A Study of Advancement of Linear Algebra for the Future Analysis

Byula Parsa[1]*, Dr. Shalu Garg[2]

[1] Research Scholar, Shridhar University

[2] Research Supervisor, Shridhar University

Abstract - Analytical applications for Big Data future analysis frequently make use of linear algebra concepts like linear equations, eigenvalue problems, principal component analysis, singular value decomposition, quadratic forms, linear inequalities, linear programming, optimization, linear differential equations, modeling, prediction, and data mining algorithms. Solving linear equations, multiplying and inverting matrices, computing and interpreting determinants, locating eigenvalues and eigenvectors, and so on are all examples of the computational techniques that make up one branch of linear algebra. Meanwhile, linear algebra also has a theoretical component that deals with things like abstract vector spaces, subspaces, linear independence, spanning sets, bases, dimension, and linear transformations. To go into several cutting-edge areas of linear algebra that reveal the various ways in which this field of study is related to other branches of mathematics. Large-scale network analysis, Leontief economic models which assume that consumption equals output on a national/regional scale and sports team rankings all make heavy use of linear equations and matrices in Future Analysis.

Keywords - Advancement, Linear Algebra, Future Analysis.

- X -

INTRODUCTION

Data in the digital cosmos doubles every two years and is expected to reach 44 zettabytes by 2020. During the last several years, there has been an explosion in the amount of digital information that can be saved. In 1986, there were just three exabytes of data, but by 2011 that number had grown to 300, and by the end of 2020 it might have reached 44,000. In addition, yearly data creation is expected to reach 165 zettabytes by 2025, as reported by the International Data Corporation (Nicolaus Henke, 2016). The proliferation of data is due to the rising expectation that we can and should conduct everything of our daily lives, from business to pleasure, online. Concurrently, we use smart gadgets that are always linked to the internet and provide steady streams of real-time data on anything from our heart rates to our precise whereabouts. Estimates suggest that by 2020, more than four billion people will be using the internet; each would generate 1.7 gigabytes of data every second; and millions of businesses will operate mostly online.

The digital world is further expanded by the millions of sensors and connected gadgets that are constantly uploading new information to the web. By 2020, it's projected that over 10 billion mobile devices will be in use, exponentially increasing the size of the digital world. Almost every industry from government and healthcare to banking and finance to manufacturing and retail to transportation and education is seeing data expansion at an exponential rate. In 2019, Facebook, for instance, managed over 30 petabytes of data from 2.3 billion active users. Around 230,000,000 tweets are sent every day, or over half a million tweets per minute. About a billion people use Google every day, 267 million transactions (4 petabytes) are recorded every day at Walmart, over a billion people watch Netflix every month, and 5 hours of video are uploaded to YouTube every second. In addition, modern telescopes are data-driven, producing vast quantities of information; the Australian Square Kilometre Array Pathfinder (ASKAP) radio telescope, for instance, streams data at a rate of 2.8 Gigabytes per second, and the proposed Large Synoptic Survey Telescope (LSST) will record 15 Terabytes of data every night. Further contributing to our understanding of the cosmos and producing 60 gigabytes of data daily is the Large Hadron collider (LHC), a particle accelerator (Jiyan Yang, 2016).

Due to the unexpected growth of digital information, new avenues for data analysis like Future Analysis have been discovered, along with a plethora of new career paths. Although just 23% of businesses had an enterprise-wide Big Data strategy in 2012, now, 97.2% are making plans to analyze the data they collect. According to a recent poll conducted by Harvard Business Review, 70% of senior executives in the Fortune 500 and government agency business and technology departments aim to recruit data

scientists in the near future. Jobs in data analytics are expected to grow by 34% between 2016 and 2026, according to the BLS's Occupational Outlook Handbook for 2018. Research from the McKinsey Global Institute suggests that there may be a significant need for big data analytical skills, leading to an increase of 50–60% in data analytic positions. Similarly, 2.7 million data science and analytics jobs will be unfilled by 2020, according to a Forbes analysis (John R. Durbin, 2008).

Linear Algebra

Analytical applications for Big Data frequently make use of linear algebra concepts like linear equations, eigenvalue problems, principal component analysis, singular value decomposition, quadratic forms, linear inequalities, linear programming, optimization, linear differential equations, modeling, prediction, and data mining algorithms. Matrix algorithms, in particular, are at the heart of contemporary Big Data analysis. This is due to the fact that matrices provide a flexible mathematical framework for representing information in many different contexts. For instance, a NxD matrix is a convenient way to represent or encode information about N objects, each of which has D attributes. All three of these methods feature extraction, grouping, and classification rely on extensive manipulations of matrices. Principal component analysis uses matrix decomposition to accomplish dimensionality reduction. Similar to how Google's PageRank makes use of eigenvectors (Gilbert Strang, 2006).

The fact that graphs may be used to depict nuances in fields as diverse as computer science, geography, language, and chemistry is testament to their adaptability. With the use of linear algebra, these graphs may be represented as matrices, which finishes the job of improving their computational features. In Big Data contexts, the graph representation of linked data is the norm. In this paper, we introduce the notion of an incidence function G that associates with each edge of G an unordered pair of (not necessarily distinct) vertices of G, and we define a graph G as the ordered pair (V (G),E(G)) consisting of a set V (G) of vertices and a set E(G), disjoint from V (G) of edges. According to this interpretation, a graph's vertices may stand in for anything from individual web pages or genes to individual pixels in a picture or even individual people engaging with one another, while the graph's edges signify the connections or relationships between the vertices (Ron Larson, 2009). Using graph analytics, we may infer properties of the graph such as centrality, shortest route, and reachability. The internet search engine is a common real-world application of huge graph analytics. Future Analysis employs a number of techniques that make use of graphs, including the visualization of enormous datasets as graphs (like the World Wide Web), computation for highly linked large graphs, and the discovery of matchings in bipartite graphs (like online advertisements). Large graph analytics have many real-world uses, such as web search engines. Since graphs might include millions of vertices, their massive

sizes provide a challenge (Roger Horn, 2012). An approach to analyzing and interpreting data based on a comparison of graphs that makes use of a low-rank approximation of the neighboring matrices or graph Laplacians.

Large-scale network analysis, Leontief economic models which assume that consumption equals output on a national/regional scale and sports team rankings all make heavy use of linear equations and matrices in Future Analysis. Google's PageRank algorithm, network clustering, and weather system modeling all make use of eigenvalues and eigenvectors; spectral decomposition, a matrix approximation technique that employs eigenvectors, is employed in spectral clustering, link prediction in social networks, recommender systems with side-information, the densest ksubgraph problem, and graph matchings. Dimension Reduction methods, including picture compression, face recognition, and El Nino, make use of methods like principal component analysis and Singular Value Decomposition to compare the structures of folded proteins. The stable marriage problem, production planning, portfolio selection, transportation problem, minimization of production costs, minimization of environmental damage, and maximization of profits are just some of the many applications of optimization, the minimization of a quadratic expression, and linear programming. Similar items and common patterns notions are being put to use in real-world applications such as facial recognition, fingerprint scanning, plagiarism detection, and Netflix movie ratings (David, 2011).

Infusing Big Data future Analytics in UG Linear Algebra Course

We used a two-part program to include Big Data and promote student engagement in a linear algebra class. The first half included the conceptual and theoretical foundations of the methodologies at hand, while the second half involved practical application to real-world data. For their projects, students are encouraged to utilize either R or Python, two popular general-purpose programming languages. In addition to Microsoft Excel, MATLAB programming is available to the students for use on their project. Our first selection of themes for instructional integration of big data analysis methodologies was based on two criteria: relevance to all computers and mathematics majors and appropriateness of content for such integration. One day, teachers may decide to include Big Data ideas into a wider variety of computer science and mathematics courses. The preexisting linear algebra course was supplemented with the following big data lectures and laboratory modules:

Lecture: The students started with a pretest designed to measure their knowledge of Future Analysis and the applications of linear algebra. To complete the module, a posttest was administered to collect additional data. The teacher introduced a notion of "Big Data" that is most appropriate from a

Byula Parsa[1]*, Dr. Shalu Garg[2]

linear algebraic perspective. Big data, in linear algebraic notation, is information that fits into a mn array where m and n are both very large. The lecture was intended to review material previously covered in the course outline and to deliver new material only if it was crucial to students' grasp of the modules' subject matter.

Linear equations and matrices, eigenvalues and eigenvectors, and singular value decomposition are just some of the subjects that have been integrated into the course thus far. The course covered fundamental applications of linear algebra on matrices, as well as techniques for collecting data and displaying it in matrix form. Data was mostly gathered from government websites like www.data.gov. In addition, the PageRank algorithm and an application were introduced to the pupils. Finally, the Leslie Matrix and demographic shift were offered as new concepts to consider. The complexity of specific algorithms or novel approaches was minimized in examples to help students grasp their significance.

Hands-on activities: Students were asked to:

- Label data sets in ways that adequately characterize their distribution. Students were instructed to apply the real-world big data approaches covered in class to the interpretation of linear equations posed by real-world commercial, tax, economic planning, and analytic issues. All of the data sets used in this section of the exercise were prepared ahead of time so that we could guarantee a few specific results and have some good talking points to share.
- Eigenvalues and eigenvectors may be used to analyze real-world data and reveal hidden patterns. Students were given a wide variety of real-world issues to work on in this lab, including those related to economic development, land analysis, structural engineering applications, control theory, vibration analysis, electric circuits, advanced dynamics, and so on. The second section of the exercise allowed teachers to choose any of the previously mentioned subjects to hone down on.

When more complex computations were required than could be accomplished with basic examples, MATLAB was the main computing program employed. As most of the class had never used MATLAB before, "cheat sheets" with detailed instructions were distributed and used to go through the examples and learn the program. After discussing possible applications of linear algebra to Future Analysis in class, students were given an independent study task to do more research on the topic and solidify their grasp of the material.

Assignment: The "The Mathematics of Google Search" assignment required students to do a

complexity study of the PageRank Algorithm. Models were built using real-world data, but were modified so as not to generate an excessive amount of iterations that would detract from the assignment's intended goal. Students were given a handout titled "Questions for Class Discussion" and instructed to use the questions within to spark classroom debate upon their return to class. The questions were designed to get a sense of the module's worth and the students' attitudes about putting classroom knowledge to practical use. Several criteria were taken into account in order to provide a comprehensive picture of the student's ability. This work did some necessary work on a number of different norms.

The expectations are that students will be able to do the following: collect data, display data graphically, interpret data as a matrix, apply techniques already included in the curriculum, create models, decide on the necessary level of precision, set up a system for storing and retrieving data, organize and present their findings, and defend their reasoning.

These measures of task and class discussion success were determined by the following criteria:

- accuracy of calculations
- accuracy of models and graphs
- usage of algorithms
- organization of calculations
- clear explanations

the culmination of the module came in the form of a posttest designed to, among other things, show if the students had a better understanding of the topic than when they began.

RESULTS

created Big Data modules to be included into preexisting Intro to Linear Algebra and Discrete Math courses. Many revisions were made to these Big Data courses, some in response to student comments and others to accommodate newer technological standards. Pre- and post-tests were used to gauge the usefulness of these modules. Furthermore, all enrolled students were required to fill out a survey on their academic experiences, attitudes about incorporating big data modules into their courses, and preferred methods of acquiring mathematical knowledge.

Student Knowledge: Each class took a pre- and post-test to compare student growth during the module's implementation. Not every student in every class was able to finish either the before or post exam. The average score on the pre-tests was just 36.63 percent, but the post-test average was 80.69 percent. Conventional standards would indicate that this difference is highly significant, with a P value for the 95% confidence interval of less than 0.0001.

Byula Parsa[1]*, Dr. Shalu Garg[2]

P value and statistical significance: P 0.0001 when using a two-tailed test. This difference is very statistically significant based on the standard measures.

Confidence interval: If we take the difference between the two tests, PreTest minus PostTest, as an example, we get a mean of -44.0588 with a 95% confidence interval of: In the range of -52.2025%35.9152E

Intermediate values used in calculations:

1. t = 10.8667
2. df = 50
3. standard error of difference = 4.054

Review of the data:

Table-1: Paired t test results

Group	PreTest	PostTest
Mean	36.6275	80.6863
SD	23.3169	15.8600
SEM	3.2650	2.2208
N	51	51

Matched Pre-Post Student Knowledge

Students who had both pre- and post-test data were included in the study, since this allowed for a more accurate depiction of the learning gains students experienced as a result of utilizing these modules. Forty-four students were counted as having taken both the pretest and posttest. The mean and standard deviation of the scores for this matched sample improved from the beginning (M=35.14, SD=23.5) to the end (M=83.61, SD=14.75). A paired-samples t-test revealed substantial improvement from the pre- to post-test (t=14.09, p 0.0001).

P value and statistical significance: P 0.0001 when using a two-tailed test. This difference is very statistically significant based on the standard measures.

Confidence interval:

- The mean of PreTest minus PostTest equals -48.4773
- 95% confidence interval of this difference: From -55.4153 to -41.5393

Intermediate values used in calculations:

- t = 14.0910

- df = 43
- standard error of difference = 3.440

Review of the data:

Table-2: Paired t test results

Group	PreTest	PostTest
Mean	35.1364	83.6136
SD	23.4963	14.7463
SEM	3.5422	2.2231
N	44	44

Confidence in using Big Data Modules in Class

Among the total survey takers, close to 80% were either juniors or seniors, and close to 30% were CS majors. The gender split was about right (52.9% female), however there wasn't much variety in terms of color, ethnicity, or disability among the participants. Almost 38% of respondents were majoring in computer science, while approximately 95% of all respondents were either juniors or seniors. Almost a third of the sample was female, and there was hardly any racial, cultural, or disability diversity. Juniors and seniors made up the majority of survey takers, with roughly 28% declaring a computer science major. There were more men than women in the sample (53.1% to 46.9%), and the majority of respondents identified as Black (87.5%) rather than Hispanic or Latino (90.6%). Students were asked to rate 31 distinct possible big data modules/applications on a scale from 1-5. Preceding the introduction of modules into mathematics education, these comments were solicited.

Student Academic Efficacy, Motivation and Learning Strategies in Math Courses

At the end of the survey, students were questioned about their confidence in their academic abilities, their interest in and success with mathematics, and the methods they employ and like while studying the subject.

- **Academic Efficacy:** In order to gauge how confident they feel in their ability to succeed in arithmetic, students were given a short survey consisting of five questions. Students had high levels of confidence in their academic ability, with an average term-to-term score of above 4. (on a 5-point scale). Students have faith in their own ability to learn if they put in the effort required. And they were certain that they could learn the

Byula Parsa[1]*, Dr. Shalu Garg[2]

material and figure out even the trickiest assignments.

- **Goals in Math:** Students ranked academic success as their top priority among all their objectives. Students enrolled in this course for a variety of reasons, including fulfilling degree requirements, enhancing their mathematical literacy, expanding their problem-solving toolset, and gaining insight into other methods of thinking about and approaching mathematical issues.

- **Preferred Learning Environments:** Students were asked to rate how much they agreed with phrases like "the teacher explains the answers to issues" and "the assignments are comparable to the examples covered in class," both of which describe effective learning settings. In addition, they detailed instances when they had to explain concepts to their peers, work in small groups, get regular feedback on their mathematical thinking, study their notes, and compare their understanding to that of their peers. Students were less enthusiastic about presenting their work to the class for feedback, taking examinations to demonstrate their knowledge, or giving oral presentations in groups.

- **General Learning Strategies used by Students:** On general, students said they try several approaches and don't give up when they're having difficulty with a math problem in class. The most common responses were developing their own methods of analysis and interpretation, as well as checking their work for errors and misunderstandings. Moreover, they said that they use techniques such as double-checking their assumptions about the problem's intent, independent research, and trusting their gut when deciding on a solution.

- **Motivation to learn Math - Task Value:** High levels of task value were reported by students, suggesting that they valued the material covered in their mathematics lessons highly. They place a high value on math literacy, and their desire to improve in this area is undeniable.

- **Learning Strategy – Critical Thinking:** Several analytical approaches to math study have been described by pupils. Students said they were able to form their own opinions in light of course material, and that they learned to critically examine claims before committing to them. They said they were also challenging what they were reading or hearing in class and were coming up with their own ideas to supplement the material.

- **Learning Strategy – Self- Regulation:** Several students' self-regulation tactics for math courses were found to be quite successful. They pay special attention to the ideas that they have trouble grasping and devote extra time to learning and revisiting them.

Internal consistency estimates for these scales were generally encouraging, ranging from 0.491 to 0.926 with a median of 0.867. Similarly, the total averages of the evaluation scales were higher than the scales' midpoints, indicating very good evaluations.

CONCLUSION

Over the course of years, we developed dozens of one-week linear algebra big data modules and integrated them into already required undergraduate mathematics courses. In-class exercises and discussions were used to demonstrate and explain the concepts covered in each module. Afterwards, they worked on projects that put their new knowledge of big data into practice. We have conducted pre- and post-tests, as well as surveys, to determine the efficacy of the big data modules. Data from a paired-samples t-test demonstrates that there is a statistically significant correlation between students' knowledge before and after a unit is taught. The answers were varied when we asked students how confident they felt utilizing big data modules in class. The total scale means were higher than the scale midpoints, indicating high levels of satisfaction among students. We found the classes to be beneficial, albeit they did highlight some areas for development in future analysis.

REFERENCES

1. David C. Lay, Linear Algebra and Its Applications (fourth ed.), Addison Wesley, Reading, MA (2011).
2. Gilbert Strang, "Linear Algebra and Its Applications," Cengage Learning; 4th edition, 2006
3. Gilbert Strang, Introduction to Linear Algebra (fourth ed.), Wellesley Cambridge Press, Wellesley, MA (2009).
4. Jean-Pierre Tignol, Galois' Theory of Algebraic Equations, World Scientific Publishing, Singapore (2001).
5. Jiyan Yang, Randomized Linear Algebra for Large-Scale Data Applications, August 2016, Stanford University.
6. John R. Durbin, Modern Algebra: An Introduction (sixth ed.), John Wiley and Sons, New York (2008).
7. Joseph J. Rotman, A First Course in Abstract Algebra (third ed.), Prentice Hall, Upper Saddle River, NJ (2005).
8. Joseph J. Rotman, Advanced Modern Algebra (second ed.), American Mathematical Society, Providence, RI (2010).
9. Kenneth Hoffman and Ray Kunze, Linear Algebra (second ed.), Prentice Hall, Upper Saddle River, NJ (1971).
10. Lloyd N. Trefethen and David Bau III, Numerical Linear Algebra, SIAM, Philadelphia (1997).

Byula Parsa[1]*, Dr. Shalu Garg[2]

11. Nicolas Bourbaki, Algebra 1, Springer-Verlag, New York (1989).

12. Nicolaus Henke, Jacques Bughin, Michael Chui, James Manyika, Tamim Saleh, Bill Wiseman and Guru Sethupathy, THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD, December 2016, McKinsey & Company.

13. Roger Horn and Charles Johnson, Matrix Analysis (second ed.), Cambridge University Press, Cambridge (2012).

14. Ron Larson and David Falvo, Elementary Linear Algebra (sixth ed.), Brooks Cole, Belmont, CA (2009).

Corresponding Author

Byula Parsa*

Research Scholar, Shridhar University