The Singular Value Decomposition

Samuel Chowning

November 2, 2020

1 Introduction

The Singular Value Decomposition (SVD) is a powerful matrix decomposition that can provide insights into key properties of, and create good (and, in fact, the best) approximations of, any real or complex matrix. One can immediately imagine the usefulness of such a decomposition when performing numerical operations on a very large matrix, as it can greatly increase the efficiency of calculations; and indeed, the SVD is an essential tool for data analysis, and is of great importance in many areas of science.

The SVD was discovered and developed independently by a number of mathematicians. Eugenio Beltrami and Camille Jordan were the first to do so, in 1873 and 1874, respectively; they were followed by James Joseph Sylvester, Erhard Schmidt, and Hermann Weyl, among others[11]. The first proof for rectangular and complex matrices was given by Carl Eckart and Gale J. Young in 1936[1], and methods for computing the SVD of a matrix continued to be refined throughout the mid-20th century, revolutionizing the field of numerical linear algebra.

The SVD has a wide and fascinating range of applications. Political scientists have used the SVD to categorize and predict the partisan voting behavior of congresspeople[8][7]. Crystallographers have used the SVD to measure crystal grain sizes[4][2]. Physicists have used the SVD to develop theory regarding quantum information and entanglement[10][9][6]. Researchers in all fields, presented with ever-increasingly abundant and complex data, have used higher-order versions of the SVD to deal with information stored in higher-dimensional objects such as tensors[3].

In this project, I aim to learn about the many-faceted SVD and its properties by investigating a range of related topics. First, I aim to study its existence and uniqueness theorems, and the Eckart-Young theorem of best k-rank approximation[1]. After gaining a solid theoretical groundwork, I aim to familiarize myself with how it is implemented in software, and finally, investigate some of its many applications.

2 Background

In this section, I will define the SVD, describe its significance as a generalization of the eigendecomposition, show how to compute it, and then discuss the Eckart-Young theorem.

The "full" Singular Value Decomposition, for any matrix $A \in \mathbb{R}^{m\times n}$, is defined as follows:

$$A = U\Sigma V^T,$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices, whose columns consist of the left and right singular vectors of A, respectively, and $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with $r = \text{rank}(A)$ leading positive diagonal entries, in non-decreasing order. The diagonal entries of Σ consist of σ_i, where $i = 1, .., p$ and $p = \min(m, n)$, and these σ_i are called the singular values of A. The SVD also exists for all $A \in \mathbb{R}^{m\times n}$ (and for all $A \in \mathbb{C}^{m\times n}$, in which case U and V are unitary). Another way to write the SVD is as follows:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T,$$

where σ_i is the i^{th} singular value, and u_i and v_i are the i^{th} columns of U and V, respectively. This sum, which in effect adds new "layers" of the form $\sigma_i u_i v_i^T$ on top of one another r times to form A, provides valuable insight into how an important application of the SVD, approximating matrices, works. I will briefly revisit this at the end of this section.

To explain the significance of the SVD, let us first recall that the regular eigendecomposition of a matrix $A \in \mathbb{R}^{n\times n}$ is given by the following:

$$A = Q\Lambda Q^{-1},$$

where $Q \in \mathbb{R}^{n\times n}$ is a square matrix whose columns consist of the eigenvectors of A, and Λ is a diagonal matrix whose diagonal entries consist of the eigenvalues of A. This decomposition is useful in a variety of contexts; imagine how it would help compute A^n for large n. However, it has some fatal limitations: to name a few, A must be square, and it must have real eigenvalues.

So, how can we generalize such a decomposition to work for any matrix? The answer is the SVD! We accomplish this by using two sets of singular vectors instead of one set of eigenvectors. These are the right orthogonal vectors $v_1, .., v_n$, which give a basis of the row space of A and go into $V \in \mathbb{R}^{n\times n}$, and the left orthogonal vectors $u_1, .., u_m$, which give a basis of the column space of A and go into $U \in \mathbb{R}^{m\times m}$. Strang[12] provides an illustration of their relationship, which is different from that between regular eigenvalues and eigenvectors (recall that $r = \text{rank}(A)$):

$$[Av_1 = \sigma_1 u_1, \ \cdots, Av_r = \sigma_r u_r] \quad [Av_{r+1} = 0, \ \cdots, Av_n = 0]$$

or, in represented in matrix form:

$$A\begin{bmatrix} v_1 & \cdots & v_r & \cdots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_r & \cdots & u_m \end{bmatrix} \left[\begin{array}{c|c} \begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{matrix} & 0 \\ \hline 0 & 0 \end{array}\right]$$

Note that only the first r equations are of interest; v_i and u_i past r fall into the nullspace of A and A^T, respectively, and the diagonal entries of Σ past σ_r are all 0. Thus, so are the associated equations. One can create a "reduced" form of the SVD by only considering the first r vectors for V and U and the first r singular values of Σ, which eliminates its zeros and turns it into a square matrix.

So, how do we go about actually computing the SVD? We want to find the sets of singular vectors and values to build U, Σ, and V. One way to do retrieve these is to examine A^TA and AA^T under the assumption that $A = U\Sigma V^T$ exists:

$$A^TA = (V\Sigma^TU^T)(U\Sigma V^T) = V\Sigma^T\Sigma V^T$$
$$AA^T = (U\Sigma V^T)(V\Sigma^TU^T) = U\Sigma\Sigma^TU^T$$

Both cases have produced symmetric matrices which can also be factorized with $Q\Lambda Q^T$, whose eigenvalues are in $\Sigma^T\Sigma$ or $\Sigma\Sigma^T$, and whose eigenvectors are in V or U. To summarize, their relationship is as follows:

- V contains orthonormal eigenvectors of A^TA.
- U contains orthonormal eigenvectors of AA^T.
- σ_i^2 to σ_r^2 are the nonzero eigenvalues of both A^TA and AA^T.

To ensure that the signs of both sides of our equations $Av_k = \sigma_k u_k$ are aligned, we define u_k to be created out of v_k, which we first find as orthonormal eigenvectors of A^TA. We then build:

$$A^TAv_k = \sigma_k^2 v_k$$

to retrieve σ_k, and then, from our $Av_k = \sigma_k u_k$ relationship:

$$u_k = \frac{Av_k}{\sigma_k}, \quad \text{for } k = 1, \dots, r$$

which gives us all of the values we need.

Finally, let us take a look at a spectacular result given by the Eckart-Young theorem. First, let us recall our earlier definition of the SVD as a sum of "skeleton" matrices: A_k is similar, except that it goes up to an arbitrary k instead of r (and will thus be rank k). It is defined as follows:

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

The Eckart-Young theorem then states the following[1]:

If B has rank k then $||A - A_k|| \leq ||A - B||$.

So, given any other matrix B also of rank k (or lower), its difference to A will be at least as big as the difference between A_k and A; in other words, no k-rank matrix is closer to A than A_k. So, when we want to create an approximation of A, we don't have to worry about whether or not there exists a better one; the SVD gives us the best!

3 Proposed Methodology

As I mentioned in the introduction, I aim to first develop my theoretical understanding of the SVD, and then move on to investigate some of its many applications. Overviews of the SVD and descriptions of its basic properties are given by Strang's Introduction to Linear Algebra (p. 363-367)[12], and further discussions (including an overview of the Eckart-Young theorem and proof) are given by his Linear Algebra and Learning From Data (p. 71-73)[13].

I plan to create the foundation of my theoretical understanding by reading and studying these texts. However, I intend to incorporate online resources into my learning, as well as materials focused exclusively on the SVD itself, such as Martin and Porter's survey paper, "The Extraordinary SVD"[5]. In the latter half of the project, I aim to focus on investigating applications such as those in image compression, crystallography, and fluid dynamics.

References

[1] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika, 1 (1936), pp. 211–218.

[2] M. D. Higgins, *Quantitative textural measurements in igneous and metamorphic petrology*, Cambridge university press, 2006.

[3] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM review, 51 (2009), pp. 455–500.

[4] B. D. Marsh, *On the interpretation of crystal size distributions in magmatic systems*, Journal of Petrology, 39 (1998), pp. 553–599.

[5] C. D. Martin and M. A. Porter, *The extraordinary svd*, The American Mathematical Monthly, 119 (2012), pp. 838–851.

[6] R. Paškauskas and L. You, *Quantum correlations in two-boson wave functions*, Physical Review A, 64 (2001), p. 042310.

[7] M. A. Porter, P. J. Mucha, M. E. Newman, and A. J. Friend, *Community structure in the united states house of representatives*, Physica A: Statistical Mechanics and its Applications, 386 (2007), pp. 414–438.

[8] M. A. Porter, P. J. Mucha, M. E. Newman, and C. M. Warmbrand, *A network analysis of committees in the us house of representatives*, Proceedings of the National Academy of Sciences, 102 (2005), pp. 7057–7062.

[9] J. Preskill, *Lecture notes for physics 229: Quantum information and computation*, California Institute of Technology, 16 (1998).

[10] J. Schliemann, D. Loss, and A. MacDonald, *Double-occupancy errors, adiabaticity, and entanglement of spin qubits in quantum dots*, Physical Review B, 63 (2001), p. 085311.

[11] G. W. Stewart, *On the early history of the singular value decomposition*, SIAM review, 35 (1993), pp. 551–566.

[12] G. Strang, *Introduction to linear algebra*, vol. 3, Wellesley-Cambridge Press Wellesley, MA, 1993.

[13] ——, *Linear algebra and learning from data*, Wellesley-Cambridge Press, 2019.