Journal of Korean
Data & Information Science Society
2007, Vol. 18, No. 4, pp. 1135 ~ 1143

The Comparison of Singular Value Decomposition and Spectral Decomposition

Yang Gyu Shin[1)]

Abstract

The singular value decomposition and the spectral decomposition are the useful methods in the area of matrix computation for multivariate techniques such as principal component analysis and multidimensional scaling. These techniques aim to find a simpler geometric structure for the data points. The singular value decomposition and the spectral decomposition are the methods being used in these techniques for this purpose. In this paper, the singular value decomposition and the spectral decomposition are compared.

Keywords : Multidimensional Scaling, Principal Component Analysis, Singular Value Decomposition, Spectral Decomposition

1. Introduction

The singular value decomposition and spectral decomposition are methods which are used to find a linear structure of reduced dimension and to give interpretation of the lower dimensional structure. Good(1969) showed the application of singular value decomposition of a matrix. Shin(1998) discussed the advantage of the singular value decomposition from the algebraic point of view. In particular, Choi and Huh(1996) derived resistant version of singular value decomposition for principal component analysis. On the other hand Kim and Park(1993) proposed an efficient algorithm for computing the orthogonal projection matrix for a balanced model using the spectral decomposition.

Multivariate techniques such as principal component analysis, factor analysis, multidimensional scaling, generalized principal component analysis, etc. are tools to analyze the multivariate data. Many statisticians (Jackson and Hearne(1975),

1) Professor, Department of Asset Management, Daegu Hanny University, Kyungsan, Korea.
E-mail : yks@dhu.ac.kr

Granadesikan(1977), Chatfield and Collins(1980), Tso(1981)) viewed these multivariate techniques as exploratory rather than inferential in that they seek to formulate hypotheses more than to test hypotheses. In this regard, the underlying theme of these techniques is to find point which would either reduce the dimensionality or suggest a possible internal relationships among units or variable.

By the singular value decomposition and the spectral decomposition of an $n\times p$ data matrix X, we can get dimensional reduction and find the linear structure in data reduction techniques. When the number of variable is large, the singular value decomposition of the data matrix is computationally far more efficient than the spectral decomposition of the sample covariance matrix.

In Section 2, we compare the singular value decomposition and the spectral decomposition of a matrix. In Section 3, we compare the way the singular value decomposition and the spectral decomposition are used in the multivariate techniques. Finally in Section 4, we discuss our results.

2. The Singular Value decomposition and the Spectral Decomposition

2.1 The Singular Value Decomposition of Matrix

Let X be a $p\times n$ matrix and consider the $p\times p$ symmetric matrix XX^t(or $n\times n$ symmetric matrix X^tX). Suppose $a_1,\cdots,a_p$ are the orthogonal eigenvectors of XX^t and that $a_1,\cdots,a_k$ correspond to the k positive eigenvalues $\lambda_1^2,\cdots,\lambda_k^2$ of XX^t. Then

$$X=(a_1a_1^t+\cdots+a_pa_p^t)X=a_1a_1^tX+\cdots+a_ka_k^tX \tag{1}$$

because for $j=k+1,\cdots,p$, $XX^ta_j=0$. Define $\lambda_jb_j=X^ta_j$. Then $X^tXb_j=\lambda_j^2b_j$. Thus $b_1,\dots,b_k$ are the orthonormal eigenvectors of X^tX correspond to the k positive eigenvalues $\lambda_1^2,\cdots,\lambda_k^2$ of X^tX. The equation (1) is written

$$X=\lambda_1a_1b_1^t+\cdots+\lambda_ka_kb_k^t \tag{2}$$

and the equation (2) is called the singular value decomposition of X.

The numbers $\lambda_1,\lambda_2,\cdots$ are the singular values of X and the vectors a_1, $a_2,\cdots$, $b_1,b_2,\cdots$, are the right and left singular vectors of X. The singular value decomposition of a matrix is a way to express a matrix as the sum of matrices of rank one.

Suppose M is a $p \times p$ real symmetric matrix of rank p. Then the right and left singular vectors are identical and reduce to eigenvectors. Thus the singular value decomposition of M is

$$M = \lambda_1 a_1 a_1^t + \cdots + \lambda_p a_p a_p^t \ , \tag{3}$$

where $\lambda_1, \cdots, \lambda_p$ are eigenvalues of M and $a_1, \cdots, a_p$ are eigenvectors of M.

The singular value decomposition of X provides an immediate analysis of the effect of X regarded as a linear transformation acting on the vectors of Euclidean n-space and p-space respectively.

We can use the singular value decomposition of X to solve p simultaneous linear equations $Xy = a$, where a is given and y is determined. Write $a = \beta_1 a_1 + \cdots + \beta_p a_p$ and $y = \gamma_1 b_1 + \cdots + \gamma_n b_n$. Then we obtain the consistency conditions $\beta_{k+1} = 0, \cdots, \beta_p = 0$ and, subject to these conditions, the solutions $\gamma_1 = \lambda_1^{-1}\beta_1, \cdots, \gamma_k = \lambda_k^{-1}\beta_k$ with $\gamma_{k+1}, \cdots, \gamma_n$ arbitrary. Thus $y = X^{-1}a$ is a solution if there is one , and the general solution is $y = X^{-1}a + (X^{-1}X - I)d$, where d is arbitrary.

2.2 The Spectral Decomposition of a Real Symmetric Matrix

Let M be a real symmetric matrix. Let λ be an eigenvalue of M and a an eigenvector associated with λ: $Ma = \lambda a (a \neq 0)$. The subspace $S_\lambda = \{a \mid Ma = \lambda a\}$ is called the subspace associated with λ. Let a_1 and a_2 be eigenvectors associated with distinct eigenvalues λ_1 and λ_2 , respectively. Then a_1 and a_2 are orthogonal. Hence, for a real symmetric matrix M, the eigensubspaces associated with distinct eigenvalues are orthogonal.

Suppose M has k distinct eigenvalues $\lambda_1, \cdots, \lambda_k$ and the corresponding eigensubspaces $S_1, \cdots, S_k$. Then any vector $y \in R^p$ may be written as

$$y = y_1 + \cdots + y_k, \tag{4}$$

where$y_i \in S_i, i = 1, \cdots, k$. From the equation (4), by the orthogonality of S_i , My and $M^2 y$ can be written as,

$$My = \lambda_1 y_1 + \cdots + \lambda_k y_k \tag{5}$$
$$M^2 y = \lambda_1^2 y_1 + \cdots + \lambda_k^2 y_k$$

Let $E_1, \cdots, E_k$ be orthogonal projection matrices on the subspaces $S_1, \cdots, S_k$ respectively, so that $E_i y_j = 0 (i \neq j)$. Because the subspaces are orthogonal, clearly we have $y_i = E_i y, i = 1, \cdots, k$. Substituting the equation (4)

$$y = E_1 y + \cdots + E_k y = (E_1 + \cdots + E_k) y.$$

Thus $I_p = E_1 + \cdots + E_k$. Similarly from the equation (5),

$$M = \lambda_1 E_1 + \cdots + \lambda_k E_k \quad (6)$$

and the equation (6) is called the spectral decomposition of the real symmetric matrix M.

Now if an eigenvalue λ of M has multiplicity m $\dim S_i = m$ and for orthonormal basis $\{\lambda_1, \cdots, \lambda_m\}$ of S_i which are eigenvectors corresponding to λ, the orthogonal projection E_λ on S_λ can be written as $E_\lambda = \lambda_1 \lambda_1^t + \cdots + \lambda_m \lambda_m^t$. Hence the equation (6) can be rewritten as

$$M = \sum_{i=1}^{p} \lambda_i a_i a_i^t \ , \quad (7)$$

where the λ_i are p eigenvalues of M and the a_i are orthonormalized eigenvectors corresponding to the λ_i.

From the equations (3) and (7), We can say that the spectral decomposition is a special case of the singular value decomposition; spectral decomposition exists only for square non-negative matrix, whereas the singular value decomposition always exists for the symmetric matrix.

3. The Singular Value Decomposition the Spectral Decomposition in Principal Component Analysis and Multidimensional Scaling

In this chapter, we illustrate and compare the way the singular value decomposition and the spectral decomposition are used in the principal component analysis and the multidimensional scaling from the geometric and algebraic point of view.

The principal component analysis concerns the recognition of lower dimensional linear subspaces which the multi responses observations may, lie. The basic idea of the principal component analysis is to describe the dispersion of an array of n points in p-dimensional space by introductory a new set of orthogonal linear coordinate so that the sample variances of the given points which respect to these derived coordinates are in decreasing order of magnitude. The first principal component is that the projection of the given points onto it have maximum variance among all possible linear coordinates; the second principal component has maximum variance subject to being orthogonal to the first; and so on. We hope that the first components will account for most of the variation in the original data so that the effective dimensionality of the data can be reduced.

Suppose $X^t = [X_1, \cdots, X_p]$ be a p-dimensional random variable with mean 0 and covariance matrix $\sum(=XX^t)$. A p by n dimensional random matrix X represents multivarite statistical sample of n observations on p variables. Denote a new set variable, $Y_1, \cdots, Y_p$, which are uncorrelated and whose variances decrease form first to last. Each Y_j is taken to be a linear combination of the X's so that

$$Y_j = a_{1j}X_1 + \cdots + a_{pj}X_p = a_j^t X \ .$$

To preserve the distances in p-space, we impose the condition that $a_j^t a_j = \sum_{k=1}^{p} a_{kj}^2 = 1$. Now, we want to find the first principal component Y_1 such that $Y_1 = a_1^t X$. Denote the eigenvalues of X by $\lambda_1, \cdots, \lambda_p$ and assume that they are distinct so that $\lambda_1 > \lambda_2 > \cdots > \lambda_p \geq 0$. Then the principal component, a_1, is the eigenvector of corresponding to the largest eigenvalue λ_1. To obtain the second principal component, $Y_2 = a_2^t X$, we have the condition such that $a_2^t a_2 = 1$ and Y_2 should be uncorrelated with Y_1. In this case, λ_2 is the second largest eigenvalue of $\sum$ and a_2 is the corresponding eigenvector. Continuing this argument, the jth principal component is the eigenvector associated with the jth largest eigenvalue λ_j. The total variance is $tr\sum = \lambda_1 + \cdots + \lambda_p$. The important of the jth component in a more parsimonious description of the system is measure by $\lambda_j / tr\sum$.

Let A denote the $(p \times p)$ matrix of eigenvectors, where $A = [a_1, \cdots, a_p]$, and Y denote the $(p \times 1)$ vector of principal components, then $Y = A^t X$. The $p \times p$ covariance matrix of Y is given by

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} = D(\lambda_i)$$

Since $var(Y) = A^t \sum A$, $D(\lambda_i) = A^t \sum A$. Since A is an orthogonal matrix with $AA^t = I$,

$$\sum = AD(\lambda_i)A^t = \sum_{i=1}^{p} \lambda_i a_i a_i^t. \tag{8}$$

Note that the rank k of $\sum$ is less than p. From the above, principal component analysis is related the spectral decomposition of the covariance matrix.

Multidimensional scaling aims to find a configuration in a much smaller number of dimensions which approximately reproduce the given dissimilarities. Suppose the data matrix X have the exact-coordinate of n points in p-dimensional Euclidean space. Let $C_{n\times n} = X^tX$, where the term of C is given by

$$c_{rs} = \sum_{j=1}^{p} x_{rj}x_{sj}.$$

Then the $(n\times n)$ matrix of squared Euclidean distance, D is

$$\begin{aligned} d_{rs} &= \text{squared Euclidean distance between point } r \text{ and } s \\ &= c_{rr} + c_{ss} - c_{rs} \end{aligned}$$

In multidimensional consider the inverse problem. Suppose we know the distance but not the co-ordinate. First we find the C matrix.

$$c_{rs} = -\frac{1}{2}(d_{rs}^2 - d_{r.}^2 - d_{.s}^2 + d_{..}^2),$$

where $d_{r.}^2$ = average term in rth row, $d_{.s}^2$ = average term in sth column, $d_{..}^{\ 2}$ = overall average squared distances.

Next we are factorizing C in the from $C = X^tX$. Since D consists of the squares of exact Euclidean distances, C is a positive symmetric matrix. Suppose C has a rank k, where $k \leq n$. Then C will have k non-zero eigenvalues which we arrange in the order of magnitude so that $\lambda_1 \geq \ldots \geq \lambda_k > 0$.

Let $\{b_i\}$ denote the corresponding eigenvectors of unit length and let B denote $[b_1, \cdots, b_k]$. To scale the eigenvectors so that their sum of squares is equal to λ_i, we set ei $= \sqrt{\lambda_i} b_i$. By the Young-Hausehold factorization theorem, a positive semidefinite matrix can be factorized into the form XX^t. Thus

$$C = XX^t = BD(\lambda_i)B^t \tag{9}$$

From the equation (9), multidimensional scaling is related to the spectral decomposition of the co-ordinate $C(= X^tX)$.

By the singular value decomposition, the data matrix X is written

$$X = \sqrt{\lambda_1}\, a_1 b_1{}^t + \ldots + \sqrt{\lambda_k}\, a_k b_k{}^t \ , \tag{10}$$

where $\lambda_1, \lambda_2, \cdots$ are the singular values of X, and the $a_1, a_2, \cdots, b_1, b_2, \cdots$ are the right and left singular vector of X.

From the equations (8) and (9),

$$XX^t = \lambda_1 a_1 a_1^t + \cdots + \lambda_k a_k a_k^t$$

$$X^t X = \lambda_1 b_1 b_1^t + \cdots + \lambda_k b_k b_k^t \tag{11}$$

where $a_1, \cdots, a_k$ and $b_1, \cdots, b_k$ are the eigenvectors of XX^t and $X^t X$ and λ_i is the eigenvalue. From the equations (10) and (11), the square of singular value of X is the eigenvalue of XX^t(or $X^t X$), And the right singular vector is the eigenvector of XX^t and the left singular vector is the eigenvector of XX^t.

Hence, in multivariate techniques such as principal component analysis and multidimensional scaling, we can find the linear structure by the singular value decomposition of a data matrix X. If the vectors $a_1, a_2, \cdots$ can be given a physical meaning then one would expect to find associated physical meanings for $b_1, b_2, \cdots$ respectively, so that it is natural to refer to a_i and b_i as a conjugate pair of vectors. Since XX^t(and $X^t X$) is a real symmetric and $b_i = \lambda_i^{-\frac{1}{2}} X^t a_i$, the linear structure in principal component analysis and multidimensional scaling is ellipsoid concerned on the origin and using as co-ordinate axes as the principal axes resulting from the principal component analysis.

On the other hand, since $m(XX^t) = m(X), m(X^t X) = m(X^t)$ and $m(X^t X)$ is the dual space of $m(XX^t)$, the linear structure of data is an ellipsoid concerned on the origin. Let y be the stationary point of that ellipsoid. Then $y = (y_1, \cdots, y_k)$ where $k = 1, \cdots, p$ is the right(or left) singular vectors of data matrix X. The stationary value of $\| y \|$ is $\frac{c}{\lambda}$, where c is a constant and λ is a singular value of X. Hence we find the linear structure of the reduced dimension and give interpretation of that space by the singular value decomposition of X.

4. Discussion

The singular value decomposition and the spectral decomposition are the useful methods in the area of matrix computation for the multivariate reduction techniques. Through section 2 and section 3, we can find certain advantage using the singular value decomposition over the spectral decomposition. Firstly, working directly with the singular value decomposition of X. we can also maintain firsthand feeling for the data which would have been at best diminished if the spectral decomposition of XX^t were used as instead. From the equations (3) and (7), we can derive that the spectral decomposition is a special case of the singular value decomposition. As a practical matter, there are certain advantages in using singular value decomposition over spectral decomposition. Working with the data matrix itself clearly involves far less computational effect than working with the sample covariance matrix or the sample coordinate matrix, and the saving of the effort becomes more apparent as the number of variable gets large. Moreover, most importantly, revealing the eigenvectors of XX^t and X^tX in the single equation, the duality between XX^t and X^tX as evidenced in the treatment of principal component analysis and the multidimensional scaling becomes more convincingly illustrated. Therefore, we can say that the singular value decomposition is more useful methods over spectral decomposition

Acknowledgement

The author thanks the editor and an anonymous referee for their comments which have led to an improved presentation of the paper.

References

1. Chatfield, C, and Collins, A. T. (1980), *Introduction to multivariate analysis*, Chapman and Hall, New York.
2. Choi, Y. and Huh, M. (1996), Resistant Singular Value Decomposition and Its Statistical Applications, *Journal of the Korean Statistical Society*, 25-1, 49-66.
3. Granadesikan, R. (1977), *Method for statistical data analysis of multivariate observations*, John Wiley & Sons, New York.
4. Good, I. C. (1969), Some applications of singular decomposition of a matrix, *Technometrics*, 11, 823-831.
5. Gower, J. C. (1966), Some distance properties of root and vector methods used in multivariate analysis, *Biometrica*, 53, 325-338.
6. Jackson, J. E. and Hearne, F. J. (1975), Relationships among coefficients of vectors used in principal components, *Technometrics*, 15, 601-610.
7. Kim, B. and Park, J. (1993), An Efficient Computing Method of the

Orthogonal Projection Matrix for the Balanced Factorial Design, *Journal of the Korean Statistical Society*, 22-2, 249-258.

8. Shin, Y. K. (1998), A study of Singular Value Decomposition in Data Reduction Techniques, *Journal of Statistical Theory & Methods*. 9-2, 63-70.
9. Tso, M. K. S. (1981), Reduced-rank regression and canonical analysis, *J. R. Statist. Soc., B*, 43-2, 183-189.

[received date : Oct. 2007, accepted date : Nov. 2007]