Journal of Sound and Vibration (2002) **254**(5), 867–876
doi:10.1006/jsvi.2001.4129, available online at http://www.idealibrary.com on IDEAL®

NON-LINEAR GENERALIZATION OF PRINCIPAL COMPONENT ANALYSIS: FROM A GLOBAL TO A LOCAL APPROACH

G. KERSCHEN AND J.-C. GOLINVAL

LTAS—Vibrations et Identification des Structures, Université de Liège, Chemin des Chevreuils 1 (*B*52), *B*-4000 *Liège, Belgium. E-mail: g.kerschen@ulg.ac.be*

(*Received* 20 *September* 2001)

Principal component analysis (PCA), also known as proper orthogonal decomposition or Karhunen–Loève transform, is commonly used to reduce the dimensionality of a data set with a large number of interdependent variables. PCA is the optimal linear transformation with respect to minimizing the mean square reconstruction error but it only considers second-order statistics. If the data have non-linear dependencies, an important issue is to develop a technique which takes higher order statistics into account and which can eliminate dependencies not removed by PCA. Recognizing the shortcomings of PCA, researchers in the field of statistics and neural networks have developed non-linear extensions of PCA. The purpose of this paper is to present a non-linear generalization of PCA, called VQPCA. This algorithm builds local linear models by combining PCA with clustering of the input space. This paper concludes by observing from two illustrative examples that VQPCA is potentially a more effective tool than conventional PCA.

1. INTRODUCTION

Principal component analysis (PCA), also known as proper orthogonal decomposition or Karhunen–Loève transform, is a ubiquitous statistical technique for data analysis. The applications of this procedure are extensive, e.g., modelling of turbulence [1] and image processing [2]. PCA is now emerging as a useful tool in the field of structural dynamics. For instance, it has been applied to study the dimensionality of a system [3], to control torsional vibrations in long strings [4], to build reduced order models [5, 6], and to identify and update non-linear systems [7–9].

The key idea of PCA is to represent a data set in terms of a minimum of variables while preserving most of the information present in the data set. PCA defines a projection matrix made up of r vectors which are the eigenvectors of the covariance matrix associated with the r largest eigenvalues. The data are then projected onto the r-dimensional subspace of their embedding space, so that the mean square distance between the original points and their projection is minimal. Accordingly, PCA is the optimal linear transformation for the reconstruction of a data set.

Despite its widespread use, the effectiveness of PCA is limited by its global linearity. PCA removes linear correlations among the data and is only sensitive to second order statistics. It is, however, very common to deal with data sets where the relations among variables are non-linear. If the data lie on a non-linear manifold, an important issue is thus

0022-460X/02/$35.00

to have a technique which considers higher order statistics and allows for elimination of dependencies not removed by PCA. This simple realization has prompted the development of non-linear alternatives to PCA.

A few global non-linear variants of PCA were proposed. Principal curves were defined by Hastie and Stuetzle [10]. A principal curve may be viewed as a non-linear generalization of a principal component and is a smooth curve that passes through the middle of the data. Finally, the notion of the principal curve was extended to that of principal surface. In the early 1990s, a neural network based generalization of PCA was introduced in the chemical engineering literature by Kramer [11]. Called non-linear principal component analysis (NLPCA), Kramer implemented the solution using an autoassociative neural network with five layers: input layer, mapping layer, bottleneck layer, demapping layer and output layer. The outputs of the bottleneck layer are the non-linear principal components. Since then, it has been applied to find low-dimensional representations of grayscale face images [12] or to analyze climate data [13]. However, NLPCA suffers from practical drawbacks: neural networks with multiple sigmoidal hidden layers are difficult to train and tend to be trapped in local optima.

PCA and NLPCA try to describe all the data using the same global features. An alternative paradigm is to capture data complexity by a combination of local linear PCA projections. A local model implementation of PCA involves a two-step procedure: a clustering of the data space into disjoint regions and the estimation of the principal axes within each region. Local PCA has been exploited to identify intrinsic dimensions of data [14, 15], for handwritten character recognition [16] and for dimension reduction of speech [17] and images [17, 18].

Nevertheless, the authors are not aware of an application of a non-linear generalization of PCA, either global or local, in the field of structural dynamics. The aim of this paper is thus two-fold. On the one hand, one wishes to introduce the structural dynamicist to a local non-linear extension of PCA, denoted as VQPCA, where VQ stands for vector quantization. On the other hand, the performance of PCA and that of VQPCA are compared using two illustrative examples.

2. PRINCIPAL COMPONENT ANALYSIS

PCA is a multivariate analysis technique that was first introduced by Pearson [19] in 1901 and developed independently by Hotelling [20] in 1933. It is also closely related to proper orthogonal decomposition, also known as the Karhunen–Loève transform, introduced in 1943 by Kosambi [21].

Given a set of observed n-dimensional data points $\mathbf{x}_i$ with $i = 1, \ldots, m$, the goal of PCA is to reduce the dimensionality of the observed vector $\mathbf{x}_i$. This is realized by finding r principal axes $\mathbf{p}_i$ with $i = 1, \ldots, r$ onto which the retained variance under projection is maximal. These axes, denoted as principal components, are given by the eigenvectors associated with the r largest eigenvalues of the covariance matrix

$$\mathbf{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})\,(\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}}] \tag{1}$$

where $E[\cdot]$ is the expectation and $\boldsymbol{\mu} = E[\mathbf{x}]$ is the mean of the data.

If the principal components are collected in a matrix $\mathbf{P} = [\mathbf{p}_1 \cdots \mathbf{p}_r]$, then $\mathbf{z}_i = \mathbf{P}^{\mathrm{T}}(\mathbf{x}_i - \boldsymbol{\mu})$ is a reduced r-dimensional representation of the observed vector $\mathbf{x}_i$. Among all linear techniques, PCA provides the optimal reconstruction $\hat{\mathbf{x}}_i = \boldsymbol{\mu} + \mathbf{P}\mathbf{z}_i$ of $\mathbf{x}_i$ in terms of the quadratic reconstruction error $\|\mathbf{x}_i^2 - \hat{\mathbf{x}}_i^2\|$.

It is worth pointing out that PCA is also closely related to singular value decomposition. If the mean is subtracted from the data and if the data are collected in a matrix $\mathbf{X}$ (n rows and m columns), then the left singular vectors of $\mathbf{X}$, as eigenvectors of $\mathbf{X}\mathbf{X}^{\mathrm{T}}$, are the principal components. The singular values indicate how the corresponding left singular vectors participate.

As underlined in the introduction, PCA linearly decorrelates the data points, i.e., diagonalize the covariance matrix. However, decorrelation only yields statistical independence under the assumption that the distribution is Gaussian.

3. A LOCAL PCA APPROACH: VQPCA

If the data have non-linear dependencies, for any given error threshold, PCA may retain more dimensions (larger r) than a non-linear technique. Naturally enough, this has motivated researchers to develop non-linear extensions of PCA.

This section aims to present a local alternative to PCA, denoted as VQPCA, where VQ stands for vector quantization. The first application of a local PCA method dates back to 1971 and is due to Fukunaga and Olsen [14]. Since then, it has been applied successfully in different areas [15–17].

A local PCA approach implies the integration of two procedures:

(1) a clustering of the data space into distinct regions. VQPCA partitions the space by vector quantization;
(2) the construction of separate low-dimensional co-ordinate systems in each local region using PCA.

3.1. VECTOR QUANTIZATION

Vector quantization (VQ) is a classical technique for signal coding and data compression [22]. Let $\mathscr{X}$ be a set of observed n-dimensional data points $\mathbf{x}_i$ with $i = 1, \ldots, m$. A q-level vector quantizer is defined by a codebook $\mathscr{C} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_q)$, a partition $\mathscr{S} = (S_1, \ldots, S_q)$ and a distortion function $d(\mathbf{x}, \boldsymbol{\mu})$. It is a mapping f that approximates each point $\mathbf{x}_i$ in the set $\mathscr{X}$ by a component $\boldsymbol{\mu}_j$ of the codebook $\mathscr{C}$: $f(\mathbf{x}_i) = \boldsymbol{\mu}_j$ if $\mathbf{x}_i \in S_j$. A q-level quantizer is said to be optimal if it minimizes the averaged distortion $D = E[d(\mathbf{x}, \boldsymbol{\mu})]$.

In his implementation of VQPCA, Kambhatla [17] makes use of VQ to define the regions for the local PCA. The algorithm to design the vector quantizer is based on an approach of Lloyd [23] and is referred to as the generalized Lloyd algorithm [22]. The codebook vectors $\boldsymbol{\mu}_j$ and the regions S_j satisfy Lloyd's optimality conditions:

(1) each region S_j (with its corresponding codebook vector $\boldsymbol{\mu}_j$) corresponds to all $\mathbf{x}_i$ that lie closer to $\boldsymbol{\mu}_j$ than to any other codebook vector (nearest-neighbor mapping). Mathematically, $S_j = \{\mathbf{x}_i \mid d(\mathbf{x}_i, \boldsymbol{\mu}_j) < d(\mathbf{x}_i, \boldsymbol{\mu}_k), \forall\, k \neq j\}$;
(2) each codebook vector $\boldsymbol{\mu}_j$ is placed at the centroid of the corresponding region S_j.

For a distortion function based on Euclidean distance, the regions are convex sets called Voronoi cells and the centroid of a region is the mean of the data points in this region $\boldsymbol{\mu}_j = E[\mathbf{x}_i \mid \mathbf{x}_i \in S_j]$.

Accordingly, the generalized Lloyd algorithm is as follows:

(1) given q a number of regions, initialize the codebook $\mathscr{C}$ from randomly selected points in the data set $\mathscr{X}$;

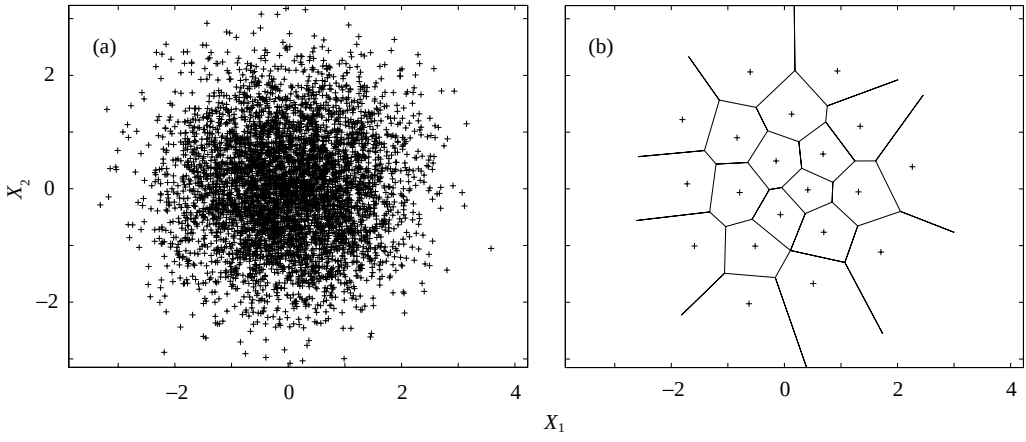

Figure 1. (a) Distribution of the data; (b) clustering of the data. —, Voronoi cells; + , centroids.

(2) compute the corresponding optimal partition following the first optimality condition;
(3) compute the corresponding optimal codebook following the second optimality condition;
(4) iterate steps 2 and 3 until convergence.

The convergence is achieved when the fractional change in the averaged distortion D between the kth and $(k+1)$th iterations is below some specified threshold. It can be argued that each iteration of the algorithm either reduces the distortion or leaves it unchanged.

It is worth noticing that several variants of this algorithm exist, e.g., tree-searched VQ and multistep VQ [22]. These variants aim to reduce the computation or memory requirements but may compromise the performance relative to what could be achieved with a standard VQ. Since the computational aspects are not an issue in this work, the basic Lloyd algorithm is considered throughout this paper.

By way of illustration, the generalized Lloyd algorithm, with a distortion function based on the Euclidean distance and the number of regions equal to 20, is applied on a set of 5000 two-dimensional random vectors chosen from a normal distribution with mean zero and variance one. The distribution of the random vectors is displayed in Figure 1(a). Figure 1(b) depicts the Voronoi cells S_j together with their corresponding centroids $\boldsymbol{\mu}_j$.

3.2. DIMENSION REDUCTION BY LOCAL LINEAR MODELS

Consider again a set of observed n-dimensional data points $\mathbf{x}_i$ with $i = 1, \ldots, m$. In order to reduce the dimensionality of the vector $\mathbf{x}_i$, one needs to determine an encoding function f: $\Re^n \to \Re^r$ such that $\mathbf{z}_i = f(\mathbf{x}_i)$ is a compact r-dimensional representation of $\mathbf{x}_i$. Similarly, a decoding function g: $\Re^r \to \Re^n$ has to be calculated such that $\hat{\mathbf{x}}_i = g(f(\mathbf{x}_i))$ is the reconstruction of the initial vector $\mathbf{x}_i$.

In the present work, the purpose is to build low-dimensional co-ordinate systems in the q local regions defined by the vector quantizer. If the local regions are small enough, the data manifold is not curved much over the extent of the region and it may be locally approximated as a hyperplane. In other words, a separate PCA model in each of the q regions should be adequate. Instead of having single encoding and decoding functions, a collection of functions $f_i(\cdot)$ and $g_i(\cdot)$ with $i = 1, \ldots, q$ is obtained.

3.3. VQPCA ALGORITHM

The VQPCA algorithm is an extension of a standard vector quantizer. VQPCA partitions the input space into a set of regions and approximates each region by a hyperplane defined by PCA, while a standard vector quantizer approximates each region by a codebook vector. The VQPCA algorithm is:

(1) Partition $\Re^n$ into q disjoint regions $S_1, \ldots, S_q$ using the generalized Lloyd algorithm with Euclidean distance as the distortion function.
(2) For each Voronoi cell S_j and its corresponding centroid $\boldsymbol{\mu}_j$, estimate the local covariance matrix

$$\boldsymbol{\Sigma}_j = \frac{1}{N_j} \sum_{\mathbf{x} \in S_j} (\mathbf{x} - \boldsymbol{\mu}_j)(\mathbf{x} - \boldsymbol{\mu}_j)^{\mathrm{T}}, \tag{2}$$

where N_j is the number of vectors mapped to S_j. Next, compute the eigenvectors $(\mathbf{p}_{j1}, \ldots, \mathbf{p}_{jr})$ of each matrix $\boldsymbol{\Sigma}_j$.
(3) To reduce dimension of any vector $\mathbf{x}_i$, determine the cell S_j which contains the vector and project $\mathbf{x}_i$ onto the r leading eigenvectors to obtain the local linear co-ordinates

$$\mathbf{z}_i = f_j(\mathbf{x}_i) = [\mathbf{p}_{j1} \cdots \mathbf{p}_{jr}]^{\mathrm{T}}(\mathbf{x}_i - \boldsymbol{\mu}_j) = \begin{bmatrix} \mathbf{p}_{j1}^{\mathrm{T}}(\mathbf{x}_i - \boldsymbol{\mu}_j) \\ \vdots \\ \mathbf{p}_{jr}^{\mathrm{T}}(\mathbf{x}_i - \boldsymbol{\mu}_j) \end{bmatrix}, \quad \text{if } \mathbf{x}_i \in S_j. \tag{3}$$

The compressed representation of $\mathbf{x}_i$ consists of the index j of the Voronoi cell in which $\mathbf{x}_i$ lies and the r-dimensional vector $\mathbf{z}_i$. The data are reconstructed from this representation according to

$$\hat{\mathbf{x}}_i = g_j(f_j(\mathbf{x}_i)) = g_j(\mathbf{z}_i) = \boldsymbol{\mu}_j + [\mathbf{p}_{j1}, \ldots, \mathbf{p}_{jr}]\mathbf{z}_i. \tag{4}$$

The accuracy of the compressed representation is assessed using the normalized mean square error (MSE)

$$MSE = E[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]/E[\|\mathbf{x} - \mathbf{E}[\mathbf{x}]\|^2]. \tag{5}$$

4. EXAMPLES

4.1. APPROXIMATION OF A NON-LINEAR NORMAL MODE

The utility of the VQPCA approach may be demonstrated with the following example. Consider 1000 samples obtained from the free response of an undamped non-linear system (Figure 2). This system has two degrees of freedom (d.o.f.) and a cubic stiffness k_{nl} between the wall and the first mass [24]. All the remaining stiffnesses k and k_c are linear. For $m = k = 1$, $k_c = k_{nl} = 15$ and initial conditions $[1; 0{\cdot}918]$ on the displacements, the motion is a single and synchronous non-linear mode, i.e., the system response is a one-dimensional curve in the co-ordinate space.

As stated in reference [25], for the case of a synchronous non-linear normal mode (NNM), *the dominant proper orthogonal mode* (*i.e., principal component*) *represents an optimal fit of a linear mode to the data on the non-linear normal mode in the sense that the distances of data from the proper orthogonal mode axis are optimized*. This is illustrated in Figure 3(a) where the synchronous NNM and the dominant principal component are compared in the co-ordinate space. The singular values normalized by their sum are equal to 0·840 and 0·160 indicating that the dominant principal component captures 84% of the energy.

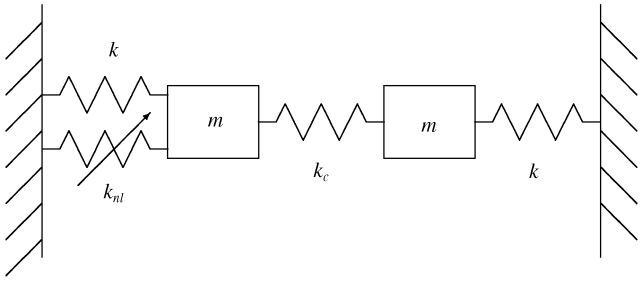

Figure 2. Model of the 2-d.o.f. example. $m = k = 1$ and $k_c = k_{nl} = 15$.

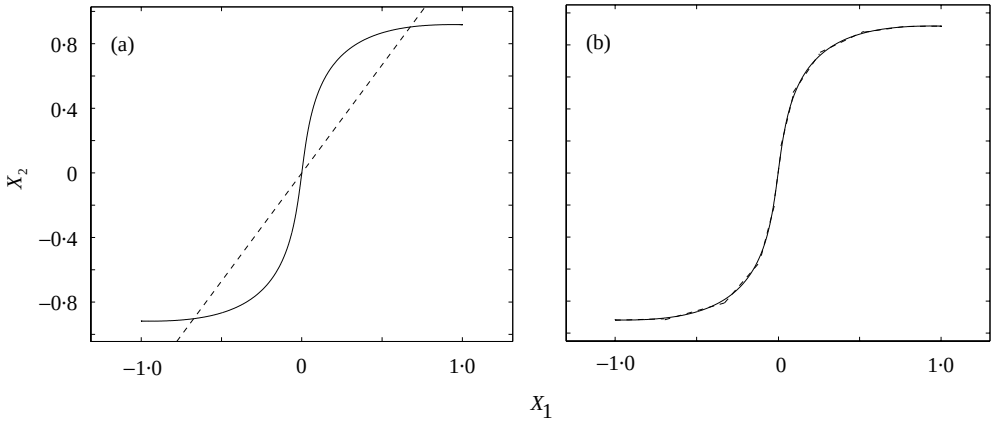

Figure 3. 2-d.o.f. example. (a) ——, Synchronous non-linear mode; – – –, dominant principal component; (b) ——, synchronous non-linear mode; – – –, dominant VQPCA mode, 10 cells.

TABLE 1

PCA *and* VQPCA *applied to the* 2-*d.o.f. example*

Number of regions	MSE (%) 1 mode	First singular value	Second singular value
1 (PCA)	3·50	0·840	0·160
3	0·21	0·900	0·100
5	$4{\cdot}04 \times 10^{-2}$	0·931	0·069
10	$2{\cdot}31 \times 10^{-3}$	0·964	0·036
15	$5{\cdot}49 \times 10^{-4}$	0·975	0·025
20	$1{\cdot}82 \times 10^{-4}$	0·981	0·019

VQPCA is then applied on the same data set. The number of regions is varied from 3 to 20. The singular values are computed in each region, normalized by their sum and averaged out over all the regions. All the results are listed in Table 1. The decrease in the MSE (5) when the number of regions grows is expected and highlights the superiority of VQPCA over PCA. It should also be noticed that the first singular value tends to approach 1, i.e.,

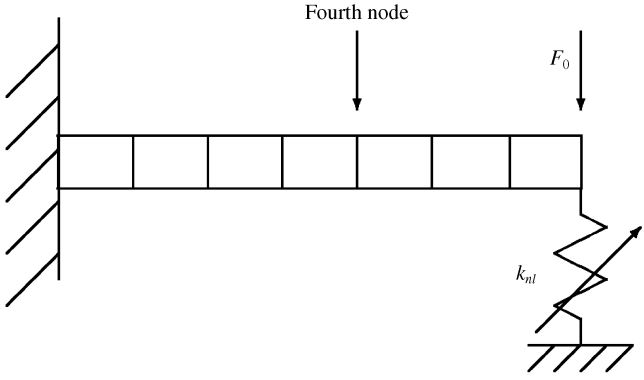

Figure 4. Model of the non-linear beam.

TABLE 2

PCA *and* VQPCA *applied to the non-linear beam example*

Number of regions	MSE (%) 1 mode	MSE (%) 2 modes	MSE (%) 3 modes	MSE (%) 4 modes
1 (PCA)	31·75	9·15	3·53	0·010
5	17·45	7·30	1·52	0·009
10	12·34	5·09	1·18	0·008
20	6·72	2·85	0·83	0·006
30	5·69	2·08	0·60	0·004

100% of the energy is captured, when the number of regions is increased. This enables one to conclude that the data are aligned on a one-dimensional manifold while PCA is not.

The first mode given by VQPCA (10 cells) is shown in Figure 3(b) and compared with the synchronous NNM. The VQPCA mode provides a close approximation to the non-linear normal mode. Finally, it must be stated that for a fixed number of regions, the VQPCA algorithm was run with different initializations of the codebook. In this way, the sensitivity of the generalized Lloyd algorithm to the starting point is tested. The results were never significantly influenced by the starting point.

4.2. RECONSTRUCTION OF DYNAMICAL RESPONSE

The second example chosen to demonstrate the application of the VQPCA algorithm is the reconstruction of the dynamical response of a non-linear beam (Figure 4). This clamped beam is modelled with seven beam elements and the local non-linearity k_{nl} is a spring that exhibits a cubic stiffness. The free vibration of the beam is simulated with an initial displacement given by a static force F_0 applied at the end of the beam.

The data set consists of seven vertical accelerations measured along the beam. PCA which is equivalent to VQPCA with a single Voronoi cell is first applied to the data. In a second step, the data are modelled with VQPCA using a number of regions varied from 5 to 30. Table 2 summarizes the relative performance of PCA and VQPCA in terms of

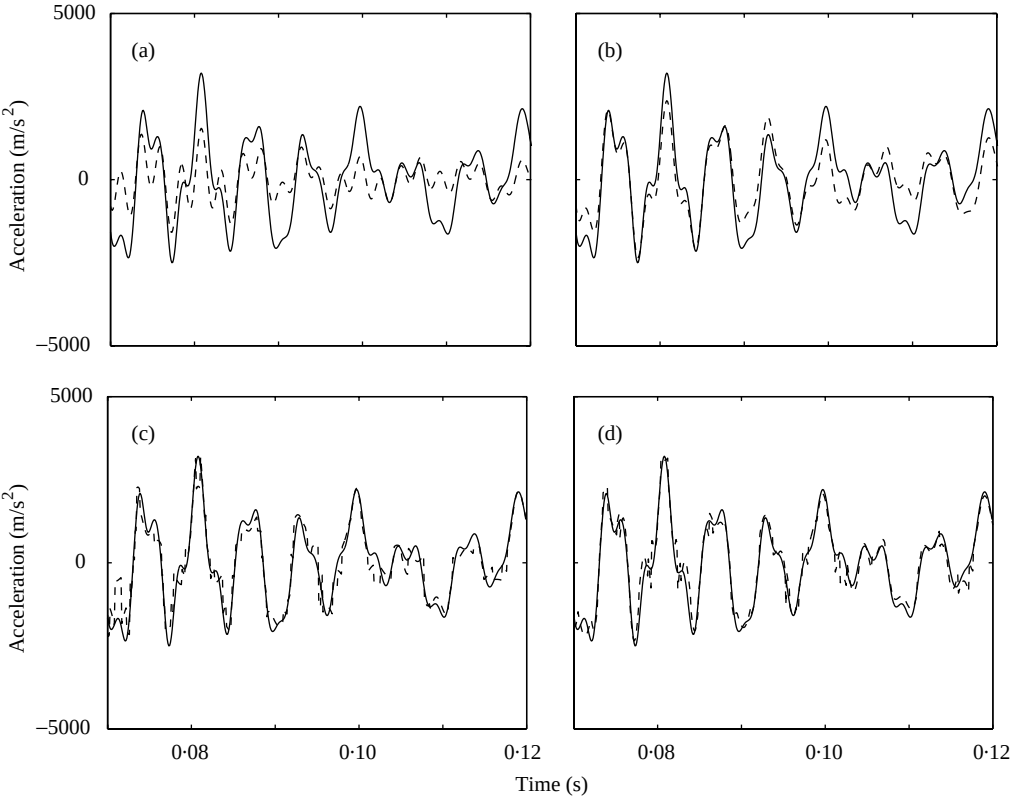

Figure 5. Reconstruction of the dynamical response of the non-linear beam: (——), system response; (– – –), reconstructed response. (a) PCA 1 mode; (b) PCA 2 modes; (c) VQPCA 1 mode, 20 cells; (d) VQPCA 2 modes, 20 cells.

compression accuracy measured by the MSE. Figure 5 shows the acceleration at the fourth node reconstructed from one- and two-dimensional representations generated by both PCA and VQPCA (20 cells).

It is clear from this figure that VQPCA provides a more accurate representation of the system response than PCA. This advantage is also reflected in the errors reported in Table 2. For instance, for a unimodal representation, VQPCA with 20 cells attains about 80% lower error than PCA and is still a better approximation than a bimodal PCA representation.

5. CONCLUSION

Modelling complexity in the data by a combination of simple linear models is an attractive paradigm. Accordingly, a local non-linear variant of PCA, denoted VQPCA, has been proposed in this paper. The method first exploits vector quantization to cluster the data space into disjoint regions. Then, a standard PCA model is built in each region defined by the vector quantizer. Thus, VQPCA approximates the data distribution with a set of local hyperplanes. The location and the distribution of this set capture the large-scale, non-linear structure of the data, while co-ordinates on the hyperplanes capture the local variations.

VQPCA provides insight into the structure of a data set that PCA could not. Its superiority over PCA has been demonstrated by two illustrative examples. It is well suited for dimensionality reduction and for estimation of the intrinsic dimensionality. There are other applications that deserve attention. PCA was used to monitor the condition of signals during manufacture [26]. The non-linear models provided by VQPCA should provide more accurate models of the undamaged data and hence improve the sensitivity and specificity for fault detection. PCA was also applied to identify and update non-linear mechanical systems [7–9]. In this context, the modes given by VQPCA may be viewed as promising features to analyze the behavior of the non-linear system. This will be studied in further work.

ACKNOWLEDGMENTS

G. Kerschen gratefully acknowledges the support of the Belgium National Fund for Scientific Research (FNRS).

REFERENCES

1. P. Holmes, J. L. Lumley and G. Berkooz 1996 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. New York: Cambridge.
2. G. Uytterhoeven 1999 *Ph.D. Thesis, Katholieke Universiteit Leuven*. Wavelets : Software and Applications.
3. J. P. Cusumano, M. T. Sharkady and B. W. Kimble 1993 *Aerospace Structures: Nonlinear Dynamics and System Response*, ASME AD-Vol. 33, 13–22. Spatial coherence measurements of a chaotic flexible-beam impact oscillator.
4. E. Kreuzer and O. Kust 1996 *Nonlinear Dynamics and Control* ASME DE-Vol. 91, 105–110. Proper orthogonal decomposition — an efficient means of controlling self-excited vibrations of long torsional strings.
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. M. F. A. Azeez and A. F. Vakakis 2001 *Journal of Sound and Vibration* **240**, 859–889. Proper orthogonal decomposition of a class of vibroimpact oscillations.
7. T. K. Hasselman, M. C. Anderson and W. G. Gan 1998 *Proceedings of the* 16*th International Modal Analysis Conference, Santa Barbara, U.S.A.*, 644–651. Principal component analysis for nonlinear model correlation, updating and uncertainty evaluation.
8. V. Lenaerts, G. Kerschen and J. C. Golinval 2001 *Mechanical Systems and Signal Processing* **15**, 31–43. Proper orthogonal decomposition for model updating of non-linear mechanical systems.
9. F. M. Hemez and S. W. Doebling 1999 *Proceedings of the* 17*th International Modal Analysis Conference, Kissimmee, U.S.A.*, 1501–1510. Test analysis correlation and finite element model updating for nonlinear transient dynamics.
10. T. Hastie and W. Stuetzle 1989 *Journal of the American Statistical Association* **84**, 502–516. Principal curves.
11. M. A. Kramer 1991 *A.I.Ch.E Journal* **37**, 233–243. Nonlinear principal component analysis using autoassociative neural networks.
12. D. Demers and G. Cottrell 1993 *Advances in Neural Information Processing Systems* 5, *San Mateo, U.S.A.*, 580–587. Nonlinear dimensionality reduction.
13. A. H. Monahan 2000. *Ph.D. Thesis, University of British Columbia*. Nonlinear Principal Component Analysis of Climate Data.
14. K. Fukunaga and D. R. Olsen 1971 *IEEE Transactions on Computers* **20**, 176–183. An algorithm for finding intrinsic dimensionality of data.
15. D. S. Broomhead 1991 *Adaptive Signal Processing, SPIE Proceedings*, Vol. 1565, 228–243. Signal processing for nonlinear systems.
16. G. E. Hinton, M. Revow and P. Dayan 1995 *Advances in Neural Information Processing Systems*, vol. 7, *Cambridge, U.S.A.*, 1015–1022. Recognizing handwritten digits using mixtures of linear models.

17. N. KAMBHATLA 1996 *Ph.D. Thesis*, *Oregon Graduate Institute of Science & Technology*. Local Models and Gaussian Mixture Models for Statistical Data Processing.
18. C. ARCHER and T. K. LEEN 1999 *Proceedings of the International Joint Conference on Neural Networks*, *Washington*, *DC*, *U.S.A*. Optimal dimension reduction and transform coding with mixture principal components.
19. K. PEARSON 1901 *Philosophical Magazine* **2**, 559–572. On lines and planes of closest fit to systems of points in space.
20. H. HOTELLING 1933 *Journal of Educational Psychology* **24**, 417–441, 498–520. Analysis of a complex of statistical variables into principal components.
21. D. KOSAMBI 1943 *Journal of Indian Mathematical Society* **7**, 76–88. Statistics in function space.
22. A. GERSHO and R. M. GRAY 1992 *Vector Quantization and Signal Compression*. Norwell: Kluwer Academic Publishers.
23. S. P. LLOYD 1957 *Bell Laboratories Technical Note*. Least squares quantization in PCM. Published in 1982 in *IEEE Transactions on Information Theory* **28**, 129–137.
24. S. W. SHAW and C. PIERRE 1993 *Journal of Sound and Vibration* **164**, 85–124. Normal modes for non-linear vibratory systems.
25. 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.
26. I. Y. TUMER, K. L. WOOD and I. J. BUSCH-VISHNIAC 2000 *Mechanical Systems and Signal Processing* **14**, 1011–1026. Monitoring signals from manufacturing processes using the Karhunen–Loève transform.