International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009

Singular Value Decomposition A Classroom Approach

Soumya T Soman, Soumya V.J, Soman K.P
Amrita Vishwa Vidyapeetham/CEN,Coimbatore,India
soumyatsoman@gmail.com
soumya.vj@gmail.com
kp_soman@ettimadai.amrita.edu

Abstract— **In this paper we describe the geometrical interpretation of SVD based on the Pythagoras Theorem and optimization theory. SVD is a technique which factorizes matrix A into three matrices U,∑,V, such that $A = U\sum V^T$. The aim of this paper is to provide an easy way to understand SVD by using Pythagoras Theorem. SVD has variety of applications in engineering, chemistry, ecology, geology, geophysics, biomedical, scientific computing, automatic control, astro physics and many other areas. Newer applications of SVD are being pursued. Here we mention some successful applications of SVD and the concept of Higher Order Singular Value Decomposition (HOSVD) and its applications.**

I. Introduction

SVD is an important concept in linear algebra. Usually the concept of SVD is rarely covered in the course of undergraduate level because it is included only in the last portion of the syllabus and often skipped over in graduate courses. Consequently relatively few mathematicians are familiar with what M.I.T. Professor Gilbert Strang calls "absolutely a high point of linear algebra."

SVD was developed in the mid 19th century but most of its applications are emerged in the 21st century. SVD for square matrices was developed by Eugenio Beltrani (1873)[1], Camille Jordan[1] (1874), James Joseph Sylvester[1] (1889) and Autonne[1] (1915). Eckart and Young developed SVD in the 1930's for rectangular matrices and its use as a computational tool dates back to the 1960's. Golub and van Loan demonstrated its usefulness and feasibility in a wide variety of applications.

Let A be an $m \times n$ matrix can be represented as the product of two orthonormal matrices (U and V) and a diagonal matrix ($\sum$).

$$A = U\sum V^T \tag{1}$$

II.PHYSICAL INTERPRETATION OF $A = U\sum V^T$

For understanding the physical interpretation first considering a matrix A that is of dimension $m \times 2$ where $m \geq 2$ .Each row vector represent a data point in R^2 ($x_1 - x_2$ plane) .The plot of the data is as shown below.

According to Pythagoras Theorem the sum of the squares of the lengths of all data points is equal to the sum of squares of all the elements of the matrix A. In another way we can express it as the sum of squares of the x_1 coordinates and x_2 coordinates. For representing the sum of elements of matrix A we consider two unit norm axes e_1 as $\begin{bmatrix}1\\0\end{bmatrix}$ and e_2 as $\begin{bmatrix}0\\1\end{bmatrix}$. Ae_1 and Ae_2 represents x_1 and x_2 coordinates of all data points. For calculating the required sum we need to perform the square and sum of each of the elements of the vectors. This is easily obtained by dot product.

Let S represent sum that is the total variation in the data.

$$S = (Ae_1)^T (Ae_1) + (Ae_2)^T (Ae_2) \tag{2}$$

$$S = e_1^T A^T Ae_1 + e_2^T A^T Ae_2 \tag{3}$$

$=$ variation along x_1 axis + variation along x_2 axis.

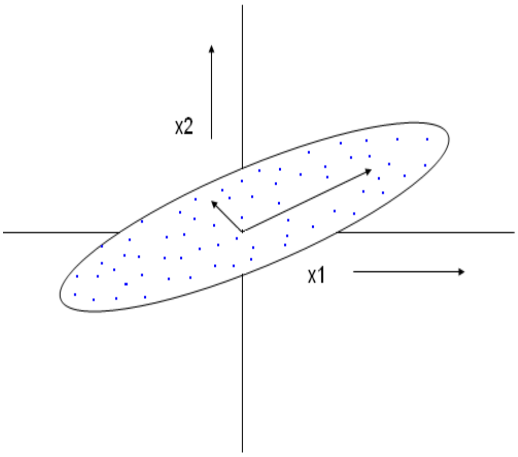

Figure 1.

A 2-D vector can be decomposed into two components in any two orthogonal directions. By using Pythagoras theorem and the data vector length remains same, we get the same sum S. The figure2 exemplifies this.

Consider two orthogonal unit vectors v_1 and v_2 and project data on to those vectors. Since the total sum of square of vector lengths do not change, S can be written as

$$S = v_1^T A^T A v_1 + v_2^T A^T A v_2 . \quad (4)$$

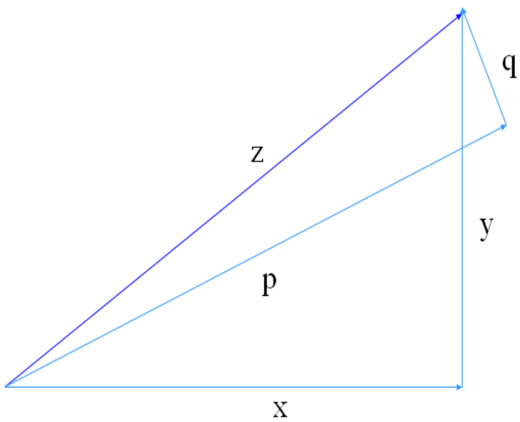

Figure 2

$$z^2 = x^2 + y^2$$

$$z^2 = p^2 + q^2$$

Choose two unit norm orthogonal vectors v_1 and v_2 such that most of the variation of the data is along one axis. Let v_1 be the unit vector along that axis – that is 45 degree to the x axis $\left(v_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right)$. The components of data points along this direction is Av_1 .So the variation along this direction is

$$(Av_1)^T (Av_1) = v_1^T A^T A v_1 . \quad (5)$$

Mathematically the direction in which the variation is maximum can be obtained from the solution of the following optimization problem

$$\max_v v^T A^T A v$$

$$subject \text{ to } v^T v = 1 \quad (6)$$

The constraint $v^T v = 1$ ensures that v is having unit norm.Taking Lagrangian and applying first order optimization condition we obtain the following

$$L(v,\lambda) = v^T A^T A v - \lambda(v^T v - 1)$$, λ is Lagrangian multiplier.

$$\frac{\partial L}{\partial v} = 2A^T A v - 2\lambda v = 0 \Rightarrow A^T A v = \lambda v$$

The direction is given by eigen vector of $A^T A$.Since $A^T A$ is 2 x 2 matrix, it has two eigen vectors. We choose the eigen vector corresponding to largest eigen value, because λ represents variation along the direction v.This follows from the fact that

$$v^T A^T A v = v^T \lambda v = \lambda v^T v = \lambda . \quad (7)$$

Let λ_1 be the eigen value, the maximum variation among all possible directions and the corresponding eigen vector be v_1. In fig 1 the principal axis of the ellipse represents the direction along which data is distributed. The remaining variation is along an axis perpendicular to the principal axis. All eigen vectors of a real symmetric matrix are orthogonal. Since $A^T A$ is symmetric, total variation S can be split by projecting data on to normalized eigen vectors of $A^T A$.These directions are v_1 and v_2 respectively and λ_1 and λ_2 are its variations. Also S= $\lambda_1 + \lambda_2$.

Let $Av_1 = a_1 u_1$. Av represent the component of each data point along the direction of vector v . Here a_1 is a scalar . Since there are m data points, there are m components and hence u_1 is of dimension mx1. Similarly we have $Av_2 = a_2 u_2$. Therefore

$$A\begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \text{ or}$$

$$AV = U\Sigma \tag{8}$$

Since the columns of V are orthonormal V is 2x2 orthonormal matrix and hence $V^T V = I$

Thus $A = U\Sigma V^T = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}$.

Observe the dimensions of each matrix.

$$A_{m\times 2} = U_{m\times 2} \Sigma_{2\times 2} V_{2\times 2}^T \tag{9}$$

SVD can be generalized to deal with matrices having r independent columns $(r \le n)$.

Any rectangular matrix A can be factorized into three matrices.

- Columns of V are eigen vectors of matrix $A^T A$.
- Columns of U are projections of A (or data points which are rows of A) on to columns of V vectors.
- $\sum$ is a diagonal matrix and diagonal elements are square root of variation of data points along the columns of V.

Using Linear Algebra we can prove the following

- Columns of U are orthonormal.
- Columns of U are eigen vectors of matrix $A\ A^T$.
- Eigen values of $A\ A^T$ and $A^T A$ are same (for normalized eigen vectors).

III.HIGHER ORDER SINGULAR VALUE DECOMPOSITION (HOSVD)

In many applications data commonly are organized according to more than two categories. The corresponding mathematical objects are usually referred to as tensors, and the area of mathematics dealing with tensors is multilinear algebra. The matrix SVD can be generalized to tensors in different ways. The generalization which is analogous to approximate principal component analysis is often referred to as Higher Order SVD (HOSVD) [2].It is used in the case of applications involving higher order tensors or multidimensional matrices, the existing framework of vector and matrix algebra appears to be insufficient.The tensor $A \in \Re^{l\times m\times n}$ can be written as

$$A = S \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^3 \quad , \tag{10}$$

where $U^{(1)} \in \Re^{l\times l}, U^{(2)} \in \Re^{m\times m}, and U^{(3)} \in \Re^{n\times n}$ are orthogonal matrices .S is a tensor of the same dimensions as A; it has the property of all – orthogonally: any two slices of S are orthogonal in the sense of the scalar product.

$$\langle S(i,:,:), S(j,:,:)\rangle = \langle S(:,i,:), S(:,j,:)\rangle = \langle S(:,:,i), S(:,:,j)\rangle = 0 \quad \text{for} \quad i \ne j$$

The 1-mode singular values are defined by $\sigma_j^{(1)} = \| S(i,:,:) \|_F , j = 1,....,l$ and they are ordered $\sigma_1^{(1)} \ge \sigma_2^{(1)} \ge \cdots \ge \sigma_l^{(1)}$. The singular values in other modes and their ordering are analogous.

Some of the applications of HOSVD are Classification of Handwritten digits, Face recognition, Blind Source Separation.

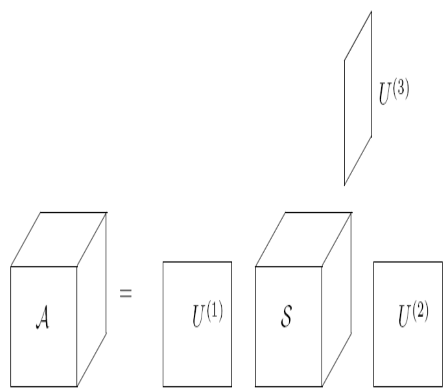

Figure 3.Visualization of the HOSVD

Some important applications of SVD are System Identification[3], Orderdetermination[4], Image Coding[5], Filter Design[6], Equalization of Fading Channels[7], Identification of Sinusoidal Components[8], Resolution of Exponential Signals[9],Direction-of-Arrival Estimate[10], Adaptive Beam Forming[11], Oriented Energy of Vector Sequences[12], Signal and Noise Resolution[13], Oriented Signal to Signal Ratio of two Vector Sequence[14], Canonical Correlation and Angles Between Subspaces[15]

IV.FURTHER RESEARCH

Some of the currently pursuing research areas are Statistical Analysis of SVD of Data Matrix in General Noisy Case, Order Threshold, SVD of Chaotic Systems, SVD of Fractal Processes, Recover From the SVD of Data Matrix, the Rank of the Signal Matrix, Recover Column Space of Signal Matrix, Adaptive SVD, Nonlinear Problems, Implementation Architectures for Computing SVD, Real-time Applications.

V. CONCLUSION

SVD is a very powerful tool. It has a lot of real world applications. SVD is considered as very sacred in the eyes of pure mathematicians. However, for the engineers who work with applied mathematics, which combines geometry with algebra, the concept of SVD becomes a little difficult to comprehend. That may be the reason why SVD is not as famous as it should be. This paper can be regarded as a small step in simplifying the concept of SVD by introducing high school level geometry using Pythagoras Theorem. Better insight gained into this technique will help to harness its power for various applications. We have already identified several applications of SVD. Here we have mentioned the concept of HOSVD that has been used in the case of applications involving higher order tensors or multidimensional matrices.

REFERENCES

[1] G.W.Stewart,"On the Early History of the Singular Value Decompositions," IMA Preprint Series #952,April 1992.

[2] Lars Elden," Matrix Mehtods in Data Mining and Pattern Recognition,"SIAM, Philadelphia, PA, 2007.

[3] Lastman, G.J Sinha," Use of Singular Value Decomposition in System Identification,"University of Waterloo, Department of Applied Mathematics, Waterloo, Canada, March 1986.

[4] Zhang, X.-D. Zhang, Y.-S,"Singular Value Decomposition-based MA order determination of non- Guassian ARMA models," Changcheng Inst.of Metrol.& Meas.,Beijing,Aug 1993.

[5] Andrews,H.Patterson,C.,III,"Singular Value Decomposition (SVD) Image Coding," Univ. of Southern California,LA,CA,Apr 1976.

[6] Fei,S.-C.;Lu,W.-S.;Tseng,C.-C,"Two-Dimensional FIR notch filter design using Singular Value Decomposition,"Circuits and Systems I:Fundamental Theory and Applications ,IEEE Transactions on Volume 45,Issue 3,Mar 1998.

[7] Tiziano Bianchi; FabrizioArgenti,"SVD-based Techniques for ZeroPadded Bloch TransmissionOver Fading Channels,"Signal Processing, IEEE Transactions on Volume 55, Issue 2,Feb 2007.

[8] Farrier,D.R,"Spectral Estmation by Damped Sinusoidal Modeling and Singular Value Decomposition[speech recognition],"Spectral Estimation Techniques for Speech Processing,IEE Colloquiun on Volume ,Issue,Feb1989.

[9] Chinghui J.Ying, Lee C.Potter,Randolph L.Moses,"On ModelOrder Determination for Complex Exponential Signals,"Performance of an FFT-Initialized ML Algorithm. Department of Electrical Engineering,The Ohio State University,2015 Neil Avenue, Columbus,OH 4310,USA ,Jun 1994.

[10] Seunghyeon Hwang; Sarkar, T.K ,"Direction of Arrival(DOA) Estimation using a Transformation Matrix through Singular Value Decomposition,"Wireless Communications and Applied Computational Electromagnetics ,2005. IEEE/ACES International Conference on Volume,Issue,April 2005.

[11] Leon H. Sibul ," Singular Value Decomposition in Adaptive Beam forming,"The Pennsylvania State University,State College,PA,Dec 1986.

[12] Li-gao Zhou;Li-yunYu,"Multi-channel FECG Signal Processing By Using Singular Value Decomposition,"Engineering in Medicine and Biology Society , 1990,Proceeding of the Twelfth Annual International Conference of the IEEE Volume,Issue ,Nov 1990.

[13] Duan Li-xiang,Zhang Lai-bin and Wang Zhao-hui ,"De-niosing of Diesel Vibration Singal using Wavelet packet and Singular Value Decomposition,"Faculty of Mechanical and Electronic Engineering,China Universityof Petroleum,Beijing ,102249,China , Oct2006

[14] DE Lathauwer,L.;De Moor ,B.; Vandewalle ,J, "SVD based Methodologies for Fetal Electrocardiogram," Extraction. Acoustics, Speech, and Signal Processing,2000.

[15] Pezeshki,A.; Scharf,L.L.;Azimi-Sadjadi,M.R,"Empirical Canonical Correlation Analyses in Subspaces ,"Signals, Systems and Computers ,2004.Conference Record of the Thirty-Eighth Asilomar Conference on Volume 1,Issue 7-10,Nov 2004.