Ethan Greene

Dr. Poll

# Math 203 Final Project

## Introduction:

In linear algebra, we have a powerful tool that is known as eigenvalues. When utilized correctly these variables can tell us many different things about a database. Examples of how they are used typically pertain to the stability of our data, the structure of our data, or the variance of our data. Though there are many use cases, in this paper we will focus specifically on how eigenvalues help us in a process known as principle component analysis commonly denoted PCA. We will then go on to discuss how PCA can be used in the realm of biomechanics to learn how groundbreaking these tools are.

The next powerful tool we will make use of is called stochastic matrices. These commonly known as probability matrices are used in order to help predict probabilities of future states given a matrix that sums to one. When used correctly we can make decisions that allow us to determine varying outcomes. In this paper, we will see how stochastic matrices allow us to form mathematical models known as Markov Chains After establishing what Markov Chains are we will see how the powerful math model is used in testing various software features.

## Eigenvalues in Studying Coordination and Variability:

In biology, there is a type of study known as biomechanics. This scientific paper takes this study and specifically delves into the area of coordination. Specifically, these researchers are interested in what features are the most important factors to drive coordination. A problem with this research is that in the field of biomechanics, there exists substantial amounts of input. So given all of this data how are we supposed to know the actual ones that matter so we can study those and learn more about the specific matter at hand?

This is where principal component analysis comes into play. The idea is to see where all of the noise in our database is and through PCA eliminate the external factors. It does this by finding what parts of our data share high correlations with a factor. Those components that share a high correlation are kept where data that is lowly correlated can be assumed to be not as useful and in this study are eliminated. Through this, we can study the useful components driving coordination, so how is this done?

While much of the math is very difficult to understand for the purpose of this paper I will speak specifically on how they made use of eigenvalues. They write about how it is important that when eliminating data not relevant we make as few errors as possible. This makes sense as eliminating data that is relevant could make solving our problem impossible. This is where they show in Figure 3 that we can make this equation in terms of an eigenvalue problem. What they do is they find all of the eigenvalues where each $\lambda_k$ tells us the "variance, deviance, or spread". This should begin to sound similar to the driving factor of PCA where we spoke of how we want to keep the data with high correlation and get rid of some of the outlying unexplained data.

The final space I would like to delve into here is how they study these eigenvalues. The rest of the article is mostly results, but eigenvalues show up in every step of these results. This is because they order the eigenvalues from largest correlations to smallest allowing them to really decide which are most important and what to get rid of. In the simulated example, they do by examining that $\lambda_1 = 97$, $\lambda_2 = 3$, and $\lambda_3 = 0$ This clearly shows that we can get rid of our third eigenvalue making our space two-dimensional allowing us to conduct our study on these two factors now.

**Stochastic matrices in Software Test Cases:**

In order to begin creating a Markov Chain we start by defining the different states that exist. These states consist of the different possibilities of a condition. In our stochastic matrix, we make it so that the columns sum to one allowing us to make sure that we clearly have the possibility to transition to another state stated. With our newly formed stochastic matrix, we are able to directly see how we can use them to form Markov Chains. This is because a Markov Chain has a property in which future transitions don't depend on previous states rather they are only dependent on the current state.

In this paper, we want to treat software testing as that of a stochastic process. We know we can use these stochastic processes to form Markov Chains, but why is this necessarily useful? Well, we can use these Markov Chains to represent finite state machines, a tool very commonly used to explore the realm of computing. This is because each Markov Chain is independent of its past and in a state machine that is critical for our definitions to stand. Overall computers are very logical, and using the realm of statistics we can logically study them. This of course is the best way to study them as we can get definitive answers.

Now going from this, they were able to modify their probabilities to refer to what we would like to test. Specifically, they state how we can develop a deeper understanding of many different performance metrics such as; traffic flows, communication networks, and genetic engineering because of these model performances gained from the creation of our Markov Chains. Through studying our transition metrics we have developed a new area of studies showing the power of stochastic matrices when used in the area of computation.

**Conclusion:** In this paper, I wrote about some of the real-life use cases of linear algebra. Through this, we can see how necessary the realm of linear algebra is. Though we have gained many answers from using PCA in biomechanics and stochastic matrices in software testing, we also now can ask more questions. Knowing that PCA was highly useful in our evaluation of the important processes that make up coordination, it begs the question would we be able to use other dimensionality-reducing techniques such as SVD or LDA, and achieve similar results?

Then in the case of stochastic matrices as a metric for testing software, we know that these Markov Chains help us a lot in the area of traditional computers. Now what if we tried to apply them to the area of quantum computing? Of course, there are many difficulties comparing the two areas, however, I believe that these are the essential questions that drive forward our knowledge of these scientific realms. Through the use of linear algebra, we make possible studies that would not be viable to even begin to examine without the magnificent tools two of which we covered in this paper; eigenvalues and stochastic matrices.

# Works Cited

Andreas Daffertshofer a, et al. “PCA in Studying Coordination and Variability: A Tutorial.” Clinical Biomechanics, Elsevier, 20 Feb. 2004, www.sciencedirect.com/science/article/pii/

Gerson Barbosa a e, et al. “A Systematic Literature Review on Prioritizing Software Test Cases Using Markov Chains.” Information and Software Technology, Elsevier, 18 Mar. 2022, www.sciencedirect.com/science/article/abs/pii/