



Singular Value Decomposition and Data Compression Techniques.

Giovanna Castello de Andrade*, Sandra Augusta Santos.

Abstract

This project addresses computational methods for obtaining singular values and singular vectors of matrices, focusing on the large-scale setting. Strategies of data compression based on the statistical technique of principal components analysis are our main motivation. At first, Lanczos method, which is a matrix-free strategy to determine a set of eigenpairs of symmetric matrices, was studied and implemented. Then, such fundamentals methods were used to obtain a partial singular value decomposition of data matrices, in order to explore practical problems by means of the principal components analysis. In particular, experimental results were performed in image compression.

Key words: Lanczos method, singular value decomposition, principal component analysis.

Introduction

Singular Value Decomposition (SVD) is an important tool for matrix analysis. Its efficient determination in large scale problems is a relevant and current research topic. According to [1], SVD sorts information contained in matrices so that its main content becomes visible. Principal Component Analysis (PCA) is a widely applied multivariate technique which is strongly based on this propriety and can be used for data compression. As theoretic foundation for this project, Lanczos Method and its limitations, such as the phenomenon of “ghost eigenvalues” [2], were studied for obtaining extreme eigenpairs of symmetric matrices. Lanczos bidiagonalization was studied and implemented for obtaining partial singular value decompositions.

Results and Discussion

Let A be an $m \times n$ rank r data matrix. Its SVD is $A = U\Sigma V^t$, where U is an $m \times m$ orthogonal matrix, V is an $n \times n$ orthogonal matrix and Σ is an $m \times n$ diagonal matrix, with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ known as singular values; or $A = \sum \sigma_i u_i v_i^t$, in which each term of the sum is a rank one matrix. The products $z_i = \sigma_i u_i$ are called principal components and are ordered by highest variance. In terms of linear algebra, principal components analysis is based on the possibility of truncating this sum after k terms to obtain the best rank k approximation for A , according to SVD proprieties. In terms of image compressing, choosing an optimal k number requires dealing with image quality measurements, such as *Peak Signal Noise Ratio (PSNR)* and *Structural Similarity Based Image Quality Assessment (SSIM)* - see [3].

When the process was applied to the widely used picture of peppers, with resolution 384 x 512 and rank 384, the variance of the principal components decreased fast, as shown in Image 1.

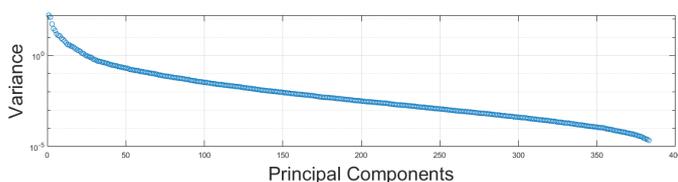


Image 1. Variance of Principal Components.

The approximations obtained with 10, 50, 80 and 150 principal components are presented in Image 2 and corresponding measures in Chart 1.

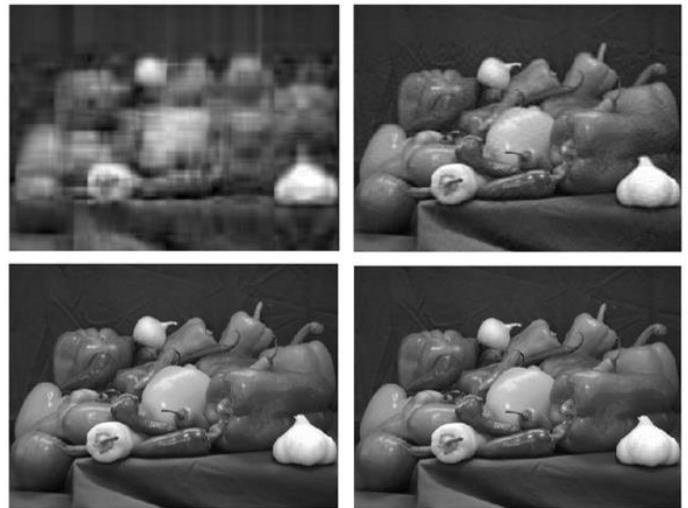


Image 2. Approximations built with 10, 50, 80 and 150 principal components.

Chart 1. Image quality measurements for image approximations.

Rank	PSNR	SSIM	Compression Factor
10	-22.92	0.3104	4.57%
50	-13.86	0.6024	22.83%
80	-9.2594	0.7295	36.54%
150	-1.9938	0.8808	68.51%

Conclusions

Experimental tests performed in different sets of images revealed that significant compression rates can be achieved when principal components variances decrease rapidly. If this doesn't occur, storage space needed to obtain a good approach to the original data might be similar or even larger, since standard deviations and means of the variables also need to be stored.

Acknowledgement

Supported by FAPESP, process 2018/07318-0.

¹ L. Eldén. *Matrix Methods in Data Mining and Pattern Recognition*. J. Korean Soc. Ind. Appl. Math., Philadelphia, **2007**.

² C.C. Paige. *The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices*, PhD thesis, University of London, **1971**.

³ Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.*, Vol. 13, No. 4, **2014**, 600-612.