Journal of Sound and Vibration (2002) **252**(3), 527–544
doi:10.1006/jsvi.2001.4041, available online at http://www.idealibrary.com on IDEAL®

PROPER ORTHOGONAL DECOMPOSITION AND ITS APPLICATIONS—PART I: THEORY

Y. C. LIANG

Department of Computer Science, Jilin University, 10 *Qian Wei Road*, *Changchun*, 130012, *People's Republic of China. E-mail*: *ycliang@public.cc.jl.cm*

H. P. LEE AND S. P. LIM

Centre for Advanced Computations in Engineering Science (*ACES*), *c/o Department of Mechanical Engineering*, *National University of Singapore*, 10 *Kent Ridge Crescent*, *Singapore* 119260

W. Z. LIN AND K. H. LEE

Institute of High Performance Computing, 89*C Science Park Drive*, #02-11/12, *the Rutherford*, *Singapore* 118261

AND

C. G. WU

Department of Mathematics, *Jilin University*, 10 *Qian Wei Road*, *Changchun*, 130012, *People's Republic of China*

(*Received* 9 *January* 2001, *and in final form* 26 *June* 2001)

In view of the increasing popularity of the application of proper orthogonal decomposition (POD) methods in engineering fields and the loose description of connections among the POD methods, the purpose of this paper is to give a summary of the POD methods and to show the connections among these methods. Firstly, the derivation and the performance of the three POD methods: Karhunen–Loève decomposition (KLD), principal component analysis (PCA), and singular value decomposition (SVD) are summarized, then the equivalence problem is discussed via a theoretical comparison among the three methods. The equivalence of the matrices for processing, the objective functions, the optimal basis vectors, the mean-square errors, and the asymptotic connections of the three methods are demonstrated and proved when the methods are used to handle the POD of discrete random vectors.

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. The POD provides a basis for the modal decomposition of an ensemble of functions, such as data obtained in the course of experiments or numerical simulations. Its properties suggest that it is the preferred basis to use in various applications. The basis functions it yields are commonly called empirical eigenfunctions, empirical basis functions, empirical orthogonal functions, proper orthogonal modes, or basis vectors. The most striking feature of the POD is its optimality: it provides the most efficient way of capturing the dominant components of an infinite-dimensional process with only a finite number of "modes", and often surprisingly few "modes" [1, 2]. In general, there are two

0022-460X/02/$35.00

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) [1–7]. 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.

In recent years, there have been many reported applications of the POD methods in engineering fields. The POD has been used in various disciplines including random variables, image processing, signal analysis, data compression, process identification and control in chemical engineering, oceanography, etc. [2]. In the bulk of these applications, the POD is used to analyze experimental data with the objective of extracting dominant features. The POD has been used to obtain approximate, low-dimensional descriptions of turbulent fluid flows [2], structural vibrations and chaotic dynamical systems [3–15], and more recently, microelectromechanical systems (MEMS) [16–18]. With the widespread applications of the POD methods, it is found that the loose description of the connection of the POD methods may confuse the researchers. Therefore, it is necessary to give a better picture of the POD methods. In view of this consideration, a summary of the equivalence of the three POD methods is made and some mathematical derivations involving them are performed in this paper.

2. THE THREE POD METHODS: PCA, KLD AND SVD

The POD was developed by several people. Lumley [19] traced the idea of the POD back to independent investigations by Kosambi (1943), Loève (1945), Karhunen (1946), Pougachev (1953) and Obukhov (1954). From the viewpoint of physical applications, only the discrete version of the POD is investigated in this paper. The three POD methods: PCA, KLD and SVD are discussed and the equivalence among them is proved and summarized in this section.

The main idea of the POD is to find a set of ordered orthonormal basis vectors in a subspace (without loss of generality, denoting the subspace as R^m) where a random vector takes its values, such that the samples in the sample space can be expressed optimally using the selected first l basis vectors. The mean square error can be used as a measure for the optimal problem, i.e.,

$$E\{\|x - x(l)\|^2\} \leqslant E\{\|x - \hat{x}(l)\|^2\}, \tag{1}$$

where $x(l)$ is the approximate expression of a random vector x using the first l basis vectors of the undetermined set of orthonormal basis vectors, and $\hat{x}(l)$ is the approximate expression of x using arbitrary l basis vectors in R^m.

The problem can be stated as follows.

Assume that $x \in R^m$ is a random vector and $\{\phi_i\}_{i=1}^m$ is a set of arbitrary orthonormal basis vectors: then x can be expressed as

$$x = \sum_{i=1}^{m} y_i \phi_i = \Phi y, \tag{2}$$

where

$$y_i = \phi_i^{\mathrm{T}} x \quad (i = 1, 2, \ldots, m),$$

$$y = (y_1, y_2, \ldots, y_m)^{\mathrm{T}}, \qquad \Phi = [\phi_1, \phi_2, \ldots, \phi_m].$$

The objective of the POD is to find a set of basis vectors that satisfies the following extreme value problem:

$$\min_{\phi_i} \varepsilon^2(l) = E\{\|x - x(l)\|^2\}$$

$$\text{s.t.} \quad \phi_i^{\mathrm{T}}\phi_j = \delta_{ij} \quad i,j = 1,2,\ldots,m, \tag{3}$$

where $x(l) = \sum_{i=1}^{l} y_i\phi_i (l \leqslant m)$. In order to obtain the same form of expressions for the mean-square errors by using the three different POD methods, the centralization on the processing data is assumed, i.e., the expectation of the random vector x is zero.

The three POD methods are introduced in the following three sections respectively.

2.1. THE PRINCIPAL COMPONENT ANALYSIS (PCA)

The PCA is a statistical technique and the idea behind the PCA is quite old. The earliest descriptions of the technique were given by Pearson (1901) and Hotelling (1933) [20]. The purpose of the PCA is to identify the dependence structure behind a multivariate stochastic observation in order to obtain a compact description of it. The PCA can be seen equivalently as either a variance maximization technique or a least-mean-squares technique.

The central idea of the PCA is to reduce the dimensionality of a data set which consists of a large number of interrelated variables, while retaining as much as possible the variation present in the data set. This is achieved by transforming the original variables to a new set of variables, the principal components, which are uncorrelated and are ordered so that the first few retain most of the variation present in all of the original variables.

There exist different versions on the description of the PCA [20–22]. In order to enable the style of the performance of the three POD approaches to be consistent, the method of realizing the POD based on the PCA is given as follows.

Suppose that $x \in R^m$ is a random vector, and $y_1, y_2, \ldots, y_m \in R$ are the 1st, 2nd, …, mth principal components respectively. In terms of the requirement of the PCA, let the first principal component y_1 be a linear combination of each element of the original random vector, i.e.,

$$y_1 = \sum_{i=1}^{m} \alpha_{i1} x_i = \alpha_1^{\mathrm{T}} x, \tag{4}$$

where $\alpha_1 = (\alpha_{11}, \alpha_{21}, \ldots, \alpha_{m1})^{\mathrm{T}}$ is a constant vector. The variance of y_1 is

$$\begin{aligned} s_{y_1}^2 &= V(y_1) = E\{(y_1 - E\{y_1\})^2\} = E\{(\alpha_1^{\mathrm{T}}x - E\{\alpha_1^{\mathrm{T}}x\})(\alpha_1^{\mathrm{T}}x - E\{\alpha_1^{\mathrm{T}}x\})^{\mathrm{T}}\} \\ &= \alpha_1^{\mathrm{T}} E\{(x - E\{x\})(x - E\{x\})^{\mathrm{T}}\}\alpha_1 . \end{aligned} \tag{5}$$

Let

$$\sum_x = E\{(x - E\{x\})(x - E\{x\})^{\mathrm{T}}\}, \tag{6}$$

where $\sum_x$ is the $m \times m$ covariance matrix corresponding to the random vector x and $E\{x\}$ is the expectation of x. From the knowledge of linear algebra, $\sum_x \in R^{m\times m}$ is a semi-definite matrix [23]. Let $\alpha_1^0 = \alpha_1/\|\alpha_1\|$, i.e., $\alpha_1^{0\mathrm{T}}\alpha_1^0 = 1$. Thus $s_{y_1}^2 = \|\alpha_1\|^2\alpha_1^{0\mathrm{T}}\sum_x\alpha_1^0$. It is apparent that the maximum of $s_{y_1}^2$ will not be achieved for a finite α_1, so a normalization constraint must

be imposed. The most convenient constraint is $\alpha_1^{\mathrm{T}}\alpha_1 = 1$. The problem of finding the first principal component is transformed to a conditional extreme value problem:

$$\begin{aligned} \max_{\alpha_1} \quad & s_{y_1}^2 = \alpha_1^{\mathrm{T}} \textstyle\sum_x \alpha_1 \\ \text{s.t.} \quad & \alpha_1^{\mathrm{T}}\alpha_1 = 1 \end{aligned} \tag{7}$$

Introducing the Lagrangian multiplier λ_1 gives

$$L(\alpha_1, \lambda_1) = \alpha_1^{\mathrm{T}} \textstyle\sum_x \alpha_1 + \lambda_1(1 - \alpha_1^{\mathrm{T}}\alpha_1).$$

Differentiating with respect to α_1 yields

$$\frac{\partial L(\alpha_1, \lambda_1)}{\partial \alpha_1} = 2(\textstyle\sum_x - \lambda_1 I)\,\alpha_1.$$

Letting the right-hand side of the above equation be zero, we have

$$\textstyle\sum_x \alpha_1 = \lambda_1 \alpha_1. \tag{8}$$

It can be seen that the solutions λ_1 and α_1 of the extreme value problem are the eigenvalue and the corresponding eigenvector of the covariance matrix $\sum_x$, respectively. Note that $s_{y_1}^2 = \alpha_1^{\mathrm{T}} \sum_x \alpha_1 = \lambda_1$, so λ_1 must be as large as possible. Thus λ_1 must be selected as the maximum eigenvalue of $\sum_x$.

Now let us find the second principal component. Let

$$y_2 = \sum_{i=1}^{m} \alpha_{i2} x_i = \alpha_2^{\mathrm{T}} x, \tag{9}$$

where $\alpha_2 = (\alpha_{12}, \alpha_{22}, \ldots, \alpha_{m2})^{\mathrm{T}}$. The variance of y_2 is

$$\begin{aligned} s_{y_2}^2 &= V(y_2) = E\{(y_2 - E\{y_2\})^2\} \\ &= E\{(\alpha_2^{\mathrm{T}} x - E\{\alpha_2^{\mathrm{T}} x\})(\alpha_2^{\mathrm{T}} x - E\{\alpha_2^{\mathrm{T}} x\})^{\mathrm{T}}\} \\ &= \alpha_2^{\mathrm{T}} E\{(x - E\{x\})(x - E\{x\})^{\mathrm{T}}\}\,\alpha_2. \end{aligned} \tag{10}$$

To find the α_2 which enables the maximum $s_{y_2}^2$ to be attained, a normalization constraint $\alpha_2^{\mathrm{T}}\alpha_2 = 1$ is necessary. The second principal component y_2 must be uncorrelated with the first principal component y_1, thus

$$0 = \operatorname{cov}(y_1, y_2) = E\{(\alpha_1^{\mathrm{T}} x - E\{\alpha_1^{\mathrm{T}} x\})(\alpha_2^{\mathrm{T}} x - E\{\alpha_2^{\mathrm{T}} x\})^{\mathrm{T}}\} = \alpha_1^{\mathrm{T}} \textstyle\sum_x \alpha_2. \tag{11}$$

Using equation (11) and the symmetry of $\sum_x$, we have $\alpha_2^{\mathrm{T}} \sum_x \alpha_1 = 0$. Note that α_1 is an eigenvector of $\sum_x$, thus $\lambda_1 \alpha_2^{\mathrm{T}} \alpha_1 = 0$.

If $\lambda_1 = 0$, because $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m \geqslant 0$, then $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$, i.e., all the eigenvalues are the same. Note that $\sum_x$ is a real symmetry matrix, therefore, there exists an orthogonal matrix $P \in R^{m \times m}$ [23], such that

$$P^{\mathrm{T}} \textstyle\sum_x P = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{pmatrix} = 0_{m \times m}. \tag{12}$$

Premultiplying equation (12) by P and postmultiplying the result by P^{T} give

$$\sum_x = 0_{m\times m}. \tag{13}$$

Thus,

$$\operatorname{cov}(x_i, x_j) = 0 \quad (i,j = 1,2,\ldots,m), \tag{14}$$

especially,

$$\operatorname{cov}(x_i, x_i) = V(x_i) = E\{(x_i - E\{x_i\})^2\} = 0. \tag{15}$$

It means that the value of each random variable $x_i (i = 1,2,\ldots,m)$ is centralized at its expectation, so it can be considered as a constant but not a random variable. The values of x_i $(i = 1,2,\ldots,m)$ can be replaced completely by their expectations.

If $\lambda_1 > 0$, then $\alpha_2^{\mathrm{T}}\alpha_1 = 0$ must hold, i.e., α_2 is orthogonal to α_1. Thus, the problem of finding the second component can be transformed into the following extreme value problem:

$$\begin{aligned} \max_{\alpha_2} \quad & s_{y_2}^2 = \alpha_2^{\mathrm{T}}\textstyle\sum_x \alpha_2 \\ \text{s.t.} \quad & \alpha_2^{\mathrm{T}}\alpha_2 = 1, \\ & \alpha_2^{\mathrm{T}}\alpha_1 = 0. \end{aligned} \tag{16}$$

To solve the conditional extreme value problem, the Lagrangian multipliers λ_2 and u are introduced and the Lagrangian function is written as

$$L(\alpha_2, \lambda_2, u) = \alpha_2^{\mathrm{T}}\textstyle\sum_x \alpha_2 + \lambda_2(1 - \alpha_2^{\mathrm{T}}\alpha_2) + u\alpha_2^{\mathrm{T}}\alpha_1.$$

Differentiation with respect to α_2 gives

$$\frac{\partial}{\partial \alpha_2} L(\alpha_2, \lambda_2, u) = 2(\textstyle\sum_x - \lambda_2 I)\alpha_2 + u\alpha_1.$$

Let the right-side of the above equation be zero, i.e.,

$$2(\textstyle\sum_x - \lambda_2 I)\alpha_2 + u\alpha_1 = 0. \tag{17}$$

Multiplying the two sides of equation (17) by α_1^{T} gives

$$2\alpha_1^{\mathrm{T}}(\textstyle\sum_x - \lambda_2 I)\alpha_2 + u = 0,$$

i.e.,

$$2\alpha_1^{\mathrm{T}}\textstyle\sum_x \alpha_2 + u = 0.$$

Because of the symmetry of $\sum_x$ and the fact that α_1 is an eigenvector of $\sum_x$, we have

$$2\lambda_1\alpha_2^{\mathrm{T}}\alpha_1 + u = 0,$$

thus $u = 0$. From equation (17) it follows that

$$\textstyle\sum_x \alpha_2 = \lambda_2\alpha_2. \tag{18}$$

Once again, α_2 is an eigenvector of $\sum_x$. Owing to the same reason that $s_{y_2}^2 = \alpha_2^{\mathrm{T}} \sum_x \alpha_2 = \lambda_2$, in order to reach the maximum variation of y_2, we can only take α_2 to be the eigenvector corresponding to the second eigenvalue of $\sum_x$. Then the variance of y_2 is the second eigenvalue of $\sum_x$.

The remaining principal components can be found in a similar manner. In general, the ith principal component of x is $y_i = \alpha_i^{\mathrm{T}} x$ and $s_{y_i}^2 = V(y_i) = \lambda_i$, where λ_i is the ith largest eigenvalue of $\sum_x$, and α_i is the corresponding eigenvector. As stated above, it can be shown that for the third, the fourth, ..., and the lth principal components, the vectors of coefficients $\alpha_3, \alpha_4, \ldots, \alpha_l$ are the eigenvectors of $\sum_x$ corresponding to $\lambda_3, \lambda_4, \ldots, \lambda_l$, the third, the fourth, ..., and the lth largest eigenvalues respectively.

To sum up, the objective function for finding the optimal basis vectors in the PCA is equivalent to

$$\begin{aligned} \max_{\alpha_i} \quad & \sum_{i=1}^{l} s_{y_i}^2 = \sum_{i=1}^{l} \alpha_i^{\mathrm{T}} \sum_x \alpha_i \\ \text{s.t.} \quad & \alpha_i^{\mathrm{T}} \alpha_j = \delta_{ij}. \end{aligned} \tag{19}$$

Then when the first l principal components are used to approximate the original random vector, the mean-square error is

$$\varepsilon^2(l) = E\{\|x - x(l)\|^2\} = E\left\{\left\| \sum_{i=l+1}^{m} y_i \alpha_i \right\|^2\right\} = \sum_{i=l+1}^{\mathrm{m}} E\{y_i^2\}, \tag{20}$$

where $x = \sum_{i=1}^{m} y_i \alpha_i$, $x(l) = \sum_{i=1}^{l} y_i \alpha_i$. Note that $E\{y_i\} = E\{\alpha_i^T x\} = \alpha_i^{\mathrm{T}} E\{x\} = 0$, therefore,

$$E\{y_i^2\} = E\{(y_i - E\{y_i\})^2\} = s_{y_i}^2. \tag{21}$$

Then the mean-square error is

$$\varepsilon^2(l) = \sum_{i=l+1}^{\mathrm{m}} s_{y_i}^2 = \sum_{i=l+1}^{m} \lambda_i. \tag{22}$$

In fact, the original random variables can be expressed exactly by all principal components. Suppose that all of the principal components y_i $(i = 1, 2, \ldots, m)$ are found, i.e., we have

$$y_i = \alpha_i^{\mathrm{T}} x \ (i = 1, 2, \ldots, m). \tag{23}$$

Premultiplying equation (23) on the two sides by α_i gives $y_i \alpha_i = \alpha_i \alpha_i^{\mathrm{T}} x$ $(i = 1, 2, \ldots, m)$. Summation of the equation on the two sides from 1 to m yields

$$\left(\sum_{i=1}^{m} \alpha_i \alpha_i^{\mathrm{T}} \right) x = \sum_{i=1}^{m} y_i \alpha_i,$$

where $\alpha_i \alpha_i^{\mathrm{T}}$ is an $m \times m$ matrix. Denoting that

$$B^{(k)} = \alpha_k \alpha_k^{\mathrm{T}} (k = 1, 2, \ldots, m)$$

the element of $B^{(k)}$ is $b_{ij}^{(k)} = \alpha_{ik} \alpha_{jk}$. Let $B = \sum_{k=1}^{m} B^{(k)}$, then the element of B is $b_{ij} = \sum_{k=1}^{m} \alpha_{ik} \alpha_{jk} = \delta_{ij}$, thus $B = I$. In fact, from

$$I = (\alpha_1, \alpha_2, \ldots, \alpha_m)(\alpha_1, \alpha_2, \ldots, \alpha_m)^{\mathrm{T}} = \sum_{k=1}^{m} \alpha_k \alpha_k^{\mathrm{T}},$$

it follows that $b_{ij} = \sum_{k=1}^{m} \alpha_{ik}\alpha_{jk} = \delta_{ij}$, thus

$$x = \sum_{i=1}^{m} y_i \alpha_i, \tag{24}$$

where $\alpha_i\,(i = 1, 2, \ldots, m)$ are the eigenvectors of $\sum_x$ corresponding to the eigenvalues of $\sum_x$ in descending order.

Now, the proper orthogonal decomposition of the sampled vector is completed using the PCA. The orthonormal basis vectors are found and the mean-square error of the approximate expression for the original random data is given.

2.2. THE KARHUNEN–LOÈVE DECOMPOSITION (KLD)

During the 1940s, Karhunen and Loève independently developed a theory regarding optimal series expansions of continuous-time stochastic processes [22, 24]. Their results extend the PCA to the case of infinite-dimensional spaces, such as the space of continuous-time functions. The KLD analysis uses single-parameter functions instead of vectors, two-parameter functions for representing autocorrelation instead of matrices. The KLD can be easily extended for discrete-time processes. In terms of optimality, the partial KLD has the same optimal properties of least-squares reconstruction and variance maximization as the PCA.

The discrete KLD is stated as follows [24]. Let $x \in R^m$ be a random vector, and $\{\phi_i\}_{i=1}^{m}$ be a set of orthonormal basis vectors in R^m, then there exist $y_i = \phi_i^{\mathrm{T}} x$such that

$$x = \sum_{i=1}^{m} y_i \phi_i = \Phi y. \tag{25}$$

Let

$$x(l) = \sum_{i=1}^{l} y_i \phi_i + \sum_{i=l+1}^{m} b_i \phi_i \,(l \leqslant m), \tag{26}$$

where $b_i\,(i = l+1, \ldots, m)$ are constants. It can be easily verified that $b_i = 0\,(i = l+1, \ldots, m)$ after the centralization to the samples, i.e., after the processing on the random vector x such that $E\{x\} = 0$. Let $\Delta x(l) = x - x(l) = \sum_{i=l+1}^{m} (y_i - b_i)\phi_i$, where x and $x(l)$ are random vectors, thus $\Delta x(l)$ is also a random vector. In order to examine the quality of the expression of x we choose the mean square error as a measure, i.e.,

$$\varepsilon^2(l) = E\{\|\Delta x(l)\|^2\} = E\left\{\left\|\sum_{i=l+1}^{m} (y_i - b_i)\phi_i\right\|^2\right\} = E\left\{\sum_{i=l+1}^{m} (y_i - b_i)^2\right\}. \tag{27}$$

To enable $\varepsilon^2(l)$ to be the minimum, the derivative of $\varepsilon^2(l)$ with respect to $b_i\,(i = l+1, l+2, \ldots, m)$ is calculated, which yields

$$\frac{\partial}{\partial b_i}\varepsilon^2(l) = -2E\{y_i - b_i\}.$$

Letting the right-hand side of the above equation be zero, we have

$$b_i = E\{y_i\}\,(i = l+1, l+2, \ldots, m). \tag{28}$$

It can be seen that $b_i = 0$ after the centralization to the samples, then $x(l) = \sum_{i=1}^{l} y_i \phi_i$ ($l \leqslant m$) is the required form of the POD. To keep the generalization for the derivation, substituting equation (28) into equation (27) gives

$$\varepsilon^2(l) = \sum_{i=l+1}^{m} E\{y_i - E\{y_i\}\}^2 = \sum_{i=1+1}^{m} \phi_i^{\mathrm{T}} E\{(x - E\{x\})(x - E\{x\})^{\mathrm{T}}\} \phi_i$$

$$= \sum_{i=l+1}^{m} \phi_i^{\mathrm{T}} \sum_x \phi_i = tr(\Phi_{m-l}^{\mathrm{T}} \sum_x \Phi_{m-l}), \tag{29}$$

where $\sum_x = E\{(x - E\{x\})(x - E\{x\})^{\mathrm{T}}\}$ is the covariance matrix of x and

$$\Phi_{m-l} = [\phi_{l+1}, \phi_{l+2}, \ldots, \phi_m] \in R^{m(m-l)}.$$

Then the KLD problem is transformed as a conditional extreme value problem:

$$\begin{aligned} \min_{\phi_i} \quad & \varepsilon^2(l) = \sum_{i=l+1}^{m} \phi_i^{\mathrm{T}} \sum_x \phi_i \\ \text{s.t.} \quad & \phi_i^{\mathrm{T}} \phi_j = \delta_{ij} \end{aligned} \tag{30}$$

where $i, j = l+1, l+2, \ldots, m$. Introducing Lagrangian multipliers u_{ij} ($i, j = l+1, l+2, \ldots, m$) gives

$$L = \sum_{i=l+1}^{m} \phi_i^{\mathrm{T}} \sum_x \phi_i - \sum_{i=l+1}^{m} \sum_{j=l+1}^{m} u_{ij}(\phi_i^{\mathrm{T}} \phi_j - \delta_{ij}).$$

Differentiation with respect to ϕ_i on the two sides of the above equation yields

$$\frac{\partial L}{\partial \phi_i} = 2(\sum_x \phi_i - \Phi_{m-l} u_i),$$

where $u_i = (u_{l+1\,i}, u_{l+2\,i}, \ldots, u_{mi})^{\mathrm{T}}$ ($i = l+1, l+2, \ldots, m$). Writing the above equation in a matrix form gives

$$\frac{\partial L}{\partial \Phi_{m-l}} = 2(\sum_x \Phi_{m-l} - \Phi_{m-l} U_{m-l}),$$

where $U_{m-l} = (u_{l+1}, u_{l+2}, \ldots, u_m)$. Letting the right-hand side of the above equation be zero gives

$$\sum_x \Phi_{m-l} = \Phi_{m-l} U_{m-l}. \tag{31}$$

It can be seen that all the orthonormal basis vectors satisfying equation (30) must satisfy equation (31), where there are no special constraints to Φ_{m-l} and U_{m-l}. Next, let us prove that all Φ_{m-l} satisfying equation (31) can be formed by the eigenvectors of $\sum_x$, and U_{m-l} is the diagonal matrix that consists of the corresponding eigenvalues of $\sum_x$.

The above conclusion is proved as follows.

Multiplying equation (31) by Φ_{m-l}^{T} yields

$$U_{m-l} = \Phi_{m-l}^{\mathrm{T}} \sum_x \Phi_{m-l}. \tag{32}$$

Note that $y_i = \phi_i^{\mathrm{T}} x$, so U_{m-l} in equation (32) is the covariance matrix of the vector formed by the last $m - l$ elements of the random vector y after the transformation $y = \Phi^{\mathrm{T}} x$. Thus U_{m-l} is a semi-definite matrix with dimensions of $(m - l) \times (m - l)$. Let the diagonal matrix formed by the eigenvalues of U_{m-l} be Λ_{m-l}, and the square matrix formed by the corresponding eigenvectors Ψ_{m-l}. Performing the transformation $z = \Psi_{m-l}^{\mathrm{T}} y$ gives

$$\Lambda_{m-l} = \Psi_{m-l}^{\mathrm{T}} U_{m-l} \Psi_{m-l}. \tag{33}$$

Substituting equation (32) into equation (33) yields

$$\Lambda_{m-l} = (\Phi_{m-l} \Psi_{m-l})^{\mathrm{T}} \Psi_x (\Phi_{m-l} \Psi_{m-l}). \tag{34}$$

It can be seen that the diagonal elements of Λ_{m-l} are the $m - l$ eigenvalues of $\sum_x$, and the eigenvectors corresponding to the eigenvalues form $(\Phi_{m-l} \Psi_{m-l})_{m \times (m-l)}$. Denote the eigenvector matrix by Φ^*_{m-l}, thus

$$\Phi_{m-l} = (\Phi^*_{m-l} \Psi^{\mathrm{T}}_{m-l})_{m \times (m-l)}.$$

Then the mean-square error is

$$\begin{aligned} \varepsilon^2(l) &= \mathrm{tr}(\Phi^{\mathrm{T}}_{m-l} \textstyle\sum_x \Phi_{m-l}) = \mathrm{tr}(\Psi_{m-l} \Phi^{*\mathrm{T}}_{m-l} \textstyle\sum_x \Phi^*_{m-l} \Phi^{\mathrm{T}}_{m-l}) \\ &= \mathrm{tr}(\Phi^{*\mathrm{T}}_{m-l} \textstyle\sum_x \Phi^*_{m-l} \Psi^{\mathrm{T}}_{m-l} \Psi_{m-l}) = \mathrm{tr}(\Phi^{*\mathrm{T}}_{m-l} \textstyle\sum_x \Phi^*_{m-l}) \\ &= \sum_{s=1}^{m-l} \lambda_{k_s}, \end{aligned} \tag{35}$$

where $\lambda_{k_s} (s = 1, 2, \ldots, m - l)$ are the eigenvalues corresponding to the columns of Φ^*_{m-1}. Once x is mapped onto the $(m - l)$-dimensional subspace spanned by $m - l$ eigenvectors of $\sum_x$, further application of an orthonormal transformation would not change the mean-square error. Therefore, Φ_{m-l} and U_{m-l} in equation (31) can be chosen simply as the matrices formed by the eigenvectors and eigenvalues of $\sum_x$ respectively. Let the descending order of the eigenvalues of $\sum_x$ be $\lambda_1, \lambda_2, \ldots, \lambda_m$, and the corresponding eigenvectors be $\phi_1, \phi_2, \ldots, \phi_m$. It can be seen that in order to enable the minimum value problem to hold, the othonormal basis vectors can be selected as the eigenvectors of $\sum_x$, and the mean-square error to approximate x by using the first l basis vectors is $\varepsilon^2(l) = \sum_{i=l+1}^{m} \lambda_i$.

2.3. THE SINGULAR-VALUE DECOMPOSITION (SVD)

Klema and Laub [25] indicated that the SVD was established for real-square matrices in the 1870s by Beltrami and Jordan, for complex square matrices in 1902 by Autonne, and for general rectangular matrices in 1939 by Eckart and Young. The SVD can be viewed as the extension of the eigenvalue decomposition for the case of non-square matrices. As far as the proper orthogonal decomposition is concerned, the SVD can also be seen as an extension for non-symmetric matrices. Because the SVD is much more general than the eigenvalue decomposition and intimately relates to the matrix rank and reduced-rank least-squares approximation, it is a very important and fundamental working tool in many areas such as matrix theory, linear systems, statistics, and signal analysis [25–29].

The third method to realize the POD is the SVD, which uses the singular-value decomposition to find the basis vectors satisfying the POD requirement in the sample space. The process for realizing the POD by using the SVD is stated as follows. The basic concept

is the same as that which appeared in most references, such as [25–29], but we try to use statements which are easy to keep the description of the three POD methods consistent.

Suppose that n samples $x_1, x_2, \ldots, x_n$ are given where $x_i \in R^m$ $(i = 1, 2, \ldots n, n)$. Consider the samples to be more than enough such that $n > m$. Let

$$X = (x_1, x_2, \ldots, x_n),$$

then $X \in R^{m \times n}$, and $XX^{\mathrm{T}} \in R^{m \times m}$ is an $m \times m$ semi-definite matrix. Let the eigenvalues of XX^{T} be arranged in decreasing order as

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > \lambda_{r+1} = \cdots = \lambda_m = 0.$$

In the SVD of matrices, $\sigma_i = \sqrt{\lambda_i}\,(i = 1,2, \ldots, m)$ are called the singular values of X^{T}. Let the eigenvectors of XX^{T} with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ be $v_1, v_2, \ldots, v_m$.

Define $V = [V_1, V_2]$ where $V_1 = (v_1, v_2 \ldots, v_r)$, $V_2 = (v_{r+1}, v_{r+2} \ldots, v_m)$ and the subscript r is the index of the smallest positive eigenvalue of XX^{T}. Then the matrix V is an $m \times m$ orthonormal matrix and we have

$$XX^{\mathrm{T}}V = V \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{pmatrix} = V \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_m^2 \end{pmatrix}. \tag{36}$$

Premultiplying equation (36) by V^{T} gives

$$[V_1, V_2]^{\mathrm{T}} XX^{\mathrm{T}} [V_1, V_2] = \begin{pmatrix} \sum_r^2 & 0 \\ 0 & 0 \end{pmatrix}, \tag{37}$$

where

$$\sum\nolimits_r^2 = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_r^2).$$

Let

$$U_1 = X^{\mathrm{T}} V_1 \sum\nolimits_r^{-1},$$

where

$$\sum\nolimits_r^{-1} = \mathrm{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_r^{-1}),$$

we have

$$U_1^{\mathrm{T}} U_1 = (X^{\mathrm{T}} V_1 \sum\nolimits_r^{-1})^{\mathrm{T}} X^{\mathrm{T}} V_1 \sum\nolimits_r^{-1} = \sum\nolimits_r^{-1} \sum\nolimits_r^2 \sum\nolimits_r^{-1} = I_r. \tag{38}$$

From equation (38) it can be seen that the columns of the matrix U_1 are mutually orthogonal. Denoting

$$U_1 = (u_1, u_2, \ldots, u_r),$$

according to the basis extension theorem in vector space, there exist $n - r$ orthonormal vectors in R^n and they are orthogonal to the columns of U_1. Let the $n - r$ orthonormal vectors be $u_{r+1}, u_{r+2}, \ldots, u_n$. In the singular-value decomposition, $u_1, u_2, \ldots, u_m$ and $v_1, v_2, \ldots, v_m$ are called left and right singular vectors of X^{T} corresponding to eigenvalues

$\sigma_1, \sigma_2, \cdots, \sigma_m$ respectively. Let $U = [U_1, U_2]$ where $U_2 = (u_{r+1}, u_{r+2}, \ldots, u_n)$, then U is an $n \times n$ orthonormal matrix. Thus

$$U^{\mathrm{T}} X^{\mathrm{T}} V = [U_1, U_2]^{\mathrm{T}} X^{\mathrm{T}} [V_1, V_2] = \begin{pmatrix} U_1^{\mathrm{T}} X^{\mathrm{T}} V_1 & U_1^{\mathrm{T}} X^{\mathrm{T}} V_2 \\ U_2^{\mathrm{T}} X^{\mathrm{T}} V_1 & U_2^{\mathrm{T}} X^{\mathrm{T}} V_2 \end{pmatrix}, \tag{39}$$

where

$$U_1^{\mathrm{T}} X^{\mathrm{T}} V_1 = \sum\nolimits_r^{-1} V_1^{\mathrm{T}} X X^{\mathrm{T}} V_1 = \sum\nolimits_r,$$

$$U_1^{\mathrm{T}} X^{\mathrm{T}} V_2 = \sum\nolimits_r^{-1} V_1^{\mathrm{T}} X X^{\mathrm{T}} V_2 = 0,$$

$$U_2^{\mathrm{T}} X^{\mathrm{T}} V_1 = U_2^{\mathrm{T}} (X^{\mathrm{T}} V_1 \sum\nolimits_r^{-1}) \sum\nolimits_r = U_2^{\mathrm{T}} U_1 \sum\nolimits_r = 0.$$

Note that

$$X X^T V_2 = 0.$$

Premultiplying the above equation by V_2^{T} gives

$$V_2^{\mathrm{T}} X X^{\mathrm{T}} V_2 = (X^{\mathrm{T}} V_2)^{\mathrm{T}} X^{\mathrm{T}} V_2 = 0. \tag{40}$$

From equation (40) it follows that

$$\mathrm{tr}((X^{\mathrm{T}} V_2)^{\mathrm{T}} (X^{\mathrm{T}} V_2)) = 0,$$

thus

$$X^{\mathrm{T}} V_2 = 0.$$

From the above equation we have

$$U_2^{\mathrm{T}} X^{\mathrm{T}} V_2 = 0,$$

therefore equation (39) can be written as

$$U^{\mathrm{T}} X^{\mathrm{T}} V = [U_1, U_2]^{\mathrm{T}} X^{\mathrm{T}} [V_1, V_2] = \begin{pmatrix} U_1^{\mathrm{T}} X^{\mathrm{T}} V_1 & U_1^{\mathrm{T}} X^{\mathrm{T}} V_2 \\ U_2^{\mathrm{T}} X^{\mathrm{T}} V_1 & U_2^{\mathrm{T}} X^{\mathrm{T}} V_2 \end{pmatrix} = \begin{pmatrix} \sum_r & 0 \\ 0 & 0 \end{pmatrix}. \tag{41}$$

Premultiplying the above equation by U and postmultiplying the result by V give

$$X^T = U \begin{pmatrix} \sum_r & 0 \\ 0 & 0 \end{pmatrix} V^{\mathrm{T}}. \tag{42}$$

Transposing equation (42) yields

$$[x_1, x_2, \ldots, x_n] = V \begin{pmatrix} \sum_r & 0 \\ 0 & 0 \end{pmatrix} U^{\mathrm{T}}. \tag{43}$$

Denote the columns of the matrix $\begin{pmatrix} \sum_r & 0 \\ 0 & 0 \end{pmatrix} U^{\mathrm{T}}$ as $d_1, d_2, \ldots, d_n$. From equation (43) it follows that

$$x_i = V d_i, \quad (i = 1, 2, \ldots, n). \tag{44}$$

Now the proper orthogonal decomposition of the sampled vectors is completed. From equations (43) and (44) it can be seen that the components $d_{r+1,i}, d_{r+2,i}, \dots, d_{m,i}$ of d_i $(i = 1,2,\dots,n)$ are equal to zero when the singular values $\sigma_{r+1}, \sigma_{r+2}, \dots, \sigma_m$ of X^{T} equal zero. Then it needs only r right singular vectors as basis to represent the samples x_i $(i = 1,2,\dots,n)$ in space R^m.

Next, let us discuss the optimality of the set of proper orthogonal basis vectors.

Suppose that a set of orthonormal basis vectors $\varphi_1, \varphi_2, \dots, \varphi_n$ is chosen arbitrarily in R^m to represent n samples, then

$$x_i = c_{1i}\varphi_1 + c_{2i}\varphi_2 + \cdots + c_{mi}\varphi_m, \tag{45}$$

where $c_{ji} = \phi_j^{\mathrm{T}} x_i$. When the first l basis vectors are selected to approximate the samples, we have

$$x_i(l) = c_{1i}\varphi_1 + c_{2i}\varphi_2 + \cdots + c_{li}\varphi_l. \tag{46}$$

What we are concerned about is the error of all the samples but not the error of an individual sample. Therefore, the following error function is considered:

$$\begin{aligned}\varepsilon^2(l) &= \sum_{i=1}^{n} \|x_i - x_i(l)\|^2 = \sum_{i=1}^{n} \left\| \sum_{j=l+1}^{m} c_{ji}\varphi_j \right\|^2 = \sum_{i=1}^{n}\sum_{j=l+1}^{m} c_{ji}^2 \\ &= \sum_{i=1}^{n}\sum_{j=l+1}^{m} \varphi_j^{\mathrm{T}} x_i x_i^{\mathrm{T}} \varphi_j = \sum_{j=l+1}^{m} \varphi_j^{\mathrm{T}} X X^{\mathrm{T}} \varphi_j = \|X^{\mathrm{T}}\Phi_{m-1}\|_F^2,\end{aligned} \tag{47}$$

where $\Phi_{m-1} = (\varphi_{l+1}, \varphi_{l+2}, \dots, \varphi_m)$. Then the problem of finding the optimal basis vectors is transformed to the following extreme value problem:

$$\begin{aligned} &\min_{\phi_j} \quad \varepsilon^2(l) = \sum_{j=l+1}^{n} \varphi_j^{\mathrm{T}} X X^{\mathrm{T}} \varphi_j \\ &\text{s.t.} \quad \varphi_i^{\mathrm{T}} \varphi_j = \delta_{ij}. \end{aligned} \tag{48}$$

In order to solve the extreme value problem we introduce a Lagrangian multiplier u_{ij} and write the Lagrangian function

$$L = \sum_{j=l+1}^{m} \varphi_j^{\mathrm{T}} X X^{\mathrm{T}} \varphi_j - \sum_{i=l+1}^{m}\sum_{j=l+1}^{m} u_{ij}(\varphi_i^{\mathrm{T}}\varphi_j - \delta_{ij}).$$

Differentiation with respect to φ_j on the two sides of the above equation yields

$$\frac{\partial L}{\partial \varphi_j} = 2\left(X X^{\mathrm{T}} \varphi_j - u_{ij} \sum_{i=l+1}^{m} \varphi_i \right) = 2 X X^{\mathrm{T}} \varphi_j - 2\Phi_{m-1} u_j \quad (j = l+1, l+2, \dots, m)$$

where $u_j = (u_{l+1,j}, u_{l+2,j}, \dots, u_{mj})^{\mathrm{T}}$. The above equation can be written in a matrix form as

$$\frac{\partial L}{\partial \Phi_{m-1}} = 2 X X^{\mathrm{T}} \Phi_{m-1} - 2\Phi_{m-1} U_{m-1},$$

where $U_{m-1} = (u_{l+1}, u_{l+2}, \dots, u_m)$.

Equating $\partial L / \partial \Phi_{m-1}$ to zero gives

$$X X^{\mathrm{T}} \Phi_{m-1} = \Phi_{m-1} U_{m-1}. \tag{49}$$

Premultiplying the two sides of equation (49) by Φ_{m-l}^{T} yields

$$U_{m-l} = \Phi_{m-l}^{\mathrm{T}} X X^{\mathrm{T}} \Phi_{m-l}. \tag{50}$$

Note that U_{m-l} is a semi-definite matrix. Then there exists an orthogonal matrix P such that

$$\Lambda = P^{\mathrm{T}} U_{m-l} P = P^{\mathrm{T}} \Phi_{m-l}^{\mathrm{T}} X X^{\mathrm{T}} \Phi_{m-l} P, \tag{51}$$

where Λ is a diagonal matrix. Postmultiplying the two sides of equation (49) by P gives

$$X X^{\mathrm{T}} \Phi_{m-l} P = \Phi_{m-l} P P^{\mathrm{T}} U_{m-l} P,$$

i.e.,

$$X X^{\mathrm{T}} \Phi_{m-l} P = \Phi_{m-l} P \Lambda. \tag{52}$$

From the above equation it can be seen that the diagonal elements of the matrix Λ are some eigenvalues λ_i of the matrix XX^{T}, and the matrix $\Phi_{m-l}P$ consists of the X eigenvectors corresponding to λ_i. That is, the diagonal elements of the matrix Λ are the squares of some singular values σ_i of the matrix X^{T}, and $\Phi_{m-l}P$ consists of the right singular vectors corresponding to σ_i.

For the next use we state the following theorem without proof [30].

Theorem. *Let $A \in R^{n\times m}$, $Q \in R^{m\times m}$ be an orthogonal matrix, and $\|\cdot\|_F$ be the Frobenius norm,* then $\|A\|_F = \|AQ\|_F$.

From equation (47) and the above theorem, it follows that

$$\varepsilon^2(l) = \|X^{\mathrm{T}}\Phi_{m-l}\|_F^2 = \|X^{\mathrm{T}}\Phi_{m-l}P\|_F^2 = \mathrm{tr}((X^{\mathrm{T}}\Phi_{m-l}P)^{\mathrm{T}} X^{\mathrm{T}}\Phi_{m-l}P) = \mathrm{tr}(\Lambda). \tag{53}$$

Note that Λ is a diagonal matrix and its diagonal elements are the squares of the singular values σ_i of the matrix X, so in order to attain the minimum error, the diagonal elements of Λ can only be the last $m-l$ singular values of the matrix X^{T}. Thus,

$$\varepsilon^2(l) = \mathrm{tr}((X^{\mathrm{T}}\Phi_{m-l}P))^{\mathrm{T}} X^{\mathrm{T}}\Phi y_{m-l}P) = \mathrm{tr}(\Lambda) = \sum_{j=1+1}^{m} \sigma_j^2. \tag{54}$$

From this it is proved not only that the optimality is attained when the right singular vectors of X^{T} are taken as basis vectors but also that the minimum error is simply the square summation of the last $m-l$ singular values of the matrix X^{T}.

3. THE EQUIVALENCE OF THE THREE METHODS

From the above discussion it can be seen that there exist close connections among the three POD methods: PCA, KLD, and SVD, although their derivations are different. The existing equivalence connections among them are understood by researchers. Some of the equivalence relationships between the three methods are summarized as follows:

(1) Mees *et al.* [31] pointed out that the connection between the KLD and the PCA was first noticed by Watanabe in 1965. Diamantaras and Kung [22], and also Ravindra [4] indicated that the difference between the KLD and PCA was that KLD typically referred to stochastic processes, whereas the PCA referred to random vectors. If the

time parameter t is a discrete variable and one has a finite collection of random variables, then the KLD reduces to the PCA.

(2) Chatterjee [1] pointed out the correspondence with the expression of the SVD and that of a finite sum of the KLD.

(3) Kunisch and Volkwein [32] described the connection between the KLD and SVD within the context of its relevance to the application to optimal control problems.

(4) Diamantaras and Kung [22] pointed out that there was an asymptotic connection between the PCA and SVD.

It is very useful for researchers to understand the equivalence connections among the three methods so as to study and apply the POD methods extensively. However, to the best of the authors' knowledge, so far there exists neither complete demonstration nor systematic and theoretical proof on the equivalence of the three methods. In this section, the equivalence of the three methods is discussed from a different point of view, and some proofs on the equivalence of the three methods are presented. The aim of the present work is to demonstrate the close connections among the three methods. It should be pointed out that in practice, the applications of the three methods may not always be the same. If the methods are actually applied in the same way they may lead to exactly the same basis functions. If the methods are applied slightly differently, yet in equivalent ways, then the equivalence is more hidden [33].

The main results on the equivalence of the three methods obtained through the above comparison are summarized as follows:

1. *The equivalence of the PCA and KLD*

(1) The same matrices for processing. Both the PCA and KLD handle problems from random vectors. For a random vector x with dimension m, the matrices used for finding basis vectors derived from the two methods are the same. The matrix for processing is the $m \times m$ covariance matrix corresponding to the random vector, i.e.,

$$\sum_x = E\{(x - E\{x\})(x - E\{x\})^{\mathrm{T}}\}. \tag{55}$$

(2) The same objective functions for finding the optimal basis vectors. In the PCA the objective function for finding the optimal basis vectors is that the variance summation of the first l ($l \leqslant l < m$) principal components is maximal, i.e.,

$$\max_{\alpha_i} \sum_{i=1}^{l} s_{y_i}^2 = \sum_{i=1}^{l} \alpha_i^{\mathrm{T}} \sum_x \alpha_i. \tag{56}$$

Obviously, equation (56) is equivalent to the fact that the variance summation of the last $m - l$ principal components is minimal, i.e.,

$$\min_{\alpha_i} \sum_{i=l+1}^{m} s_{y_i}^2 = \sum_{i=l+1}^{m} \alpha_i^{\mathrm{T}} \sum_x \alpha_i. \tag{57}$$

In the KLD the objective function for finding the optimal basis vectors is that the error is minimal after removing the last $m - l$ basis vectors, i.e.,

$$\min_{\phi_i} \varepsilon^2(l) = \sum_{i=l+1}^{m} \phi_i^{\mathrm{T}} \sum_x \phi_i. \tag{58}$$

Making a comparison between equations (57) and (58) it can be seen that the objective functions for finding the optimal basis vectors by using the two methods possess the same form.

(3) The same or equivalent optimal basis vectors. The basis vectors found using the two methods are the eigenvectors of a covariance matrix corresponding to the same random vector. In fact, the covariance matrix $\sum_x$ here is an $m \times m$ linear transformation in a real field. Let the linear transformation be σ. Because σ is semi-definite, it has m non-negative real eigenvalues. Let the eigenvalues be $\lambda_1, \lambda_2, \ldots, \lambda_m$, and let them be arranged in decreasing order.

If all $\lambda_i (i = 1, 2, \ldots, m)$ are distinct, then each eigen-subspace of σ: $\sigma(\lambda_1), \sigma(\lambda_2), \ldots, \sigma(\lambda_m)$ has only one basis vector respectively. Let them be $v_1, v_2, \ldots, v_m$. The difference among the elements in each subspace $\sigma(\lambda_1), \sigma(\lambda_2), \ldots \sigma(\lambda_m)$ is only a constant factor, i.e., if $v_i^{(1)}$ and $v_i^{(2)}$ are the elements of the eigen-subspace $\sigma(\lambda_i)$ of σ, then there exist a real number α such that $v_i^{(1)} = \alpha v_i^{(2)}$. Because the basis vectors are required to be normal, the basis vectors belonging to the λ_i obtained by using the PCA and KLD must be completely the same.

If some eigenvalues are multiple, without loss of generality, we let λ_1 have multiplicity n_1. Then in the eigen-subspace $\sigma(\lambda_1)$ there exist n_1 orthonormal vectors that can be selected as basis vectors. The basis vector α_1 belonging to the λ_1 selected by using the PCA may not be the same as the basis vector ϕ_1 belonging to the λ_1 selected by using the KLD. But both of them are basis vectors of the eigen-subspace $\sigma(\lambda_1)$ of σ. If we select n_1 optimal basis vectors by using PCA and KLD, respectively, to approximate the original random vector x, they may be two different basis vectors of the eigen-subspace $\sigma(\lambda_1)$. However, they are equivalent obviously. In fact, they can be expressed mutually, i.e., there exist constants β_{ij} such that $\alpha_i = \sum_{j=1}^{n_1} \beta_{ij} \phi_j \ (i = 1, 2, \ldots, n_1)$. Because the n_1 basis vectors selected, respectively, satisfy the orthonormal condition, we need only to make an orthogonal transformation to enable the orthonormal basis vectors selected by using the PCA and KLD to be completely the same.

(4) The same approximate matrices processed in practical calculation. Because variables such as the probability and the expectation associated with the covariance matrix are not known *a priori*, the estimate of the covariance matrix is needed in order to obtain the approximate covariance matrix. In the PCA and KLD, the expression $(1/n)(X - \bar{X})(X - \bar{X})^{\mathrm{T}}$ is used as the approximation of $\sum_x$. In general, the data are centralized before the proper orthogonal decomposition is performed, i.e., $\bar{X} = 0$. Therefore in the two methods the approximate matrix processed in practical calculation is $(1/n) XX^{\mathrm{T}}$, where X is a matrix whose columns are formed by the given samples.

2. *The equivalence of the PCA (KLD) and SVD*

(1) The equivalence of the eigenvalue problems of the PCA (KLD) and SVD. From the above discussion it can be seen that the SVD is to perform the singular value decomposition to the transposed matrix X^{T} of the matrix X. The singular values obtained are the arithmetic square roots of the eigenvalues of the matrix XX^{T} and the right singular vectors selected as the basis vectors are the eigenvectors of the XX^{T}. Thus if we make the transformation $\tilde{X} = (1/\sqrt{n}) X$, performing the singular-value decomposition to the matrix $\tilde{X}^{\mathrm{T}}$ is equivalent to searching the eigenvalues and eigenvectors of the matrix $(1/n) XX^{\mathrm{T}}$. Note that both the matrices XX^{T} and $(1/n) XX^{\mathrm{T}}$ possess the same eigenvectors. Therefore, the basis vectors obtained using the SVD to the X^{T} are the same as those obtained using the PCA (KLD) to the $(1/n) XX^{\mathrm{T}}$.

(2) The asymptotic connection between the PCA (KLD) and SVD. The asymptotic connection between the PCA (KLD) and SVD can be obtained directly by using the eigenvalue problems of the SVD and PCA (KLD) and the asymptotic connection between the matrices $(1/n) XX^{\mathrm{T}}$ and $\sum_x$. Now let us prove the asymptotic connection between the matrices $(1/n) XX^{\mathrm{T}}$ and $\sum_x$.

Denote the elements of the covariance matrix $\sum_x$ as σ_{ij}. From equation (55), the definition of the covariance matrix, it follows that

$$\sigma_{ij} = \mathrm{cov}(x_i, x_j) = E\{(x_i - E\{x_i\})(x_j - E\{x_j\})\}. \tag{59}$$

Let the values of the ith component x_i of the random vector x be $x_i^1, x_i^2, \ldots$ $(i = 1, 2, \cdots, m)$, which represent some events of the component x_i. Let the expectation of the x_i be $E\{x_i\} = u_i$ $(i = 1, 2, \ldots, m)$ and the probability of the event $(x_i^p - u_i)(x_j^q - u_j)$ be P_{ij}^{pq}, then

$$\sigma_{ij} = \sum_{p,q} (x_i^p - u_i)(x_j^q - u_j) P_{ij}^{pq}. \tag{60}$$

Because in most cases the value, expectation u_i and probability P_{ij}^{pq} of a random variable are not known *a priori*, we can only obtain their approximate values using a large number of samples.

Assume that n samples of a random vector x are selected, which are $x^{(1)}, x^{(2)}, \ldots, x^{(n)} \in R^m$ $(n > m)$. Let

$$X = (x^{(1)}, x^{(2)}, \ldots, x^{(n)}) \in R^{m \times n}.$$

Firstly, let us count the number of events of x_i, i.e., the number of times that different values of the ith components of all n samples appear. Let the number be n_i $(i = 1, 2, \ldots, m)$ and the expectation of x_i be

$$u_i = \frac{1}{n} \sum_{j=1}^{n} x_i^{(j)} \quad (i = 1, 2, \ldots, m). \tag{61}$$

Then let us count the number of events of $(x_i^p - u_i)(x_j^q - u_j)$ $(i, j = 1, 2, \ldots, m, p = 1, 2, \ldots, n_i, q = 1, 2, \ldots, n_j)$ where x_i^p and x_j^q represent the p and q kinds of values for the ith and jth components of the random vector x respectively. Let the number of appearances of $(x_i^p - u_i)(x_j^q - u_j)$ in the n samples be n_{ij}^{pq}. Define the probability

$$\tilde{P}_{ij}^{pq} = n_{ij}^{pq}/n, \tag{62}$$

where $\sum_{p,q} n_{ij}^{pq} = n$ because both x_i^p and x_j^q are in the same sample merely appear in the ith and jth places respectively.

The larger the number n of the samples is, the closer the probability $\tilde{P}_{ij}^{pq}$ defined by equation (62) will tend to the true probability P_{ij}^{pq} of the random event $(x_i^p - u_i)(x_j^q - u_j)$, i.e.,

$$\lim_{n \to 4} \tilde{P}_{ij}^{pq} = P_{ij}^{pq}. \tag{63}$$

Let

$$\tilde{\sum}_x = \frac{1}{n}(X - \bar{X})(X - \bar{X})^{\mathrm{T}} = \frac{1}{n} \sum_{i=1}^{n} (x^{(i)} - u)(x^{(i)} - u)^{\mathrm{T}}, \tag{64}$$

where $u = (u_1, u_2, \ldots, u_m)^{\mathrm{T}} = (1/n)\sum_{i=1}^{n} x^{(i)}$ and $\bar{X} = \underbrace{(u, u, \cdots, u)}_{n}$. Let the element of $\tilde{\sum}_x$ be $\tilde{\sigma}_{ij}$, then

$$\tilde{\sigma}_{ij} = \frac{1}{n} \sum_{l=1}^{n} (x_i^{(l)} - u_i)(x_j^{(l)} - u_j) = \sum_{p,q} \frac{n_{ij}^{pq}}{n} (x_i^p - u_i)(x_j^q - u_j). \tag{65}$$

From equation (59), the definition of σ_{ij}, and equation (65) it follows that $\lim_{n\to 4} \tilde{\sigma}_{ij} = \sigma_{ij}$. Then the limitation $\lim_{n\to 4} \tilde{\sum}_x = \sum_x$ holds. In general, the centralization for the samples is performed after the n samples are selected, i.e., the expectation $(1/n)\sum_{i=1}^{n} x^{(i)}$ is subtracted from each sample. Thus $\tilde{\sum}_x = (1/n) XX^{\mathrm{T}}$ and we have

$$\sum_x = \lim_{n\to\infty} \tilde{\sum}_x = \lim_{n\to\infty} \frac{1}{n} XX^{\mathrm{T}}. \tag{66}$$

Equation (66) shows the asymptotic connection between the covariance (because of the centralization, the covariance is now the correlation) matrix $\sum_x$ in the PCA (KLD) and its approximate matrix $(1/n) XX^{\mathrm{T}}$. The asymptotic connection of the PCA (KLD) and the SVD can be obtained theoretically from the combination of equation (66) with the equivalence of the eigenvalue problems of the SVD and the PCA (KLD) mentioned above.

We have completed the demonstration on the equivalence connections of the three POD methods. The above discussion is performed for the discrete cases. It should be noted that the KLD can also be employed to handle the POD of continuous random variables, whereas the PCA and SVD can only be used to deal with discrete random variables.

4. CONCLUSION

In this paper, we summarized the derivation and the performance of the three POD methods: PCA, KLD and SVD, then discussed the equivalence problem via the theoretical comparison among the three methods and presented some proofs on their equivalence of them. We demonstrated and proved the equivalence of the matrices for processing, the objective functions, the optimal basis vectors and the mean square errors of the different POD methods, and the asymptotic connections among the three methods when they are used to handle the POD of discrete random vectors.

ACKNOWLEDGMENTS

The authors would like to acknowledge the reviewers for helpful comments and suggestions. The first author is grateful for the support of the NSFC (Grant No. 19872027) and the Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education of PRC.

REFERENCES

1. A. CHATTERJEE 2000 *Current Science* **78**, 808–817. An introduction to the proper orthogonal decomposition.
2. P. HOLMES, J. L. LUMLEY and G. BERKOOZ 1996 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge: Cambridge University Press.
3. B. F. FEENY and R. KAPPAGANTU 1998 *Journal of Sound and Vibration* **211**, 607–616. On the physical interpretation of proper orthogonal modes in vibrations.
4. B. RAVINDRA 1999 *Journal of Sound and Vibration* **219**, 189–192. Comments on "On the physical interpretation of proper orthogonal modes in vibrations".
5. R. KAPPAGANTU and B. F. FEENY 1999 *Journal of Sound and Vibration* **224**, 863–877. An "optimal" modal reduction of a system with frictional excitation.
6. R. KAPPAGANTU and B. F. FEENY 2000 *Nonlinear Dynamics* **22**, 317–333. Part 1: dynamical characterization of a frictionally excited beam.
7. R. KAPPAGANTU and B. F. FEENY 2000 *Nonlinear Dynamics* **23**, 1–11. Part 2: proper orthogonal modal modeling of a frictionally excited beam.

8. M. F. A. AZEEZ and A. F. VAKAKIS 2001 *Journal of Sound and Vibration* **240**, 859–889. Proper orthogonal decomposition (POD) of a class of vibroimpact oscillations.
9. X. H. MA, A. F. VAKAKIS and L. A. BERGMAN 2001 *American Institute of Aeronautics and Astronautics Journal* **39**, 687–696. Karhunen–Loève modes of a truss: transient response reconstruction and experimental verification.
10. X. H. MA and A. F. VAKAKIS 2001 *Journal of Vibration and Acoustics-Transactions of the American Society of Mechanical Engineers* **123**, 36–44. Nonlinear transient localization and beat phenomena due to backlash in a coupled flexible system.
11. X. H. MA, M. F. A. AZEEZ and A. F. VAKAKIS 2000 *Mechanical Systems and Signal Processing* **14**, 37–48. Non-linear normal modes and non-parametric system identification of non-linear oscillators.
12. X. H. MA and A. F. VAKAKIS 1999 *American Institute of Aeronautics and Astronautics Journal* **37**, 939–946. Karhunen–Loève decomposition of the transient dynamics of a multibay truss.
13. I. T. GEORGIOU, I. B. SCHWARTZ, E. EMACI and A. F. VAKAKis 1999 *Journal of Applied Mechanics. Transactions of the American Society of Mechanical Engineers* **66**, 448–459. Interaction between slow and fast oscillations in an infinite degree-of-freedom linear system coupled to a nonlinear subsystem: theory and experiment.
14. I. T. GEORGIOU and I. B. SCHWARTZ 1999 *SIAM Journal on Applied Mathematics* **59**, 1178–1207. Dynamics of large scale coupled structural mechanical systems: a singular perturbation proper orthogonal decomposition approach.
15. J. P. CUSUMANO, M. T. SHARKADY and B. W. KIMBLE 1994 *Philosophical Transactions of the Royal Society of London. Physical Sciences and Engineering* **347**, 421–438. Experimental measurements of dimensionality and spatial coherence in the dynamics of a flexible-beam impact oscillator.
16. E. S. HUNG and S. D. SENTURIA 1999 *Journal of Microelectromechanical Systems* **8**, 280–289. Generating efficient dynamical models for microelectromechanical systems from a few finite-element simulation runs.
17. W. Z. LIN, K. H. LEE, S. P. LIM and P. LU. 2000 *Dynamoics, Acoustics and Simulations American Society of Mechanical Engineers, DSC* **68**, 155–162. A method of model reduction for dynamic simulation of microelectromechanical systems.
18. Y. C. LIANG, W. Z. LIN, H. P. LEE, S. P. LIM, K. H. LEE and D. P. FENG 2001 *Journal of Micromechanics and Microengineering* **11**, 226–233. A neural-network-based method of model reduction for dynamic simulation of MEMS.
19. J. L. LUMLEY 1970 *Stochastic Tools in Turbulence*. New York: Academic Press.
20. I. T. JOLLIFFE 1986 *Principal Component Analysis*. New York: Springer-Verlag Inc.
21. G. H. DUNTEMAN 1989 *Principal Components Analysis*. California: Sage Publications Inc.
22. K. I. DIAMANTARAS and S. Y. KUNG 1996 *Principal Component Neural Networks: Theory and Applications*, New York: John Wiley & Sons Inc.
23. D. C. MURDOCH 1970 *Linear Algebra*. New York: John Wiley & Sons Inc.
24. K. FUKUNAGA 1990 *Introduction to Statistical Pattern Recognition*. San Diego: Academic Press Inc.
25. V. C. KLEMA and A. J. LAUB 1980 *IEEE Transactions on Automatic Control* **25**, 164–176. The singular value decomposition: its computation and some applications.
26. S. L. MARPLE JR. 1987 *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall Inc.
27. G. E. FORSYTHE, M. A. MALCOLM and C. B. MOLER 1977 *Computer Methods for Mathematical Computations*. Englewood Cliffs, NJ: Prentice-Hall Inc.
28. C. L. LAWSON and R. J. HANSON 1974 *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall Inc.
29. E. BIGLIERI and K. YAO 1989 *Signal Processing* **18**, 277–289. Some properties of singular value decomposition and their applications to digital signal processing.
30. G. H. GOLUB and C. F. VAN LOAN 1989 *Matrix Computations*. Baltimore, MD: The Johns Hopkins University Press.
31. A. I. MEES, P. E. RAPP and L. S. JENNINGS 1987 *Physical Review A* **36**, 340–346. Singular-value decomposition and embedding dimension.
32. K. KUNISCH and S. VOLKWEIN 1999 *Journal of Optimization Theory and Applications* **102**, 345–371. Control of the Burgers equation by a reduced-order approach using proper orthogonal decomposition.
33. Y. CAO and D. N. LEVIN 1995 *Magnetic Resonance in Medicine* **33**, 140–142. On the relationship between feature-recognizing MRI encoded by singular value decomposition.