively over a broad range of research categories. Established in many fields, future implementation holds the promise of increased human understanding concerning everyday events. Future research marks the doorway into even more examples of principal component analysis's potential.

VI. Fundamental Linear Algebra Concepts

Prior to beginning a discussion of linear algebra topics and their relatedness to principal component analysis, it is imperative that one understand the basic definitions and principles of the field itself. This section is structured to provide general bases for various linear algebra concepts that are further applied throughout this thesis. It is also important to note that while this section provides an overview of general forms for the various mathematical objects, more extensive definitions may be found in any entry-level linear algebra textbook, including Elementary Linear Algebra: Applications Version.

Square matrices are bracketed sets of data that can be used to represent linear systems. They are of the size $n \times n$, are denoted with uppercase letters, and are represented using the following notation:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

The subscripts associated with each entry represent the entry's location in the matrix by row and column, respectively. The dots symbolize additional rows and columns for forms that extend beyond the typical 2×2 matrix. Although the above figure represents a square matrix, matrices may have other dimensions, $m \times n$, with m rows and n columns where m and n are not equal.

An **identity matrix** is any $n \times n$ matrix with 1's along the main diagonal and 0's in all other positions. The main diagonal of a matrix extends from the uppermost, left-hand corner to the lower, right-hand corner. It takes the following form:

$$I = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

A **diagonal matrix** is any $n \times n$ matrix in which the only entries that are values other than zero are along the main diagonal. Its general form is as follows:

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \vdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

A **triangular matrix** is any $n \times n$ matrix in which either the entries above or entries below the main diagonal are all zeroes. An **upper triangular matrix** has zeroes for all entries below the main diagonal. A **lower triangular matrix** has zeroes for all entries above the main diagonal. The general forms for both matrices with 4×4 dimension are provided below:

$$\text{Upper triangular} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

$$\text{Lower triangular} = \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

Matrix multiplication is performed to produce an individual product matrix as long as the dimensions of the two matrices A and B are $m \times r$ and $r \times n$, respectively. Once this stipulation has been met, the resulting matrix will have dimension $m \times n$. Matrix multiplication can be described by the following definition and may be repeated for each entry within the matrix:

> *To find the entry in row i and column j of AB, single out row i from the matrix A and column j from the matrix B. Multiply the corresponding entries from the row and column together, and then add up the resulting products.* (Anton and Rores 29)

The **determinant of a square matrix** is a number assigned to the matrix according to certain rules. For a general 2×2 matrix, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the determinant may be evaluated by multiplying entries $a = a_{11}$ and $d = a_{22}$ and multiplying $b = a_{12}$ and $c = a_{21}$, then finding the difference of the products. It may be denoted using either of the following two notations interchangeably:

$$\det(A) = ad - bc$$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

For matrices of dimensions larger than 2×2, a variety of other methods may be used to evaluate their determinants. These methods include cofactor expansion, row reduction, geometric interpretation and inspection. For diagonal and triangular matrices, the determinant may be calculated by multiplying all values along the main diagonal.

The **inverse of a matrix $\boldsymbol{A}$** is denoted A^{-1}. If the determinant of a matrix is equal to zero, the matrix is not invertible. For matrices where $ad - bc \neq 0$, the inverse of a given matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, may be computed by the following formula:

$$A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

The **transpose of a matrix $\boldsymbol{A}$**, denoted A^T, represents the alteration of a matrix through the interchanging of rows and columns. It may be calculated for any $m \times n$ matrix. An example of this for a 2×3 matrix is exhibited below:

$$A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

$$A^T = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}$$

The **trace of a square matrix A**, denoted $tr(A)$, is the sum of the entries on A's main diagonal. For matrices of dimensions $m \times n$, where $m \neq n$, the trace of A is undefined. According to Elementary Linear Algebra: Applications Version, its general form for a matrix of 3×3 dimension is as follows in Figure 1:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$\mathrm{tr}(A) = a_{11} + a_{22} + a_{33}$$

Figure 1: Formula for Trace of a Matrix; Anton, Howard, and Chris Rorres. *Elementary Linear Algebra: Applications Version*. 11th ed., Wiley, 2013, p. 36.

Vectors are objects with both length and direction. They are often denoted as arrows, with the tail of the arrow being the initial point and the tip of the arrow being the terminal point. A visual representation of a vector may is depicted in Figure 2 as follows:

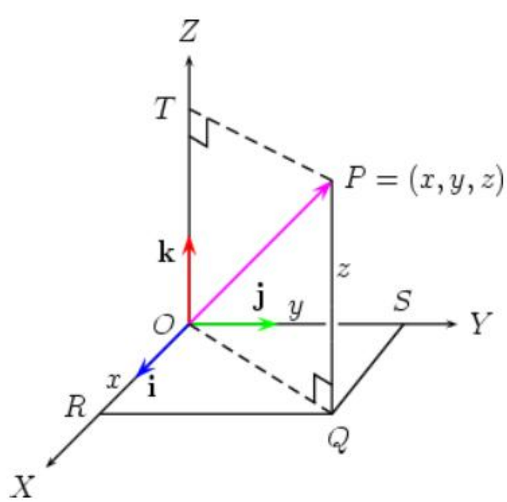

Figure 2: Image of a Vector in 3-D Space; "Vectors." *The University of Sydney - School of Mathematics and Statistics*, The University of Sydney, 9 Nov. 2009, www.maths.usyd.edu.au/u/MOW/vectors/vectors-7/v-7-2.html.

While this indicates a vector in the positive direction for all components, it is important to note that the direction of the arrow may be in any orientation. Notation for a vector with initial point A and terminal point B is as a lower-case letter and is represented as follows:

$$\mathbf{v} = \overrightarrow{AB}$$

It may be represented as a parenthetical list of its components as follows:

$$\mathbf{v} = (v_1, v_2, \ldots, v_n)$$

It can also be denoted using a column matrix with $n \times 1$ dimension as demonstrated below:

$$\mathbf{v} = \begin{bmatrix} w_{11} \\ \vdots \\ w_{n1} \end{bmatrix}$$

The **norm** of a vector **v**, also known as the length or magnitude of that vector is denoted $\|\mathbf{v}\|$. For a vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, it may be computed by taking the square root of the sum of its squared components, using the following formula:

$$\|\mathbf{v}\| = \sqrt{{v_1}^2 + {v_2}^2 + \cdots + {v_n}^2}$$

Unit vectors are vectors that have a norm value of one. They are denoted with a lower-case **u**. The unit vector for a given vector **v** extends in the same direction as the original vector and may be computed using the following formula:

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$$

VII. Implementing Linear Algebra

Matrices

Within any mathematical process that takes place, the first order of business is establishing a method of recording and organizing one's data. Generally, tables and graphs provide sufficient structure for mathematical data. However, linear algebra relies on the heavy implementation of matrices. As it has already been mentioned above, a matrix is a form of representing linear equations or systems. In most cases, matrices only contain the coefficients of their respective data.

For principal component analysis, the usage of matrices is an integral part of the process. Before data that has been collected can be properly analyzed, it must first be incorporated into matrices. For each item being assessed within the data system, multiple factors are often assessed. Each item may therefore be represented as a vector in matrix form with entries being designated for specific factors for consistency purposes (Jauregui 1). Since matrices are row and column specific, this allows computations to be executed with ease instead of exhaustion over sifting through each individual record.

To illustrate this visual representation and use of linear algebra in principal component analysis, consider the following example of data taken from a practice problem in Linear Algebra and Its Applications (Lay 489). As a practical application, this is similar to collecting the heights and weights of six different people. Table 1 indicates the two variables categorized for six different individuals. For mathematical manipulation for principal component analysis, the data from Table 1 was placed in Matrix 1, with each column representing a different individual and each row representing height and weight, respectively.

	Individual #1	Individual #2	Individual #3	Individual #4	Individual #5	Individual #6
Height	19	22	6	3	2	20
Weight	12	6	9	15	13	5

Table 1: Data taken from Linear Algebra and Its Applications, p. 489, #2.

$$\begin{bmatrix} 19 & 22 & 6 & 3 & 2 & 20 \\ 12 & 6 & 9 & 15 & 13 & 5 \end{bmatrix}$$

Matrix 1: Data from Table 1

Once data has been organized in entry-specific matrices, it must be transformed into **mean-deviation form**. Mean-deviation form is executed by calculating the sample mean, **M**, of the data to represent the "'center' of the [data's] scatter plot" (Lay 484). This may be done by the following procedure where $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ are the **observation vectors**, or data taken from each individual. First, the sample mean is calculated:

$$\mathbf{M} = \frac{1}{N}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_N)$$

$$\mathbf{M} = \tfrac{1}{6}(\begin{bmatrix} 19 \\ 12 \end{bmatrix} + \begin{bmatrix} 22 \\ 6 \end{bmatrix} + \begin{bmatrix} 6 \\ 9 \end{bmatrix} + \begin{bmatrix} 3 \\ 15 \end{bmatrix} + \begin{bmatrix} 2 \\ 13 \end{bmatrix} + \begin{bmatrix} 20 \\ 5 \end{bmatrix})$$

$$\mathbf{M} = \begin{bmatrix} 12 \\ 10 \end{bmatrix}$$

Then mean-deviation form depicts the data in a $p \times N$ matrix of the following form

$$B = [\widehat{\boldsymbol{X}_1} \quad \widehat{\boldsymbol{X}_2} \quad \cdots \quad \widehat{\boldsymbol{X}_N}]$$

where $\widehat{\boldsymbol{X}_k} = \boldsymbol{X}_k - \mathbf{M}$ when $k = 1, \dots, N$ (Lay 484). In this example, mean-deviation form is as follows:

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

The versatility of matrices themselves allows manipulation while retaining the integrity of the data it contains. This is heavily relied upon as large data sets are reduced or rotated in principal component analysis (Meyers et al. 428). Covariance matrices and orthogonality each provide useful structures for representing principal component analysis procedures and results. As a part of extraction, principal component analysis generates ones along the main diagonal of the correlation matrix (Meyers et al. 423).

Using the described mean-deviation form, another linear algebra matrix form known as a **covariance matrix** should be computed. The covariance matrix, S, is a $p \times p$ matrix in which the mean-deviation form is transformed by the following definition:

$$S = \frac{1}{N-1} BB^T$$

In this example of Olympic running finalists, the approximate covariance matrix is as follows:

$$S = \frac{1}{6-1} \begin{bmatrix} 7 & 10 & -6 & -9 & -10 & 8 \\ 2 & -4 & -1 & 5 & 3 & -5 \end{bmatrix} \begin{bmatrix} 7 & 2 \\ 10 & -4 \\ -6 & -1 \\ -9 & 5 \\ -10 & 3 \\ 8 & -5 \end{bmatrix} = \begin{bmatrix} 86 & -27 \\ -27 & 16 \end{bmatrix}$$

Eigenvalues, Trace and Eigenvectors

The field of linear algebra defines an eigenvector of a given square matrix A as a non-zero vector that when multiplied by the matrix A is equal to the product of itself by some scalar. This may be formulated as an equation with matrix A, eigenvector **x** and scalar λ as in Figure 3:

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

Figure 3: Formula for Relationship between an Eigenvector, an Eigenvalue, and the Original Matrix; Anton, Howard, and Chris Rorres. *Elementary Linear Algebra: Applications Version*. 11th ed., Wiley, 2013.

This scalar λ is called an eigenvalue of A. In linear algebra, eigenvalues may be computed through the development of a characteristic equation. Formally, the characteristic equation may be computed using the following formula:

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

For principal component analysis, eigenvalues are also valuable for the interpretation of data. Also known as characteristic roots, the eigenvalues of a data set are the "sum of squared correlations for each component over the full set of variables" (Meyers et al. 420). The squared values within this sum are the distances between variables within a given principal component and may also be identified as **component loadings** (*Introduction to Principal Components and Factor Analysis*). Therefore, each principal component has an eigenvalue that reflects the variation present within it.

For example, using the covariance matrix S, one may find the eigenvalues using the following procedure:

$$\det(\lambda I - S) = \det\left(\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} 86 & 27 \\ 27 & 16 \end{bmatrix}\right) = \begin{vmatrix} \lambda - 86 & 27 \\ 27 & \lambda - 16 \end{vmatrix}$$

$$= (\lambda - 86)(\lambda - 16) - (27)(27)$$

$$= \lambda^2 - 102\lambda + 647$$

By setting this quadratic equation equal to zero and implementing the quadratic formula, the eigenvalues of this data were computed to be $\lambda_1 = 95.2$ and $\lambda_2 = 6.8$.

Additionally, the sum of all eigenvalues equals the total variance. Thus, the presence of eigenvalues in principal component analysis is significant because they reflect a part of the whole. They allow statisticians to numerically comprehend the distances between any entry in the set of data and its principal component (Meyers et al. 420).

The total variance of a data set may be computed by calculating the trace of its covariance matrix. This is accomplished by adding all values along the main diagonal. The total variance of S is as follows:

$$\operatorname{tr}(S) = 86 + 16$$

$$\operatorname{tr}(S) = 102$$

Eigenvectors, therefore, are used to represent the principal components themselves within principal component analysis. From a linear algebra standpoint, eigenvectors represent linear transformations of a given square matrix. In a similar manner, the principal components are linear combinations that represent the variance in the original data set

(*Introduction to Principal Components and Analysis*). They are modified by weighting based on their orthonormal contribution to the total variance present (Jauregui 6).

In the example illustrated above, principal components, or eigenvectors, may be calculated as follows:

For $\lambda = 95.2$:

$$\begin{bmatrix} 9.2 & 27 \\ 27 & 79.2 \end{bmatrix} \mathbf{x} = \vec{\mathbf{0}}$$

By interpreting this equation using row reduction, one may calculate the following linear combination, $9.2\mathbf{x}_1 + 27\mathbf{x}_2 = \vec{\mathbf{0}}$, which simplifies to $\mathbf{x} \sim \begin{bmatrix} 2.93 \\ -1 \end{bmatrix}$. To convert it to a unit vector length, it may be divided by its norm: $||\mathbf{x}|| = \sqrt{2.93^2 + 1^2} = 3.10$. Thus, the corresponding eigenvector for eigenvalue $\lambda = 9.52$ is as follows:

$$\mathbf{x} = \frac{1}{3.10} \begin{bmatrix} 2.93 \\ -1 \end{bmatrix} = \begin{bmatrix} .95 \\ -.32 \end{bmatrix}$$

For $\lambda = 6.8$:

$$\begin{bmatrix} -79.2 & 27 \\ 27 & -9.2 \end{bmatrix} \mathbf{x} = \vec{\mathbf{0}}$$

By interpreting this equation using row reduction, one may calculate the following linear combination, $-79.2\mathbf{x}_1 + 27\mathbf{x}_2 = \vec{\mathbf{0}}$, which simplifies to $\mathbf{x} \sim \begin{bmatrix} .34 \\ 1 \end{bmatrix}$. To convert it to a unit vector length, it may be divided by its norm: $||\mathbf{x}|| = \sqrt{.34^2 + 1^2} = 1.06$. Thus, the corresponding eigenvector for eigenvalue $\lambda = 6.8$ is as follows:

$$\mathbf{x} = \frac{1}{1.06}\begin{bmatrix} .34 \\ 1 \end{bmatrix} = \begin{bmatrix} .32 \\ .94 \end{bmatrix}$$

VIII. Importance in Biomechanics

In biomechanics, the importance of principal component analysis is evident in its application concerning data and its subsequent interpretation. While it has been established in various biomechanical branches in previous research, its application regarding running form and technique is still expanding. This is likely due to the fact that running is one of few sports that requires one to use nearly all of his or her muscles at one point or another. Unlike simpler movements, the data sets may be excessively large. Because of the vast capabilities of principal component analysis regarding data set size, this is all the more a reason to use it to quantify running biomechanics. Therefore, the purpose of this portion of this thesis is to synthesize past research studies utilizing principal component analysis on running to establish its potential in the future development of the sport. General kinematic background material necessary for understanding biomechanics is also provided.

As defined by Roberston et al., three-dimensional kinematics is the "description of motion 3-D space without regard to the forces that cause the motion" (35). Coordinate systems and segments are concepts used to mathematically quantify anatomical body structures and positions. Most often, joints are expressed as the terminal points with vectors extending in the directions of the limb and options for movement. Range of motion (ROM) encompasses the range of angles in which a joint may be flexed, extended or rotated. Biomechanics is similar to linear algebra and principal component analysis in that it also implements linear transformations which may be interpreted.

While raw computations may be used to evaluate these angles and rotations, obtaining biomechanical data has been expedited by technological development. This has been deemed particularly advantageous in collecting and interpreting large data sets. If movement is expressed as a function of time, it is often generated as waveform data (Robertson et al. 317). In cases with many participants, a separate waveform may be displayed for each. Their visual similarities and overlaps depict correlations and foreshadow relationship development.

Although visual interpretations are beneficial, they are not all-inclusive. Having a plethora of data present may make it difficult to mathematically analyze. Principal component analysis provides a method of data reduction that develops the most important components while maintaining the shape and integrity of the original data set. It can also be used to quantify discrimination between groups and in the development of future hypotheses (Robertson et al. 319). In the world of running, being able to discern between good technique and bad technique is essential if one wants to improve. Deciding which part of one's form is counterproductive is inefficient may be difficult because of how much of one's body is involved. However, with research implementing principal component analysis, optimizing one's personal performance is no longer a fantasy.

In a study entitled "Motor Patterns in Human Walking and Running," scientists did an elementary study on just that. The objective of their study was to use principal component analysis to compare and discriminate between the kinematic sequences of walking and running (Cappellini et al. 3427). Using nine cameras and infrared reflective markers for data collection, scientists recorded participants' exercise at a variety of speeds on a treadmill. Results indicated that both walking and running have the same five principal components.

However, they may be uniquely identified by the differences in foot-strike timing (Cappellini et al. 3434). Through this implementation, principal component analysis was effective in "standardizing" two seemingly different sets of data.

While many studies are structured to interpret qualitative information like those above, other studies seek to quantify the differences between athletic performances based on age, gender, experience level, training background and many other variables. From a competition standpoint, these studies yield a practical application with which most athletes are primarily concerned. Based on the amount that one athlete trains, is there really a difference that sets one apart as a more proficient athlete or is it all based on luck and natural abilities? Does the fact that I am a female naturally dispose me to certain techniques that are indicative of less efficient movement than those of my male teammates?

Although it may seem impossible to ask these qualitative questions, principal component analysis may be used to quantify them. This may be examined through a study that chose to use a tri-axial accelerometer to examine the differences between soccer players, new runners and experienced marathoners (Kobsar et al. 2509). In this study, scientists used a single tri-axial accelerometer to capture data pertaining to the treadmill running patterns of its participants. From the forty-four variables that were classified, principal component analysis was used to retain eight principal components (Kobsar et al. 2509). These results indicated that the accelerometer was capable of correctly classifying 85% of the runners based on their training backgrounds (Kobsar et al. 2508). While it could not differentiate between experience levels, principal component analysis proved to be a sufficient means of organizing other biomechanical measurements for assessment.

Outside of the athlete's mentality of winning, principal component analysis may be used in other ways to assess one's running. It may also be used to analyze the influences of various efforts, conditions and other elements on one's performance. In a study focused on evaluating the influence of exertion on leg joint mechanics while running, scientists were purposed with defining its relationship to athletic injury (Benson and O'Connor 250). As a long-distance runner who broke my leg racing in a collegiate track and field 5,000 meter race from simply overexerting on a preexisting stress fracture, I can personally attest to the usefulness of this application.

In this study, female runners with no preexisting injuries underwent pre-run, examination and post-run series observations. The pre-run and post-run assessments were generated using retroflective markers at the major joints in order to quantify the differences between pre- and post-workout. Afterwards, principal component analysis was used to depict significant, mathematical changes in joint angles and movements. Upon observation, results indicated that there were discrepancies in movements of the ankle, knee, and hip (Benson and O'Connor 253). Ankle eversion was greater after participants ran and can be associated with unnecessary tibial rotation. These connections imply that principal component analysis's reduction to conclusions regarding technique should be used to design programs that protect athletes from naturally occurring injuries.

While each of these studies depicts a different application within the field of biomechanics, each is representative of a body of studies that have been developed on their respective topics. Each study demonstrates principal component analysis's usefulness in evaluating biomechanical data within running. However, in order to achieve a more complete understanding of the concepts quantified in each, more research is needed. Principal component analysis offers a promising future in understanding the inner workings of running biomechanics in a more holistic way. It provides coaches and athletes alike with mathematical representations to fill the gaps within their visual observations. All that is necessary is more detailed exploration.

IX. Connections to Other Branches of Mathematics

With respect to the overall purpose of this thesis, principal component analysis was depicted through the perspectives of linear algebra and biomechanics. In previous sections, concepts native to linear algebra were outlined as significant contributors to this tool for data evaluation. Applications were demonstrated for the field of athletic performance. As mentioned earlier, these extensions are not solely of academic reward. Principal component analysis also provides athletes and coaches with necessary tools to improve movements vital to their respective sports. In addition to this interdisciplinary illustration, principal component analysis can also be found extremely useful in other areas in the field of mathematics.

In its most basic form, principal component analysis is rooted in statistics. While linear algebra has been demonstrated as its foundation, its usefulness is still primarily as a purely statistical measure. Aside from analysis of biomechanical data, principal component analysis can also be used to statistically represent other fields. Several fields include, but are not limited to, finances, the stock market, systematic risk, and environmental conditions. Not only can principal component analysis be used to identify patterns, but it may be further used to make predictions in each field.

In addition to explicitly statistical fields, principal component analysis has also been acknowledged as a vital contributor for computations and interpretations of data within the field of physics. Even coined as "quantum principal component analysis," principal component analysis is specifically used to provide physicists with a quantitative method for evaluating an unknown quantum state. This unknown quantum state is primarily denoted ρ. Quantum tomography is a widely used practice for "discovering features of an unknown quantum state" (Lloyd et al. 631). Because the state is often composed of multiple copies or dimensions, it is hypothesized to be an aid in its own analysis.

Once eigenvalues and eigenvectors have been established, the formation of a density matrix as well as the extraction phase of principal component analysis are useful in representing the data in a concise way. Through extraction, key features of the unknown state may be identified. Upon further development, quantum principal component analysis (qPCA) may lead to the construction of principal components "in time $O(\log d)$, an exponential speed-up over existing algorithms" where d denotes the number of dimensions in a Hilbert space (Lloyd et al. 631). Moreover, qPCA opens the door for future physical understanding of the unknown state as well as other established states, such as the Choi-Jamiolkowski state (Lloyd et al. 632).

Another application within the realm of physics may be found in the study of turbulence. Turbulence is acknowledged by the scientific community as the irregular flow of a fluid in which its speed undergoes drastic changes (Serway et al. 339). In a study entitled "Principal Component Analysis studies of turbulence in optically thick gas," scientists used several data analysis methods to evaluate the sensitivity of the velocity power spectrum in opaque gaseous forms (Correia et al, 1). Using Position-Position-Velocity (PPV) cubes, they investigated both fractional Brownian motion (fBm) and magnetohydrodynamics (MHD) simulations for changes and components of PPV cubes in a wide range of opacities.

Using principal component analysis, Correia et al. found that principal component analysis retained its effectiveness in detecting velocity and density spectra changes even within high opacity environments (5). In doing so, they concluded that principal component analysis could be a "valuable tool for studies of turbulence at high opacities provided that the proper gauging of the PCA index is made" (Correia et al. 1). In essence, this study's results indicated that principal component analysis's usefulness supersedes other methods in some areas of physics. While there is still much research to be done to address inconsistences and irregularity in some results, physicists still are beginning to quantify turbulence in ways they had not been able to before.

X. Future Significance

As illustrated in the various sections throughout this thesis, the implementation of principal component analysis has a profound impact on the field of biomechanics. Through its extensive use of linear algebra and statistics, it has been deemed an effective method for interpreting and analyzing data, reducing data into primary variables and providing a platform for further hypotheses regarding the nature of the research being evaluated. Not only has it extended into the mathematical field of physics and personal movement and athletic motion, but it permeates a variety of other practical platforms. Therefore, current research results yield its success across a broad spectrum. However, principal component analysis still has a plethora of untapped potential yet to be discovered and observed.

Within the large field of mathematics, principal component analysis has been observed through the perspectives of statistics, linear algebra and physics. Concepts in all three platforms are fundamental to one's understanding of principal component analysis computations. In turn, principal component analysis is established as a meaningful method of interpreting everyday applications of these fields. Moving forward, more extensive understanding of these applications as well as the discovery of other applications are to come.

Personally, I hope to continue developing my conceptual understanding of the linear algebra inner workings of principal component analysis. As a statistical tool that is primarily used through the assistance of technology, it is challenging to exhaustively understand the procedural progressions that principal component analysis utilizes. While I feel that I have a solid understanding of how covariance matrices, orthogonality, eigenvalues and eigenvectors are critical components of the process, there are still some areas where I must turn to technology, software and research to fill the gaps. A goal of my future research in this field will be to discover other linear algebra concepts that are useful in bridging these gaps.

Pertaining to the field of biomechanics, there are also a several kinematic opportunities for the development of principal component analysis. While it has been used to analyze data regarding the mechanics of walking, athletic movements and physical therapy related conditions, its results for future improvement have mildly been explored. More specifically, the population of coaches and athletes that review this data as if it were "game-time film" remain in the vast minority. In the future, the application of principal component analysis's progress may have a lasting impression on performance improvement on the lives of those who take it seriously.

As a runner, I plan to continue studying principal component analysis from this much more personal perspective. Throughout my athletic career prior to college, I was always referred to as “T-Rex” for the way I carried my arms. Upon coming to college, Coach Brian Baker has helped me to identify the cause of this as well as other motions that hinder my speed and comfort when I race. While I have improved throughout my time at Gardner-Webb, I plan to pursue future opportunities where I can use principal component analysis to further characterize my own athleticism. As an aspiring teacher and occupational therapist for the special needs’ population, I conjecture that there will also be ways through which I can also apply principal component analysis for my students in both areas. Not only will this improve my athletic performance and the biomechanics of my students’ movements, but it will solidify my conceptual understanding of this application of principal component analysis.

In conclusion, the future implementation of principal component analysis is bright and unlimited. Although sometimes viewed as primitive and basic in a world full of statistical analysis methods, it contains foundational mathematics that may be universally effective in positively impacting the average individual. While its efficiency in the fields of mathematics and biomechanics have been emphasized within this thesis, practical use on everyday events are budding. Mathematicians, coaches, and actuaries alike will soon be able to investigate their respective fields and make accurate hypotheses that shape and improve the world population’s daily living experiences.

XI. Works Cited

Almonroeder, Thomas G., et al. "The Effect of a Prefabricated Foot Orthotic on Frontal Plane Joint Mechanics in Healthy Runners." *Journal of Applied Biomechanics*, vol. 31, no. 3, June 2015, p. 149. EBSCO*host*, ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=103119659&site=eds-live.

Anton, Howard, and Chris Rorres. *Elementary Linear Algebra: Applications Version*. 11th ed., Wiley, 2013.

Benson, Lauren C. and Kristian M. O'Connor. "The Effect of Exertion on Joint Kinematics and Kinetics during Running Using a Waveform Analysis Approach." *Journal of Applied Biomechanics*, vol. 31, no. 4, Aug. 2015, pp. 250-257. EBSCO*host*, doi:10.1123/jab.2014-0138.

Cappellini, G, et al. "Motor Patterns in Human Walking and Running." *Journal of Neurophysiology*, vol. 95, no. 6, June 2006, pp. 3426-3437. EBSCO*host*, ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=22223461&site=eds-live.

Correia, C., et al. “Principal Component Analysis Studies Of Turbulence In Optically Thick Gas.” *The Astrophysical Journal*, vol. 818, no. 2, 11 Nov. 2015, p. 118., doi:10.3847/0004-637x/818/2/118.

Daffertshofer, Andreas, et al. "PCA in Studying Coordination and Variability: A Tutorial." *Clinical Biomechanics*, vol. 19, no. 4, May 2004, p. 415. EBSCO*host*, doi:10.1016/j.clinbiomech.2004.01.005.

De Leeuw, Jan. “History of Nonlinear Principal Component Analysis.” *Compstat 82*, 1982.

Deluzio, K.J. and J.L. Astephen. "Biomechanical Features of Gait Waveform Data Associated with Knee Osteoarthritis. An Application of Principal Component Analysis." *Gait & Posture*, vol. 25, 01 Jan. 2007, pp. 86-93. EBSCO*host*, doi:10.1016/j.gaitpost.2006.01.007.

Federolf, P., et al. "The Application of Principal Component Analysis to Quantify Technique in Sports." *Scandinavian Journal of Medicine & Science in Sports*, vol. 24, no. 3, June 2014, pp. 491-499. EBSCO*host*, doi:10.1111/j.1600-0838.2012.01455.x.

Foch, Eric, and Clare E. Milner. “The Influence of Iliotibial Band Syndrome History on Running Biomechanics Examined via Principal Components Analysis.” *Journal of Biomechanics*, vol. 47, no. 1, 2014, pp. 81–86., doi:10.1016/j.jbiomech.2013.10.008.

Introduction to Principal Components and Factor Analysis.

ftp://statgen.ncsu.edu/pub/thorne/molevoclass/AtchleyOct19.pdf.

Jauregui, Jeff. “Principal Component Analysis with Linear Algebra.” 31 Aug. 2012.

Jia, Renfu, et al. "Driven Factors Analysis of China’s Irrigation Water Use Efficiency by Stepwise Regression and Principal Component Analysis." *Discrete Dynamics in Nature & Society*, 28 Mar. 2016, pp. 1-12. EBSCO*host*, doi:10.1155/2016/8957530.

Kobsar, Dylan, et al. "Classification Accuracy of a Single Tri-Axial Accelerometer for Training Background and Experience Level in Runners." *Journal of Biomechanics*, vol. 47, no. 10, July 2014, pp. 2508-2511. EBSCO*host*, doi:10.1016/j.jbiomech.2014.04.017.

Kuo, Li-Chieh, et al. "The Force Synergy of Human Digits in Static and Dynamic Cylindrical Grasps." *Plos ONE*, vol. 8, no. 3, Mar. 2013, p. 1. EBSCO*host*, ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=87681988&site=eds-live.

Lay, David C. *Linear Algebra and Its Applications*. 3rd ed., Addison Wesley, 2003.

Lloyd, Seth, et al. “Quantum Principal Component Analysis.” *Nature Physics*, vol. 10, 27 July 2014, pp. 631–633., doi:doi:10.1038/nphys3029.

Meyers, Lawrence S., et al. *Applied Multivariate Research: Design and Interpretation*. 3rd ed., SAGE, 2017.

Millsap, Roger E. "Review: A User's Guide to Principal Components J. Edward Jackson." *Journal of Educational and Behavioral Statistics*, no. 1, 1995, p. 105-107. EBSCO*host*, ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.1165392&site=eds-live.

Robertson, D. Gordon E., et al. *Research Methods in Biomechanics*. 2nd ed., Human Kinetics, 2014.

Sato, Kimitake, et al. "Preliminary Study: Interpretation of Barbell Back Squat Kinematics Using Principal Component Analysis." *International Symposium on Biomechanics in Sports: Conference Proceedings Archive*, vol. 28, Jan. 2010, p. 1. EBSCO*host*, ezproxy.gardner-webb.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=59696201&site=eds-live.

Schmitz, Randy J., et al. "Lower-Extremity Biomechanics and Maintenance of Vertical-Jump Height during Prolonged Intermittent Exercise." *Journal of Sport Rehabilitation*, vol.

23, no. 4, Nov. 2014, pp. 319-329. EBSCO*host*, doi:10.1123/JSR.2013-0065.

Serway, Raymond A., et al. *Physics for Scientists and Engineers*. 9th ed., Cengage, 2014.

Suryanarayana, T. M. V., and P. B. Mistry. "Principal Component Analysis in Transfer Function." *SpringerBriefs in Applied Sciences and Technology Principal Component Regression for Crop Yield Estimation*, 2016, pp. 17–25., doi:10.1007/978-981-10-0663-0_2.

"Vectors." *The University of Sydney - School of Mathematics and Statistics*, The University of Sydney, 9 Nov. 2009, www.maths.usyd.edu.au/u/MOW/vectors/vectors-7/v-7-2.html.

Warner, Rebecca M. *Applied Statistics: From Bivariate Through Multivariate Techniques*. SAGE Publications, 2008.