Teaching the Singular Value Decomposition of Matrices: A Computational Approach

Zoltan Lazar

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master in the Teaching of Mathematics at

Concordia University

Montreal, Quebec, Canada

August 2012

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Zoltan Lazar

Entitled: An algorithmic approach to teaching the Singular Value Decomposition of matrices

and submitted in partial fulfillment of the requirements for the degree of

Master in the Teaching of Mathematics

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____Nadia Hardy__________________________Chair

_____Nadia Hardy__________________________Examiner

_____Fred E. Szabo_________________________Examiner

_____Anna Sierpinska_______________________Supervisor

Approved by _____Yogendra Chaubey____________________

Chair of Department or Graduate Program Director

September 10, 2012 ______Brian Lewis__________________________

Dean of Faculty

Abstract

In this thesis, I present a small experiment of teaching the singular value decomposition (SVD) of matrices using a computational approach. The goal of the experiment was to check if students who completed the first two undergraduate linear algebra courses are prepared for this topic and if they would be satisfied with the computational approach.

The experiment took place in the summer of 2011 and consisted in two sessions of lectures of four hours each, in a computer lab, on the premises of Concordia University. The same four students attended both sessions.

The approach consisted in first introducing students to the general ideas and then gradually zooming into the details of the theory and the computational techniques and algorithms. In the instructional sessions, lecturing by the teacher alternated with participants' exploring the theoretical results and algorithms using prepared Maple worksheets.

Before the sessions started, participants were asked to respond to a questionnaire (Pre-test) that verified their knowledge of basic linear algebra concepts necessary for understanding the SVD theory. After the session, participants were asked to respond to another questionnaire (Post-test) addressing their understanding of SVD and their opinions about the teaching approach and the teaching of SVD in an undergraduate program. Participants' responses to test questions were collected and analyzed.

One of the immediate conclusions is that without a good understanding of the fundamental concepts of linear algebra the topic of singular value decomposition of matrices could prove challenging for even the top achieving undergraduate students.

The participants showed interest in the teaching method, but mentioned that more time would be required to really benefit from learning about the numerical advantages and the vast applications of the singular value decomposition.

Acknowledgements

I would like to express my sincerest gratitude to my supervisor, Dr. Anna Sierpinska for her patience and guidance throughout writing this thesis.

I would also like to thank Dr. Armen Atoyan for recommending three of his top students, thus making the teaching experiment possible.

I should not forget the participants to the teaching experiment, who volunteered to attend my lectures, during their summer break. I really hope that it was worth their while.

Last but not least, I would like to thank my family who supported and believed in me.

Table of Contents

Abstract iii

Acknowledgements iv

Table of Contents v

List of Figures vii

List of Tables vii

Chapter 1. Introduction 1

Chapter 2. Literature review 4

2.1 Teaching Mathematics with Computers 4

2.2 Teaching Linear Algebra with Computers 6

Chapter 3. Singular Value Decomposition of Matrices as a Mathematical Result and a Topic in a Linear Algebra Course 9

3.1 Singular Value Decomposition as a mathematical result 9

3.2 Singular Value Decomposition as an object of teaching 11

3.2.1 An introduction to Linear algebra with Maple by F. Szabo 12

3.2.2 A “modern introduction” to linear algebra by D. Poole 17

3.2.3 Theory and applications of linear algebra, by W. Cheney and D. Kincaid 19

Chapter 4. Trying a Computational Approach to Teaching the Singular Value Decomposition of Matrices with a Small Group of Students 21

4.1 The teaching plan 21

4.1.1 Session 1, titled “The Singular Value Decomposition of a matrix, theory and applications” 22

4.1.2 Session 2, titled “Computational aspects of SVD” 52

4.2 Recruitment of participants in the experiment 65

4.3 Results of the teaching experiment 65

4.3.1 Participants’ responses to the Pre-test 66

4.3.2 Expectations about participants’ understanding of the subject of SVD based on their performance on the Pre-test 74

4.3.3 Participants' behavior during the sessions 77
4.3.4 Participants' responses to the Post-test, Part I – Linear Algebra questions ... 78
4.3.5 Participants' responses to Post-test, Part II – Views about the teaching approach 87
Chapter 5. Discussion and Recommendations for Teaching the Singular Value Decomposition of Matrices.. 92
References 96
Appendix 101
The recruitment letter 101
The Consent form 102

List of Figures

Figure 1. Visualization of the full SVD with matrices as rectangles 36
Figure 2. Visualization of the reduced SVD 36
Figure 3. More detailed diagrammatic representation of SVD 37
Figure 4. A breakdown of the action of a matrix on a unit circle into actions of its SVD components 37
Figure 5. Visualization of SVD in terms of orthogonal grid transformations and stretching 38
Figure 6. Transformation of a sphere into an ellipsoid 39
Figure 7. Example of obtaining the SVD of a matrix using Maple 40
Figure 8. A visualization of the notion of orthogonal matrix 41
Figure 9. Validating the property of singular values of A as square roots of eigenvectors of the associated symmetric matrix A^TA 43
Figure 10. The "Fundamental Theorem of Linear Algebra" 43
Figure 11. Showing that the eigenvalues computation is ill-conditioned on the example of a special 5 x 5 diagonal matrix 47
Figure 12. Representation of the steps to compute SVD of a matrix 53
Figure 13. A visual representation of the full iteration cycle of reducing superdiagonal entries of a matrix through multiplications by Givens rotation matrices 60
Figure 14. Diagonalization program by means of Givens rotation matrices 61
Figure 15. The output of the diagonalization program for the bidiagonal matrix A and EPS = 1/10 62

List of Tables

Table 1. Participants' performance on the Pre-test 76
Table 2. Participants' performance on the first part of the Post-test 87

Chapter 1. Introduction

At Concordia University, mathematics students have the option to study linear algebra in a computer environment. The courses MAST 234 Linear Algebra with Applications I and MAST 235 Linear Algebra with Applications II, intended for mathematics majors, take place in a computer lab equipped with the Computer Algebra System "Maple". A class in those courses usually starts with a brief instructor's theoretical introduction to a topic, after which students solve exercises that illustrate the theory using Maple at individual computer stations. The instructor circulates among the stations, helping to solve technical problems, answering questions and pointing to those aspects of the output that illustrate the theoretical concepts introduced in the lecture. Exercises are prepared by the instructor in a Maple worksheet before class. The worksheet is posted on the course website. The worksheet serves also as a place for students to write class notes. Students are expected to come to class a few minutes earlier to access the course website and download the worksheet for the day.

The courses include several examples of applications and algorithms (e.g., iterative methods for solving systems of linear equations such as the Jacobi and Gauss-Seidel methods). Overall, however, the approach remains "structural"[1]: Theory is presented in the language of the structural algebraic properties of systems of linear equations, matrices, vector spaces and linear transformations. Maple affords only an illustration of the structural theory, and gives students an opportunity to get a better "feel" of the abstract objects of the theory. We will call this approach "structural with computer illustrations".

We believe that the computational power of Maple could be used to better highlight the algorithmic aspects of linear algebra in those courses. In this thesis, we describe one modest attempt at such "computational approach" to teaching one topic, namely the topic of the singular value decomposition (SVD) of matrices.

---

[1] The notion of "structural approach" to linear algebra is understood here as an approach based on "analytic-structural thinking" about linear algebra concepts in the sense of Sierpinska (2000).

The topic of SVD is sometimes taught by the end of the MAST 235 course, as a culmination of the theory that brings all the concepts introduced in the linear algebra courses together. The approach is usually structural, as presented in Poole's "Linear Algebra" textbook (Poole, 2006), but students are shown the Maple commands to produce the SVD of a matrix. It is the highlight of the leading thread of the courses, namely the problem of the factorization of matrices into simpler components, aimed at directly providing important information about the matrix and the linear transformation it defines. The notions of eigenvalue and eigenvector, introduced by the end of the first course (MAST 234), are presented as tools of diagonalization of matrices. This is possible for a very restricted class of matrices only (square matrices with a sufficient number of linearly independent eigenvectors). Orthogonal diagonalization, presented in the second course (MAST 235), is applicable to an even more restricted class of matrices (symmetric matrices). SVD is applicable to all matrices.

The advantages of the SVD factorization, however, go beyond the theoretical fact that it is the most generally applicable tool. These advantages can be appreciated when a computational rather than analytic-structural point of view is taken. In particular, SVD is the most numerically stable technique for determining the rank-deficiency and nearness to singularity of a matrix.

Therefore, SVD would be a very natural crowning topic in a computer-based linear algebra course. This thesis is a result of our reflection on this idea, supported by a small scale implementation of a plan of teaching SVD to a group of students who have previously successfully completed the MAST 234 and 235 courses in a year where the topic was omitted. The "teaching experiment", especially the resulting participants' understanding of SVD and their opinions about the instruction, demonstrated the many shortcomings of the devised plan and indicated the directions for its improvement. We hope that this thesis will help future instructors wishing to take a computational point of view to devise better teaching plans and avoid the pedagogical and didactical mistakes that we have made.

This thesis is organized as follows. Chapter 2, which follows the Introduction directly, is a review of the literature concerning the teaching of mathematics in general, and linear algebra in particular, with computers. Chapter 3 presents the singular value decomposition as, on the one hand, a mathematical result, and, on the other, an object of teaching, particularly as a topic in one of Concordia University's linear algebra courses. Three textbooks used in the Concordia University linear algebra courses run in a computer lab are reviewed here. In Chapter 4, a small-scale teaching experiment on teaching the singular value decomposition is described, and an analysis of the participating students' reactions to it is presented. A discussion of the results of the experiment, and recommendations on teaching the SVD drawn form it constitute the content of the final Chapter 5. The thesis contains a list of references and an appendix. The appendix contains the Recruitment Letter and the Consent Form used in the experiment.

Chapter 2. Literature review

In this chapter, I will review concepts and ideas found in a selection of research papers in the area of teaching mathematics, particularly linear algebra, with computers.

2.1 Teaching Mathematics with Computers

In this day and age electronic devices in general and computers in particular are an integral part of our lives. They came to be widely used in education, mainly for administrative purposes but also, to some extent, for the presentation and study of the content. Universities, Concordia included, started blending regular classes and online communication (Garrison & Vaughan, 2008), or offering fully online courses. At Concordia University, for example, a college level “Vectors and matrices” course is offered as an online alternative to a regular classroom course. Computer Algebra Systems such as *Maple* (Maplesoft. A Cybernet Group Company, 2012), *Matlab* (MathWorks, 2012) or *Mathematica* (Wolfram, 2012), and statistical software packages such as SAS/STAT (SAS Institute Inc., 2008) are used in some university mathematics courses. The use of such proprietary software in secondary schools does not seem to be widespread (Boileau, 2012). Those advanced software packages are more often used by mathematicians, in research and various applications; they are valued for their powerful functions that allow users to get precise results of very complicated calculations; have facilities for image processing, graphics and two and three–dimensional data visualization. Most of these mathematical software packages allow for data acquisition and analysis, symbolic calculations and application development, simulation and prototyping.

Some mathematics education researchers advocate the use of such software also in the teaching of mathematics, saying that the mathematical software may be used as a problem solving assistant, as a tool for visualization and validation of mathematical results and as a tool for discovery and pattern recognition (Lagrange, 2005); (Thomas & Hong, 2004); (Berry, Graham, & Watkins, 1994). Computer Algebra

Systems are claimed to facilitate the conversion and coordination between the algebraic and graphical registers (Winsløw, 2003). A common argument is that teaching mathematics with computers can reinforce concepts and motivate learning by alleviating the burden of procedural manipulation. In other words, students can follow the logical reasoning without being trapped in the algorithm. The impressive graphical possibilities of mathematical software incite some to say that its use in teaching is likely to motivate students' geometric intuition. Certainly, lecturers can use those graphical possibilities to show students some mathematical aspects that would not be as accessible only with pen and paper: higher dimensional graphs, trajectories, solutions of linear equations, effects of matrix transformations. Once the teacher is armed with computers, he or she can do things that are not otherwise possible in the math class. For example, if a student proposes an inappropriate solution to a problem, it is much easier, with the use of mathematical software, to show that his or her approach is incorrect. In a traditional blackboard and prepared calculations environment, it is much more time consuming to prove the student wrong and he or she might be left unconvinced by only verbal explanations. The interaction between students and their teacher need not be diminished: a good way to raise students' attention and to make them more willing to participate would be to ask them to anticipate what will happen when the "enter" key will be pushed. One of the most important roles of a teacher is to demonstrate thought processes. This can be achieved by presenting computer algorithms and by showing their intermediate steps.

There exist experiments in teaching mathematics with mathematical software (Trouche, 2005); (Drijvers & Gravemeijer, 2005) where some gain in conceptual understanding has been achieved. However, these experiments have not been widely applied (Artigue, 2005). Indeed, the financial, temporal and institutional (curricula) constraints of mathematics teaching may constitute obstacles to a widespread use of (proprietary) mathematical software in education (Boileau, 2012). So far, research has not convincingly demonstrated that the benefits of teaching and learning mathematics with mathematical software justify the expense of money, and time and effort required to learn the syntax of the software and prepare and organize the learning process (Pruncut, 2007). It is an agreed upon fact that the presence of technology creates new challenges for students. They have to understand how a certain “instrumented technique” (Lagrange, 2005) relates to their prior knowledge acquired in the traditional environment. They have to learn to interpret and anticipate the returned results and also to distinguish when it is worthwhile to use the tool instead of solving the mathematical problem on paper. This additional degree of complexity is perceived by some as a burden. Only when the student overcomes the constraints (syntactic, organizational) of the mathematical software can this perceived burden be alleviated.

There are researchers who worry that using computers in a mathematics course can turn students into button-pushers and that they will rely solely on mathematical software to assist them in solving the problem (Thomas & Hong, 2004); (Crowe & Zand, 2000). Also, in the particular case of teaching Linear Algebra, such a concrete approach by using visualization and examples in low dimensions may lead to irrelevant interpretations and misunderstandings (Sierpinska, Dreyfus, & Hillel, 1999).

2.2 Teaching Linear Algebra with Computers

The vast majority of researchers and undergraduate students agree that Linear Algebra is a difficult course. Students complain about its apparent disconnection from other areas of mathematics and its formalism (Dorier & Sierpinska, 2001). These observations have been done both in lecture-driven, paper-and-pencil based theoretical courses (Dorier, Robert, Robinet, & Rogalski, 2000), and in experiments with hands-on, computer-based approaches (Sierpinska, Dreyfus, & Hillel, 1999); (Pruncut, 2007). Undergraduate Linear Algebra courses that go beyond techniques of solving systems of linear equations in the language of vectors and matrices and geometrical applications cannot avoid using analytic-structural thinking (Sierpinska, 2000) and reasoning in terms of axiomatic definitions and properties, which are well known to be difficult for students. Such Linear Algebra courses are rich in concepts and are characterized by frequent transitions between representation modes (analytic, algebraic, and geometric) (Hillel, 2000).

Most students master the algorithmic skills involved in linear algebra, but many fail to achieve a “conceptual understanding of the subject” (Sabella & Redish, 1995). They may know the algorithms but have difficulty in choosing the appropriate one in solving a given problem (Dubinsky, 1997). Carlson states that solving systems of linear equations and calculating products of matrices is easy for students, but they become confused and disoriented with subspaces, linear independence, spanning and various aspects of bases and dimension (Carlson, 1993). Harel (1989) singles out the problem that students have with the linear-algebraic notations and explains that for them, abstract concepts arrive too quickly and without a firm intuitive base. Geometric interpretations of linear transformations, expected to make the concept more intuitively meaningful for students, do not necessarily facilitate their understanding because this requires a good spatial sense and a conceptual connection between geometric and algebraic objects which are both cognitively demanding.

Mathematics education researchers have proposed and tried various ways of helping students overcome their difficulties with the subject. Some of them involved the use of technology, particularly mathematical software. One has already been mentioned above: the teaching of the notion of linear transformation in the dynamic geometry environment Cabri (Cabrilog, 2009) has been tried, not very successfully, by Sierpinska, Dreyfus and Hillel (1999); (Sierpinska, 2000). Another was proposed by Dubinsky and his associates, based on the use of the computer programming language ISETL (Dubinsky, 1995).

Dubinsky (1997) named three sources of students’ difficulties. The first is that teachers “succumb to the student demand that we first show them how to solve a certain kind of problem and then ask them to solve many instances of this same problem.” (Dubinsky, 1997, p. 93) This causes students to have difficulties in understanding the concepts because they never get the chance to build their own thoughts about them. Second, he affirms that students lack the background concepts essential to learning linear algebra. The third source, Dubinsky states, is the fact that the teachers do not use efficient pedagogical strategies to entice students to construct their own ideas about important concepts. Dubinsky proposes that pedagogical strategies based on cooperative learning and programming computers may have a better effect. He puts the emphasis on the idea of “having students construct implementations of mathematical concepts on computers, essentially by writing programs.” (Dubinsky, 1997, p. 99) In the author’s view, programming in ISETL can help students visualize linear transformations dynamically. For example, in the initial stages of learning linear transformations, students could program a computer to transform a given vector, or a figure such as a square or a circle, by the action of a 2x2 matrix. Later on, students could be engaged in discussing the linear-algebraic aspects of the solution space of a linear system of differential equations. The author considers that applying Euler’s method to approximate the solution for a given initial value has a nice computer graphics potential and would lead students to enhance their understanding.

It must be noted that using computers in linear algebra is not just a matter of choosing a particular pedagogical strategy. It is a fact that there is no important application of linear algebra that does not require a computer. A good example is the standard way of computing the eigenvalues of a matrix: first find the characteristic polynomial of the matrix and then find the roots of this polynomial. For large matrices this approach is hopeless, but using mathematical software such as *Maple*, *Matlab* or *Mathematica* that have been programmed with analytic or numerical methods, these eigenvalues or their approximations can always be computed.

Chapter 3. Singular Value Decomposition of Matrices as a Mathematical Result and a Topic in a Linear Algebra Course

In this chapter, the significance of the singular value decomposition of matrices (SVD) as a mathematical result will be demonstrated and some approaches to teaching this topic in an undergraduate linear algebra course will be discussed.

3.1 Singular Value Decomposition as a mathematical result

The brief account of the history of SVD in this section is based on the following sources: (Stewart, 1992) and (Wikipedia, 2012).

The singular value decomposition of a matrix is a relatively new result in the history of mathematics. It was developed by mathematicians who tried to determine if two real bilinear forms could be made equal by orthogonal transformations.

In the 1870s, Eugenio Beltrami and Camille Jordan discovered that singular values form a complete set of invariants under orthogonal substitutions. It was only in 1910 that the term "valeurs singulières" of a matrix A was coined (by Emile Picard) for the square roots of the eigenvalues of the associated symmetric matrix $A^T A$ which is what we call singular values today. The first proof of the SVD for any complex matrix (not necessarily square) was done by Carl Eckart and Gale Young less than one hundred years ago, in 1936. By finding a first computational algorithm, Gene H. Golub and William M. Kahan introduced the SVD into numerical analysis (Golub & Kahan, 1965). However, it was Golub and Christian Reinsch who later developed the improved version that is used in most mathematical software today (Golub & Reinsch, 1970).

The singular value decomposition is an extension of the diagonalization of a matrix. The diagonalization of a matrix is applicable only to square matrices and only to those that satisfy a quite demanding condition. The matrix must have a sufficient number of linearly independent eigenvectors; n for an $n \times n$ matrix. For an $n \times n$ square matrix A, if diagonalizable, there exists an invertible $n \times n$ matrix P and an $n \times n$

diagonal matrix D such that $A = PDP^{-1}$. The diagonal entries of D are the eigenvalues of A, and the column vectors of P are the corresponding eigenvectors of A. It is said that the matrix P "diagonalizes" A.

SVD lifts the assumptions of squareness and existence of a basis of eigenvectors; it is applicable to any type of matrices, even rectangular ones. It is founded on orthogonality theory and especially the theorem ("Spectral Theorem") that any real symmetric matrix can be orthogonally diagonalized, i.e., that there exists an orthogonal matrix that diagonalizes it; the converse is also true. The theorem can be generalized to complex matrices that are Hermitian (those that are identical with their conjugate transpose – a complex analog of the real symmetric matrices). SVD holds in general for complex matrices.

Let A be an $m \times n$ complex matrix. The singular value decomposition of A is the factorization $A = U.\Sigma.V^*$ where U is an $m \times m$ unitary matrix that holds the left singular vectors of A; Σ is an $m \times n$ "pseudo-diagonal" matrix that holds the singular values of A, and V is an $n \times n$ unitary matrix holding the right singular vectors of A. The concept of unitary matrix is the complex analog of the real orthogonal matrix.

Over the years, mathematicians found several key applications to SVD. Some of them are in numerical methods related to linear algebra. A list of these include: the general pseudo-inverse of a matrix (the Moore-Penrose inverse), the computation of the four fundamental subspaces associated with a matrix (column space, row space, null space of A and null space of A^T), estimation of the rank, computation of the inverse, perturbation theory (sensitivity of linear equations to data errors) and solving linear equations with inequality constraints.

Among important applications in other fields we can name: noise reduction, image compression, data analysis and prediction.

3.2 Singular Value Decomposition as an object of teaching

This section contains an overview of the presentation of the topic of the Singular Value Decomposition of matrices in the textbooks that have been used in the MAST 234 and MAST 235 Linear Algebra with Applications courses.

Since the inception of these courses at Concordia University in 1997, three textbooks have been used:

2001-2005

Szabo, F. (2002). *Linear algebra. An introduction using Maple*. Boston: Harcourt/Academic Press.

2006-2008

Poole, D. (2006). *Linear algebra. A modern introduction*. Boston: Thomson Brooks/Cole

2009-2012

Cheney, W. & Kincaid, D. (2009). Linear algebra. Theory and applications. Boston: Jones and Bartlett.

Before 2001, lecture notes written by the instructor have been used, titled, "Linear algebra and its applications" (1997) and "Linear algebra with Scientific Notebook" (1998-2000).

The above-listed three textbooks contain chapters on SVD and the topic was taught until 2008. Then, for some reason, it was abandoned. In fact, it appears that, in general, SVD is not often taught in undergraduate linear algebra courses. A search of textbooks available in the Concordia libraries revealed only a few linear algebra textbooks with the notion of singular values or SVD theory. Among those that include at least one chapter about the topic, some are applied linear algebra books (Gelbaum, 1988), others are books about the use of computers and numerical methods in teaching linear algebra (Bau & Trefethen, 2000); (Natt, 2010). Following this trend, one might expect the MAST 234 and 235 courses to include SVD, since they have been designed to teach linear algebra in a computer environment. Yet, between 2009 and 2012, the topic did not appear in the course outlines.

The prominent mathematician Gilbert Strang considers that SVD deserves being taught in linear algebra courses no matter whether the approach is computational or structural. In his Linear Algebra video lectures (Strang, 2010), he called SVD "a highlight of linear algebra" and emphasized that it is the best possible matrix factorization, with "especially good" matrices: orthogonal and diagonal. He added that SVD gained more exposure lately and that it is "bringing together everything in the linear algebra course". Other mathematicians and mathematics instructors have caught on to the idea and there are now interesting and dynamically illustrated online expositions of SVD and its applications; e.g., (Davis & Uhl, 2011), (Will, 2003).

3.2.1 An introduction to Linear algebra with Maple by F. Szabo

SVD occupies the last chapter of the textbook by F. Szabo (2002), and appears, therefore, as the crowning of the whole course. It comes after chapters on vector spaces, linear transformations, eigenvalues and eigenvectors, norms and inner products, and orthogonality. The chapter is titled "Singular values and singular vectors" and the topic is introduced by a brief statement of the main ideas it is based on:

> *In this chapter, we show that every rectangular real matrix A can be decomposed into a product UDV^T of two orthogonal matrices U and V and a generalized diagonal matrix D. The product UDV^T is called the singular value decomposition of A.*
>
> *The construction of UDV^T is based on the fact that for all real matrices A, the matrix A^TA is symmetric and that… therefore exists an orthogonal matrix Q and a diagonal matrix D for which $A^TA = QDQ^T$. We know from our earlier work that the diagonal entries of D are the eigenvalues of A^TA . We now show that they are nonnegative in all cases and that their square roots, called the singular values of A, can be used to construct UDV^T. Due to the nature of the singular value decomposition algorithm, all numerical results in this chapter are approximations. (Szabo, 2002, p. 619)*

In the chapter, the reader is given Maple commands to produce the SVD of a matrix, but the algorithm mentioned in the introduction above is not presented or discussed.

The Author begins with purely algebraic definitions and results and then illustrates them with concrete numerical examples in Maple. He first proves that the eigenvalues of the matrix A^TA are nonnegative using the Spectral Theorem. Then, he uses this result to define the singular values of a real matrix A as the square roots of the eigenvalues of the matrix $A^T A$.

Two numerical examples follow, one for a square and one for a rectangular matrix. The first example asks students to find the singular values of the matrix $\begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$. This is a good opportunity for the teacher to reinforce the definition of singular values, by correcting the possible first-impulse answers that might consider eigenvalues of the matrix A instead of the matrix $A^T A$, which is all the more likely since the matrix is triangular. In addition, the discrepancy between the results of finding the singular values based on the Maple output of the command "Eigenvalues(Transpose(A).A)" ($\sqrt{7+3\sqrt{10}}, \sqrt{7-3\sqrt{10}}$) and obtaining them directly using the command "SingularValues(A)" ($\frac{3}{2}\sqrt{2} + \frac{1}{2}\sqrt{10}, \frac{3}{2}\sqrt{2} - \frac{1}{2}\sqrt{10}$) could lead students into thinking that they obtained the wrong result. The teacher could then challenge the students to show that both results are in fact correct. In the second example, the singular values of a 3 x 5 rectangular matrix and of its transpose are to be computed. Here, the instructor could highlight the fact that the singular values exist even for rectangular matrices and that the matrix and its transpose have the same positive singular values.

The examples are followed by relating singular values with matrix norms. First, the theorem that the largest singular value of a matrix is identical with its two-norm is given and proved. The definition of the two-norm used in the proof is $\|A\|_2 = \max\left\{\frac{\|Av\|_2}{\|v\|_2}: \ \|v\|_2 \neq 0\right\}$. This theorem is intended to give more meaning to the concept of singular values by relating it to previously learned notions – matrix norms, in this case. Norms are measures, real numbers, associated with vectors and matrices, computable using Maple commands, and they are given much prominence in the textbook. This approach differs from the next textbook that will be discussed here, by Poole (2012) , where the same meaning of the concept of the maximal singular value is represented in geometric terms and given a graphical representation in the context of the image of the unit circle under the transformation by the matrix: the image is an ellipse and the length of its major half-axis is shown to be equal to the maximal singular value of the matrix.

In Szabo's textbook, the theme of relations between singular values and matrix norms is pursued in an example following the previous theorem. In the example, students are asked to confirm that the Frobenius norm of a matrix A, defined previously as $\|A\|_F = \sqrt{Trace(A^T.A)}$, is the square root of the sum of the squares of its singular values, for a 4 x 4 matrix with integer entries.

At this point, the instructor could highlight the use of the Frobenius norm, by talking about the notion of distance between matrices. He could find concrete examples of matrices that are "close" to singular matrices and show that in such cases the distance expressed in terms of Frobenius norm is small.

Before proving the SVD theorem, the Author lists theoretical results that make the singular value decomposition valuable. These "properties" constitute good starting points for in-class discussions, quizzes and validation based on Maple numerical examples. For instance, a challenging exercise would be to ask students to come up with an example of an ill-conditioned matrix, using the fact that the condition number can be written as $\frac{max(\sigma_i)}{\min(\sigma_i)}$.

By devising a session of questions and answers, followed by validating these answer with concrete numerical examples in Maple, the teacher could showcase how much of information about the matrix itself is held by the singular values. As an example, consider the following exercise: given that the square matrix A has singular values 2, 1 and 1, what can be deduced about the matrix?

We can claim that the rank of the matrix is equal to three, because there are three non-zero singular values. Thus, the matrix has three linearly independent columns.

We can compute the determinant of the matrix, based on the identity $\det(A) = \prod \sigma_i$. Given that the determinant is non-zero, we know that A is invertible. Its inverse A^{-1} will have singular values 1, 1, and 1/2.

We can state that the matrix is well-conditioned (because $\frac{max(\sigma_i)}{\min(\sigma_i)} = 2$) so it is relatively far from a singular matrix.

We can immediately find the 2-norm of A, $\|A\|_2 = 2$, as the largest singular value of the matrix and the Frobenius norm of A , $\|A\|_F = \sqrt{2^2 + 1^2 + 1^2}$ = $\sqrt{6}$.

It can be easily shown that the transpose matrix A^T has the same singular values as A.

Using a concrete numerical example of a positive definite symmetric matrix, the instructor could show that, in this case, the eigenvalues and singular values coincide. For all other matrix types he should emphasize that eigenvalues relate directly to the original matrix A, while the singular values are computed from $A^T.A$.

Applications of SVD are also included: the computation of the four fundamental subspaces, of the pseudo-inverse and of the least square solutions of a system of linear equations. A section is dedicated to a practical example of how SVD, by way of the Kronecker product expansion of a matrix, is used in. The Author mentions the remarkable result that the error in approximating an image to the k^{th} singular value of the corresponding matrix is the next singular value, σ_{k+1}. He expresses this in terms of distance to the original matrix, using the 2-norm: $\|A - A_k\|_2 = \sigma_{k+1}$. I think that the aim of the author is to entice students to discover the usefulness of Linear Algebra by learning about these applications.

The approach used in the textbook can be described as structural with Maple illustrations, but certain computational notions are also introduced. An example of such notion is the condition number of a matrix and the underlying idea of well-conditionedness of a matrix in relation to the perturbation theory. When defining the effective rank of a matrix, the Author mentions that one can use the computer version of the Gaussian elimination to reduce a matrix to its row echelon form. However, he states that this method is not numerically stable, due to the accumulation of round-off errors. Instead, "a more reliable method is to find the singular value decomposition of the matrix and then discard the small singular values." (Szabo, 2002, p. 644).

Compared with other books or articles on singular value decomposition, the textbook under discussion here, “Linear Algebra. An introduction using Maple” (Szabo, 2002) does not include any geometrical representations of the matrix transformation that maps x to Ax. The book lacks the classical example of how the image of the unit sphere in $\mathbb{R}^n$ becomes the surface of an ellipsoid in $\mathbb{R}^m$ if the transformation matrix is of maximum rank.

An aspect worth mentioning is the fact that mathematics textbooks that incorporate important reference to the use of particular mathematical software in teaching need frequent revisions because of the ageing of the software. The textbook under discussion was published in 2002, and Maple versions 5 and 6 were used for the numerical examples. Most of these examples use the now “deprecated” Maple package *linalg*. Using a newer version of the software together with the textbook could be confusing for students. Another problem with using particular software within a textbook is that the software’s representation of the mathematical objects could be based on a different conceptualization of those objects than is usually assumed in the theory. In particular, the *linalg* package represents vectors as lists or one-row matrices and matrices as lists of rows. The proofs in the textbook under discussion (Szabo, 2002) are based on the representation of vectors as one-column matrices. For example, the proof that, for all real $m \times n$ matrices A, the eigenvalues of the matrix $A^T A$ are nonnegative, uses a representation of the dot product of two vectors v and w as $v^t . w$ which conceives of vectors as one-column matrices. The same conceptualization appears in viewing the system $Ax = b$ as the question of whether b belongs to the column space of A. In the example that directly follows this proof, a matrix is defined in *linalg* as a list of rows, and in further examples vectors are declared as rows. This inconsistency of representation can be confusing for the students.

3.2.2 A “modern introduction” to linear algebra by D. Poole

Poole’s textbook (2012) represents a structural approach to linear algebra, with meaning sought mainly in the geometric intuitions underlying the concepts, and a few suggestions of “exploration” using a computer algebra system (CAS).

The text, however, comes in a package with supplementary materials and resources some of which allow more extensive use of CAS. Besides the Student Solutions Manual that accompanied also Szabo’s textbook, the package included an Instructor’s Guide, a test Bank and three internet resources. One was an electronic version of the textbook. Another was an online course management system complete with a large bank of problems for weekly assignments and periodic tests. The problems were programmed for electronic grading and the instructor could easily modify the problems and the programming of the grading to suit his or her purposes (iLrn Testing, 2003). The instructor of the MAST 234-5 courses who used Poole’s textbook made important use of the *iLrn Testing* system, both for class management and assignments[2]. Finally, the textbook was accompanied by a CD-ROM with

> *… data sets for more than 800 problems in Maple, MATLAB, and Mathematica, as well as data sets for selected examples. Also contains CAS enhancements to the vignettes and explorations that appear in the text and includes manuals for using Maple, MATLAB, and Mathematica. (Poole, 2012, p. xv)*

The instructor used some of the Maple-based problems on the CD-ROM in preparation of classroom activities.

It has to be mentioned, that neither the textbook, nor the “ancillaries” accompanying it present the algorithms for computing SVD, and the computational advantages of using SVD are not highlighted in the materials.

Here, I will describe only the presentation of the SVD topic in the hard copy of the textbook.

[2] This was before the Moodle class management system was imposed on instructors in the Mathematics and Statistics Department.

Similarly to Szabo's textbook, this one also presents SVD as the last topic, but it does not devote a separate chapter to it. "The Singular Value Decomposition" section is part of Chapter 7, "Distance and Approximation" and comes after sections on "Inner Product Spaces", "Norms and Distances" and "Least Square Approximation". The Author introduces SVD referring to results about diagonalization of matrices from previous chapters. He starts by recalling the Spectral Theorem. Then, he writes that certain non-symmetric matrices are still diagonalizable but in this case the diagonalizing invertible matrix cannot be orthogonal. However, while not every matrix is diagonalizable, every matrix, regardless of its shape, has a factorization of the form $A = P.D.Q^T$, where P and Q are orthogonal and D is pseudo-diagonal, and the factorization is called "the singular value decomposition of the matrix A". He calls this a "remarkable result" and deems SVD to be one of the most important matrix factorizations.

After showing that $A^T.A$ is symmetric and that this implies that all eigenvalues of this matrix are real and nonnegative, he defines the singular values of *A* as the square roots of the eigenvalues of $A^T.A$.

The Author gives a geometric interpretation for the above definition writing that the singular values of *A* are the lengths of the vectors Av_1 , Av_2 ,... , Av_n , where { v_1 , v_2 ,... , v_n} is an orthonormal basis of eigenvectors of $A^T.A$. Using a concrete numerical example of a 2 x 2 matrix, he shows that the image of the unit circle under the transformation is an ellipse and that the singular values σ_1 and σ_2 are the lengths of the major and minor half-axes of this ellipse.

The proof of the SVD theorem follows and the Author underlines the idea that the matrices *U* (whose columns are left singular vectors of *A*) and *V* (whose columns are the right singular values of A) are not uniquely determined by *A*. Using a 2x3 and a 3x2 matrix, he shows the steps to compute their singular values, without the use of mathematical software.

Next, the Kronecker (or dyadic) decomposition of a matrix is presented as "the outer product form" of the SVD, and the Author mentions that this expansion is very useful in applications. One of these applications, the digital image compression, is showcased at the end of the section.

The Author proves next that the matrices *U* and *V* contain the orthonormal bases for the four fundamental subspaces (col (A), null (A^T), row (A) and null (A)).

Consistent with his geometric perspective, he continues by stating that SVD provides a "new geometric insight into the effect of matrix transformations". The fact that an *m* x *n* matrix transforms the unit sphere in $\mathbb{R}^n$ into an ellipsoid in $\mathbb{R}^m$, noted in previous sections, is stated as a theorem and proved. A nice graphical visualization of this transformation for a 2 x 3 matrix is included.

The steps involved in this transformation are presented: the orthogonal matrix V^T maps the unit sphere onto itself. Then, the *m* x *n* pseudo-diagonal matrix D collapses $n - r$ of the dimensions of the unit sphere (*r* is the rank of the matrix) and the other *r* nonzero diagonal entries distort it into an ellipsoid. Finally, the orthogonal matrix *U* aligns the axes of the ellipsoid with the orthonormal vectors u_i in $\mathbb{R}^m$. Once again, a graphical representation of these steps is included.

Among the applications of the singular value decomposition of a matrix, the Author lists: the computation of the rank as the number of nonzero singular values; expressing the matrix norms and the condition number of the matrix in terms of its singular values; the matrix pseudo-inverse ($A^+ = V.D^+.U^T$, where D^+ holds on its diagonal the inverses of the nonzero singular values of *A*) and the least squares approximation (the system $Ax = b$ has a unique least squares solution of minimal length: $\bar{x}$ = $A^+.b$). Every application is supported by a numerical example.

3.2.3 Theory and applications of linear algebra, by W. Cheney and D. Kincaid

In Cheney and Kincaid's Linear Algebra textbook (2009), the topic of SVD occupies a little more than a page. The text is part of a section devoted to "Matrix Factorizations and Block Matrices" in Chapter 8 "Additional Topics".

The Authors affirm that SVD has "many uses, among them a way to produce a reliable estimate of the rank of a matrix." (p. 501) The SVD theorem is stated and proved next. The steps involved in the proof are similar to the ones in the other two books. In a footnote to the proof, the Authors mention the creators of the most well-known algorithm for computing SVD: W. Kahan and G. Golub. They also refer to the applications of this algorithm: signal processing, data analysis and Internet search engines.

In my opinion, this textbook does not apportion SVD the space that it deserves. The Authors adopt a purely structural-algebraic approach, with no geometrical or computational insights. The computational advantages of SVD and its wide range of applications are mentioned just in passing, with no concrete examples.

Chapter 4. Trying a Computational Approach to Teaching the Singular Value Decomposition of Matrices with a Small Group of Students

This chapter presents the content of the teaching experiment and participants' responses to a Pre- and a Post-test. The Pre-test measured the participants' preparedness for the SVD lectures in terms of their memory and understanding of basic linear algebra concepts. The Post-test looked at some aspects of participants' understanding of the SVD lectures and their opinions about the teaching approach used in the lectures.

The Pre-test was administered right before the beginning of the first session of the lectures. The lectures were prepared before the Pre-test and the plan was not changed to accommodate the shortcomings of the participants' understanding of the basic notions of linear algebra. The instructor (the author of the thesis) only allowed time for questions during the lectures to clarify any issues that might arise. Therefore the teaching plan and the actual content of the teaching experiment were essentially the same. In section 4.1, the description of the teaching plan will be worded partly in the past tense, stating what the instructor actually did, but without describing the individual participants' reactions. Only their general activity as expected will be described.

4.1 The teaching plan

The plan for teaching SVD from a computational point of view will be presented here using the rubrics of the Concordia University "lesson plan template"[3].

There were two four-hour classroom sessions planned for the "teaching experiment". More were initially hoped for, but it was a challenge to recruit volunteer students and to synchronize their availability. A separate session for revising the basic

[3] http://teaching.concordia.ca/resources/lesson-plan-template/

linear algebra notions used in the lectures was initially planned to ensure a better understanding, but it had to be given up because of the little available time. In a questionnaire administered at the end of the experiment, some of the participants stated that they would have liked more time allocated to the experiment, especially to the section presenting the numerical algorithms.

4.1.1 Session 1, titled "The Singular Value Decomposition of a matrix, theory and applications"

Purpose

The goals of this first session are: to inform the participants about the purpose and plan of the experiment, and what they will be expected to do; to collect some data about their knowledge of basic linear algebra concepts (Pre-test), and to introduce the topic of SVD using a PowerPoint presentation with an overview of the history of its discovery, the fundamental theoretical results and a glimpse of its applications, interrupted by brief illustrations of the concepts on prepared Maple worksheets.

Learning outcomes

At the end of this lesson students should be able to remember the properties of the matrices involved in the singular value decomposition and the underlying geometric representation. In addition, they should grasp the advantages of SVD: the fact that it can be applied to any matrix and the wide range of its applications.

Bridge-in

Participants will modify the Maple worksheets, run the examples with new input data, and actively interact during class. They will build new plots to visualize the effects of matrix transformations. They will comment on the difference in results and remark the computer round-off errors. They will be enticed to use specific matrices to validate previous theoretical results (for example for symmetric, positive-definite matrices).

Pre-test

Before the session started, participants were asked to respond to a questionnaire about their mathematical background and answer 15 linear algebra questions. The linear algebra questions were asked not so much as a guide for the instructor to adjust his teaching to the level of students' knowledge – the lectures were prepared before the participants' responses were known to the instructor – but as part of a measure of the impact of participating in the experimental SVD sessions on students' understanding of linear algebra. The second part of the measure was the post-test. It was assumed that if there was improvement between the pre-test and post-test for a student, then the impact of the sessions on his or her understanding of linear algebra was positive.

The Pre-test questions

The pre-test was called "Preliminary questionnaire" and it contained the following questions:

INITIALS: Please print your initials

PROGRAM: What program are you registered in?

LINEAR ALGEBRA KNOWLEDGE

> What was the latest Linear Algebra course that you took?
>
> How long ago?
>
> What was your final mark?

Please answer all questions below to the best of your knowledge. When not sure about an answer, write "N.S." next to the question

- Question 1: Orthogonal vectors in a vector space
  - 1a. Given two orthogonal vectors in a vector space: what is their dot product?
  - 1b. What is the angle between them?
- Question 2: Multiplication of a row vector by a matrix
  - Complete the following sentence:
  - Multiplying a row vector by a matrix A is a linear combination of A's ….

Question 3: Orthogonal matrices

What properties of vectors remain unchanged (are preserved) when multiplied by orthogonal matrices?

Question 4: Fundamental subspaces of a vector space

4a. What are the row space and the null space of a matrix A?

4b. What is the relationship between the row space and the null space of a matrix?

Question 5: Multiplication of a square matrix by a vector

Complete the sentence:

The non-zero vectors ($n\ x\ 1$) that, after being multiplied by the square matrix A ($n\ x\ n$) remain in the same direction as the original vector, are…

Question 6: Roots of the characteristic polynomial of a matrix

Given a matrix A, we know that the roots of its characteristic polynomial are -1 and +1. How one can go about computing A^7?

Question 7: Symmetric matrices

Given A and B, real square matrices, next to each of the sentences below write True or False (if False, give a counter-example):

7a. $A + B$ is symmetric

7b. A^n is symmetric

7c. If A^{-1} exists, then A^{-1} is symmetric

7d. AB is symmetric

7e. $A^T A$ is symmetric (does A have to be square?)

Question 8: Positive definite matrices

8a. What is the definition of a positive definite matrix?

8b. Complete the sentence: A is positive definite if and only if all its eigenvalues are…

Question 9: Matrix rank

9a. How would you define the rank of a matrix?

9b. How would you go about computing it?

Question 10: Elementary row operations

What properties of the original matrix are not modified when Elementary Row Operations are applied?

Question 11: Matrix decompositions

11a. Give 3 examples of matrix decompositions.

11b. In what area of mathematics do you think that these decompositions are useful? Explain.

Question 12: Matrix determinant

If all the eigenvalues of a matrix are known, how can its determinant be computed?

Question 13: Linear independence

13a. Given 3 distinct, non-zero vectors in $\mathbb{R}^3$, how can we prove that they are linearly independent?

13b. Can a set of 8 distinct, non-zero vectors in $\mathbb{R}^7$ be linear independent? Explain.

Question 14: Diagonalizable matrices

When is an $n\ x\ n$ matrix, with real entries, diagonalizable?

Question 15: Basis

Complete the sentence:

Given a basis of a finite vector space, … element of the vector space can be expressed … as a … of basis vectors.

Expected answers to the Pre-test questions

Below, the Linear Algebra questions are justified and expectations about the participants' responses are presented. Some notes about how the teaching experiment could alleviate some of the possible students' misconceptions are also made.

Question 1: Orthogonal vectors in a vector space

The question was: "1a. Given two orthogonal vectors in a vector space: what is their dot product? 1b. What is the angle between them?"

While the theory of eigenvalues and eigenvectors and diagonalization of matrices abstracts from the magnitude of vectors (norms) and relations between their positions (angles), SVD looks at matrices from a much more geometrical perspective. The concept of orthogonality is essential in grasping the decomposition. This justifies asking a question about the meaning of orthogonality first in the Pre-test.

The perspective is geometrical in a very general sense, not just in the sense of shapes in the Euclidean plane or space. Linear algebra has developed a language that allows to speak about geometric relations in any number of dimensions, thus making the theory applicable not only to two and three-dimensional geometric problems but to any situations that can be modeled in terms of vector spaces and linear transformations between them. The language represents the mutual position between vectors by means of an operation on these vectors, the inner product, of which the dot product is a classic example. Instead of saying that two vectors are orthogonal if the angle between them is the right angle or measures 90 degrees, linear algebra *defines* the expression "vectors *v* and *w* are orthogonal relative to a given inner product" as equivalent to " the inner product of *v* and *w* is equal to 0". The angle *t* between vectors *v* and *w* is *defined* as represented by the number $\frac{<u,w>}{\|u\|\cdot\|v\|}$, and it applies to vectors in any inner product space, including spaces of real functions.

In the hope that participants gained a little of this generalized geometric perspective in their undergraduate Linear Algebra courses, we expected them to know that "orthogonal vectors" translates into "dot product is equal to 0" in case the dot product is the chosen inner product, which is the case in the SVD theorem presented in the lectures. We expected them to say that the angle between orthogonal vectors is $\frac{\pi}{2}$, rather than, say, "90 degrees", because the measurement of angles in terms of degrees reveals thinking of the cosine as a relation between the sides of a right-angled triangle (characteristics of school mathematics) rather than as a real function, as it is understood in linear algebra.

Question 2: Multiplication of a row vector by a matrix

The question was: "Complete the following sentence: Multiplying a row vector by a matrix A is a linear combination of A's …."

The expected correct answer was: multiplying a row vector by a matrix A is a linear combination of the rows of A. Participants could hesitate between linear combination of rows versus columns. Participants could use Maple to verify these possibilities. Moreover, they could reject the linear combination of columns conjecture by observing that multiplication of a row vector by a matrix produces a row vector, whereas a linear combination of columns would produce a column vector.

Question 3: Orthogonal matrices

The question was: "What properties of vectors remain unchanged (are preserved) when multiplied by orthogonal matrices?"

We expected that participants would mention the preservation of norms of vectors and angles between vectors, since they have been taught that orthogonal matrices preserve inner products: $\langle Qx\ ,\ Qy\rangle = \langle x\ ,y\rangle$.

This question was asked because orthogonal matrices and the fact that they preserve norms and angles play a pivotal role in the SVD. In the planned teaching sessions, their computational advantages would be highlighted as well: their inverse can be easily computed by transposing the matrix.

Question 4: Fundamental subspaces of a vector space

The question was: "4a. What are the row space and the null space of a matrix A?
4b. What is the relationship between the row space and the null space of a matrix?"

The notions of row space and nullspace of a matrix are taught early in the MAST 234 course. They are some of the standard examples of subspaces when the notion of subspace is introduced. We expected, therefore, that participants who achieved highly in the course would be able to answer question 4a easily. Regarding question 4b, they could recall the relation between the dimensions of these subspaces (as adding up to the number of columns of the matrix), the fact that the nullspace is the orthogonal complement of the rowspace and therefore the space $\mathbb{R}^n$where n is the number of columns of A can be decomposed into a direct sum of these subspaces.

In the lecture, in the section dedicated to the immediate SVD applications, the computation of the four fundamental subspaces was to have a prominent role. The relationships that exist between them will be highlighted and shown in a graphical manner. Therefore, if participants did not remember these notions well, the lecture was planned to clarify these notions for them.

Question 5: Multiplication of a square matrix by a vector

The question was: "Complete the sentence: The non-zero vectors (n x 1) that, after being multiplied by the square matrix A (n x n) remain in the same direction as the original vector, are…"

This question aimed at testing the capacity of the participants to move from the graphical thinking "register" to the more frequently used algebraic one. The notion of vector direction might be challenging for some students. Rewriting the question in the form: “Given a matrix A and a scalar λ, what is name of non-zero vectors x satisfying $Ax = \lambda x$?”, would probably get the correct answer, eigenvectors, most of the time.

Question 6: Roots of the characteristic polynomial of a matrix

The question was: “Given a matrix A, we know that the roots of its characteristic polynomial are -1 and +1. How one can go about computing A^7?”

The question doesn’t say anything about the size of the matrix or the multiplicity of the eigenvalues. It cannot be taken for granted, therefore, that the matrix is diagonalizable. However, questions about calculating a power greater than 2 of a matrix in the MAST 234-5 Linear Algebra courses rarely if ever involve non-diagonalizable matrices. Students who completed the more theoretically-oriented Linear Algebra course MATH 251-2 and had seen the Jordan and Rational Canonical Forms of matrices might be perhaps more likely to consider the case of a non-diagonalizable matrix. This would be the sign of theoretical thinking in linear algebra. However, in view of the existing research – (Sierpinska, Nnadozie, & Oktaç, 2002); (Dorier & Sierpinska, 2001) – this way of thinking is not common even among the best MATH 251-2 students.

Therefore, we expect that participants will assume that the matrix can be diagonalized, and that therefore there exists a transformation matrix P and a diagonal matrix D, with values -1 and 1 on the diagonal, such that $A = P^{-1}DP$. Raising this to power seven, we get $A^7 = P^{-1}.D^7.P$. The matrix P is obtained by computing the eigenvectors corresponding to the eigenvalues – 1 and 1. Students could notice that, with 1 and – 1 on the diagonal, $D^7 = D$, and therefore, $A^7 = A$. The question did not ask about the outcome of raising the matrix to power 7, however, but only about ways of calculating it.

To answer this question, participants do not have to make a mental link between eigenvalues and the roots of the characteristic polynomial. However, this link will be necessary in understanding the lecture where the numerically unstable aspect of computing the eigenvalues as roots of the characteristic polynomial will be brought to the fore. The complexity of computing the singular values by hand will help students strengthen their belief in the usefulness of mathematical software.

Question 7: Symmetric matrices

The question was: "Given A and B, real square matrices, next to each of the sentences below write True or False (if False, give a counter-example): 7a. $A + B$ is symmetric; 7b. A^n is symmetric; 7c. If A^{-1} exists, then A^{-1} is symmetric; 7d. AB is symmetric; 7e. $A^T A$ is symmetric (does A have to be square?)

It will be interesting to see if the participants will attempt to prove any of the parts of this question in an analytic-structural way (by using the fact that if A is symmetric then $A^T = A$).

A good understanding of the properties of symmetric matrices is at the heart of understanding SVD. Sub-question 7e, in particular, is very important, because it helps in gaining the knowledge of how singular values can be computed (as square roots of the eigenvalues of the symmetric matrix $A^T A$. During the lecture, the fact that regardless of the shape of matrix A, the matrix $A^T A$ is always symmetric will be emphasized.

Question 8: Positive definite matrices

The question was: "8a. What is the definition of a positive definite matrix? 8b. Complete the sentence: A is positive definite if and only if all its eigenvalues are…"

We expected that the definition of a positive definite matrix: an $n \times n$ symmetric real matrix A is positive definite if the number $v^T.A.v$ is positive for all non-zero column vectors $v \in \mathbb{R}^n$, would be known by the students. If they did not remember this definition, we expected them to at least know the test of positive-definiteness hinted at in question 8b.

The concept of positive definite matrices and the attributes of their eigenvalues will be part of the theoretical examples accompanying the SVD theory section in the lecture. Participants will have a chance to get to refresh their memory of this concept by using Maple commands and by validating their own examples. In one of the exercises, participants will have a chance to verify that if a matrix is positive definite then its eigenvalues are equal to its singular values. This might help consolidate their understanding of positive definite matrices.

Question 9: Matrix rank

The question was: "9a. How would you define the rank of a matrix? 9b. How would you go about computing it?"

Participants can be expected to define the rank of a matrix in terms of the number of pivots in the reduced row echelon form of the matrix, or the number of non-zero rows in that form. Row reduction is the expected answer to question 9b.

While teaching SVD, the idea that the rank of a matrix is equal to the number of its non-zero singular values and is more numerically stable than row reduction will be introduced. Students will build examples of matrices that are not of maximum rank, apply SVD and validate that there is at least one zero singular value in the decomposition.

Question 10: Elementary row operations

The question was: "What properties of the original matrix are not modified when Elementary Row Operations are applied?"

Students are used to elementary row operations because in the Linear Algebra courses the technique of row reduction is used in many situations. One of the situations is the topic of elementary matrices and the effect of multiplying by such matrices on other matrices. Multiplying a matrix by another simple one to obtain a certain effect will appear in the lectures in the context of multiplication by Givens rotation matrices.

Question 11: Matrix decompositions

The question was: "11a. Give 3 examples of matrix decompositions. 11b. In what area of mathematics do you think that these decompositions are useful? Explain."

This question tested the students' knowledge of matrix factorizations. It is expected that they list at least diagonalization and the QR decomposition.

Question 12: Matrix determinant

The question was: “If all the eigenvalues of a matrix are known, how can its determinant be computed?”

In the Linear Algebra courses, students learn that similar matrices have the same determinants. This could suggest to them the idea that the absolute value of the determinant can be computed by multiplying the eigenvalues.

In the SVD lecture, a similar result will be introduced and tested using Maple: the absolute value of a determinant is equal to the product of its singular values. At that time, references will be made to the relation between eigenvalues and the matrix determinant. Also, it will be stressed that if any singular value is zero the matrix becomes singular. These conclusions should help students clarify the role of the determinant and its relationship to both eigenvalues and singular values.

Question 13: Linear independence

The question was: “13a. Given 3 distinct, non-zero vectors in $\mathbb{R}^3$, how can we prove that they are linearly independent? 13b. Can a set of 8 distinct, non-zero vectors in $\mathbb{R}^7$ be linear independent? Explain.”

It was expected that participants would know the structural definition of linearly independent vectors. Some of the students, however, might formulate definitions based on creating a matrix with the given vectors and finding its rank, reflecting their tendency to use the analytic-arithmetic mode of thinking in linear algebra.

In the SVD lectures, the fact that the number of linearly independent rows (or columns) of the matrix to be decomposed is equal to the number of non-zero singular values of the given matrix will be highlighted. In addition, it will be stated that the matrices holding the left and right singular vectors (U and V respectively) have linearly independent rows (or columns) since they are orthogonal. Linear independence of the rows (columns) of an $n\ x\ n$ matrix, perceived as vectors in $\mathbb{R}^n$, in relation to matrix invertibility and matrix orthogonality will be explained.

Question 14: Diagonalizable matrices

The question was: "When is an $n \; x \; n$ matrix, with real entries, diagonalizable?"

Participants are expected to name at least some sufficient conditions for diagonalization: for example, if a nxn matrix has n distinct eigenvalues then it is diagonalizable. Ideally, they would mention the theorem that an *n x n square matrix is diagonalizable if and only if the sum of the dimensions of its eigenspaces is equal to n.*

In the lectures, SVD will be introduced as an extension of the diagonalization of a matrix. A comparison between the conditions for matrix diagonalization and for matrix singular value decomposition will be made in the first session, concerned with the more theoretical and structural aspects of the decomposition. The fact that SVD can always be applied, even for rectangular matrices, will be underlined.

Question 15: Basis

The question was: "Complete the sentence: Given a basis of a finite vector space, … element of the vector space can be expressed … as a … of basis vectors".

A fundamental notion of linear algebra is that of the basis of a vector space. At the completion of a linear algebra course, any student should know that there are many possible bases in a vector space, but they all have the same number of vectors. Also, it should be clear in their mind that any vector in a vector space can be expressed, uniquely, as a linear combination of the basis vectors, which act as a coordinate system. During the SVD lecture these definitions will be reinforced. Moreover, it will be clearly stated that, as opposed to diagonalization, in SVD one has to find a basis in the domain and usually another one in the range to create a pseudo-diagonal matrix and that such bases always exist.

Instructor's input in the first session

The instructor introduced the concepts and results about SVD using the lecture format interrupted from time to time by questions from students and exercises for them. The session can be divided into six parts or phases. First the theoretical results were introduced, and then students were directed to modify and execute the corresponding numerical examples. Using the projection screen, the teacher went through the slides underlining the theoretical aspects, the advantages, the geometrical representations and the applications of SVD.

Session I, Part 1– Historical introduction

The session started with historical information about the invention of SVD, in the context of differential geometry first with further developments in the domain of numerical methods. The intention was to present the topic as a relatively recent invention in mathematics, and so convey the notion that mathematics is a living area of activity: that it is a “hot” topic. This was expected to make the subject more exciting for the participants. This part of the lecture was meant also to familiarize participants with the key words of the topic – Singular Value Decomposition, singular values, and orthogonal transformations – without defining them. The introduction served thus also as an advanced organizer.

Below I reproduce, in a concise form, the content of the slides presented to the participants.

> SVD (Singular Value Decomposition) was first developed by differential geometers who wanted to determine whether a real bilinear form could be made equal to another by orthogonal transformations of the two spaces it acts on.
>
> In 1873 and 1874 respectively, E. Beltrami and C. Jordan discovered that the singular values form a complete set of invariants under orthogonal substitutions. In 1889, Sylvester also arrived at the SVD for real square matrices.
>
> In 1910, Emile Picard was the first to call the numbers σ_k singular values (or rather, “valeurs singulières”)
>
> In 1915, Autonne used the polar decomposition to arrive at the SVD.
>
> The first proof of the SVD for rectangular and complex matrices was done by Carl Eckart and Gale Young in 1936.
>
> J. E. Schmidt and H. Weyl took part in the final developments of the SVD in the mid-1900s.
>
> Methods for computing the SVD date back to Kogbetlianz in 1954, 1955 and Hestenes in 1958, resembling closely the Jacoby eigenvalue algorithm, which uses plane rotations or Givens rotations.
>
> These were replaced by the Golub-Kahan method published in 1965, which uses Householder transformations or reflections.

> In 1970, Golub and Rensch published a variant of the Golub-Kahan method that is still the one most used today.
>
> The Golub-Rensch algorithm was later improved by using different flavors of the Lanczos bidiagonalization algorithm. (Wikipedia, 2012).

The introduction ended with a first mention of the usefulness of SVD:

> From a numerical point of view SVD is more stable than the eigenvalue decomposition because the multiplying matrices are orthogonal (no inverses to be computed).
>
> Also, SVD is less prone to data perturbations.

The usefulness of SVD in computations aimed at deciding about basic properties of a matrix was highlighted at many points in the session.

Session I, Part 2 – Informal descriptions of the main idea of the SVD theorem

The general idea of the lecture was to go from a very informal presentation of the SVD theorem and gradually make the formulation more and more detailed and precise. The informal presentations were going from verbal, to diagrammatic, to geometric. The basis of the presentations was the view of matrices as transformations:

> A matrix represents a linear transformation from one vector space, the domain, to another, the range. Compared to eigenvalues, which are relevant only when the matrix is regarded as a transformation from one space onto itself, singular values are relevant when the matrix is regarded as a transformation from one space to a possibly different space of not necessarily the same dimension.
>
> The SVD states that for any linear transformation it is possible to choose an orthonormal basis for the domain and a possibly different orthonormal basis for the range.

Behind this formulation was a view of diagonalization of a matrix transformation as the result of a process of choosing a basis (or changing the coordinate system) in which the transformation simply stretches the space along the vectors of the basis (or axes) by factors equal to eigenvalues. SVD is likewise viewed as a product of choosing a basis so that the transformation is seen as a “stretcher” (Will, 2003), except that in the general case of rectangular matrices, two changes of basis may be needed, one for the domain and one for the range of the transformation. The SVD theorem states that even in such general case, particularly convenient bases can be found, namely orthonormal bases.

Before making the formulation more precise, it was necessary to introduce the notion of singular value. The first informal description of singular value was referring to its relationship with the singularity of a matrix. This was meant to justify the name "singular" for singular values: they are indicators of the singularity of a matrix.

> The term singular value relates to the distance of the given matrix to a singular matrix (a matrix that has at least one column linearly dependent on the other columns).
>
> A square diagonal matrix is nonsingular if and only if its diagonal elements are nonzero.
>
> The SVD implies that any square matrix is nonsingular if, and only if, its singular values are nonzero.

In the MAST 234-5 courses, given the simplicity of the matrices in the examples, row reduction and calculating determinants were the most commonly used techniques for determining the singularity of a matrix. It was therefore necessary to add a comment on the relevance of the new technique based on SVD:

> The most numerically reliable way to determine if a matrix is singular is to test its singular values. This is far better than trying to compute determinants, which have very bad numerical properties.
>
> Traditionally, courses in linear algebra use the reduced row echelon form (RREF), but the RREF is an unreliable tool for computation in the face of inexact data and arithmetic. Therefore, SVD can be regarded as a modern, computationally powerful replacement for the RREF.

The next step in the description of SVD was to visualize the matrices in the decomposition as rectangles (Figure 1, (AI Access)). Immediately, two forms of the decomposition were visualized: the full and the reduced SVD. The visualization and its description were displayed on the PowerPoint slide, where it appeared in the general form of decomposition of complex matrices, and in the Maple worksheet for students, where only the version for real matrices was given. Here, we reproduce the version for real matrices only.

> Any $m \times n$ matrix has a singular value decomposition. The factorization can be performed in two distinct ways:
> (1) the full SVD
> $A = U\Sigma V^T$ where Σ is an $m \times n$ rectangular, generalized diagonal matrix; U is an $m \times m$ square, orthogonal matrix, V is an $n \times n$ square, orthogonal matrix.
> U holds the left singular vectors of A; Σ has the singular values as its non-zero diagonal entries; and V holds the right singular vectors of A.

This text was followed by a diagram, as shown in Figure 1.

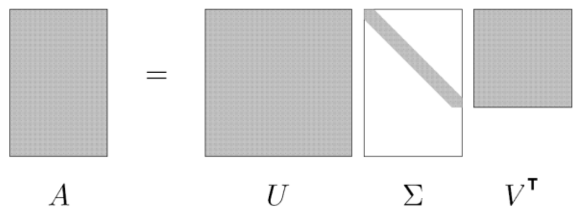


Figure 1. Visualization of the full SVD with matrices as rectangles

Next, the reduced SVD was presented and visualized:

> (2) the reduced SVD (for $m \geq n$)
> $A = U_r \Sigma_r {V_r}^T$ where Σ_r is an $r \times r$ diagonal matrix; U_r is an $m \times r$ rectangular matrix whose columns are orthogonal and norm are 1 , V_r is an $n \times r$ rectangular matrix whose columns are orthogonal and norms are 1.

A similar diagram was displayed for the reduced SVD (Figure 2, (AI Access)).

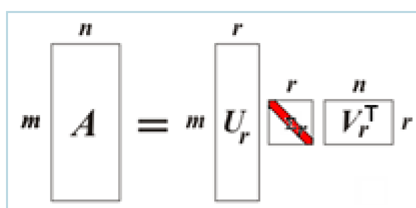


Figure 2. Visualization of the reduced SVD

The diagram was explained:

> This decomposition can be obtained by removing the $m - r$ rightmost columns of U, the $n - r$ lowest rows of V^T and by keeping the square upper-left part of Σ (containing the strictly positive singular values), and discarding the rest.

Two comments were given, one concerning the uniqueness of the reduced SVD and another highlighting the computational advantages of using the reduced SVD:

It can be shown that the reduced Singular Value Decomposition is unique (up to the signs of the singular vectors) if and only if all the positive singular values are distinct.

From a numerical point of view the reduced decomposition of SVD requires less storage (m+5n compared to 3m+3n, where m ≥ n). This can become significant when m >> n.

In a third step of gradual formalization of the presentation of the SVD, the contents of the matrices was represented in slightly more detail (Figure 3 (Strang, 2010)). The columns of U and the rows of V^T were drawn as thin rectangles and labeled as u_i and v_j^T respectively; the singular values were labeled as well as σ_k. Moreover, the relation between the left and right singular vectors and the fundamental subspaces of a matrix was already mentioned at this point.

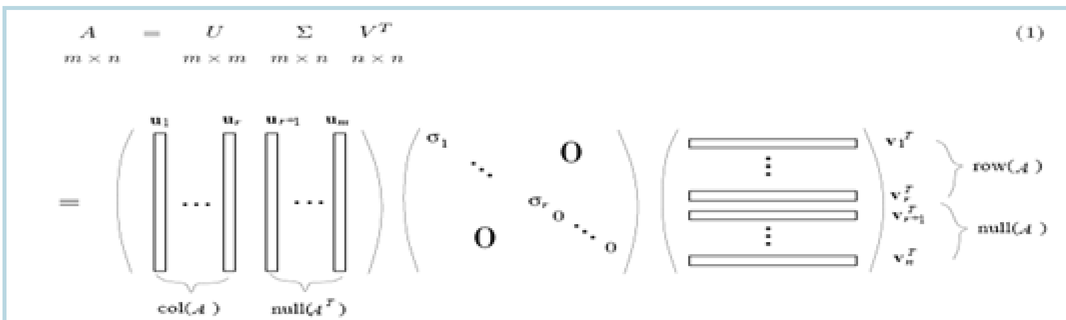


Figure 3. More detailed diagrammatic representation of SVD

In a fourth step of the informal presentation of SVD, a graphical representation of the action of a matrix on the unit circle was displayed, in terms of the combination of actions of the matrices V^T, Σ and U (Figure 4 (Strang, 2010)).

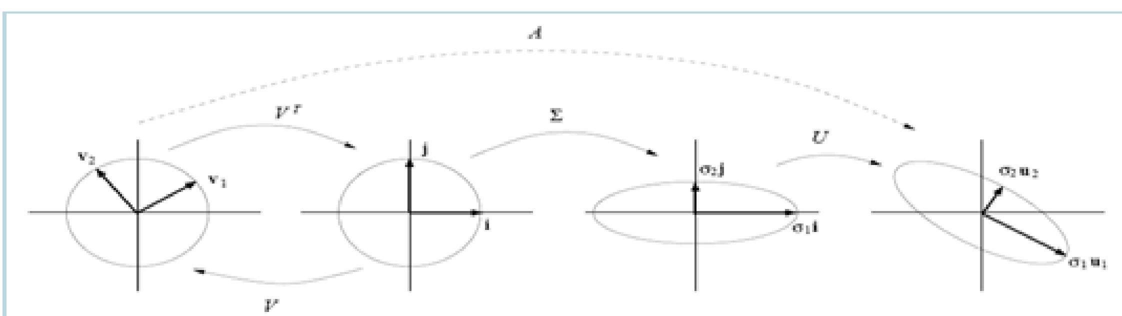


Figure 4. A breakdown of the action of a matrix on a unit circle into actions of its SVD components

This figure was commented upon as follows, suggesting viewing the general in the particular geometric instance:

As we have already seen an $m \times n$ matrix A can be regarded as the matrix representation of a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$.

Consider the SVD of A :

As U is an orthogonal $m \times m$ matrix, its columns form an orthonormal basis of $\mathbb{R}^m$.

As V is an orthogonal $n \times n$ matrix, its columns form an orthonormal basis of $\mathbb{R}^n$.

Let $\boldsymbol{x}$ be a vector in $\mathbb{R}^n$.

For simplicity, we assume A to be full rank, say, $n, n \leq m$.

Then to obtain Ax, we take the following steps :

First, represent $\boldsymbol{x}$ in the orthonormal basis of $\mathbb{R}^n$ made of the columns of V (the right singular vectors).

Then multiply ("stretch") each of the n coordinates of $\boldsymbol{x}$ in the orthonormal basis of $\mathbb{R}^n$ by the corresponding singular value of A.

Use these numbers as the first n coordinates of an $\mathbb{R}^m$ vector $\boldsymbol{y}$ in a basis made of the left singular vectors of A (i.e. the columns of U).

Set the remaining $m - n$ coordinates of $\boldsymbol{y}$ to 0.

The vector $\boldsymbol{y}$ obtained in this way is equal to $A\boldsymbol{x}$.

This more detailed geometrical explanation, which highlights the view of the action of matrices U and V^T as changes of the coordinate system was further visualized in terms of transformations of the orthogonal grid in the plane with the comment, “The image of the unit circle through the matrix M is an ellipse whose major and minor axes define the orthogonal grid in the range”, and a diagram shown in Figure 5 (Austin, 2012).

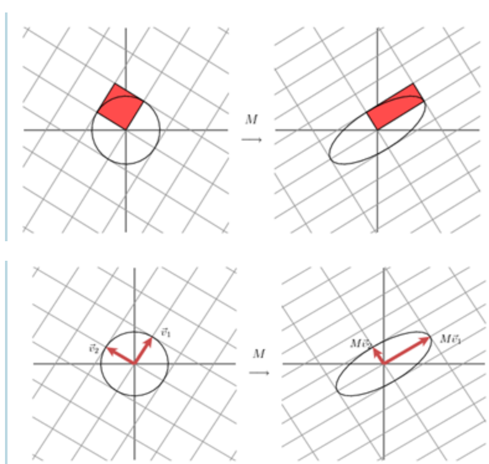


Figure 5. Visualization of SVD in terms of orthogonal grid transformations and stretching

This visualization was then generalized to higher dimensions.

In $\mathbb{R}^n$, consider the set of all unit vectors: their tips form the n-dimensional unit sphere. It can be shown that the tips of the A-transformed vectors are on an r-dimensional ellipsoid. Moreover, the directions of the principal axes of this ellipsoid are the columns of U_r (whose antecedents are the columns of V_r). The half-lengths of these principal axes are the singular values of A.

A drawing represented in Figure 6 (AI Access) was simultaneously shown, visualizing the generalization in three dimensions.

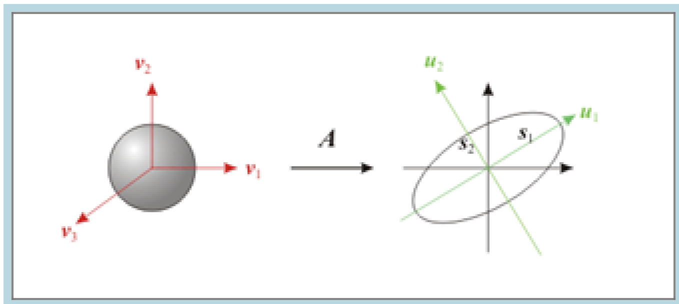


Figure 6. Transformation of a sphere into an ellipsoid

Session I, Part 3 – Theoretical results about SVD

The above informal introduction of SVD was concluded by stating, again, the SVD theorem, and listing properties of a matrix that can be read off its SVD:

1. Every matrix has a singular value decomposition.
2. The singular values σ_j are uniquely determined.
3. If A is square and σ_j are distinct, the left and right singular vectors u_j, v_j are uniquely determined up to complex signs, and, if A is real, they are uniquely determined up to signs.
4. The rank of A is r, the number of nonzero singular values.
5. $Range(A) = Span(u_1, \dots, u_r)$ and $Nullspace(A) = Span(v_{r+1}, \dots, v_n)$
6. The two-norm of A is its largest singular value: $\|A\|_2 = \sigma_1$
7. Nonzero eigenvalues of A^TA are nonzero σ_i^2; eigenvectors are v_i. Nonzero eigenvalues of AA^T are nonzero σ_i^2, eigenvectors are u_i.
8. Equivalent forms of SVD are:

$$A^TAv_j = \sigma_j^2 v_j$$

$$AA^Tu_j = \sigma_j^2 u_j$$

9. For a square matrix A, $\det(A) = \sigma_1\sigma_2 \cdot \sigma_m$.
10. For symmetric positive definite matrices A, the eigenvalue decomposition and SVD are identical.
11. The Kronecker expansion of a matrix.

The presentation of these theoretical results led to discussions with the students, who had to be reminded certain linear algebra definitions and notations. They were told that no formal proof will be attempted, but most of the above results will be illustrated in the next Maple worksheets.

Session 1, Part 4 – Numerical examples illustrating SVD, using Maple

Some of the above listed results (1, 4, 6, 7, 8, 5, 9, 10 and 11, in this order) were then illustrated with a few simple, small (2x2, 2x4) matrices with integer entries, using Maple (not by calculating the singular values and vectors from their defining properties).

Illustrating Property 4: The rank of A is r, the number of nonzero singular values

The first example shown to students was the SVD of the matrix

$$A = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$$

This triangular 2x2 matrix, with integer entries, was chosen because students could compute its singular values without the use of mathematical software. Thus, they could become reassured of the validity of the theoretical results presented to them.

The SVD of this matrix was then produced using the Maple command "SingularValues" (Figure 7).

$with(LinearAlgebra):$

$A := \langle\langle 4, 3\rangle | \langle 0, -5\rangle\rangle;$

$$\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$$

$U, S, Vt := SingularValues(A, output = ['U','S','Vt']);$

$$\begin{bmatrix} -0.447213595499958 & -0.894427190999916 \\ -0.894427190999916 & 0.447213595499958 \end{bmatrix}, \begin{bmatrix} 6.32455532033676 \\ 3.16227766016838 \end{bmatrix}, \begin{bmatrix} -0.707106781186547 & 0.707106781186547 \\ -0.707106781186547 & -0.707106781186547 \end{bmatrix}$$

Figure 7. Example of obtaining the SVD of a matrix using Maple

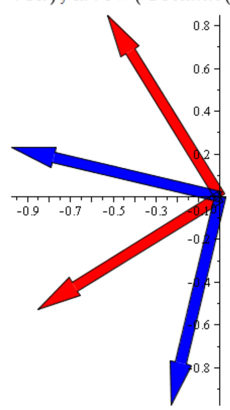
Students were shown how to interpret the output of the command and that they could verify that the matrices U and Vt were orthogonal (using the command "IsOrthogonal([matrix name])), and that the rank of A produced by the command "Rank(A)" is the same as the number of non-zero singular vectors.

In relation with the same matrix A , the use of the command SingularValues with output restricted to the list of singular values only – “SingularValues(A, output=[‘S’])” – produced the exact singular values ($2\sqrt{10}, \sqrt{10}$) and not their decimal approximations as in Figure 7. Students were made aware that different versions of the “SingularValue” Maple command will generate singular values in either exact or decimal format.

Students were also shown a figure containing the column vectors of the matrices *U* and *Vt* and attention was drawn to the perpendicularity of these vectors, thus offering a visual support to the notion of orthogonal matrix (Figure 8).

with(plots) :

display(arrow(Column(U, 1), color = blue), arrow(Column(U, 2), color = blue), arrow(Column(Vt, 1), color = red), arrow(Column(Vt, 2), color = red));



The above picture shows that the corresponding column vectors of matrices U and V are orthogonal.

Figure 8. A visualization of the notion of orthogonal matrix

Students could then change the entries of the matrix *A* and observe the properties of rank and orthogonality of the matrices *U* and *Vt* for themselves.

The second example was a singular matrix (still 2x2), to highlight the role of the assumption that it is the number of non-zero singular values that represents the rank of the matrix, and not the number of all singular values.

The third example was a rectangular matrix (2x4, integer entries) to highlight the applicability of SVD to not necessarily square matrices.

Illustrating Property 6. The two-norm of A is its largest singular value: $\|A\|_2 = \sigma_1$

The discussion of the property was extended to relations between singular values and two matrix norms: the two-norm and the Frobenius norm. The students were reminded the definitions of the vector norm and of the induced matrix norm. For a 2x4 matrix *A*, students were given to observe that the output of the matrix two-norm command "MatrixNorm(*A*,2)" and the maximal singular value of the matrix coincide and that the Frobenius norm command "MatrixNorm(*A*, Frobenius)", the computation of "sqrt(Trace(Transpose(*A*).A))" and of the square root of the sum of the squares of the singular values of *A* all produce the same result.

Illustrating Properties 7 and 8 for real matrices: Nonzero eigenvalues of $A^T A$ *are nonzero* σ_i^2; *eigenvectors are* v_i. *Nonzero eigenvalues of* AA^T *are nonzero* σ_i^2, *eigenvectors are* u_i. *Equivalent forms of SVD are:* $A^T A v_j = \sigma_j^2 v_j$, $AA^T u_j = \sigma_j^2 u_j$

It is only at this point that singular values of a matrix A are presented as square roots of the eigenvalues of the matrix $A^T A$. Textbook presentations of the SVD topic usually start by defining singular values this way. In this lecture, the starting point was an overview of the general shape of the matrix components of the SVD of a matrix; the lecturer was then only gradually "zooming in" to look at the entries of the component matrices.

As in the illustration of the previous property, students were invited to validate them by calculating (in Maple) objects in the two ways deemed equivalent by a property and comparing the results. For example – applying the "SingularValues(A,output=[…])" and "Eigenvectors($A^T A$)" commands and comparing the results, for the matrix:

$$A = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$$

Students' worksheets contained the commands allowing them to execute the comparison (using "evalb" – to evaluate the "Boolean value" of an equality as true or false) and try it also on a different matrix of their choice. We include an excerpt of the worksheet in Figure 9.

Illustration of Properties 9 and 10 was done in a similar fashion. The determinant of the same matrix A as above was calculated in two ways, using the "Determinant" command and the product of singular values. For illustration of the Property 10, regarding SVDs of positive definite matrices, the matrix $\begin{bmatrix} 4 & 3 \\ 3 & 5 \end{bmatrix}$ was used (Figure 9).

$A;$

$$\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}$$

$S := SingularValues(A);$

$$\begin{bmatrix} 2\sqrt{10} \\ \sqrt{10} \end{bmatrix}$$

$v, e := Eigenvectors(Transpose(A).A);$

$$\begin{bmatrix} 40 \\ 10 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$$

$evalb(S[1] = \mathrm{sqrt}(v[1]));$

true

$evalb(S[2] = \mathrm{sqrt}(v[2]));$

true

Figure 9. Validating the property of singular values of A as square roots of eigenvectors of the associated symmetric matrix A^TA

Illustration of Property 5: Relations between matrices U and V and the fundamental subspaces of the matrix A

This property was awarded the most attention in the sequence of examples. It was named "The fundamental theorem of linear algebra" or "SVD vector basis theorem".

The diagram in Figure 3 was reproduced in the students' worksheet, and statements about the relationship between column vectors of the matrices U and V and bases of the fundamental spaces of the matrix were listed Figure 10.

A basis for Col(A) (Range of A or ColSpace) is: $[u_1, u_2 ... u_r]$
A basis for Null(A) (Kernel of A) is: $[v_{r+1}, v_{r+2}, ... v_n]$
A basis for Col(A^T) (Range of A^T or RowSpace) is: $[v_1, v_2, ... v_r]$
A basis for Null(A^T) (Kernel of A^T) is: $[u_{r+1}, u_{r+2}, ... u_m]$

And we have $\mathrm{Col}(A) \oplus \mathrm{Null}(A^T) = \mathbb{R}^m$ and $\mathrm{Col}(A^T) \oplus \mathrm{Null}(A) = \mathbb{R}^n$

Figure 10. The "Fundamental Theorem of Linear Algebra"

The relationships were illustrated using the 2x4 matrix $A = \begin{bmatrix} 0 & 1 & 4 & 0 \\ 0 & 2 & 8 & 0 \end{bmatrix}$. The decimal approximations of the matrices U, Σ, V^T were found using the command "$U, \Sigma, Vt := SingularValues(A, output = ['U', 'S', 'Vt'])$", but the exact values of the singular values and vectors were also calculated using the “Eigenvectors” command. This produced the SVD component matrices:

$$U = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}, \Sigma = \begin{bmatrix} \sqrt{85} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } V = \begin{bmatrix} 0 & 0 & 0 & 1 \\ \frac{1}{\sqrt{17}} & 0 & -\frac{4}{\sqrt{17}} & 0 \\ \frac{4}{\sqrt{17}} & 0 & \frac{1}{\sqrt{17}} & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Multiplying the matrix A by the matrix V, the result was:

$$A.V = \begin{bmatrix} \sqrt{17} & 0 & 0 & 0 \\ 2\sqrt{17} & 0 & 0 & 0 \end{bmatrix}$$

The students were told to observe that “because the 3 rightmost columns of *V* were mapped to zero by *A* , they are a basis for the nullspace of *A*”. Next, students were asked to “compute NullSpace(A) to validate” this statement. The command “$NullSpace(A)$” produced a basis that was slightly different from the last three columns of V. The students were asked to explain why these bases are in reality the same.

Illustration of Property 11: The Kronecker decomposition of a matrix

First, the teacher presented, in its algebraic form, the theoretical result that any matrix A can be written as a sum of products of norm 1 vectors resulting from the singular values decomposition of A:

> If $A = U.\Sigma.V^T$, then A can be written as
> $$A = \sigma_1.u_1.v_1^T + \sigma_2.u_2.v_2^T + \cdots + \sigma_r.u_r.v_r^T,$$
> where r is the rank of A, and u_i and v_i are the columns of U and V^T respectively.

Then, he directs students to a numerical example of a 2 x 4 matrix of rang 1,

$$A = \begin{bmatrix} 0 & 1 & 4 & 0 \\ 0 & 2 & 8 & 0 \end{bmatrix}$$

The participants are asked to confirm that the rank of A is indeed 1 and to obtain the decomposition of A, using the SingularValues command in a Maple worksheet:

$U, \Sigma, Vt := SingularValues(A, output = ['U', 'S', 'Vt'])$.

The next step is to confirm, for this particular matrix, that the Kronecker identity holds. To achieve this, students will compute the 2 x 4 matrix $\sigma_1.u_1.v_1^T$ using the “Multiply” Maple command, and will compare the result with the original matrix A. This comparison is another good opportunity for the teacher to underline the fact that round-off errors exist, due to the algorithmic nature of SVD.

Next, students are encouraged to repeat the above procedure for matrices of different shape and rank and confirm that the identity still holds (barring inherent round-off errors).

Session 1, Part 5: Singular value and Eigenvalue decomposition comparison

In this part of the lecture, the following points were made:

1) Eigenvalues are important in situations where the matrix is a linear transformation from one vector space onto itself (for example: systems of linear differential equations). Singular values are essential where the matrix is a transformation from one vector space to a different vector space, possibly of different dimensions (for example: systems of equations that are over or underdetermined).

2) As such, eigenvalues can be computed only for square matrices. Singular values exist for any m x n rectangular matrix (due to the fact that $A^T.A$ is always a symmetric matrix).

3) The eigenvalue decomposition uses the same basis for row and column spaces and this basis is not necessarily orthonormal. SVD uses different orthonormal matrices (U and V) for the decomposition.

4) The reason for the eigenvalues decomposition is to find a basis for the space in which the linear transformation can be represented by a diagonal matrix. This basis may be complex even if the matrix A is real. Moreover, if the number of linearly independent eigenvectors is not equal to the dimension of the space, such a basis does not even exist.

In the SVD case, we try to find one change of basis in the domain and usually a different one in the range, so that the matrix becomes diagonal. Such bases always exist and, if A is real, they are real.

5) In the case of SVD, the fact that the change-of-basis matrices are orthonormal means that they preserve lengths and angles, so they do not magnify errors. As such, perturbations of any size in the original matrix cause perturbations of roughly the same size in the singular values. On the other hand, the eigenvalues of certain matrices are sensitive to perturbations.

6) The condition number of a matrix A is defined as: $cond(A) = \|A\|_2 \|A^{-1}\|_2$ if A is nonsingular and +∞ if A is singular. This number expresses how sensitive the matrix A is to perturbations in its elements. Using SVD the above number can be easily computed as: $(A) = \frac{\sigma_1}{\sigma_r}$, and it is always ≥ 1.

Matrices with condition numbers close to 1 are well-conditioned. If $cond(A)$ is very high for a given matrix, the matrix is said to be ill-conditioned.

> The computation of eigenvalues of symmetric matrices is numerically stable. On the other hand, the computation of eigenvalues as roots of the characteristic polynomial: $\det(A - \lambda.\mathrm{I}) = 0$, is extremely sensitive to perturbations, which means it does not provide for a robust software algorithm. The most numerically reliable way to determine whether matrices are singular is to test their singular values. This is far better than trying to compute determinants, which have atrocious scaling properties. Also, because reducing a matrix to its reduced row echelon form (RREF) is an unreliable tool for computation due to its sensitivity to inexact data, SVD can be seen as a modern, computationally stable replacement for RREF.

At this point, the example of the 20 x 20 diagonal matrix with diagonal entries equal to the consecutive numbers from 1 to 20 was presented (Wilkinson, 1963). This matrix was used to illustrate the fact that the computation of the eigenvalues is ill-conditioned: a small perturbation in the coefficients of the characteristic polynomial could dramatically change its eigenvalues.

Next, the students were encouraged to slightly modify the coefficients of different monomials of the given characteristic polynomial and to observe the effects of these changes on the computed eigenvalues.

In retrospect, I think I could have made the same point using a smaller matrix. Maple compacts square matrices larger than 10×10, thus students had difficulties in visualizing the initial matrix given in the lecture. Using the 5×5 matrix as in Figure 11 , I could have shown that by modifying the coefficient of λ^4 from 15 to 15.01, the eigenvalues "shift" from the expected $\{1,2,3,4,5\}$ to $\{1.0004, 1.9751, 3.4147 - 0.0851\mathrm{i}, 3.4147 + 0.0851\mathrm{i}, 5.2049\}$.

Commenting on these results, I could have highlighted the fact that the third and fourth eigenvalues became complex numbers and that the error in the fifth eigenvalue is significant (a perturbation of magnitude 10^{-2} in the coefficient introduced a delta of $2 * 10^{-1}$ in the fifth eigenvalue computation: from 5 to 5.2049).

$with(LinearAlgebra):$

$$A := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix};$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

$$Id := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix};$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$Determinant(A - \lambda \cdot Id);$

$$-(1-\lambda)\,(2-\lambda)\,(\lambda-3)\,(\lambda-4)\,(\lambda-5)$$

$solve((\lambda-1)\cdot(\lambda-2)\cdot(\lambda-3)\cdot(\lambda-4)\cdot(\lambda-5)=0, \lambda);$

$$1, 2, 3, 4, 5$$

$expand((\lambda-1)\cdot(\lambda-2)\cdot(\lambda-3)\cdot(\lambda-4)\cdot(\lambda-5));$

$$\lambda^5 - 15\,\lambda^4 + 85\,\lambda^3 - 225\,\lambda^2 + 274\,\lambda - 120$$

$solve(\lambda^5 - 15\,\lambda^4 + 85\,\lambda^3 - 225\,\lambda^2 + 274\,\lambda - 120 = 0, \lambda);$

$$1, 2, 3, 4, 5$$

$solve(\lambda^5 - 15.01\,\lambda^4 + 85\,\lambda^3 - 225\,\lambda^2 + 274\,\lambda - 120 = 0, \lambda);$

$$1.000417727,\ 1.975136863,\ 5.204976521,\ 3.414734445 + 0.08516012844\,I,\ 3.414734445 - 0.08516012844\,I$$

Figure 11. Showing that the eigenvalues computation is ill-conditioned on the example of a special 5 x 5 diagonal matrix

Session 1, Part 6: Applications of SVD

The following applications of SVD were briefly presented in the lecture:

- The pseudo-inverse of a matrix
- Calculating the inverse of an invertible square matrix
- Low-rank approximation
- Closest orthonormal matrix
- Signal and image processing
- Data compression
- Data analysis

The pseudo-inverse of a matrix

This part of the session started with a presentation of the definitions of the exact pseudo-inverse of a matrix of full rank. Participants were informed that when the rows are linearly independent, $A.A^T$ is invertible the exact pseudo-inverse is then defined as $A^+ = A^T.(A.A^T)^{-1}$. When the columns are linearly independent, $A^T.A$ is invertible; and the exact pseudo-inverse is defined as $A^+ = (A^T.A)^{-1}.A^T$. Then, the students were presented with the following four criteria met by the exact pseudo-inverse (Wikipedia, 2012):

1. $A.A^+.A = A$
2. $A^+.A.A^+ = A^+$
3. $(A.A^+)^T = A.A^+$
4. $(A^+.A)^T = A^+.A$

Participants were also told about the following basic properties of the pseudo-inverse:

1. If the matrix A is invertible, then $A^+ = A^{-1}$.
2. The pseudo-inverse of the pseudo-inverse is the original matrix: $(A^+)^+ = A$.
3. The pseudo-inverse commutes with transposition: $(A^T)^+ = (A^+)^T$.

If the singular value decomposition of A is $A = U.\Sigma.V^T$, then the exact pseudo-inverse of A can be easily obtained from $A^+ = V.\Sigma^+.U^T$, where Σ^+ has $1/\sigma_i$ on its diagonal for non-zero singular values and 0 everywhere else.

Next, the notion of approximative pseudo-inverse was introduced: $A^+ = V_r.\Sigma_r{}^{-1}.U_r{}^T$, where the matrices are size r appropriate submatrices of U, V, Σ.

Only now, the participants are told that the main application of the pseudo-inverse is solving systems of linear equations with more equations that variables. The general solution of the system $A.x = b$ is given in the form $x = A^+.b + (I - A^+.A)z$, where z is an arbitrary vector and I is the identity matrix. Thus, the closest solution in the least-square sense is obtained when $z = 0$ and it is $x = A^+.b$ (A^+ is the exact pseudo-inverse if A is of full rank and the approximative pseudo-inverse if not).

A numerical example of a system of equations with the coefficient matrix not of full rank was given. Participants were asked to compute the pseudo-inverse of the matrix associated to the system, following its singular value decomposition. Then, they were asked to validate that $x = A^{+}.b$ is the closest solution to b in the least-squares sense.

Calculating the inverse of an invertible square matrix

For an invertible square matrix A, with the singular value decomposition $A = U.\Sigma.V^T$, its inverse can be determined without tedious determinants computations as $A^{-1} = V.\Sigma^{-1}.U^T$.

Students were presented with a Maple worksheet in which the above method of computing the inverse of A was employed. Then, they were asked to compare this result for the inverse to the one returned by the Maple command A^{-1} and to validate that they are equal, barring round-off errors. The participants were encouraged to modify the initial square matrix, to validate that it is invertible and then observe that computing the matrix inverse starting from its singular value decomposition or by using Maple commands (A^{-1} or $se(A)$) generate the same results.

Low-rank approximation

From the Kronecker expansion $A = \sigma_1.u_1.v_1^T + \sigma_2.u_2.v_2^T + \cdots + \sigma_r.u_r.v_r^T$, where r is the rank of A, we observe that if we set all but the first k singular values to zero and we use only the first k columns of U and V we get the best low-rank approximation for A, i.e. the matrix given by $A_k = \sigma_1.u_1.v_1^T + \sigma_2.u_2.v_2^T + \cdots + \sigma_k.u_k.v_k^T$.

The theoretical results that $\|A - A_k\|_2 = \sigma_{k+1}$ and $\|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \cdots\ \sigma_r^2}$ were also mentioned.

Closest orthonormal matrix

The notion that distances between matrices can be determined using matrix norms was revisited. The question addressed here was how to determine the closest orthonormal matrix to a given matrix A.

If A has the singular value decomposition $= U.\Sigma.V^T$, then it can be shown that Q, the closest orthonormal matrix to A, can be obtained by replacing all the singular values of A with 1. Thus, $Q = U.V^T$. Participants were asked to notice that Q is indeed orthonormal, being the product of two orthonormal matrices. Participants were asked to compute Q with the above formula and determine the distance to A (in terms of Frobenius norm). Then, they were required to construct another random orthonormal matrix, compute the distance to it and compare it to the distance to Q.

Signal and image processing

The fact that SVD can be used in digital signal and image processing was highlighted. An example of a scanned image with slight imperfections (Austin, 2012) was given to students in a Maple worksheet. The instructor explained that by removing the small singular values and applying a low rank approximation to the original image matrix, the resulting image looks improved, because the noise was eliminated. The students were asked to comment on the choice of selecting only the first three significant singular values in the low rank approximation of the original matrix.

Data compression

The teacher made clear fact that a picture can be stored into a large matrix in which each pixel is represented by a number that records the light intensity. To demonstrate this, an image with associated matrix of rank 200 was displayed. Several images, obtained by taking rank approximations – via Kronecker expansion – of ranks 1, 2, 5, 15 and 50 respectively were displayed in succession. Students observed that the quality of the image improved with every increase in rank and that the image corresponding to the rank 50 approximation was already very close to the original. They were asked to comment on the reduction in storage space if the rank 50 approximation of the original image is saved.

A second example, in which the original 25 x 15 image can be broken down into small, repetitive patterns, was also shown. The instructor highlighted the fact that the original matrix had only three non-zero singular values, thus only three vectors v_i with

15 entries each, three vectors u_i with 25 entries each and three singular values σ_i need to be stored. Students observed that this makes up for 123 total numbers stored compared to the original 375, a 67% compression rate.

Data analysis

The instructor guided students through an example where noise arose during a data collection process. The example displayed 10 points representing collected data given by their x, y coordinates. (Austin, 2012) Then the corresponding 2 x 10 matrix was built and the Maple *SingularValues* command was applied to it. Now, the participants observed that the second singular value was much smaller than the leading one, thus they could assume that this was due to noise in the collected data. They were asked to make this second singular value 0 and notice the changes that this entails: the points representing the coordinates were closer to the line defined by u_i.

The teacher mentioned that this example introduces the modern field of Principal Component Analysis, a set of techniques that use singular values to detect dependencies and redundancies in data.

Guided practice

The teacher elicited the students' opinion on the outcome of the numerical examples. The participants could interrupt at any time to ask questions. There was time allocated for exercises and the teacher commented on the results. The students were asked to work on a short home assignment.

Closure

At the end of this session, the instructor recapitulated the key-points of the lecture and announced the main themes of the next session.

4.1.2 Session 2, titled "Computational aspects of SVD"

Purpose

The purpose of this second session was to familiarize the participants with the computational aspects of SVD.

Learning outcomes

In the proposed computational approach to teaching SVD, participants were expected to become acquainted with the numerical algorithms behind SVD and their advantages with regard to numerical stability to data perturbations, complexity and algorithm performance.

Bridge-in

Students were asked to change initial values of algorithms and confirm certain theoretical aspects (for example the non-uniqueness of the decomposition) on examples. The teacher encouraged students to modify the Maple scripts by trying the algorithms on new matrices, or with new initial vector values.

Instructor's input

The instructor taught numerical procedures behind the singular value decomposition, using the lecture format interrupted from time to time by questions from students and exercises for them. The session had three parts:

1) General outline of the procedure;
2) Detailed presentation of the first step of the procedure, i.e., bidiagonalization;
3) Detailed presentation of the second step of the procedure, i.e., diagonalization.

Session 2, Part 1– General outline of a procedure to compute the SVD

This session started by a general description of the steps involved in the algorithm of computing SVD of a matrix, supported with a visual representation (Figure 12 (Eiland, 2011)).

Students’ Maple worksheets contained a verbal description of the steps in terms of their outcome, and the visual representation.

> The algorithm of computing the SVD of a matrix involves two major steps:
>
> 1) Reduce the initial matrix to a bidiagonal form
>
> 2) Diagonalize the bidiagonal form

$$\underset{\text{Initial Matrix}}{\begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{bmatrix}} \rightarrow \underset{\text{Bidiagonal Form}}{\begin{bmatrix} x & x & & & \\ & x & x & & \\ & & x & x & \\ & & & x & x \\ & & & & x \end{bmatrix}} \rightarrow \underset{\text{Diagonal Form}}{\begin{bmatrix} x & & & & \\ & x & & & \\ & & x & & \\ & & & x & \\ & & & & x \end{bmatrix}}$$

Figure 12. Representation of the steps to compute SVD of a matrix

The same text was displayed on the projector screen and the instructor explained the text by pointing to relevant aspects of the visual representation. The instructor said that the first step is usually done using Householder transformations, but in this lecture a different method will be shown, namely a double recursion that computes only the diagonal and the superdiagonal elements, called the Golub-Kahan-Lanczos algorithm. The instructor stressed that this bidiagonalization algorithm is applicable to matrices of any shape. He also said that the second step can be done using QR transformations or, as he will demonstrate later, by Givens rotations.

Session 2, Part 2 – Bidiagonalization algorithm (Golub-Kahan-Lanczos)

This part of Session 2 was devoted to the first step of the algorithm, i.e., the bidiagonalization procedure. The source for this algorithm was Demmel (2000). Below is an approximate script of the lecture.

> Given an nxn square matrix A, the Golub-Kahan-Lanczos bidiagonalization procedure computes a matrix that has non-zero values only on the diagonal and the superdiagonal (bidiagonal form), by using orthonormal transformation matrices. In other words, orthogonal nxn matrices U and V and a bidiagonal nxn matrix B can be computed, such that
>
> $A = U.B.V^T$ (1)

The Golub-Kahan-Lanczos bidiagonalization procedure can also be applied to $m\ x\ n$ rectangular matrices (with $m > n$).

Rewriting (1) and using the information that B is bidiagonal:

$$U^T.A.\mathrm{V} = B = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & 0 & 0 & 0 \\ 0 & \alpha_2 & \beta_2 & 0 & 0 & 0 \\ 0 & 0 & \alpha_3 & \beta_3 & 0 & 0 \\ 0 & 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & 0 & 0 & \alpha_n - 1 & \beta_n - 1 \\ 0 & 0 & 0 & 0 & 0 & \alpha_n \end{bmatrix}, \textit{where } \alpha_i, \beta_i \textit{ are real numbers } (2)$$

As a side note, the indices in the accompanying figures were 1-based, but the algorithm used 0-based indices (a convention used in many object-oriented programming languages).

The constants α_k and β_k are given by: $\alpha_k = u_k^T A v_k$ and $\beta_k = u_k^T A v_{k+1}$, where u_k and v_k are columns of the matrices U and V, respectively.

From $A.V = U.B$ we have:

$$A[v_1\, v_2 \dots v_n] = [u_1\, u_2\ \dots\ u_n] \begin{bmatrix} \alpha_1 & \beta_1 & 0 & 0 & 0 & 0 \\ 0 & \alpha_2 & \beta_2 & 0 & 0 & 0 \\ 0 & 0 & \alpha_3 & \beta_3 & 0 & 0 \\ 0 & 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & 0 & 0 & \alpha_n - 1 & \beta_n - 1 \\ 0 & 0 & 0 & 0 & 0 & \alpha_n \end{bmatrix} (3)$$

The columns of U and V can be derived from a double recursion as described below.

Equating the k^{th} columns in (3) we get:

$$A.v_k = \beta_{k-1}\,.u_{k-1} + \alpha_k\,.u_k \quad (4)$$

At this point understanding of the product of a matrix by a column vector as a linear combination of the columns of the matrix is necessary. It is also necessary to understand the concise notation in (4) as representing any of the columns of the matrices of the left and right of the equality sign. Knowing from the Pre-test results that participants had trouble with multiplication of matrices and vectors, this could have been a difficult point for them. Some may have been lost at this point.

The equality (4) is equivalent to:

$\alpha_k \,.u_k = A.v_k - \beta_{k-1}\,.u_{k-1}$ (4`)

On the other hand, from $U^T.A.V = B$ and using the fact that $V^{-1} = V^T$, since V is orthogonal, we get: $U^T A = BV^T$. By transposing both sides, we obtain,

$A^T U = VB^T$ (5)

Now (5) can be re-written as:

$$A^T[u_1\ u_2\ ..\ u_n] = [v_1\ v_2\ ..\ v_n] \begin{bmatrix} \alpha_1 & 0 & 0 & 0 & & 0 & 0 \\ \beta_1 & \alpha_2 & 0 & 0 & & 0 & 0 \\ 0 & \beta_2 & \alpha_3 & 0 & & 0 & 0 \\ 0 & 0 & \beta_3 & \dots & & 0 & 0 \\ 0 & 0 & 0 & \dots & \alpha_{n-1} & & 0 \\ 0 & 0 & 0 & 0 & \beta_{n-1} & & \alpha_n \end{bmatrix} \quad (6)$$

Equating the k^{th} columns in (6) we get:

$A^T u_k = \alpha_k\,.v_k + \beta_k\,.v_{k+1}$ (7)

The equality (7) is equivalent to

$\beta_k\,.v_{k+1} = A^T u_k - \alpha_k\,.v_k$ (7`)

The double recursion consists of the equations (4') and (7').

Also, since all columns of U and V are normalized we have, from (4') and (7'):

$\alpha_k = \|A.v_k - \beta_{k-1}\,.u_{k-1}\|,\ \beta_k = \|A^T u_k - \alpha_k\,.v_k\|$ (8)

If the participants were not very familiar with recursive processes – although some may have seen iterative methods in the context of solving linear systems (e.g. the Jacobi iteration method) or discrete dynamical systems (e.g. Markov chains) – the idea of how the equations (4') and (7') represented a recipe for computing the entries alpha and beta of the bidiagonal form of the given matrix could have been somewhat obscure for them. It was hoped, however, that the upcoming program of calculation and the concrete numerical examples would clarify the notion.

The Golub-Kahan-Lanczos bidiagonalization procedure has the following steps:

Step 1: Choose an initial normalized vector v_0 and set $\beta_{-1} = 0$ and $u_{-1} = \theta$

Step 2: For $k = 0, 1, 2, \ldots, n$, do:

Compute a vector in the direction of the next u_k, from 4`:

$$u'_k = A.v_k - \beta_{k-1} . u_{k-1}$$

Calculate its 2-norm, which, by (8), is the corresponding alpha entry of the bidiagonal form of the given matrix A:

$$\alpha_k = \|u'_k\|$$

Normalize u'_k to obtain the next column of the matrix U:

$$u_k = \frac{u'_k}{\alpha_k}$$

If $k < n$, compute a vector in the direction of the next v_k, from 7`:

$$v'_{k+1} = A^T u_k - \alpha_k . v_k$$

Calculate its 2-norm, which, by (8), is the corresponding beta entry of the bidiagonal form of the given matrix A:

$$\beta_k = \|v'_{k+1}\|$$

Normalize v'_{k+1} to obtain the next column of the matrix V:

$$v_{k+1} = \frac{1}{\beta_k} v'_{k+1}$$

Step 3: Form the matrices U, V and B using the obtained coefficients alpha, beta, and vectors u and v: The matrix U is obtained by concatenating the vectors u_k.

The matrix B has α_k on the diagonal, β_k on the superdiagonal and zeros everywhere else; this is the bidiagonal form of the matrix A. The matrix V is obtained by concatenating the vectors v_k.

At this juncture the participants were directed to an example in their worksheets. They were given the following 3×3 matrix with integer entries:

$$A = \begin{bmatrix} -9 & 1 & -4 \\ -5 & 6 & -10 \\ -10 & -4 & -4 \end{bmatrix}$$

They were instructed to consider the initial, normalized vector,

$$v_0 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

take its decimal approximation (using the "evalf" command in Maple) and to follow the steps described in the above bidiagonalization procedure. The commands for the steps were already written for them in the worksheet and participants only had to press ENTER at each command. They could see that after three steps an approximation of the bidiagonal form of the initial matrix was computed (we name it "B"):

$$B = \begin{bmatrix} 12.12435565 & 8.602325266 & 0 \\ 0 & 5.89850262029624250 & 2.239894303 \\ 0 & 0 & 11.4101069395020982 \end{bmatrix}$$

Then, students were asked to repeat the same steps, starting with different initial vectors and notice that the resulting bidiagonal forms differ. In the next part of the session, they were shown that the diagonalization procedure does not depend on the choice of the initial vector: all those different bidiagonal forms lead nevertheless to the same set of singular values.

Session 2, Part 3 – The diagonalization algorithm (with Givens rotations)

In this part of the session, the diagonalization step of the singular value decomposition of a matrix was presented. This step takes the bidiagonal form of the matrix obtained at the end of the bidiagonalization step and, following a series of transformations using Givens rotations, reduces the superdiagonal values to values lesser than a preset threshold (Eiland, 2011). The resulting quasi-diagonal matrix holds approximations of the singular values on the diagonal, with the assumed accuracy. This part of the session was introduced by a return to the visual representation displayed in Figure 12.

Next, participants were directed to their worksheets which contained the following procedure (Demmel & Kahan, 1990) to calculate a Givens rotation matrix for a vector:

```
rotate:=proc(f:numeric, g:numeric)
    module()
    option package;
    local t,tt;
    export cx,sx,rx;
    if f=0 then
      cx:=0;sx:=1;rx:=g
    elif evalf(abs(g)) <evalf(abs(f)) then
      t:=g/f;tt:=sqrt(1+t*t);cx:=1/tt;sx:=t*cx;rx:=g*tt;
    else
      t:=f/g;tt:=sqrt(1+t*t);sx:=1/tt;cx:=t*sx;rx:=g*tt;
    end if
    end module
    end proc
```

The Maple command "*GivensRotationMatrix(V,i,j)"* was not used because the point was to show participants the procedures behind the Maple command, even if they were not explained in detail. Participants were told that the Givens rotation matrices are used to turn an entry in a row or a column of a matrix into 0. To create a zero in a row of the original matrix, it is enough to post-multiply it with the transpose of an appropriate Givens rotation matrix. To create a zero in a column, the original matrix is pre-multiplied by a Givens rotation matrix. After a sufficient number of such multiplications, the original transformed into a matrix that is still bidiagonal, but the superdiagonal entries are reduced.

Due to time constraints, the Givens rotation matrices were not given the attention that they deserve. This would have been a very good opportunity to show graphically, in $\mathbb{R}^2$, the effect of multiplying a Givens rotation matrix with a given vector. Participants could have computed the angle that rotates the given vector onto the y-axis, for example. Then, they could have constructed the corresponding Givens rotation matrix and use Maple to multiply it with the original vector and graphically display the resulting vector.

At this point, the instructor directed participants to Figure 13 (Eiland, 2011), and explained in general terms the main idea of a full iteration cycle of transforming a given bidiagonal matrix with one having smaller superdiagonal entries. In the diagram, the plus sign (+) represents the target entry, or the entry to be annihilated. The participants were told that the diagram depicts the full cycle of reducing the superdiagonal entries of a 5 x 5 matrix.

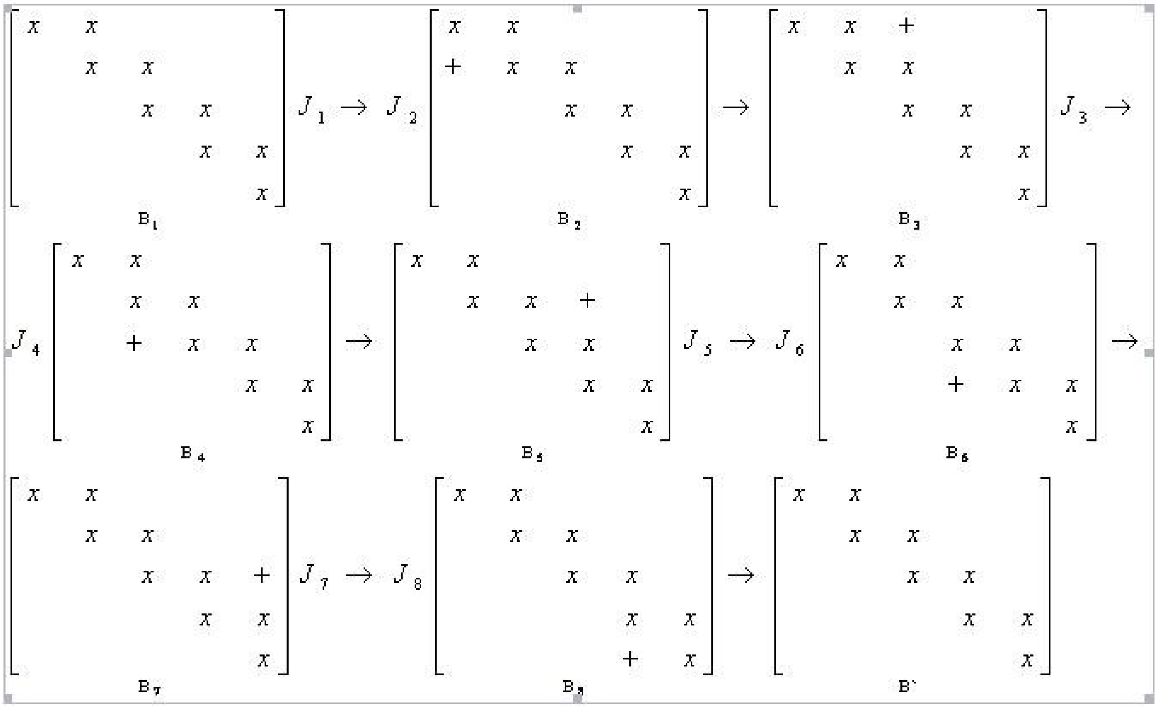


Figure 13. A visual representation of the full iteration cycle of reducing superdiagonal entries of a matrix through multiplications by Givens rotation matrices

Next, participants were directed to their worksheets and asked to perform a cycle of 2(n-1) multiplications (n being the size of the original square matrix) by appropriate Givens rotation matrices of the bidiagonal matrix B obtained at the end of the second part of this session, alternating between post-multiplication by the transpose of a Givens matrix and pre-multiplication by such a matrix. The initial matrix thus was:

$$B = \begin{bmatrix} 12.12435565 & 8.602325266 & 0 \\ 0 & 5.89850262029624250 & 2.239894303 \\ 0 & 0 & 11.4101069395020982 \end{bmatrix}$$

The resulting matrix was:

$$B_1 = \begin{bmatrix} 15.30885040 & 0.3571210470 & -1.136080339 * 10^{-10} \\ 0 & 11.30809957 & 2.533749742 \\ 0 & 1 * 10^{-9} & 4.713656767 \end{bmatrix}$$

Participants were asked to note that at least one of the values on the superdiagonal of B_1 is much smaller than the corresponding value from the original matrix B.

Next, participants were told that we may still not be satisfied with the reduction. Normally, we decide on how small we want the superdiagonal entries to be by establishing a threshold value (we name it “EPS”, short for “epsilon”). We then iterate the cycles of pre- and post-multiplications by appropriate Givens rotation matrices until the superdiagonal entries are less than EPS. The end product of this process is a matrix that is almost diagonal, with the required degree of accuracy. The diagonal entries are satisfactory approximations of the singular values of the original matrix. Participants were directed to their worksheets, where they were given a program that repetitively performs the cycle described above and stops only when the entries located above the diagonal are less than a given EPS value. The program, written by the author of this thesis, is presented in Figure 14.

```
EPS:=evalhf(1.0*((10)^(-1)));
while evalb(EPS<evalhf(A[1,2])) do
 AT:=Transpose(A):
        v:=AT[1..3,1]:
        rot:=rotate(v[1],v[2]):
        J:= Matrix([[rot:-cx,rot:-sx,0], [-rot:-sx,rot:-cx,0], [0, 0, 1]]):
        JT:=Transpose(J):
        A:=MatrixMatrixMultiply(A,JT):

        v:=A[1..3,1]:
        rot:=rotate(v[1],v[2]):
        J:= Matrix([[rot:-cx,rot:-sx,0], [-rot:-sx,rot:-cx,0], [0, 0, 1]]):
        A:=MatrixMatrixMultiply(J,A):

        AT:=Transpose(A):
        v:=AT[1..3,2]:
        rot:=rotate(v[2],v[3]):
        J:= Matrix([[1,0,0],[0,rot:-cx,rot:-sx], [0,-rot:-sx,rot:-cx]]):
        JT:=Transpose(J):
        A:=MatrixMatrixMultiply(A,JT):

        v:=A[1..3,2]:
        rot:=rotate(v[2],v[3]):
        J:= Matrix([[1,0,0],[0,rot:-cx,rot:-sx], [0,-rot:-sx,rot:-cx]]):
        A:=MatrixMatrixMultiply(J,A):

        A;
end do;
```

Figure 14. Diagonalization program by means of Givens rotation matrices

Participants were asked to run this program for the bidiagonal matrix obtained at the end of Golub-Kahan-Lanczos bidiagonalization algorithm, namely:

$$A = \begin{bmatrix} 12.12435565 & 8.602325266 & 0 \\ 0 & 5.89850262029624250 & 2.23989395020982 \\ 0 & 0 & 11.4101069395020982 \end{bmatrix}$$

assuming the “epsilon” value to be 10^{-1}. The resulting matrix is displayed in Figure 15.

$$\begin{bmatrix} 15.3180157787614420 & 0.0663996415453415034 & 3.13748799793758755\ 10^{-10} \\ 2.91447242887321383\ 10^{-11} & 11.6359248368426034 & 0.00977835907744034616 \\ -2.32483935249342854\ 10^{-10} & 5.00086803463192526\ 10^{-12} & 4.57811543299781220 \end{bmatrix}$$

Figure 15. The output of the diagonalization program for the bidiagonal matrix A and EPS = 1/10

Participants were asked to compare the diagonal entries of the resulting matrix with the output of the command “SingularValues” applied to the matrix A and observe that the outputs agree on at least three places after the decimal point.

The instructor concluded by saying that the accuracy of the result can be improved by decreasing the threshold value (EPS) but highlighted the fact that a smaller EPS implies more computation cycles, execution time and memory consumption.

Guided practice

Guided practice in this session consisted in participants’ performing the examples programmed in their worksheets, observing the outputs and comment on the complexity, execution time and memory consumption of the presented algorithms. They could also interrupt the instructor at any time and ask for clarifications.

Closure

The instructor closed the study session by summarizing the advantages of SVD from theoretical and numerical standpoints. He also encouraged students to widen their interest about the subject of singular value decomposition. Students were informed about web sites, articles and books for further study. They were also provided with all materials from the sessions. Finally, they were asked to stay after the session to fill out a final questionnaire.

Post-test: Check for understanding

Students were asked to fill a questionnaire composed of questions about SVD and questions about the teaching approach. The purpose of the first part of this questionnaire was to find if the computational teaching approach helped them in getting a clearer picture of key linear algebra concepts in general and of SVD in particular. The second part of the questionnaire addressed their views regarding the instruction style and material. The questions are reproduced below.

Post-test questions, Part I

1. To what type of matrices is SVD applicable? Why?
2. Is the SVD decomposition unique (i.e. given an $m\ x\ n$ rectangular matrix A, such that $A = USV^T$ with orthogonal matrices U $(m\ x\ m)$, V $(n\ x\ n)$ and S $(m\ x\ n)$ pseudo-diagonal; are the matrices U, V, S unique with this property)?
3. Why do you think the term “singular” was attached to the singular values of a matrix?
4. Name three applications of the singular value decomposition.
5. What are the left singular vectors of a matrix A? What are the right singular vectors of a matrix A? Do they belong to the same vector space?
6. What is the benefit of having the matrices U and V in the SVD orthogonal? Bases of which subspaces can be obtained by partitioning the matrices U and V of SVD with regards to the rank r of A?
7. What matrix norm can be expressed only in terms of singular values?

8. What is the condition number of a matrix and how can it be expressed in terms of singular values?
9. In which step of the presented computer algorithm for computing SVD are the Givens rotation matrices used? Why are Givens rotation matrices used in the SVD algorithm?
10. How can one compute the rank of a matrix by knowing its singular values?
11. For a real matrix A, what type of numbers are its singular values? Hint: Two characteristics are needed: First, select one of: positive; negative; positive or zero; negative or zero; positive or negative; positive or negative or zero. Second, select one of: complex; real; rational; integer; natural.
12. For what type of matrices are the singular values and the eigenvalues equal?
13. What theoretical result allows using SVD in data compression?
14. The reason for the eigenvalues decomposition is to find a basis for the space so that the matrix becomes diagonal. Does this basis always exist? Does this basis always have real entries? SVD tries to find one change of basis in the domain and usually another one in the range so that the matrix becomes diagonal. Do these bases always exist? Do these bases always have real entries?
15. Given the SVD decomposition of the matrix A: $A = U.\Sigma.V^T$, how can the exact pseudo-inverse (A^+) be written?

Post-test questions, Part II

1. What type of classes would you prefer: A. “Chalk and talk”: teacher gives a lecture, students take notes. B. Interactive lecture: students can interrupt the instructor by asking questions, discussing among themselves and the teacher. C. Computer lab classes: students solve problems on the computer; no lecture and no teacher guidance. D. Interactive lectures in a computer environment. Explain.
2. Which type of class do you think the SVD classes you participated in belong to?
3. In your opinion, does using mathematical software in teaching helps students better understand new concepts?
4. What did you like in the teaching approach that was used in this experiment?

5. What did you NOT like in the teaching approach that was used in this experiment?
6. Do you have any suggestions about how the teaching approach could be improved?
7. Do you consider SVD a subject appropriate to be taught at the undergraduate level?
8. Do you think that learning SVD could improve the overall understanding of linear algebra concepts?

4.2 Recruitment of participants in the experiment

Four students volunteered to participate in the teaching experiment. We label them "Nat", "Desse", "Chal" and "Carrie". The first three were recommended by the most recent instructor of the MAST 235 Linear Algebra with Applications II course as his top students in the past term. Their final grades in the course were A+, A+ and A, respectively. They were all young people, in the age range of 20-25. Carrie was older. She already had one BSc degree and was close to completing a second one, in mathematics. Her final grade in the Linear Algebra course was A+.

All participants were already familiar with Maple. Nat, Desse and Chal studied the Linear Algebra course in a Maple environment. Carrie took her Linear Algebra course in the more traditional "chalk-and-talk" environment, but was familiar with the LinearAlgebra package in Maple and the software in general.

4.3 Results of the teaching experiment

In this section, I present the participants' responses to the Pre-test, my expectations as to their understanding of my lectures based on those responses, their behavior during the sessions, and their responses to the Post-test.

4.3.1 Participants' responses to the Pre-test

First, participants' responses to the individual questions on linear algebra of the Pre-test will be presented. Next, based on a summary of the responses, expectations about their potential to understand and benefit from the lectures will be described.

For the purpose of assessment of the participants' background knowledge of basic concepts of linear algebra before the teaching experiment and later the comparison of participants' performance on the Post-test and the Pre-test, we assign points to their answers: 1 point for a correct answer, 0.5 points for a partly correct answer and 0 points to an incorrect answer or no answer. The number of no answer or "don't remember" or "don't know" responses will be counted for each student. Based on the accumulated points, each student will be assigned a "pre-grade" calculated as the percentage of their points out of the maximum number of points they could obtain in answering the linear algebra questions. The maximum number of points was 25: there were 15 "questions" but some of them had sub-questions and so there were 25 individual questions in total.

Question 1: Orthogonal vectors in a vector space

1a. Given two orthogonal vectors in a vector space: what is their dot product?

1b. What is the angle between them?

All students answered "0" to the first question. Desse's answer was a bit more elaborate. She gave a typical example of a pair of orthogonal vectors, namely the pair of two-dimensional unit vectors, and stated: "*u* and *v* are orthogonal; by definition their product is equal to zero". Therefore, while she appeared to treat the example as an illustration of the definition of orthogonality only, the meaning she attached to the concept was associated with such typical examples of it. All students answered "90 degrees" to the second question, with Chal adding "or Pi/2 radians" to it.

Therefore, although orthogonality is firmly associated with the zero dot product, it could also be associated with the measure of angles in degrees, as in high school geometry.

We consider the answer “0” to Question 1a as correct (1 point), the answer “90 degrees” to Question 1b as partly correct (0.5 points) and the answer containing “Pi/2 radians” as correct (1 point).

Question 2: Multiplication of a row vector by a matrix

> Complete the following sentence: multiplying a row vector with a matrix A is a linear combination of A’s….

This question generated the same, incorrect response from all participants: “columns”. The students might have confused the actual question with something that they have learned: multiplying the matrix A with a column vector is a linear combination of A’s columns.

None of the students considered using an example to validate their claim. With the “LinearAlgebra” package of Maple, they could have easily obtained the product of, say, a row vector $Transpose(< a, b, c >)$ and a matrix, say, $\ll a11, a21, a31 > | < a12, a22, a33 \gg$, and observed the output to check if they get a combination of rows or columns. They could have also reasoned that, if we multiply a row vector by a matrix, the result is a row vector, whereas a linear combination of the columns of a matrix gives a column vector. Obviously, however, the participants did not reflect on their answers and responded automatically to what appeared to them as trivial question.

Question 3: Orthogonal matrices

> What properties of vectors remain unchanged (are preserved) when multiplied by orthogonal matrices?

Again, the participants’ responses were either incorrect, or, in Nat’s case, there was no answer. Desse and Chal confused orthogonal matrix transformations with dilations and claimed that the resulting vector will preserve direction and it could either stay the same or be “scaled”. Carrie wrote that the property that is preserved is invertibility although this property applies to square matrices and not to vectors.

Question 4: Fundamental subspaces of a vector space

> 4a. What are the row space and the null space of a matrix A?

4b. What is the relationship between the row space and the null space of a matrix?

Only Nat and Carrie provided correct answers to question 4a. Chal appeared to answer the question about the rank of a matrix: “the number of pivot elements on the reduced form of that matrix”. Desse also appeared to associate the question with the concept of rank. Her answer was: “Row space: collection of rows of A. Nullspace: linearly dependent rows (If rank of 3 by 3 matrix is less than 3)”. Desse may have been thinking about the theorem stating that the dimension of the nullspace of an $m \times n$ matrix is equal to $n - r$, where r is the number of linearly independent columns or rows. But she seems to be thinking about the number $m - r$ rather than $n - r$.

There was only one correct answer to question 4b, by Nat, who said that “the nullspace is orthogonal to the rowspace”. Carrie gave no answer and the other students formulated relationships that did not hold. Chal offered the relationship $RowSpace(A) = NullSpace(A^T)$ which he could have confused with $RowSpace(A) = ColSpace(A^T)$. Desse displayed an even deeper confusion, using expressions such as “span of row space” and claiming that “span of row space is in the null space”.

Question 5: Multiplication of a square matrix by a vector

Complete the sentence:

The non-zero vectors (*n* x 1) that, after being multiplied by the square matrix *A* (*n* x *n*) remain in the same direction as the original vector, are…

Only Carrie provided the expected answer, “eigenvectors”. Nat provided no answer, and Desse and Chal answered “linearly dependent”. They appeared to interpret the question as concerned with relations between vectors that have the same direction, and, if this is the case, their answers were not incorrect, even if not what we expected. Therefore, we assign 0.5 points to Desse and Chal for this question. This unexpected interpretation could have been avoided, had the question been formulated using the word “vector” in singular rather than plural.

Question 6: Roots of the characteristic polynomial of a matrix

> Given a matrix *A*, we know that the roots of its characteristic polynomial are -1 and +1. How one can go about computing A^7?

Only Carrie's response was approximately correct and we assign it 1 point. She described the procedure that is commonly taught in undergraduate linear algebra courses in the context of this type of problems (with concrete numeric matrix A), without worrying about the possibility of +1 or -1 being multiple roots in which case the matrix might be non-diagonalizable: "We have two eigenvalues. We find the corresponding eigenvectors in order to diagonalize A. From $A = P^{-1}DP$, rewrite, raise D to the 7th power then multiply".

Chal's and Nat's responses started similarly as Carrie's, but Chal's falls short of saying what one would do after having diagonalized the matrix, and Nat's response ends after stating that +1 and -1 are the eigenvalues of the matrix *A*. We assign 0.5 points to Chal's answer and 0 points to Nat's.

Desse said she was "not sure" (she wasn't sure in questions 3 and 4b, either), but proposed finding the eigenvalues although they are given in the problem. It is quite possible that she failed to recall that roots of the characteristic polynomial are the eigenvalues; in a Maple environment, students rarely have to calculate eigenvalues manually or using Maple to solve the characteristic equation. Usually the command "Eigenvalues" is used. Her response suggests also that she could be a procedural learner: she would be able to carry out the procedure described by Carrie but only if she were given a concrete numerical matrix to start with. Then the first step would be to find the eigenvalues. She cannot imagine what she would do in a hypothetical situation.

Question 7: Symmetric matrices

> Given A and B, real square matrices, next to each of the sentences below write True or False (if False, give a counter-example):
>
> 7a. $A + B$ is symmetric
> 7b. A^n is symmetric
> 7c. If A^{-1} exists, then A^{-1} is symmetric
> 7d. AB is symmetric

7e. $A^T A$ is symmetric (does A have to be square?)

All responses to question 7a were correct, “True”. Nat included a numerical example with two symmetric 2 x 2 matrices.

All but Chal’s responses were correct in question 7b. In question 7c, Carrie, Chal and Desse had correct answers, and Nat did not answer at all.

In question 7d, Carrie responded correctly, “false”, but gave an incomplete reason why the product of two symmetric matrices is not necessarily symmetric: “*AB* is not possible unless *A* and *B* are of the same size”. With a numeric example in Maple she showed the error message that is fired when this condition is not met. She failed to consider the case where the matrices are of the same size, in which case the product is a symmetric matrix if and only the matrices commute. She obtains 0.5 points for this answer. Nat did not answer and Chal’s and Desse’s answers were incorrect (“True”).

In question 7e, Carrie and Nat gave correct and complete answers. Desse’s answer is partly correct (“True”) but she did not respond to the second part of the question. Chal’s answer was incorrect: “False, because $A^T A = A^2$ and it is not necessa[ri]ly symmetric”.

Question 8: Positive definite matrices

8a. What is the definition of a positive definite matrix?

8b. Complete the sentence: A is positive definite if and only if all its eigenvalues are…

No one remembered the definition of a positive definite matrix in question 8a. Desse did not even recall having “covered such material”, although the concept has certainly appeared in MAST 235. Chal was the only student who answered question 8b correctly. Nat completed question 8b by “real” and Carrie and Desse did not answer the question.

Question 9: Matrix rank

9a. How would you define the rank of a matrix?

9b. How would you go about computing it?

All answers to question 9a were correct, but there were two types of responses. One type could be called “analytic-structural” since it referred to a property of a set, namely linear independence of the columns of the matrix: Carrie and Nat wrote that rank is “the number of linearly independent columns of the matrix”. The other type could be called “analytic-arithmetic” because it referred to the particular entries of the matrix after row reduction: “the number of pivots on the reduced row echelon form of that matrix” (Desse and Chal). Of course, it is not necessary to obtain a “reduced” row echelon form.

In question 9b, two answers were correct: Carrie and Chal referred to row reducing the matrix and counting the number of leading 1’s or the number of columns with leading 1’s. Nat answered “Gauss-Jordan Elimination” which refers to a procedure for solving systems of linear equations which may include row reduction of the coefficient matrix. The concept of rank is valid also outside of the context of solving linear systems. Nat either did not abstract the notion of rank from this context, or just confused the terminology. We gave him the benefit of the doubt and assigned 1 point to this answer. Desse’s answer was rather sloppy: “By performing several elementary row operations or calling the command Rank(A) in Maple”. But her answer to question 9b shows that she is aware that the “several” row operations must be sufficient to obtain an echelon form, so, again, with the benefit of the doubt, 1 point is assigned.

Question 10: Elementary row operations

> What properties of the original matrix are not modified when Elementary Row Operations are applied?

No answer was completely satisfactory. Carrie’s was the best, however. She was correct in enumerating the row and null spaces, and the solution set. She was not correct in adding the column space. She could have also listed the fact that the columns preserve linear independence under elementary row operations or just said that the rank of the matrix is preserved. She was awarded 0.5 points for her answer. Chal and Nat each mentioned one correct property preserved under elementary row operations: linear independence of columns and the solution space, respectively. Their answers were also assigned 0.5 points. Desse mentioned the determinant, which is preserved intact only through the “AddRow” elementary operation. Her answer was assigned 0 points.

Question 11: Matrix decompositions

> 11a. Give 3 examples of matrix decompositions.
>
> 11b. In what area of mathematics do you think that these decompositions are useful? Explain.

Only Carrie’s answer in question 11a was correct. She mentioned QR, diagonalization and LU. Chal and Nat claimed they “never studied” matrix decompositions. Chal’s and Desse’s answers suggest that, for them, “matrix decomposition” refers to the components of a matrix such as row, column or minors. However, in his answer, Chal wonders if the question refers to “decomposition of a matrix in diagonal matrix with main diagonal with eigenvalues of the original matrix?” For this, his answer was assigned 0.5 points.

Carrie was the only one to attempt question 11b and she offered a satisfactory answer: “Decomposing a matrix allows data to be manipulated more efficiently. I have been told there are applications in computer science. I would like to know what other applications there are.”

This question tested the students’ familiarity with matrix factorizations. From their answers it follows that this topic was not covered in the linear algebra courses that they have attended. With the teaching of the numerical algorithms of computing SVD, that rely heavily on matrix factorization and QR decompositions, the use of matrix decomposition will be exhibited.

Question 12: Matrix determinant

> If all the eigenvalues of a matrix are known, how can its determinant be computed?

Carrie, Chal and Nat did not remember or know the relationship between the matrix eigenvalues and its determinant. Surprisingly, Desse was the only one to offer an answer and it was correct: “By multiplying eigenvalues”. She used Maple to illustrate or validate her answer. She took the matrix

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 4 & 5 & 6 \\ 6 & 3 & 3 \end{bmatrix}$$

She asked Maple to calculate $v \coloneqq Eigenvalues(A)$, obtained a list of rather complicated expressions, but, not bothered by the output, she calculated $evalf(v[1] \cdot v[2] \cdot v[3])$, obtained -33.0000000 – 3.467703671 $\cdot\ 10^{-8} I$, and compared it with the output of $Determinant(A)$, which was -33. Desse appears to have developed quite a skill in using Maple and perhaps a habit of validating theoretical results about equivalence of two objects or processes by obtaining them in two different ways in Maple and then comparing the results, probably practiced in the MAST 234 and 235 courses by the instructors.

Question 13: Linear independence

> 13a. Given 3 distinct, non-zero vectors in $\mathbb{R}^3$, how can we prove that they are linearly independent?
>
> 13b. Can a set of 8 distinct, non-zero vectors in $\mathbb{R}^7$ be linearly independent? Explain.

Question 13a was answered correctly by all but one student (Chal). The correct answers were all different:

> Carrie*: "We prove that there is not solution other than the trivial one to AX = 0 (c1v1 + c2v2 +… = 0 and solve for constants"*
>
> Desse*: "Compute a matrix with the three vectors (3 by 3 matrix), find the rank, and if the rank is 3, the three vectors are linearly independent"*
>
> Nat*: "The three vectors and put in row and then the process of Gauss-Jordan elimination will give you the Identity Matrix"*

We note that the answers by Carrie and Nat are general enough to apply to vectors of any dimension, and Carrie's answer is analytic-structural in that it could apply to vectors in any vector space. Desse's answer is specific to the particular dimension and vector space given in the question.

Chal's answer described a method of verifying if a given set of vectors spans the space and not a method of verifying if the set is independent.

Question 13b elicited all correct answers with acceptable explanations.

Question 14: Diagonalizable matrices

> When is an n x n matrix, with real entries, diagonalizable?

None of the answers was correct. Carrie said she did not remember the conditions. Chal asserted that the matrix must “not be a singular matrix”. Desse wrote that the matrix is diagonalizable “when the transition matrix exists and is invertible”, which additionally shows that she does not seem to be aware that a transition matrix (from one basis to another) is necessarily invertible. Nat required that the matrix must have n distinct eigenvalues, which is a sufficient condition but not a necessary one.

A general observation of Chal’s answers so far is that the linear algebra notions are not well crystallized in his mind, therefore he often jumps to incorrect conclusions.

Question 15: Basis

> Complete the sentence: Given a basis of a finite vector space, … element of the vector space can be expressed … as a … of basis vectors.

Only Carrie’s answer was completely correct and as expected. Chal and Nat gave partly correct answers. Chal filled the second blank with “if not zero vector”, thus incorrectly excluding the zero vector from being spanned by the basis vectors. He filled the other blanks correctly. Nat modified the sentence by adding “n” after “vector space”; then he put “the n+1” in place of the first blank; he ignored the second blank and filled the last blank correctly with “linear combination”. Desse said she doesn’t know.

4.3.2 Expectations about participants’ understanding of the subject of SVD based on their performance on the Pre-test

Below is a summary of the four students’ performance on the Pre-test in terms of the numbers of points they were assigned in each question and in total (Table 1).

As we can see from Table 1, there were five questions that none of the participants answered correctly (questions 2, 3, 8a, 8b, 14). The linear algebra topics covered by these questions were: orthogonal matrices, positive definiteness, and necessary and sufficient conditions to diagonalize a square matrix.

On the other hand, all students answered correctly questions 1, 7a, 9a, 9b and 13b. This suggests that they have better understood and remembered the concepts of orthogonal vectors, symmetric matrix properties, rank, and linear independence of vectors.

Nat showed strength in the area of fundamental subspaces of a vector space. He was the only participant to answer correctly both questions regarding this topic. Not answering questions 11 and 12 is an indication that he did not remember and/or understand well concepts such as matrix decompositions and the relationship between eigenvalues and the determinant of a square matrix.

Desse was the only participant to work out the relationship between eigenvalues and the determinant. She did not initially remember it, but using concrete examples in Maple, she was able to come up with the correct answer. She exhibited weakness in the areas of fundamental subspaces of a vector space and elementary row operations. The latter is a little surprising, because she seemed to be familiar with algorithms and numerical methods. On the other hand, in the MAST 234-5 courses, elementary row operations do not occupy much space; row reduction is left to Maple.

Chal obtained the poorest overall results. Surprisingly, he was the only participant to give a correct answer on the questions regarding positive definiteness, but this could be an effect of the randomness of his answers.

The only student to get full marks on questions 5, 6 and 11a was Carrie. She appeared to have acquired a good understanding of the concepts of eigenvalues, characteristic polynomial and matrix decompositions. She got the best overall score.

PRE-TEST Question	Nat	Desse	Chal	Carrie
1a	1	1	1	1
1b	0.5	0.5	1	0.5
2	0	0	0	0
3	0	0	0	0
4a	1	0	0	1
4b	1	0	0	0
5	0	0.5	0.5	1
6	0	0	0.5	1
7a	1	1	1	1
7b	1	1	0	1
7c	0	1	1	1
7d	0	0	0	0.5
7e	1	0.5	0	1
8a	0	0	0	0
8b	0	0	1	0
9a	1	1	1	1
9b	1	1	1	1
10	0.5	0	0.5	0.5
11a	0	0	0	1
11b	0	0	0.5	1
12	0	1	0	0
13a	1	1	0	1
13b	1	1	1	1
14	0	0	0	0
15	0.5	0.5	0	1
TOTALS	**11.5**	**11**	**10**	**16.5**
"PRE-GRADE" (% out of 25)	**46**	**44**	**40**	**66**
Number of questions left unanswered	8	4	4	5

Table 1. Participants' performance on the Pre-test

Based on the participants' performance on the Pre-test the following conjectures can be drawn about their potential to understand the SVD lectures:

Nat and Carrie seem to have a better understanding of linear algebra notions. Overall, they remembered the definitions of important concepts, such as fundamental subspaces, better than Desse and Chal. Chal's answers suggest that he mixes up definitions and that his understanding of linear algebra concepts is rather superficial. Desse exhibited poor memory, but when in doubt, she tried to validate her "conjectures" with numerical examples in Maple.

Looking at the pre-grades of the four participants, only Carrie can be said to have "passed the test", with 66%. The pre-grades of the other participants were all below

50%. These results are surprising, given that Nat, Desse and Chal completed their last Linear Algebra course three months before the experiment, and Carrie completed hers one year before. Carrie took a more theoretically-oriented course, without CAS support, which required students to engage their analytic-structural thinking for solving assignment and test problems. Signs of the analytic-structural mode of thinking could be observed in several of Carrie's responses to the Pre-test. The Pre-test may have been more favorable to this mode of thinking than to the analytic-arithmetic mode which is usually fostered in the Maple lab Linear Algebra courses. As mentioned above, Desse sometimes quite successfully took advantage of Maple affordances to put her analytic-arithmetic mode to use in solving theoretical questions. Nat and Chal did not appear to have used such opportunities. They had poor memory of the theoretical concepts of the course and did not think of using Maple to turn the questions in the Pre-test into opportunities for numerical exploration or even search for definitions using the Maple "Help" utility. Nat had a tendency to give up answering a question if he was not sure or did not remember the concepts: he was the participant with the highest number of non-attempted questions. Chal appeared to rush into an answer without taking time to reflect on the question: his answers were sometimes chaotic. He did, however, spend time thinking about the questions and shared some of his thoughts and questions in the responses.

4.3.3 Participants' behavior during the sessions

Most of the time students listened to instructor's presentation of the lecture. Whenever they had to clarify certain theoretical aspects, they interrupted the flow of the lecture with their questions. Carrie showed the most interest in asking theoretical questions.

There were times when the instructor asked questions to refresh the students' knowledge of linear algebra notions or to quiz their understanding of the singular value decomposition related topics.

During the lectures, students executed the prepared numerical examples included in the Maple worksheets. There was very minimal need in assisting them with their tasks in the Maple environment. The teacher explained the steps involved and encouraged participants to repeat the execution with modified initial data and to observe the results. Then, students were free to comment on their findings. Desse was the most inclined to validate conjectures using numerical examples.

Students were given assignments to work on at home between the two sessions. For example, they were asked to use Maple commands to determine the Kronecker decomposition of a 2×2 matrix.

4.3.4 Participants' responses to the Post-test, Part I – Linear Algebra questions

As in our analysis of the Pre-test, points in the Post-test were assigned to students' answers as follows: 1 point for a correct answer, 0.5 points for a partly correct answer and 0 points to an incorrect answer or no answer. The number of no answer or "don't remember" or "don't know" responses will be counted for each student. Each student will be assigned a "post-grade" calculated as the percentage of their points out of the maximum number of points they could obtain in answering the questions. The maximum number of points was 23: there were 15 "questions" but some of them had sub-questions and so there were 23 individual questions in total.

Question 1: SVD applicability

1a. To what type of matrices is SVD applicable?

1b. Why?

Carrie, Desse and Nat captured correctly the essence of SVD: the fact that it is applicable to matrices of any shape. They all got 1 point for their answers to 1a. Chal did not answer correctly: he wrote that SVD is applicable only to square matrices and was given 0 points.

The expected answer to question 1b was that SVD is applicable even for rectangular matrices because the matrix $A^T.A$ is square and symmetric and, as such, its eigenvalues are always positive, real numbers. Carrie and Chal provided answers that correctly referred to the fact that the matrix $A^T.A$ is square and symmetric. Chal, however, appeared to believe that this implies that the matrix A itself must be square.

> Carrie: The procedure uses $At * A$ ($A * At$), square and symmetric.
>
> Chal: Because to compute singular values, we must have a symmetric matrix $A^T A$ from a matrix. Therefore, the matrix should be square.

Both appeared to be associating SVD with the computational procedures of calculating it rather than with the structural properties of the decomposition, which is not surprising in view of the importance given to the former in the lectures. Both were awarded 1 point for 1b.

Desse's response to 1b was merely an elaboration of her answer to question 1a: "Can be square or rectangular, with different ranks. That is the main point of SVD". Nat gave no answer. Both were awarded 0 points.

Question 2: SVD uniqueness

> Is the SVD decomposition unique (i.e. given an $m\ x\ n$ rectangular matrix A, such that $A = USV^T$ with orthogonal matrices U $(m\ x\ m)$, V $(n\ x\ n)$ and S $(m\ x\ n)$ pseudo-diagonal; are the matrices U, V, S unique with this property)?

During the first SVD session, students were presented with this result regarding the uniqueness of the singular value decomposition: the singular values are uniquely determined and if A is square and all singular values are distinct, the left and right singular vectors are uniquely determined up to complex signs (if A has real entries, up to sign). For the same singular values ordering, it follows that if A is square and all singular values are distinct, the matrices U and V are unique up to complex signs.

Carrie only mentioned the uniqueness of the singular values without referring to their order and did not talk about the matrices U and V that hold the left and right singular vectors respectively. She was awarded 0.5 points.

Chal and Desse captured the non-unique aspect of SVD, but did not speak about the uniqueness of the singular values and about the conditional uniqueness of the singular vectors. Chal mentioned the fact that by reordering the singular values on the diagonal of matrix Σ we obtain a different decomposition. Desse stated that "there can be a little variation in the matrix Σ". They were also given 0.5 points. Nat provided no answer.

Question 3: SVD naming

> Why do you think the term "singular" was attached to the singular values of a matrix?

Carrie remembered the fact that the "values serve as an indication of how close the matrix A is to being 'singular'", which was the expected answer.

Chal tried to express the idea that the name comes from the distance to a singular matrix but his explanation was very hard to follow:

> Chal: A singular matrix is a matrix such that determinant of that matrix is null. Therefore, I guess that singular values of a matrix refer to the distance between a matrix and the decompositions of that matrix such that we get a 0 determinant in the decomposition matrices (except the diagonal one)."

He was assigned 0.5 points for his response.

Desse answered that the name "comes from the singularity of the matrix", but mentioned nothing about "distance to a singular matrix". We awarded her 0.5 point.

Nat did not provide the right answer. He stated that SVD could be used to compute the pseudo-inverse of a given matrix, which is correct but irrelevant in this question. Next, he said that the term "singular" comes from the fact that a pseudo-inverse can be computed for singular matrices. He was assigned 0 points for this response.

Question 4: SVD applications

> Name three applications of the singular value decomposition.

Responses to this question were as expected in the case of Carrie, Chal and Desse. Nat mentioned only two applications: image processing and "data correction (if one singular value is dominant and the other is not)". It is not clear what Nat meant in the second example; perhaps he referred to the numerical stability to "data perturbations". Carrie, Chal and Desse all mentioned data compression. Noise reduction was mentioned by Carrie and Desse. Other applications mentioned by one student each were: pseudo-inverse, determinant computation, rank computation, image processing. All students (Nat included) were awarded 1 full point for their responses.

Question 5: Singular vectors

5a. What are the left singular vectors of a matrix A?

5b. What are the right singular vectors of a matrix A?

5c. Do they belong to the same vector space?

Carrie, Chal, and Desse provided the expected correct answer to 5a that the left singular vectors are the columns of U. Nat did not provide an answer to this. In question 5b, only Carrie and Chal correctly wrote that the right singular vectors are the rows of V^T. Desse confused the right singular vectors with the columns of V^T. She was assigned 0.5 point for this. Nat provided no answers to 5a and 5b. As for question 5c, the only correct answer was Carrie's. Desse and Chal incorrectly claimed that the vectors do belong to the same vector space and Nat, again, provided no answer.

Question 6: Matrices U and V in SVD

6a. What is the benefit of having the matrices U and V in the SVD orthogonal?

6b. Bases of which subspaces can be obtained by partitioning the matrices U and V of SVD with regards to the rank r of A?

Only Carrie and Desse attempted the question 6a. However, only Carrie's answer was correct. She captured the most important point: that there is no need to compute matrix inverses since the inverse of an orthogonal matrix is equal with its transpose. Desse wrote that the orthogonality is preserved when multiplying orthogonal matrices. While this statement is correct, the singular value decomposition of A does not involve multiplication of orthogonal matrices (matrix Σ is pseudo-diagonal and not orthogonal). Therefore she was assigned 0 points for her answer. Chal wrote that he doesn't know and Nat left a blank.

In 6b there were two correct or almost correct responses: Carrie's and Desse's. Carrie correctly listed the four fundamental subspaces (row space, column space, null space and null space of the transpose matrix). Desse listed: null space, row space, column space and the “null space of column”. Although she did not remember the correct name of one of the subspaces, she was assigned 1 point on this question. Nat gave an incomplete answer listing only the column space and the null space. He was awarded 0.5 points. Chal left a blank.

Question 7: Matrix norm in terms of singular values

> What matrix norms can be expressed only in terms of singular values?

Carrie and Nat did not provide an answer to this question. By listing rank, norm and diagonal, Desse probably mistakenly answered the question: which attributes of a matrix can be computed using singular values? The only response that came close to the expected answer was Chal’s. Chal remembered that one norm is equal with the largest singular value of the matrix, without naming the norm (2-norm). He was given 1 point for his response.

Question 8: Condition number of a matrix in terms of singular values

> What is the condition number of a matrix and how can it be expressed in terms of singular values?

The only response that could count as correct was Chal’s who stated that the condition number of a matrix is the ratio between its largest and its smallest singular value. Carrie wrote that the condition number of a matrix is “the product of its singular values and serves as an indication of how ‘large’ the matrix is”, confusing condition number with information about the relationship between singular values and the determinant of a matrix. Desse was even further off: “Singular matrix. Can be expressed through U, Σ and $V^T$by multiplying these matrices”. It is very difficult to make sense of her reasoning. It is obvious that she mixed up several definitions. Nat gave no answer.

Question 9: Givens rotations usage in the SVD algorithm

> 9a. In which step of the presented computer algorithm for computing SVD are the Givens rotation matrices used?
>
> 9b. Why are Givens rotation matrices used in the SVD algorithm?

Only Carrie and Desse provided answers to 9a. Carrie's response to 9a was: "when converting a bi-diagonal matrix to a diagonal matrix" which was the expected answer. Desse's answer was: "Bidiagonalization part 2". It is not clear if she was aware of the purpose of this second part, i.e., diagonalization. Giving her the benefit of the doubt, we awarded her 1 point as well for this question. Chal said he didn't know and Nat left a blank.

Except Chal, who said he did not know, all the other participants handled question 9b correctly. They used different words to express that the Givens rotation are used to minimize the elements on the superdiagonal of the matrix. Carrie wrote "reduce the size"; Desse – "minimize the elements", and Nat – "to eliminate". Even if he wrote "to eliminate", Nat probably meant "to bring close to 0", so he was given the benefit of the doubt and assigned 1 point for his answer.

Question 10: Rank and singular values

> How can one compute the rank of a matrix by knowing its singular values?

All students except Nat correctly remembered that the rank of a matrix equals the number of its non-zero singular values. Nat mentioned the number of singular values, but failed to write that only the non-zero singular values count towards computing the rank. He was assigned only 0.5 points.

Question 11: Singular values type

> How can one compute the rank of a matrix by knowing its singular values?
> Hint: Two characteristics are needed:
> First, select one of: positive; negative; positive or zero; negative or zero; positive or negative; positive or negative or zero.
> Second, select one of: complex; real; rational; integer; natural.

Carrie and Chal answered correctly: Positive or zero; real. Desse answered that the singular values could be negative. She forgot that the singular values are the square roots of the eigenvalues of a matrix, therefore they cannot be negative, and was awarded 0 points. Nat gave no answer.

Question 12: Equality of singular values and eigenvalues

For what type of matrices are the singular values and the eigenvalues equal?

Only Desse gave the expected correct answer: "symmetric positive definite matrix". Nat required that to have identical eigenvalues and singular values, the matrix must be symmetric and diagonally dominant. The latter condition guarantees only that the matrix is positive semi-definite, but we accepted his answer as worth 1 point nevertheless.

Carrie's response was: "Square, invertible. Positive definite". She did not mention that the positive definite matrix must be symmetric. However, in the linear algebra courses she took, a positive definite matrix is usually defined as a symmetric matrix A such that $v^T A v > 0$ for any vector v of appropriate size. Positive definite matrices are indeed square and invertible, but it seems that she believes that any invertible matrix has identical eigenvalues and singular values. Therefore, she is awarded 0.5 points.

Chal required that, to have identical eigenvalues and singular values, the matrix must be symmetric and idempotent. The answer was considered incorrect.

Question 13: SVD and data compression

What theoretical result allows using SVD in data compression?

The expected correct responses to this question were: Kronecker expansion or low-rank approximation. Only Nat provided a correct answer: "Lowest Rank approximation". Chal wrote that he didn't know. Neither Carrie nor Desse named any theoretical results. Their answers pointed to some aspects associated with data compression: "If the singular values are small, the data is not very relevant" (Carrie); "Instead of using a lot of space, with SVD can be found a minimum amount of data to be stored" (Desse). They were given 0 points.

Question 14: Singular values and eigenvalues

> The reason for the eigenvalues decomposition is to find a basis for the space so that the matrix becomes diagonal.
> 14a. Does this basis always exist?
> 14b. Does this basis always have real entries?
>
> SVD tries to find one change of basis in the domain and usually another one in the range so that the matrix becomes diagonal.
> 14c. Do these bases always exist?
> 14d. Do these bases always have real entries?

All participants' answers to question 14a were correct. They were brief and straightforward ("No"), except for Chal who added "If we have a matrix with just the value 0 as eigenvalues", which reveals certain some misconception of the conditions for diagonalization.

Responses to 14b and 14c were all correct and brief ("No" for 14b and "yes" for 14c).

In question 14d, only Carrie and Chal provided correct answers ("Yes"). Desse incorrectly said "No", and Nat introduced an unnecessary constraint; he required that the matrix be positive definite for the entries to be real.

Question 15: Exact pseudo-inverse and SVD

> Given the SVD decomposition of the matrix A: $A = U.\Sigma.V^T$, how can the exact pseudo-inverse (A^+) be written?

Carrie and Desse provided the expected correct answer $A^+ = V.\Sigma^+.U^T$, and Nat came close with $A^+ = V.\Sigma^{-1}.U^T$. All three were given 1 point for their answers. Chal wrote that he didn't know.

Summary of participants' performance on Post-test, Part I

We present the points scored by the participants in the first part of the Post-test in Table 2.

The Pre-test provided us with a measure of the participants' "preparedness" for the lectures on SVD. The Pre-test scores were a better predictor of participants' performance on the Post-test than their grades from Linear Algebra courses, although also not a very accurate one. The expectations were verified in the case of Carrie, Chal and but not Nat. Carrie appeared to be best prepared: she remembered most of the important fundamental notions from her previous linear algebra courses and her knowledge of them was analytic-structural and theoretical rather than analytic-arithmetic and procedural. It was therefore no surprise that she scored the highest on the Post-test.

Desse scored only a little higher than Chal in the Pre-test and this relation was maintained in the Post-test. Desse's interest in numerical methods and verifying theoretical ideas on concrete examples was probably satisfied in the SVD lectures and this could have helped her to remember more from them than she could remember of the theoretical knowledge addressed in the Pre-test. Her ratio of good versus incorrect answers was improved in the Post-test relative to the Pre-test.

Chal still answered many questions on first-impulse but, since the number of un-attempted ones increased from four (out of 25) to six (out of 23), we can say that he became a little less likely to offer any association that came to his mind. He also provided fewer incorrect answers.

Nat's poorest score in the Post-test was surprising both because of his high performance in the linear algebra course and the fact that he scored better than Desse and Chal on the Pre-test. He left blanks in many questions. Perhaps he preferred not to give an answer rather than risk an incorrect one. It was also noticeable that he would sometimes get stuck in giving the same explanation even in incorrect contexts. He appeared to like the idea of dominant values (as in "diagonally dominant matrix", question 12, or "dominant singular value" question 4).

Some of the predictions regarding the participants' challenges in understanding certain topics of the lectures were confirmed. The lack of a good understanding of the properties of symmetric matrices, of eigenvalues or of the conditions for matrix diagonalization led them to give incorrect answers in the post-test. Participants found that the time allocated to describe the algorithms was not sufficient. Those not used to the compact representation of computer algorithms had to be explained the meaning of the notations.

POST-TEST Question	Nat	Desse	Chal	Carrie
1a	1	1	0	1
1b	0	0	1	1
2	0	0.5	0.5	0.5
3	0	0.5	0.5	1
4	1	1	1	1
5a	0	1	1	1
5b	0	0.5	1	1
5c	0	0	0	1
6a	0	0	0	1
6b	0.5	1	0	1
7	0	0	1	0
8	0	0	1	0
9a	0	1	0	1
9b	1	1	0	1
10	0.5	1	1	1
11	0	0	1	1
12	1	1	0	0.5
13	1	0	0	0
14a	1	1	1	1
14b	1	1	1	1
14c	1	1	1	1
14d	0	0	1	1
15	1	1	0	1
TOTALS	10	13.5	13	19
"POST-GRADE" (% out of 23)	43	59	57	83
Number of questions left unanswered	10	0	6	1

Table 2. Participants' performance on the first part of the Post-test

4.3.5 Participants' responses to Post-test, Part II – Views about the teaching approach

The set of questions in Part II of the Post-test was aimed at finding the participants' views about the teaching approach and obtaining their suggestions for improvement. Below are these questions and a brief analysis of the students' answers.

Question 1: Preference for type of classes

What type of classes would you prefer:
A. "Chalk and talk": teacher gives a lecture, students take notes.

B. Interactive lecture: students can interrupt the instructor by asking questions, discussing among themselves and the teacher.
C. Computer lab classes: students solve problems on the computer; no lecture and no teacher guidance.
D. Interactive lectures in a computer environment.
Explain.

Two of the four participants expressed their preference for interactive lectures in a computer environment (Chal, Desse). Chal explained his preference by stressing “mix between theory and practice” and interest: “not boring and much more interesting”. Desse stressed that “a computer helps a lot to compute the answer and to understand the material”.

Carrie selected both A and D and remarked that “each type of class has its advantages; the type and goal of the course do make a difference”. She also noted that a small-sized class “with an approachable teacher and a computer to play with and practice on” would be the ideal way to learn. Desse expressed a similarly conditional preference in Question 3, where she said that “it also depends on the subject: calculus and statistics are better when taught with A”.

Only Nat chose option B, which mentions “lecture” and teacher-student interactions but not the computer. This preference explains perhaps his low performance on the Post-test; he may have felt uncomfortable in the SVD sessions where the computer played a very important role. To some extent, it may explain, his low performance on the Pre-test; he did not seem to remember much from a linear algebra course conducted in the form of interactive lectures in a computer environment.

Question 2: Identification of the approach used in the SVD lectures

Which type of class do you think the SVD classes you participated in belong to?

All students concurred that SVD lectures fits best in the “interactive lecture in a computer environment” type of class. Even Nat agreed with that and explained that SVD is best learned in this type of class because it alleviates the burden of tedious computations and allows focusing on the theoretical aspects.

Question 3. Opinion about usefulness of mathematical software

> In your opinion, does using mathematical software in teaching helps students better understand new concepts?

Participants did agree with the statement that mathematical software helps in better understanding new concepts and that explained that it allows focusing on concepts and applications by relieving us from tedious calculations. However, Carrie's and Desse's views were not unconditional. Carrie added that this is true as long as the software used works well and does not hinder the process of learning. Desse emphasized the fact that the use of computers in class depends on the subject.

Question 4: Aspects of the lectures that participants liked

> What did you like in the teaching approach that was used in this experiment?

In order to avoid misrepresentation of the participants' answers, their responses will be quoted verbatim.

> Nat: "The teacher stopped to allow us to try the different applications of Singular Value Decomposition and he had some examples that we could try and he showed us some practical examples like with the rank approximation and image reduction."
>
> Desse: "Better to understand and to practice by changing the matrices and vectors to see what happens."
>
> Chal: "The theory was well written."
>
> Carrie: "Good explanations, instructor was sensitive to questions, very knowledgeable."

Thus, participants stressed the clarity of teacher's explanations and presentation, the possibility to ask questions, try examples and see applications. Desse appreciated the fact that she could freely change initial input values and visualize the effects of these changes and said that in helped her understand.

Question 5: Aspects that the participants did not like

> What did you NOT like in the teaching approach that was used in the experiment?

Again, we quote participants' views verbatim:

> Nat: "The time was not distributed well. 8 hours in two days probably did not allow you to have a good understanding of the material. If it was broken down too smaller sessions the learning would have be better."
>
> Desse: "The last part of the material was too complicated, needed more explanation."
>
> Chal: "The practice section was much like change the data in a previously done exercise in Maple. I prefer when we have all the exercise to do by ourselves; we learn more..."
>
> Carrie: "There was a lot of material for 2 days and 4 hours was a bit long for learning new concepts."

Participants felt that the material was vast and the time allotted to it was not sufficient. There was both not enough time overall and the sessions were too long.

Unlike Desse, Chal was not satisfied with just running examples using different numbers in matrices. He would have preferred more freedom in the exercises.

Question 6: Suggestions for improvement

Do you have any suggestions about how the teaching approach could be improved?

Participants reiterated the opinion that there should be more time for the material. Carrie joined Chal in proposing that students should be given more autonomy in the exercises. Desse proposed that the topic be taught right after the MAST 235 course so that one remembers the material related to SVD better.

Question 7. Appropriateness of SVD in an undergraduate program

Do you consider SVD a subject appropriate to be taught at the undergraduate level?

On this topic, the opinions were very divided. Carrie was the only one to answer a resounding "yes". Desse thought that SVD is more suitable as an optional course at the undergraduate level. Nat answered that it should be taught as a 300-level course and with pre-requisites classes that teach the basics of linear algebra and the use of Maple software.

Chal considered that adding SVD to a linear algebra course would make it much harder, especially that they don't learn much even about eigentheory at this level.

Question 8: Usefulness of the SVD topic in understanding linear algebra

> Do you think that learning SVD could improve the overall understanding of linear algebra concepts?

We quote participants' responses verbatim below.

> Nat: "There are some interesting applications in SVD which might show the usefulness of linear algebra."
>
> Desse: "Yes, it better helps to understand the usage and utility of linear algebra, its applications."
>
> Chal: "Yes, because it was a good reminder of many notions of linear algebra and I think that it was a good complement to eigentheory; and I liked that."
>
> Carrie: "Yes, SVD seems like a culmination of many linear algebra concepts. Learning the concepts without SVD is like travelling a road and never reaching the destination. It gives some meaning to the concepts."

Only Chal gave a straightforward positive answer to the question. The other participants looked at SVD as giving more meaning to linear algebra. Carrie stressed the fact that SVD gives more meaning to the concepts. Nat and Desse only highlighted the possibility of becoming more aware of the usefulness of linear algebra outside of mathematics. They did not mention that it helps understanding it. In fact, Desse suggested before that understanding linear algebra is a prerequisite to understanding SVD.

Chapter 5. Discussion and Recommendations for Teaching the Singular Value Decomposition of Matrices

There were useful lessons learned from the teaching experiment. One important lesson was related to the students' preparedness for the study of the SVD. Participants' responses to the tests made us aware that even students who were at the top of their class in previous linear algebra courses would need a substantial revision of the basic concepts. It may be true what Gilbert Strang said (2010) that SVD is a culmination of linear algebra in that it brings together all its fundamental concepts, but if students haven't mastered these concepts to some degree, there is nothing to bring together and the SVD topic may be lost on them.

Three of the participants completed two undergraduate courses of linear algebra with Maple. In these courses, new concepts were introduced in a structural way; then Maple was used to illustrate them and to confirm conjectures. The pre-test scores imply that even top students in those courses retain little of the theory three months after the end of the courses; they don't seem to make the necessary links between linear algebra notions. They may acquire some skills in using the computer software, but they are not always able to use Maple to explore or verify theoretical relationships when they are not sure about them. In view of their scores on the Post-test, which, after all, required only a very general grasp of ideas in the lectures, these three participants must have had hard time following the lectures.

The fourth participant attended undergraduate linear algebra courses taught without Maple, using a strictly structural approach. She scored somewhat better on the Pre-test and had a deeper understanding of the concepts, but was hungry for knowledge of applications of the linear algebra theory which these courses hardly even mention. The lectures, she admitted, satisfied this hunger to some extent, and she did benefit from the sessions more than the other participants in terms of a conceptual understanding of the SVD.

Another lesson learned from the experiment may be concerned with the mathematical organization of the lectures. Some topics could have been introduced differently and some were not emphasized enough.

One of the neglected topics was the concept of Givens rotation matrices. In the teaching experiment, the matrices were very quickly introduced as an operator acting on a matrix; indeed, as a mere module in a larger computer program for diagonalizing a bidiagonal matrix. All the participants learned about the Givens matrices was their numerical effect. The concept, however, is valuable well beyond the SVD context. In particular, it allows learners to make links between the geometric and numerical meanings of multiplication by a matrix. From the numerical point of view, we may be interested in turning one entry of a vector to 0. Geometrically, this can be obtained by rotating the vector so that it becomes aligned with one of the axes. Students have enough knowledge of the standard matrix of rotation in two dimensions to compute the 2×2 Givens rotation matrix that reduces a required entry of a given vector to 0. Their knowledge of the definition of the cosine of an angle between two vectors in terms of dot product and norms would be made use of here as well. This experience could then be extended to 3×3 matrices and then to matrices of higher dimensions.

A truly computational approach to teaching SVD would require more attention paid to numerical stability. In particular, the difference between ill- and well-conditioned matrices could be studied in more depth. Systems of linear equations with ill- and well-conditioned coefficient matrices could be considered. Then, students could observe, in the two cases, how small perturbations in the coefficients propagate into the solutions of these systems.

In showing that the computation of the eigenvalues of a matrix as roots of its characteristic polynomial is not numerically stable, a matrix of smaller size could have been used. Wilkinson's classical 20 x 20 matrix example is not well suited for Maple, because its elements are not visually displayed, leading students to doubt that the correct entries were assigned to it. The pedagogical idea, both here and in preceding examples, is to let students explore general concepts in situations that are accessible for them through operations that can be executed manually or with only small help of the mathematical software and only then generalize and formalize them so that they can be encapsulated in concise algebraic notation or a computer program.

The second session of the teaching experiment revealed that presenting the numerical algorithm behind a linear algebra concept raised the participants' interest. However, it requires more time. Students have to get familiar with the compact notations that typically accompany numerical algorithms. They have to understand how such a construct can be "unpacked" and to get a feel about how the algorithm operates by executing a number of its initial steps.

One more lesson that will be mentioned here is that teacher-student interactions need to be less authoritarian on the part of the teacher and students must be left more freedom in the exercises. As one of the participants suggested, one does not learn much from just running teacher-prepared examples with different initial input. Students should be allowed to experiment with their own examples and conjectures with the teacher acting only as an on-demand advisor.

The mathematical significance of SVD is highlighted by many authors. Austin affirms that "[b]esides having a rather simple geometric explanation, the SVD offers extremely effective techniques for putting linear algebraic ideas into practice." (Austin, 2012). Certainly, owing to its algorithmic and well-conditioned nature, SVD is an excellent candidate for numerical or applied linear algebra courses, taught to more advanced undergraduate students. Notions such as algorithm stability to input data perturbations could be taught in comparison with eigenvalues' computation.

> *SVD definitely deserves a place in courses with a numerical or applied emphasis. Golub and Van Loan ascribe a central significance to the SVD in their definitive explication of numerical matrix methods. (Kalman, 1996).*

In spite of its significance, SVD is a topic rarely reached in undergraduate linear algebra courses and often skipped in graduate courses (Will, 2003). Our teaching experiment suggests some reasons for this state of affairs: if we want students to really benefit from SVD in terms of an improved conceptual understanding of fundamental linear algebra concepts, of becoming aware of the numerical advantages of SVD, and of the vast applications of linear algebra, then we need a lot of time, a month or more. The curriculum is already packed so it is difficult to squeeze in one more topic.

I hope that the results of this modest teaching experiment will help in making the right decision concerning the topic of singular value decomposition in the linear algebra curriculum at Concordia University.

References

AI Access. (n.d.). *Singular Value Decomposition*. Retrieved 05 24, 2011, from AI Access: http://www.aiaccess.net/English/Glossaries/GlosMod/e_gm_singular_value_decomposition.htm

Artigue, M. (2005). The integration of symbolic calculators into secondary education: Some lessons from didactical engineering. In K. R. D. Guin, *The Didactical Challenge of Symbolic Calculators. Turning a Computational Device into a Mathematical Instrument.* (pp. 231-294). New York: Springer.

Austin, D. (2012). *We Recommend a Singular Value Decomposition*. Retrieved June 17, 2012, from American Mathematical Society: http://www.ams.org/samplings/feature-column/fcarc-svd

Bau, D., & Trefethen, L. N. (2000). *Numerical Linear Algebra.* SIAM.

Berry, J., Graham, T., & Watkins, A. (1994). Integrating the Derive Program into the Teaching of Mathematics. *International Derive Journal*, 83-96.

Bindel, D., Demmel, J., Kahan, W., & Marques, O. (n.d.). *On computing Givens rotations reliably and efficiently .* Retrieved from LAWNs (LAPACK Working Notes): www.netlib.org/lapack/lawnspdf/lawn148.pdf

Boileau, N. (2012). *Two secondary school mathematics teachers' use of technology through the lens of Instrumentation Theory. A thesis for the degree of Master in the Teaching of Mathematics.* Montreal: Concordia University.

Cabrilog. (2009). *Cabri II Plus*. Retrieved July 12, 2012, from Cabrilog. Innovative Math Tools: http://www.cabri.com/cabri-2-plus.html

Carlson, D. (1993). Teaching linear algebra: must the fog always roll in? *College Mathematics Journal, 12(1)*, 29-40.

Cheney, W., & Kincaid, D. (2009). *Linear algebra. Theory and applications.* Boston: Jones and Bartlett.

Crowe, D., & Zand, H. (2000). Computers and undergraduate mathematics 1: setting the scene. *Computers & Education, 35*, 95-121.

Davis, B., & Uhl, J. (2011). *Matrices, Geometry & Mathematica*. Retrieved July 10, 2012, from Math Everywhere, Inc.: http://www.matheverywhere.com/mei/courseware/mgm/

Demmel, J. (2000, November 20). *Golub-Kahan-Lanczos Bidiagonalization Procedure*. Retrieved 08 25, 2011, from Templates for the Solution of Algebraic Eigenvalue Problems: a Practical Guide: http://web.eecs.utk.edu/~dongarra/etemplates/node198.html

Demmel, J., & Kahan, W. (1990, November). Accurate singular values of bidiagonal matrices. *SIAM Journal on Scientific Computing*, 873-912.

Dorier, J.-L., & Sierpinska, A. (2001). Research into the teaching and learning of linear algebra. In D. Holton, *The teaching and learning of mathematics at the university level. An ICMI Study* (pp. 255-274). Dortrecht: Kluwer.

Dorier, J.-L., Robert, A., Robinet, J., & Rogalski, M. (2000). The obstacle of formalism in Linear Algebra. In D. J.-L., *On the Teaching of Linear Algebra* (pp. 85-124). Dordrecht: Kluwer.

Drijvers, P., & Gravemeijer, K. (2005). Computer algebra as an instrument: Examples of 10 algebraic schemes. In K. R. D. Guin, *The Didactical Challenge of Symbolic Calculators. Turning a Computational Device into a Mathematical Instrument* (pp. 163-196). New York: Springer.

Dubinsky, E. (1995). ISETL: A programming language for learning mathematics. *Communications on Pure and Applied Mathematics, 48*(9), pp. 1027-1051.

Dubinsky, E. (1997). Some thoughts on a first course in linear algebra at the college level. In D. Carlson, C. R. Johnson, D. C. Lay, A. D. Porter, A. Watkins, & W. Watkins, *Resources for teaching Linear Algebra. MAA Notes Volume 42* (pp. 85-105). Washington, DC: The Mathematical Association of America.

Eiland, D. (2011, January 31). *SVD-Presentation-Updated 2.* Retrieved May 25, 2011, from Scientific Computing: Special Topic in CSE 5810 Florida Institute of Technology (Instructor Debasis Mitra): http://cs.fit.edu/~dmitra/SciComp/11Spr/SVD-Presentation-Updated2.ppt

Garrison, D., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles and guidelines.* San Francisco, CA: Jossey-Bass. A Wiley Imprint.

Gelbaum, B. R. (1988). *Linear Algebra. Basics, Practice and Theory.*

Golub, G. H., & Kahan, W. (1965). Calculating the singular values and the pseudo-inverse of a matrix. *SIAM Journal on Numerical Analysis, 2*, 205-224.

Golub, G. H., & Reinsch, C. (1970). Singular value decomposition and least square solutions. *Numerische Mathematik, 14*, 403-420.

Harel, G. (1989). Learning and teaching linear algebra: Difficulties and an alternative approach to visualizing concepts and processes. *Focus on Learning Problems in Mathematics, 11*, 139-148.

Hillel, J. (2000). Modes of description and the problem of representation in Linear Algebra. In J.-L. Dorier, *On the Teaching of Linear Algebra* (pp. 191-208). Dordrecht: Kluwer.

iLrn Testing. (2003). *What is iLrn Testing?* Retrieved July 26, 2012, from http://www.ilrn.com/fp-what.html

Kalman, D. (1996). A Singularly Valuable Decomposition: The SVD of a Matrix. *The College Mathematics Journal*, Vol. 27, No. 1.

Lagrange, J.-B. (2005). *Using symbolic calculators to study mathematics.*

Maplesoft. A Cybernet Group Company. (2012). *Maple 16*. Retrieved July 10, 2012, from Maplesoft: http://www.maplesoft.com/products/Maple/index.aspx

MathWorks. (2012). *Product documentation: Matlab*. Retrieved July 11, 2012, from MathWorks: http://www.mathworks.com/help/techdoc/

Natt, B. D. (2010). *Numerical Linear Algebra and Applications - 2nd Ed.* SIAM.

Poole, D. (2012). *Linear algebra. A modern introduction. Second edition.* Boston: Thomson Brooks/Cole.

Pruncut, A. (2007). *A study of students' theoretical thinking in a technology-assisted environment.* Montreal: Concordia University.

Sabella, M. S., & Redish, E. F. (1995). Student understanding of topics in linear algebra. *Physics Education Research Group Unıversity of Maryland Physics Department, College Park*, 1-6.

SAS Institute Inc. (2008). *SAS/STAT(R) 9.2 User's Guide.* Cary, NC: SAS Institute Inc.

Sierpinska, A. (2000). On some aspects of students' thinking in linear algebra. In J.-L. Dorier, *On the teaching of linear algebra* (pp. 209-246). Dortrecht: Kluwer.

Sierpinska, A., Dreyfus, T., & Hillel, J. (1999). Evaluation of a teaching design in Linear Algebra: The case of linear transformations. *Recherches en Didactique des Mathématiques, 19*(1), 7-40.

Sierpinska, A., Nnadozie, A., & Oktaç, A. (2002). *A study of relationships between theoretical thinking and high achievement in Linear Algebra.* Montreal: Concordia University.

Stewart, G. W. (1992). *On the early history of the Singular Value Decomposition.*

Strang, G. (2010). *Linear Algebra, as taught in 2010*. Retrieved July 10, 2012, from MIT Open Coursware: http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/

Szabo, F. (2002). *Linear algebra. An introduction using Maple.* Boston: Harcourt/Academic Press.

Thomas, M. O., & Hong, Y. Y. (2004). Integrating CAS Calculators into Mathematics: Issues of Partnership. *28th Conference of the International Group for the Psychology of Mathematics Education*, (pp. 297-304). Bergen, Norway.

Trouche, L. (2005). An instrumental approach to mathematics learning in symbolic calculators environments. In K. R. D. Guin, *The Didactical Challenge of Symbolic Calculators. Turning a Computational Device into a Mathematical Instrument.* (pp. 137-162). New York: Springer.

Wikipedia. (2012, July 30). *Moore-Penrose pseudoinverse*. Retrieved June 15, 2012, from Wikipedia: http://en.wikipedia.org/wiki/Moore%E2%80%93Penrose_pseudoinverse

Wikipedia. (2012). *Singular value decomposition*. Retrieved July 23, 2012, from Wikipedia: http://en.wikipedia.org/wiki/Singular_value_decomposition

Wilkinson, J. (1963). *Rounding errors in algebraic processes.* Englewood Cliffs, NJ: Prentice Hall.

Will, T. (2003). *Introduction to the singular value decomposition*. Retrieved July 10, 2012, from University of Wisconsin - La Crosse, Faculty, Todd Will: http://www.uwlax.edu/faculty/will/svd/index.html

Winsløw, C. (2003). Semiotic and Discursive Variables in CAS – Based Didactical Engineering. *Educational Studies in Mathematics*, pp. 271-288.

Wolfram. (2012). *Wolfram Mathematica 8*. Retrieved July 10, 2012, from Wolfram: http://www.wolfram.com/mathematica/

Appendix

The recruitment letter

Dear student,

[Your instructor] recommended you as a student with strong performance, seriously interested in mathematics, with both the required potential and the motivation to learn additional topics of applied linear algebra. My name is Zoltan Lazar; I am a student in the MTM (Master of Teaching Mathematics) program and part of my graduation thesis is to conduct an experiment of teaching the concept of singular values decomposition of a matrix using Maple software.

The participation will consist in spending 2 consecutive days (4 hours each) in a computer assisted environment on the premises of Concordia University. Tentatively (depending on students' and rooms availability), the target days are either on the last week of July or the first week of August. Participation in the teaching experiment will have no adverse effect on the mathematics courses you are taking at Concordia University. On the contrary, they may improve your understanding of linear algebra and of computer algorithms.

Also, by taking part in this study, the researchers could better understand how students feel about this teaching methodology. This understanding can ultimately contribute to devising better approaches to introducing new linear algebra notions.

If interested, please fill and print the attached consent form and confirm your participation (and the preferred days among July 27, 28, 29; August 1, 2, 3) by replying to this email.

Kind regards,

Zoltan Lazar

The Consent form

CONSENT TO PARTICIPATE IN A TEACHING EXPERIMENT INTRODUCING THE CONCEPT OF SINGULAR VALUE DECOMPOSITION (SVD) IN A COMPUTER ASSISTED ENVIRONMENT

This is to state that I agree to participate in a teaching experiment conducted by Zoltan Lazar, graduate student in the Master of Teaching Mathematics (MTM) program (zoltan.lazar@sympatico.ca). This teaching experiment is part of his graduation thesis.

A. PURPOSE

I have been informed that the purpose of the teaching experiment is to study how students respond to the introduction of a new mathematical subject (Singular Value Decomposition, or SVD), using computer software (Maple). The goal is to analyze if this teaching approach facilitates students' understanding.

B. PROCEDURES

I have been informed that I will participate in the study in the role of a student. The participation will involve spending 2 days (4 hours each) with the researcher on the following activities:

1. Filling out a short preliminary questionnaire on linear algebra notions ;
2. Viewing a PowerPoint presentation of SVD and a brief list of its applications ;
3. Using Maple to interactively learn the concept of SVD;
4. Listening to an interactive lecture on the numerical algorithms used in the singular value decomposition of a matrix;
5. Filling a final questionnaire to assess my understanding of SVD.I was informed that the aim of the final questionnaire is NOT to assess my knowledge and ability to do mathematics but to evaluate the effectiveness of the teaching approach. I was also assured that all my reactions and responses will be kept strictly confidential and treated as anonymous in any publications that may result from this study. My participation in the teaching experiment will have no adverse effect on my performance in the

mathematics courses I am taking at Concordia University. On the contrary, they may improve my understanding of linear algebra.

C. RISKS AND BENEFITS

I am aware that I may experience some discomfort in the session due to frustration with the lecture or the problems to solve. But I am also aware that I can help researchers to better understand how students feel about this teaching methodology. This understanding can contribute to devising better approaches to introducing new linear algebra notions and to teaching mathematics in general.

D. CONDITIONS OF PARTICIPATION

I understand that I am free to withdraw my consent and discontinue my participation at any time without negative consequences.

I understand that my participation in this teaching experiment is confidential (the researcher will know, but will not disclose my identity).

I understand that the data collected from this study may be published.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND VOLUNTARILY AGREE TO PARTICIPATE IN THIS STUDY.

Name (please print) ____

SIGNATURE ____

If at any time you gave questions about your rights as a research participant, please contact Adela Reid, Research Ethics and Compliance Officer, Concordia University, at (514)848-2424x7481 or by email at areid@alcor.concordia.ca.

If you have any specific doubts or questions about this research, you can contact Dr. Anna Sierpinska, the supervisor of Zoltan Lazar (sierpins@mathstat.concordia.ca; tel 514-848 2424 ext. 3239).