Available online at www.sciencedirect.com
SCIENCE DIRECT

Journal of Sound and Vibration 265 (2003) 1103–1110

JOURNAL OF SOUND AND VIBRATION

www.elsevier.com/locate/jsvi

Letter to the Editor

A note on equivalence of proper orthogonal decomposition methods

C.G. Wu[a], Y.C. Liang[a,b,*], W.Z. Lin[b], H.P. Lee[b,c], S.P. Lim[c]

[a] *College of Computer Science and Technology, Jilin University, 10 Qian Wei Road, Changchun 130012, People's Republic of China*

[b] *Institute of High Performance Computing, 1 Science Park Road, #01-01 The Capricorn, Singapore Science Park II, Singapore 117528, Singapore*

[c] *Department of Mechanical Engineering, National University of Singapore, 9 Engineering Drive 1, Singapore 119260, Singapore*

Received 30 September 2002; accepted 2 January 2003

1. Introduction

The proper orthogonal decomposition (POD) is a powerful and elegant method for data analysis aimed at obtaining low-dimensional approximate descriptions of a high-dimensional process. It is an important and essential technique for data reduction and feature extraction, and has been widely used in various disciplines including image processing, signal analysis, data compression, process identification, adaptive control, and many others. In general, there are two different interpretations for the POD. The first interpretation regards the POD as the Karhunen–Loève decomposition (KLD) and the second one considers that the POD consists of three methods: the KLD, the principal component analysis (PCA), and the singular value decomposition (SVD) [1]. The first interpretation appears in many engineering literatures related to the POD. Because of the close connections and the equivalence of the three methods, the authors prefer the second interpretation for the POD, that is, the POD includes the KLD, PCA and SVD.

The widespread applications of the POD methods enable the POD to be a popular tool in many fields. In recent years, there have been many reported applications of the POD methods in engineering fields. More recently, the POD methods have also been successfully used in structural vibrations, such as the physical interpretation of the proper orthogonal modes [2–4], applications of the KLD [3–6], the SVD [2,7] and the PCA [5,8–11] for system identification, dynamic analysis, and many others. With the increasing applications of the POD methods, it has been found that the loose description on the connection of the POD methods may confuse the applications. Therefore,

*Corresponding author: College of Computer Science and Technology, Jilin University, Changchun 130012, People's Republic of China.

E-mail addresses: wuchunguo@sina.com.cn (C.G. Wu), liangyc@ihpc.a-star.edu.sg (Y.C. Liang).

doi:10.1016/S0022-460X(03)00032-4

a summary of the equivalence of the three POD methods has been made and some mathematical derivations about them have been performed in Ref. [1]. In this paper, on the basis of Ref. [1] from the authors, the PCA of a random vector is obtained by discretizing the KLD of a continuous stochastic process, and then the equivalence of the KLD and the PCA is proven. A novel proof on the proper orthogonal basis vectors of the SVD satisfying the optimality of the POD is presented and the equivalence of the SVD and the PCA (KLD) is expounded.

2. The equivalence of KLD and PCA

Karhunen and Loève independently developed a theory regarding optimal series expansions of continuous-time stochastic processes [12]. The theory is called Karhunen–Loève decomposition and has been used extensively in the fields such as image processing, digital communication, and many others. Their results extend the PCA to the case of infinite-dimensional spaces, such as the space of continuous-time functions. Inversely, if discretizing the KLD for the continuous-time stochastic process and taking the finite discrete time to consider problems we may use the PCA for processing random vectors.

Let $x(t) \in R$ be a continuous-time stochastic process with a zeromean, where $t \in [a, b]$ (a, b finite), and the autocorrelation function be

$$r_x(t, t') = E\{x(t)x(t')\} \quad (a \leqslant t, t' \leqslant b). \tag{1}$$

From $E\{x(t)x(t')\} = E\{x(t')x(t)\}$ and the continuity of $x(t)$ it follows that $r_x(t, t')$ is symmetric and continuous.

For all $\psi(\cdot)$ satisfying

$$\int_a^b \psi^2(t)\,\mathrm{d}t < \infty, \tag{2}$$

we have

$$\begin{aligned}\int_a^b \int_a^b r_x(t, t')\psi(t)\psi(t')\,\mathrm{d}t\,\mathrm{d}t' &= E\left\{\int_a^b \int_a^b x(t)x(t')\psi(t)\psi(t')\,\mathrm{d}t\,\mathrm{d}t'\right\} \\ &= E\left\{\left(\int_a^b x(t)\psi(t)\,\mathrm{d}t\right)^2\right\} \geqslant 0.\end{aligned} \tag{3}$$

Therefore, from Mercer's theorem [13], it follows that the continuous symmetric function $r_x(t, t')$ satisfies the condition for the series expansion, that is, we have

$$r_x(t, t') = \sum_{i=1}^{\infty} \lambda_i e_i(t) e_i(t'), \tag{4}$$

where the functions $e_i(t)$ are called eigenfunctions of the expansion and the numbers λ_i are called eigenvalues, which satisfy

$$\int_a^b r_x(t, t') e_i(t')\,\mathrm{d}t' = \lambda_i e_i(t). \tag{5}$$

For the $e_i(t)$ and λ_i ($i = 1, 2, \ldots$) in the expansion, we have the following theorem.

Theorem 1 (Diamantaras and Kung [12]). *Let $x(t)$ be a zero-mean continuous-time stochastic process with correlation $r_x(t, t') = E\{x(t)x(t')\}$. Then*

$$x(t) = \sum_{i=1}^{\infty} y_i e_i(t), \tag{6}$$

where

$$y_i = \int_a^b x(t) e_i(t)\,\mathrm{d}t \tag{7}$$

are uncorrelated random variables with zero mean and variance λ_i, that is,

$$E\{y_i y_j\} = \begin{cases} 0, & i \neq j, \\ \lambda_i & i = j. \end{cases} \tag{8}$$

Eq. (7) in Theorem 1 is called the Karhunen–Loève transform of $x(t)$, and Eq. (6) is called the Karhunen–Loève decomposition of $x(t)$. In order to extend the continuous-time process into the discrete-time process and expound the equivalent relationship between the KLD and PCA, we present the following theorem.

Theorem 2. *If the time is taken as finite discrete values and integrals are replaced by sums, then the KLD of the continuous stochastic process $x(t)$ is equivalent to the PCA of the random vector.*

Proof. Let $x(t)$ be a given continuous stochastic process. Discretizing the time in $[a, b]$ infinitely, denoting the discrete-time points in order from left to right as $t_1, t_2, \ldots$ and denoting $x_k = x(t_k)$ $(k = 1, 2, \ldots)$, then we have that $x_1, x_2, \ldots$ is a discrete stochastic process satisfying

$$E\{x_k\} = 0, \quad r_x^{(k,l)} \equiv r_x(x_k, x_l) = E\{x_k x_l\}.$$

From Eq. (6), which is used to perform the KLD for the original continuous stochastic process, we have

$$x_k = \sum_{i=1}^{\infty} y_i e_i^k, \tag{9}$$

where

$$y_i = \sum_{k=1}^{\infty} x_k e_i^k \tag{10}$$

and $e_i^k = e_i(t_k)$ $(k = 1, 2, \ldots,)$ is the kth component of the infinitely dimensional eigenvector $\mathbf{e}_i$ of $r_x^{(k,l)}$, which satisfies that

$$\sum_{l=1}^{\infty} r_x^{(k,l)} e_i^l = \lambda_i e_i^k. \tag{11}$$

If we take the time as finite discrete values $x_k = x(t_k)$ $(k = 1, 2, \ldots, m)$ and denote that

$$\mathbf{x} = (x_1, x_2, \ldots, x_m)^{\mathrm{T}}, \tag{12}$$

$$\mathrm{R}_x = \begin{pmatrix} r_x^{(1,1)} & r_x^{(1,2)} & \cdots & r_x^{(1,m)} \\ r_x^{(2,1)} & r_x^{(2,2)} & \cdots & r_x^{(2,m)} \\ \vdots & \vdots & \ddots & \\ r_x^{(m,1)} & r_x^{(m,2)} & \cdots & r_x^{(m,m)} \end{pmatrix}, \tag{13}$$

$$\mathbf{e}_i = (e_i^1, e_i^2, \ldots, e_i^m)^{\mathrm{T}}, \tag{14}$$

$$\mathbf{y} = (y_1, y_2, \ldots, y_m)^{\mathrm{T}}, \tag{15}$$

then Eqs. (9), (10) and (11) can be written in matrix forms, respectively, as

$$\mathbf{x} = \sum_{i=1}^{m} y_i \mathbf{e}_i, \tag{16}$$

$$y_i = \mathbf{e}_i^{\mathrm{T}} \mathbf{x}, \ (i = 1, 2, \ldots, m), \tag{17}$$

$$\mathbf{R}_x \mathbf{e}_i = \lambda_i \mathbf{e}_i, \ (i = 1, 2, \ldots, m). \tag{18}$$

It can be seen that Eqs. (16)–(18) have the identical forms with Eqs. (24), (23) and (18) in Ref. [1]. Therefore, it is the principal component analysis of a random vector to perform KLD for a continuous stochastic process and to take finite discrete values. This completes the proof of the equivalence of the KLD and PCA. □

3. The equivalence of SVD and PCA (KLD)

The SVD was established for real square matrices at first, and was thereafter extended for complex square matrices and general rectangular matrices. Similar to the eigenvalue decomposition the SVD is a very important and fundamental working tool in matrix analysis. Because the SVD intimately relates to the matrix rank and reduced-rank least-squares approximation, it has been extensively applied to many areas such as matrix theory, statistics, and signal analysis [12].

Suppose that n samples $\{\mathbf{x}_i\}_{i=1}^{n}$ are given where $\mathbf{x}_i \in R^m$. Let the sample matrix be $\mathbf{X}$ ($\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$). The equivalence of the SVD and PCA (KLD) has been demonstrated and proven in Ref. [1] by using eigenvalue problems of matrices and the asymptotic connection of the samples. It has also been proven that the proper orthogonal basis vectors of the SVD satisfy the optimality of the POD in terms of Lagrangian function. In this paper, we present a novel proof on the optimality of the proper orthogonal basis vectors of the SVD by using the Frobenius norm of the matrix. In the following theorems $\|\bullet\|_2$ represents the Euclidian norm and $\|\bullet\|_F$ the Frobenius norm.

Theorem 3 (Diamantaras and Kung [12]). *Let* $\mathbf{U}\sum\mathbf{V}^{\mathrm{T}}$ *be the SVD of an* $n \times m$ *matrix* $\mathbf{A}$ *with rank* r. *Assume that the singular values are arranged in decreasing order* $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > \sigma_{r+1} = \cdots =$

$\sigma_m = 0$, then for any $l \leqslant r$,

$$\min_{rank(\mathbf{B})=l} ||\mathbf{A} - \mathbf{B}||_F^2 = ||\mathbf{A} - \mathbf{A}_l||_F^2 = \sum_{i=l+1}^{r} \sigma_i^2, \tag{19}$$

where

$$\mathbf{A}_l = \sum_{i=1}^{l} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}.$$

Next, we present a theorem and its proof on that the proper orthogonal basis vectors of the SVD satisfy the optimality of the POD.

Theorem 4. *Let the sample matrix be $\mathbf{X}$, the matrix formed by the proper orthogonal basis vectors obtained from the SVD be $\mathbf{V}$ and the matrix formed by any other set of orthogonal basis vectors be Φ. If the first l basis vectors of $\mathbf{V}$ and Φ are taken to perform the reconstruction for the sample matrix, respectively, and the errors of the reconstruction are denoted as $\varepsilon^2(\mathbf{V}_l)$ for the former and $\varepsilon^2(\Phi_l)$ for the latter, then we have*

$$\varepsilon^2(\Phi_l) \geqslant \varepsilon^2(\mathbf{V}_l). \tag{20}$$

Proof. Let $\hat{\mathbf{x}}_j$ $(j = 1, 2, \ldots, n)$ represent the reconstruction for the original samples using the first l basis vectors of $\mathbf{V}$.

$$\hat{\mathbf{x}}_j = \sum_{i=1}^{l} \alpha_{ij} \mathbf{v}_i \quad (j = 1, 2, \ldots, n), \tag{21}$$

where

$$\alpha_{ij} = \mathbf{v}_i^{\mathrm{T}} \mathbf{x}_j. \tag{22}$$

Let $\tilde{\mathbf{x}}_j$ $(j = 1, 2, \ldots, n)$ represent the reconstruction for the original samples using the first l basis vectors of Φ

$$\tilde{\mathbf{x}}_j = \sum_{i=1}^{l} \beta_{ij} \varphi_i \quad (j = 1, 2, \ldots, n), \tag{23}$$

where

$$\beta_{ij} = \varphi_i^{\mathrm{T}} \mathbf{x}_j. \tag{24}$$

The error of the reconstruction of the former is

$$\varepsilon^2(\mathbf{V}_l) = \sum_{j=1}^{n} ||\mathbf{x}_j - \hat{\mathbf{x}}_j||_2^2 = ||\mathbf{X} - \hat{\mathbf{X}}||_F^2, \tag{25}$$

where $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_n]$. The error of the reconstruction of the latter is

$$\varepsilon^2(\Phi_l) = \sum_{j=1}^{n} ||\mathbf{x}_j - \tilde{\mathbf{x}}_j||_2^2 = ||\mathbf{X} - \tilde{\mathbf{X}}||_F^2, \tag{26}$$

where $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_n]$.

From Theorem 3, if there exists a matrix $\mathbf{C} \in R^{n\times m}$ with $rank(C) = s \leqslant l$, then we have

$$\|\mathbf{A} - \mathbf{C}\|_F^2 \geqslant \min_{rank(\mathbf{B})=s} \|\mathbf{A} - \mathbf{B}\|_F^2 = \|\mathbf{A} - \mathbf{A}_s\|_F^2 = \sum_{i=s+1}^{r} \sigma_i^2 \geqslant \sum_{i=l+1}^{r} \sigma_i^2 = \|\mathbf{A} - \mathbf{A}_l\|_F^2. \tag{27}$$

In fact, since $rank(\mathbf{C}) = s$ and $\min_{rank(\mathbf{B})=s} \|\mathbf{A} - \mathbf{B}\|_F^2$ is the minimum Frobenius norm of the difference of all matrices that have the same rank s with $\mathbf{A}$, we have that $\|\mathbf{A} - \mathbf{C}\|_F^2 \geqslant \min_{rank(\mathbf{B})=s} \|\mathbf{A} - \mathbf{B}\|_F^2$ holds. Note that

$$\varepsilon^2(\mathbf{V}_l) = \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 = \|(\mathbf{X} - \hat{\mathbf{X}})^{\mathrm{T}}\|_F^2 = \|\mathbf{X}^{\mathrm{T}} - \hat{\mathbf{X}}^{\mathrm{T}}\|_F^2 \tag{28}$$

and

$$\varepsilon^2(\Phi_l) = \|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2 = \|(\mathbf{X} - \tilde{\mathbf{X}})^{\mathrm{T}}\|_F^2 = \|\mathbf{X}^{\mathrm{T}} - \tilde{\mathbf{X}}^{\mathrm{T}}\|_F^2. \tag{29}$$

Replacing $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{A}_l$ in Eq. (27) with, $\mathbf{X}^{\mathrm{T}}$, $\tilde{\mathbf{X}}^{\mathrm{T}}$ and $\hat{\mathbf{X}}^{\mathrm{T}}$, respectively, it can be seen that to prove

$$\varepsilon^2(\Phi_l) \geqslant \varepsilon^2(\mathbf{V}_l),$$

we need only to prove that

$$rank(\tilde{\mathbf{X}}^{\mathrm{T}}) \leqslant l \tag{30}$$

and

$$\hat{\mathbf{X}}^{\mathrm{T}} = \sum_{i=1}^{l} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}. \tag{31}$$

Since $\tilde{\mathbf{x}}_j \in span\,(\varphi_1, \varphi_2, \ldots, \varphi_l)$ $(j = 1, 2, \ldots, n)$, we have obviously that inequality (30) holds. In order to prove Eq. (31), using the SVD on the sample matrix $\mathbf{X}$ and letting

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > \sigma_{r+1} = \cdots = \sigma_m = 0,$$

we have

$$\mathbf{U} = \mathbf{X}^{\mathrm{T}} \mathbf{V} \Sigma_r^{-1}, \tag{32}$$

that is,

$$\mathbf{U} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^{\mathrm{T}} [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r] \begin{pmatrix} \sigma_1^{-1} & & & \\ & \sigma_2^{-1} & & \\ & & \ddots & \\ & & & \sigma_r^{-1} \end{pmatrix}. \tag{33}$$

Post-multiplying the two sides of Eq. (33) by $\mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$ gives

$$[\mathbf{u}_1\sigma_1, \mathbf{u}_2\sigma_2, \cdots, \mathbf{u}_r\sigma_r] = \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{bmatrix} [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r]. \tag{34}$$

From Eqs. (34) and (22) we have $\sigma_i \mathbf{u}_{ji} = \mathbf{x}_j^{\mathrm{T}} \mathbf{v}_i = \mathbf{v}_i^{\mathrm{T}} \mathbf{x}_j = \alpha_{ij}$. Using Eq. (21) yields

$$[\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \cdots, \hat{\mathbf{x}}_n] = \left[\sum_{i=1}^{l} \sigma_i \mathbf{u}_{1i} \mathbf{v}_i, \sum_{i=1}^{l} \sigma_i \mathbf{u}_{2i} \mathbf{v}_i, \ldots, \sum_{i=1}^{l} \sigma_i \mathbf{u}_{ni} \mathbf{v}_i \right] = \sum_{i=1}^{l} \sigma_i \mathbf{v}_i \mathbf{u}_i^{\mathrm{T}}. \tag{35}$$

Transposing Eq. (35) gives Eq. (31).

Summing up the above proof, we have

$$\|\mathbf{X}^{\mathrm{T}} - \tilde{\mathbf{X}}^{\mathrm{T}}\|_F^2 \geqslant \min_{rank(\mathbf{B})=l} \|\mathbf{X}^{\mathrm{T}} - \mathbf{B}\|_F^2 = \|\mathbf{X}^{\mathrm{T}} - \hat{\mathbf{X}}^{\mathrm{T}}\|_F^2 = \sum_{i=l+1}^{r} \sigma_i^2, \tag{36}$$

that is

$$\varepsilon^2(\Phi_l) \geqslant \varepsilon^2(\mathbf{V}_l).$$

This completes the proof of Theorem 4. □

Using a different approach from that used in Ref. [1] we have made the proof on that the proper orthogonal basis vectors of the SVD satisfy the optimality of the POD. Combining this with the relevant results in Ref. [1], it follows that the matrix used to solve the eigenvalue problem is $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ in the SVD and the correlation matrix used to solve the eigenvalue problem is $\mathbf{R}_x$ in the PCA (KLD). In general, $\mathbf{R}_x$ cannot be determined exactly but we can use the sample matrix to form its approximation $(1/n)\, \mathbf{X}\mathbf{X}^{\mathrm{T}}$. When the number n of the samples increases we have $\mathbf{R}_x = \lim_{n\to\infty}(1/n)\mathbf{X}\mathbf{X}^{\mathrm{T}}$. Note that both the matrices $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ and $(1/n)\, \mathbf{X}\mathbf{X}^{\mathrm{T}}$ possess the same eigenvalues and eigenvectors. Hence the SVD and PCA (KLD) possess the asymptotic connection. Since the squares of the singular values are the eigenvalues of the original eigenproblem, it can be seen from the above proof on the optimality of the SVD that the errors are also identical when both the SVD and PCA (KLD) take the first l proper orthogonal bases to approximately reconstruct the original samples. Therefore, the SVD and PCA (KLD) possess the equivalence.

4. Conclusion

In this paper, we discussed two approaches in the study of equivalence on the PCA, KLD, and SVD. Firstly, proceeding from finitely discretizing the continuous-time variables we prove the equivalence of the KLD and PCA. Secondly, using the Frobenius norm of a matrix we present a novel proof showing that the proper orthogonal basis vectors of the SVD satisfy the optimality of the POD and as well as demonstrating the equivalence of the SVD and PCA (KLD). These provide different approaches to study the equivalence of the three POD methods.

Acknowledgements

The first two authors are grateful to the support of the NSFC.

References

[1] Y.C. Liang, H.P. Lee, S.P. Lim, W.Z. Lin, K.H. Lee, C.G. Wu, Proper orthogonal decomposition and its application—Part I: theory, Journal of Sound and Vibration 252 (2002) 527–544.

[2] G. Kerschen, J.C. Golinval, Physical interpretation of the proper orthogonal modes using the singular value decomposition, Journal of Sound and Vibration 249 (2002) 849–865.

[3] W.Z. Lin, K.H. Lee, P. Lu, S.P. Lim, Y.C. Liang, The relationship between eigenfunctions of Karhunen–Loève decomposition and the modes of distributed parameter vibration system, Journal of Sound and Vibration 256 (2002) 791–799.

[4] B.F. Feeny, R. Kappagantu, On the physical interpretation of proper orthogonal modes in vibrations, Journal of Sound and Vibration 211 (1998) 607–616.

[5] B.F. Feeny, On proper orthogonal co-ordinates as indicators of modal activity, Journal of Sound and Vibration 255 (2002) 805–817.

[6] M.F.A. Azeez, A.F. Vakakis, Proper orthogonal decomposition (POD) of a class of vibroimpact oscillations, Journal of Sound and Vibration 240 (2001) 859–889.

[7] R. Ruotolo, C. Surace, Using SVD to detect damage in structures with different operational conditions, Journal of Sound and Vibration 226 (1999) 425–439.

[8] Y.C. Liang, W.Z. Lin, H.P. Lee, S.P. Lim, K.H. Lee, D.P. Feng, A neural-network-based method of model reduction for dynamic simulation of MEMS, Journal of Micromechanics and Microengineering 11 (2001) 226–233.

[9] Y.C. Liang, W.Z. Lin, H.P. Lee, S.P. Lim, K.H. Lee, H. Sun, Proper orthogonal decomposition and its application —Part II: model reduction for MEMS dynamical analysis, Journal of Sound and Vibration 256 (2002) 515–532.

[10] C. Zang, M. Imregun, Structural damage detection using artificial neural networks and measured FRF data reduced via principal component protection, Journal of Sound and Vibration 242 (2001) 813–827.

[11] A.S. Al-Dmour, K.S. Mohammad, Active control of flexible structures using principal component analysis in the time domain, Journal of Sound and Vibration 253 (2002) 545–569.

[12] K.I. Diamantaras, S.Y. Kung, Principal Component Neural Networks: Theory and Applications, Wiley, New York, 1996.

[13] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice-Hall, Englewood Cliffs, NJ, 1999.