Project Proposal: Understanding, implementing and using SVD

Dylan Laurianti, Yilan Dong

Overview

The singular value decomposition(SVD) of a matrix A is a way to split a $m \times n$ matrix into three interpretable components, two unitary matrices U, V and one diagonal matrix Σ such that $A = U\Sigma V^T$. The column vectors of U, V provide insight into patterns in the columns and rows of A, respectively. The diagonal matrix Σ has non-increasing singular values σ_i, which scale the importance of these patterns. Common applications of the SVD are to solve $Ax = b$, to approximate or compress a matrix, to recognize patterns in data, to deflate the size of a problem, and to find the pseudoinverse of a matrix. The SVD is a more generally applicable matrix decomposition than most others because it can be used on both real and complex rectangular matrices with no caveats. The general algorithm for computing the SVD first converts A to upper bidiagonal form using Householder reflections, so that then $A^T A$ is symmetric and in tridiagonal form. Then the symmetric QR algorithm is implicitly applied to $A^T A$ by alternating row and column Givens rotations until it is reduced to a diagonal matrix, which is Σ. Then U is recovered as the accumulation of the Householder reflections that pre-multiplied A, and the row rotations on $A^T A$, and V is recovered likewise from the reflections that post-multipied A as well as the column rotations on $A^T A$.

The generalizablity of SVD on non-square matrices makes it an ideal candidate in solving real-world problems. In fact, the use of SVD is prevalent in various fields related to Computer Science, such as Network Science, Natural Language Processing, Information Retrieval, Computer Vision and Bioinformatics. In particular, SVD is often used during data prepossessing to reduce the size and dimension of the data, as well as to minimize the effects of noisy data. In doing literature review, we find that models and methods that incorporate SVD often demonstrate better performance than their counterparts that do not use SVD. These observations motivate us to learn more about SVD: its theoretical foundations, invention, evolution, applications, as well as its various branches.

Goals

Our goal is to understand, implement and explore the use cases of SVD. In particular, we hope to implement in Julia the original SVD, the reduced SVD, as well as at least one other variation. We hope to measure the correctness of our algorithms by running a comprehensive range of tests; we will measure the robustness and efficiency of our implementation by comparing them with the existing SVD methods in Julia.

Literature Review

A) Understanding SVD

1. Stewart, G. W. On the early history of the singular value decomposition. *SIAM Review 35*, 4 (1993), 551–566

Stewart provides an overview of the work of five influential mathematicians who worked on developing the original theory of the singular value decomposition. In 1873, Beltrami discovered the SVD for real, square, nonsingular matrices. Independently in 1874, Jordan produced the same results more elegantly, leading to further insight into problem size reduction and characterizing the largest singular value as a maximum of a function. More progress was made in 1889 when Sylvester demonstrated that a matrix can be iteratively diagonalized. The SVD was brought to the field of integrals by Schmidt in 1907, where he also created an approximation theorem. Finally, in 1912 Weyl found how to determine the rank of a matrix in the presence of error, expanding the usefulness of Schmidt's approximation theorem.

2. Willems, P. R., Lang, B., and Vömel, C. Computing the bidiagonal SVD using multiple relatively robust representations. *SIAM Journal on Matrix Analysis and Applications 28*, 4 (Dec. 2006), 907–926

This paper presents the effects of the development of coupling relations on the Multiple Relatively Robust Representations algorithm. These improvements were then adapted and implemented by the research team in a bidiagonal SVD algorithm as part of the LAPACK library. To apply the MRRR, the bSVD can be reduced to the tridiagonal symmetric eigenproblem, which must be done using the normal equations. The new algorithm was shown to be an improvement over existing methods, with a cost of $O(nk)$ for k singular values and vectors.

3. Vasudevan, V., and Ramakrishna, M. A hierarchical singular value decomposition algorithm for low rank matrices. *CoRR abs/1710.02812* (2017)

This paper begins from the idea of how a truncated SVD can save runtime compared to the full SVD, and develops an even more cost-cutting algorithm for matrices of low rank. This new algorithm's usefulness is not restricted just to the tall and skinny matrices that the truncated method is. Using simultaneous row and column splits to reduce the size of computations needed, the researchers created a merge and truncate algorithm. They found that their algorithm is both faster on single-core processors, and more parallelizable. The algorithm provided even better improvements when working on tall and skinny matrices, and had typical error between 1-2%.

4. Huang, F., and Wilks, Y. Clustered sub-matrix singular value decomposition. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, NAACL-Short '07, Association for Computational Linguistics (USA, 2007), 69–72

Motivated by the problem of the SVD overlooking the importance of outliers and smaller sub-patterns in their datasets, these researchers looked to create a modified SVD algorithm that would use clustering techniques first to preserve the patterns that only appear in a subset of the data. They used a single-link algorithm to create clusters, then split the clusters into submatrices, ranked the submatrices in order of their size, and then repeatedly calculated a basis vector from the submatrix that is determined through residual ratios to have the most data not yet represented in the existing basis vectors. The application they were concerned with was summarizing the topics across a group of multiple text documents. They found that the unmodified SVD algorithm would heavily favor the documents that only focused on one topic, where documents with multiple ideas would have all of their data relatively ignored. In their testing of the algorithm, they found that it was not only more precise for the type of data they were working with, but also produced a much lower dimension solution than the original SVD.

5. Feldman, D., Volkov, M., and Rus, D. Dimensionality reduction of massive sparse datasets using coresets. *CoRR abs/1503.01663* (2015)

These researchers saw that none the existing dimensionality reduction algorithms were specialized for sparse matrices, and sought to find a new algorithm tailored to that common kind of data. They used the idea of a coreset, a weighting of a subset of the rows in the matrix such that the sum of squared distances from a point to the coreset approximates the sum of squared distances from the point to the original data. The researchers showed that for every matrix A there exists a (k, ϵ)-coreset of cardinality $|C| = O(k/\epsilon^2)$, where k is the dimensionality reduction, and ϵ is the error ratio. These coresets can then be used to approximate the matrix, while having much lower dimensionality.

6. Li, H., Kluger, Y., and Tygert, M. Randomized algorithms for distributed computation of principal component analysis and singular value decomposition. *CoRR abs/1612.08709* (2016)

This paper presents randomized algorithms for PCA and SVD calculation. They found that the deterministic SVD algorithms can end up returning left singular vectors that are far from orthonormal, and use randomized algorithms to generate the correct orthonormal singular vectors. They focused on algorithms for two specific applications of the SVD, the reduced SVD and low-rank approximation. They found that their algorithms at least matched the performance and accuracy of the deterministic algorithms, and in the types of cases that caused the deterministic algorithms trouble, the randomized algorithms far outperformed them.

B) Using SVD

7. Harrag, F., Hamdi-Cherif, A., and El-Qawasmeh, E. Performance of MLP and RBF neural networks on Arabic text categorization using SVD. *Neural Network World 20* (Jan. 2010), 441–459

This paper proposes a new text classifier in the form of Artificial Neural Network (ANN) for Arabic text. SVD is used in data preprocessing to reduce the size and the dimension of the data. The authors found the use of SVD decreases the time and space complexity, and increases the performance. In particular, SVD-supported ANN classifier outperform their counterparts without the use of SVD, when evaluated using an Arabic corpus.

Text categorization is a fundamental component of natural language processing. This research is especially important as it uses an Arabic corpus to measure the performance of various tools, which is something relatively scant yet invaluable in the status quo of the Euro-centric academia.

8. Elayaraja, E., Thangel, K., Chitralegha, M., and Chandrasekhar, T. Extraction of motif patterns from protein sequences using SVD with rough K-means algorithm. ***International Journal of Computer Science Issues 9*****, 2 (Nov. 2012), 350–356**

This paper looks at how to extract recurring motif patterns from protein sequences in the Protein Sequence Culling Server (PISCES) dataset through clustering algorithms. The PISCES dataset contains more than 660,000 180-dimensional protein segments, rendering the use of any clustering algorithm that has an undesirable time-complexity impossible, without further data processing. SVD is therefore chosen among a pool of similar algorithms to reduce segments, making it possible to generate meaning clusters using rough K-means algorithm within reasonable amount of time. The authors further proposes that SVD be used in other bioinformatics research.

Efficient and accurate identification of recurring patterns in protein sequences can provide invaluable information on protein interactions, a phenomena that affects our quotidian life.

9. Gorfte, Z. E., Cherkaoui, N., Bouzid, A., and Roukhe, A. Multi-data embedding in to RGB image with using SVD method. ***International Journal of Computer Science Issues 10*****, 2 (Sept. 2013), 190–193**

This paper proposes a new method of watermarking RBG image. SVD is used as an image compression tool and data embedding tool. The researchers find that the proposed method works especially well with JPEG compressions and GIF color reductions. They observe no visible difference before and after watermark applied using this method, and state that the watermarks created are robust in facing conventional attacks.

10. Zhang, W., Xiao, F., Li, B., and Zhang, S. Using SVD on clusters to improve precision of interdocument similarity measure. ***Computational Intelligence and Neuroscience*** **(Aug. 2016)**

This paper proposes the use of SVD on clusters in SVD as a new indexing method in lexical matching. The proposed method is motivated by the observation that, although SVD based Latent Semantic Indexing (LSI) demonstrates good representative quality, and seems to overcome the problems of polysemy and homonym in traditional lexical matching, it shows a limited inter-document discriminitive power. Through the evaluation of an English and a Chinese corpus, the authors claim that the proposed method, compared to other SVD-based LSIs, demonstrates improved precision in inter-document similarity measurement.

Lexical matching is an important data-mining tool. Polysemy and homonym, both of which inherent to interpretation to human speakers of natural languages, can be hard for algorithms. The method proposed by the authors takes lexical m matching one step further, reconciling the difficulties posed by polysemy and homonym, and the lack of discriminatory abilities of other SVD based LSIs.

11. Liu, B., and Liu, Q. Random noise reduction using SVD in the frequency domain. *Journal of Petroleum Exploration and Production Technology 10*, 7 (Oct. 2020), 3081–3089

The authors of the paper presents a novel random noise reduction method using SVD, based on the observation that there is a high-degree of trace-to-trace correlation in the frequency spectrum of effective signals. More specifically, SVD is incorporated to decompose the frequency spectra into eigenimages. The authors then compare the result of filtering through the proposed method with that of traditional filter methods on both synthetic and real-life seismic data, and state that their method demonstrates better performance both in the removal of background noise, as well as in the preservation of the effective signals.

Seismic frequencies provide important information on the underlying geological structures, and effective denoising techniques aid in the discovery of such information.