Available online at www.sciencedirect.com
ScienceDirect
Journal of Sound and Vibration 325 (2009) 224–239
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Damage diagnosis under environmental and operational variations using unsupervised support vector machine

Chang Kook Oh[a], Hoon Sohn[b,*]

[a]Department of Civil and Environmental Engineering, KAIST, Daejeon, Republic of Korea
[b]Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, USA

Received 1 September 2008; received in revised form 13 March 2009; accepted 16 March 2009
Handling Editor: J. Lam
Available online 18 April 2009

Abstract

The goal of structural health monitoring is to provide reliable information regarding damage states that include damage presence, location, and severity. Damage diagnosis is performed by utilizing measurements that are obtained from a structure being monitored. However, time-varying environmental and operational conditions such as temperature and external loading may produce an adverse effect on damage detection within the structure exposed to these changes. Therefore, in order to achieve successful structural health monitoring goals, it is necessary to develop data normalization techniques which distinguish the effects of damage from those caused by environmental and operational variations. In this study, nonlinear principal component analysis based on unsupervised support vector machine is introduced and incorporated with a discrete-time prediction model and a hypothesis test for data normalization. The proposed nonlinear principal component analysis characterizes the nonlinear relationship between extracted damage-sensitive features and unmeasured environmental and operational parameters by employing kernel functions and by solving a simple eigenvalue problem. The performance of the proposed method is compared with that of another nonlinear principal component analysis realized by auto-associative neural network. It is demonstrated that the proposed method is a promising data normalization tool that is capable of detecting damage in the presence of environmental and operational variations.
© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

The primary goal of structural health monitoring is to provide reliable information regarding damage identification, damage location, and damage severity by using measurements obtained from a structure being monitored. Once damage is detected, subsequent damage assessment is performed in order to locate damage and estimate its severity. The basic premise is that damage alters the dynamic characteristics of the structure when damage occurs.

In reality, time-varying environmental and operational conditions also affect the measured signals, so that they may blur slight changes induced by damage or cause an adverse effect on detection of damage within the structure exposed to these changes. Changes of temperature, humidity and wind are categorized as

*Corresponding author. Tel.: +1 412 268 2077; fax: +1 412 268 7813.
E-mail address: hoonsohn@kaist.ac.kr (H. Sohn).

0022-460X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2009.03.014

environmental variations, while operational speed and traffic loading changes are in the category of operational variations [1].

Numerous researchers demonstrated the changes of structural dynamic properties are related with environmental and operational variations. Temperature, for example, is reported to change stiffness [2] as well as boundary conditions [3]. Wind-induced vibration is known to have a significant influence on dynamics of long span bridges by changing their damping characteristics [4]. Traffic loadings are also reported to alter the measured natural frequencies [5] and damping ratios [6]. Therefore, in order to develop a robust structural health monitoring system, it is important to distinguish the effects of damage from those caused by environmental and operational variations.

Data normalization is a procedure to separate the signal changes caused by environmental and operational variations from those by damage [7]. Data normalization can be performed in three different ways. When it is possible to measure the environmental and operational parameters, various regression analyses can be performed to relate the measured parameters with the environmental and operational conditions [8–10]. On the other hand, when this direct measurement is not feasible, data normalization is performed by characterizing the relationship between the environmental and operational parameters and damage-sensitive features, and this case is the primary interest of this study. In Kullaa [11], factor analysis was used to eliminate the effects of environmental variations on the measured features. The linear and nonlinear models were attempted to reveal the relationship between temperature and four natural frequencies without measuring environmental quantities. In Ruotolo and Surace [12], singular value decomposition was used to detect damage when the structure was under two different working conditions, (e.g., with or without a concentrated mass at the end of a cantilever beam). Damage was identified when the number of singular values exceeds a threshold. In Chelidze and Liu [13], experimental data were used to separate the changes caused by operating conditions from those caused by damage, and smooth orthogonal decomposition was employed to identify damage. Finally, there is an ongoing effort to develop ''reference-free'' structural health monitoring techniques that do not require direct comparison with previously obtained baseline data [14].

In this paper, a nonlinear principal component analysis is integrated with a discrete-time prediction model and a hypothesis test and applied to structural health monitoring with an emphasis on data normalization. The proposed nonlinear principal component analysis is based on unsupervised least-squares support vector machine, and it characterizes the nonlinear relationship between extracted damage-sensitive features and unmeasured environmental and operational parameters by employing kernel functions and by solving a simple eigenvalue problem.

In the proposed method, an autoregressive and autoregressive with exogenous inputs (AR-ARX) model [15] is first constructed to extract the damage-sensitive features from measured time signals. Then, nonlinear principal component analysis is applied to characterize the hidden relationship between unmeasured environmental and operational parameters and extracted damage-sensitive features. Finally a hypothesis test named a sequential probability ratio test [16] is performed on the extracted features to evaluate the damage state of the structure (see Fig. 1).

The proposed method is compared with another data normalization technique previously studied by Sohn et al. [17]. In this previous study, auto-associative neural network was employed for data normalization. It was demonstrated that the incorporation of auto-associative neural network with time series analysis and statistical inference allowed detecting damage in the presence of operational and environmental variability. In this study, advantages of the proposed least-squares support vector machine technique over auto-associative neural network are discussed including reduced misclassification, better generalization, and assurance of a

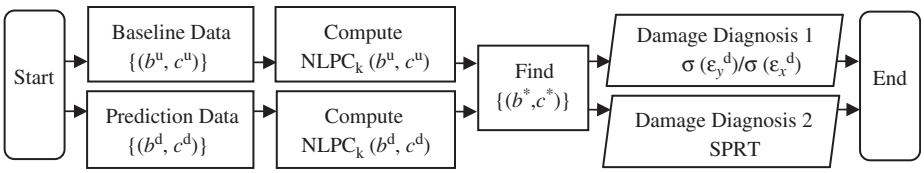

Fig. 1. A damage diagnosis procedure proposed in this study.

unique and global solution. For meaningful comparison, the same data set that was used by Sohn et al. [17] is utilized in this study.

This paper is organized as follows. In Section 2, a discrete-time prediction AR-ARX model used to extract damage-sensitive features is briefly reviewed. Section 3 presents the theoretical formation of the proposed nonlinear principal component analysis based on unsupervised least-squares support vector machine. First, a supervised least-squares support vector machine algorithm is introduced and compared with a classical supervised support vector machine. Then, an unsupervised least-squares support vector machine is described and extended in order to realize nonlinear principal component analysis. In Section 4, a statistical damage classification method named a sequential probability ratio test is presented for damage diagnosis. In Section 5, the proposed method is applied to an experimental example of an eight-dof mass–spring system under various input excitation levels. Finally, the conclusion and discussions are provided in Section 6.

2. Discrete-time prediction model

In this section, a main concept of an AR-ARX model is briefly outlined. More details can be found in Sohn et al. [17]. A combination of autoregressive and autoregressive with exogenous inputs models is used for linear discrete-time prediction of measured time signals. First, every measured time signal is standardized to have zero mean and unit variance.

For a standardized time signal, $x(t)$, an autoregressive model is constructed with r number of autoregressive terms [18]:

$$x(t) = \sum_{i=1}^{r} a_i x(t-i) + e_x(t), \tag{1}$$

where $e_x(t)$ represents an error between measurement and prediction, and it is assumed that the error is caused primarily by the unknown external input [17]. Note that linear discrete-time prediction using this autoregressive model is applied to a single time signal, i.e., $x(t)$ in Eq. (1) is a scalar. The r number of coefficients, a_i, can be estimated by various parameter estimation techniques [18].

Next, the error term, $e_x(t)$, in Eq. (1) is employed as an input for an autoregressive with exogenous inputs model to construct the input/output relationship between $e_x(t)$ and $x(t)$:

$$x(t) = \sum_{i=1}^{p} b_i x(t-i) + \sum_{j=1}^{q} c_j e_x(t-j) + \varepsilon_x(t), \tag{2}$$

where p and q are the numbers of b_i and c_j coefficients, and $\varepsilon_x(t)$ is the residual error. Similar to r parameters in the autoregressive model, the p and q parameters are also estimated by various parameter estimation methods.

This two-stage time prediction using a combined AR-ARX model is similar to an autoregressive moving average model described in Ljung [19], and it was suggested to keep the sum of p and q smaller than r, i.e., $p+q \leqslant r$. For a proper comparison, the numbers of the a_i, b_i, and c_j coefficients are set equal to the ones reported in Sohn et al. [17] ($r = 30$, $p = 5$, and $q = 5$).

3. Principal component analysis

Principal component analysis is an orthogonal transformation of a coordinate system in which given data can be described as a combination of new variables [20]. Principal component analysis is known as an efficient way in reducing dimensionality of data, since a small fraction of the entire principal components can often account for most of the data structure [21].

Linear principal component analysis aims to search for a transformed coordinate in the form of *straight* lines in such a way that maximizes the variance of original variables. Linear principal component analysis is realized by eigenvector decomposition of the data covariance matrix. The first eigenvector corresponding to the largest eigenvalue of the covariance matrix represents the direction into which the variance of the projected variable is maximized. This new variable projected onto the first eigenvector is called the first principal

component, and the subsequent principal components associated with the remaining eigenvalues can be computed by projecting the data onto the successive eigenvectors.

Linear principal component analysis can be generalized to nonlinear principal component analysis in order to reveal the nonlinear correlations inherent in the data. Auto-associative neural network is one such method [22]. Auto-associative neural network is a five-layer network that consists of input, mapping, bottle-neck, de-mapping, and output layers as shown in Fig. 2(a). When auto-associative neural network is trained, given data are used both as network inputs and outputs, i.e., in an auto-associative mode. Thus, the dimensions of the input and output layers, namely, the number of neurons in the corresponding layers, are identical. On the other hand, the dimension of the bottle-neck layer is designed to be less than those of the input and output layers. Nonlinear principal component analysis is achieved via the bottle-neck layer with a reduced dimension. However, training auto-associative neural network requires solving a complex optimization problem and using several different initial values to avoid getting trapped in possible local minima. It is also known that auto-associative neural network may over-fit data since the training continues until it minimizes the difference between target outputs and network outputs. Several approaches were suggested to overcome this over-fitting problem such as regularization and early stopping schemes [22]. Besides auto-associative neural network, there are a number of nonlinear principal component analysis methods such as self-organizing mapping [22], principal curves [23], and Hebbian networks [24].

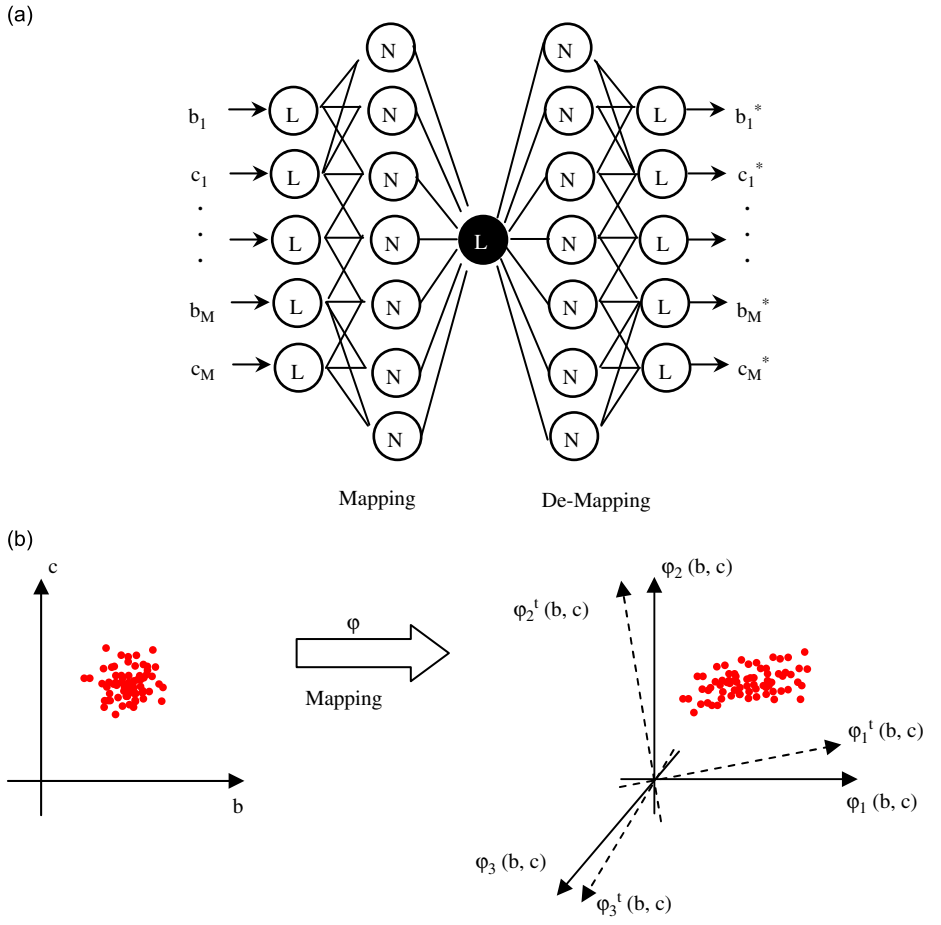

Fig. 2. Comparison between (a) auto-associative neural network and (b) kernel principal component analysis. Nonlinear principal component analysis is performed (a) via a bottle-neck layer (shown as black circle) and (b) via linear principal component analysis after mapping input data into a higher dimensional space.

In this study, nonlinear principal component analysis is achieved using unsupervised least-squares support vector machine, also known as kernel principal component analysis [25], instead of auto-associative neural network as shown in Fig. 2. It will be shown shortly after that kernel principal component analysis achieves nonlinear principal component analysis in a relatively easy way by solving a simple eigenvalue problem in a transformed space. Because a regularization term is explicitly included in the formation of the unsupervised least-squares support vector machine, it avoids data over-fitting. Note that a different formulation of kernel principal component analysis is also available as in Oh and Sohn [26] where emphasis is laid on performing linear principal component analysis in the transformed coordinate system.

In the next subsections, a main concept of linear principal component analysis is first introduced. Then, two kinds of supervised support vector machine such as classical support vector machine and its modification, least-squares support vector machine, are presented in a way of introducing unsupervised least-squares support vector machine. Finally, unsupervised least-squares support vector machine is presented to perform nonlinear principal component analysis.

3.1. Linear principal component analysis

Let $\mathscr{D}_N = \{\mathbf{x}_j \in \mathrm{R}^{m\times 1}: j = 1,\ldots,N\}$ denote a set of N number of centered, i.e., zero mean, m-dimensional feature vectors extracted from measurements and $\mathbf{v}$ be the unknown normalized eigenvector in the transformed coordinate system. (Another case that deals with non-centered data will be discussed shortly.) Then, linear principal component analysis is performed to find $\mathbf{v}$ that maximizes the variance of the projected variables $\mathbf{v}^{\mathrm{T}}\mathbf{x}$:

$$\max_{\mathbf{v}}[\mathrm{Var}(\mathbf{v}^{\mathrm{T}}\mathbf{x})] = \max_{\mathbf{v}}[\mathbf{v}^{\mathrm{T}}\mathbf{C}_1\mathbf{v}], \quad \text{subject to } \mathbf{v}^{\mathrm{T}}\mathbf{v} - 1 = 0, \tag{3}$$

where $\mathrm{Var}(\cdot)$ is a variance operator and $\mathbf{C}_1 = (1/N)\Sigma_j \mathbf{x}_j \mathbf{x}_j^{\mathrm{T}}$.

The constrained optimization problem can be re-stated using Lagrangian, $L_1(\mathbf{v},\lambda^*)$:

$$L_1(\mathbf{v},\lambda^*) = \tfrac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{C}_1\mathbf{v} - \lambda^*(\mathbf{v}^{\mathrm{T}}\mathbf{v} - 1), \tag{4}$$

where λ^* is a Lagrange multiplier. The solution of this optimization problem can be calculated from $\partial L_1/\partial \mathbf{v} = \mathbf{0}$ and $\partial L_1/\partial \lambda^* = 0$, which leads to the following eigenvalue problem:

$$\lambda^*\mathbf{v} = \mathbf{C}_1\mathbf{v} \tag{5}$$

where λ^* and $\mathbf{v}$ are eigenvalues and corresponding eigenvetors, respectively.

From Eq. (5), m number of eigenvalues and eigenvectors are calculated. The kth principal component for a feature vector $\mathbf{x}_j$ is computed as an inner product between $\mathbf{x}_j$ and the corresponding kth eigenvector $\mathbf{v}_k$:

$$\mathrm{PC}_k(\mathbf{x}_j) = (\mathbf{v}_k)^{\mathrm{T}}\mathbf{x}_j, \tag{6}$$

where PC_k $(\mathbf{x}_j)$ represents the kth principal component of a feature vector $\mathbf{x}_j$.

3.2. Support vector machine

Support vector machine is a recently developed machine learning algorithm that has been applied to various supervised classification problems [27]. Supervised learning is distinguished from unsupervised learning in its employment of the pre-defined class labels. For supervised learning, a known class label is assigned to each distinct class. For example, -1 is assigned to the data measured from an undamaged structure, and 1 to the data measured from a damaged structure. Then, support vector machine is trained to estimate the boundary between these two classes. When a new data set arrives, a suitable class label decided by the estimated boundary is assigned to each data point.

Let $\mathscr{F}_N = \{(\mathbf{x}_j, y_j): j = 1,\ldots,N\}$ denote a data set consisting of N number of m-dimensional feature vectors, $\mathbf{x}_j \in \mathbb{R}^{m\times 1}$, extracted from measurements and the corresponding labels, $y_j \in \mathbb{R}$. For simplicity, a linear and binary classification problem, i.e., a problem to estimate a linear separating boundary between two labels $y_j \in \{-1, 1\}$, is illustrated. Nonlinear and multi-class classification problems can be found in Schölkopf and Smola [28].

Support vector machine determines the separating boundary between two class data by maximizing the margin and minimizing the misclassification. The maximization of the margin is achieved by minimizing $\|\mathbf{w}\|$, leading to the following constrained optimization problem:

$$\min_{\mathbf{w},\boldsymbol{\xi}} \tau_1(\mathbf{w},\boldsymbol{\xi}) = \min_{\mathbf{w},\boldsymbol{\xi}} \left[\frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + c\sum_{i=1}^{N}\xi_i\right], \tag{7}$$

subject to

$$y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geqslant 1 - \xi_i, \quad \xi_i \geqslant 0 \text{ and } c \geqslant 0, \ i = 1,\ldots,N, \tag{8}$$

where $\tau_1(\cdot)$, ξ_i, and c are an objective function, a slack variable, and an unknown constant, respectively. $\boldsymbol{\xi} = [\xi_1,\ldots,\xi_N]^{\mathrm{T}}$, $\mathbf{w}$ is a vector that defines the direction of the separating boundary, and b is a threshold. Note that $y_j \in \{-1,1\}$, so ± 1 in an inequality constraint of Eq. (8) can be interpreted as threshold values.

The first term in Eq. (7) plays an important role for regularization by discouraging $||\mathbf{w}||$ from becoming a large value. A slack variable, ξ_i, is adopted to consider a data point that is not properly classified by the estimated boundary. Here, the misclassification is reduced by minimizing the linear summation of these non-negative slack variables in Eq. (7) [27]. Finally, an unknown positive constant, c, is a regularization factor to control the trade-off between the two terms in Eq. (7) [29].

3.3. Least-squares support vector machine

For the same data set, $\mathscr{F}_N = \{(\mathbf{x}_j, y_j): j = 1,\ldots,N\}$, least-squares support vector machine is developed to simplify classical support vector machine without compromising its advantages [30]. The objective function is reformulated:

$$\min_{\mathbf{w},\mathbf{e}} \tau_2(\mathbf{w},\mathbf{e}) = \min_{\mathbf{w},\mathbf{e}} \left[\frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \gamma\frac{1}{2}\sum_{i=1}^{N}e_i^2\right], \quad \text{subject to } y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) = 1 - e_i \text{ and } \gamma \geqslant 0, \tag{9}$$

where γ is an unknown positive constant, and $\mathbf{e} = [e_1,\ldots,e_N]^{\mathrm{T}}$. Contrary to an inequality constraint of Eq. (8), least-squares support vector machine adopts an equality constraint of Eq. (9), and ± 1 represent two target values. For example, if $y_i = -1$, e_i measures the error of $\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b$ to a target value of -1. Therefore, e_i represents the degree of scattering of within-class data, i.e., the data that have the same class label, with respect to the corresponding target values of either 1 or –1 [30].

This modified least-squares support vector machine with an equality constraint tries to maximize the margin and minimize within-class scattering. Least-squares support vector machine differs from classical support vector machine primarily in the following two ways:

(1) Least-squares support vector machine uses an equality constraint instead of inequality one. In classical support vector machine, the value of 1 in Eq. (8) is considered as a threshold rather than a target value.
(2) Least-squares support vector machine includes a linear summation of squared error terms in an objective function instead of a linear summation of positive slack variables.

Due to these differences, least-squares support vector machine provides a set of *linear* equations to solve, while classical support vector machine produces a *quadratic* optimization problem. Note that both methods guarantee a global solution although it may be different with each other. More details about least-squares support vector machine and comparison with conventional support vector machine can be found in Suykens et al. [30].

3.4. Application of lease-squares support vector machine to linear principal component analysis

Principal component analysis, whether it is linear or nonlinear, is one of the unsupervised learning methods, since labels, y_i, for given data are not available. Therefore, a single target value of zero, instead of two target

values of ± 1, is used when least-squares support vector machine is applied to principal component analysis. The objective function is reformulated from Eq. (3) that is an established form for linear principal component analysis:

$$\max_{\mathbf{w},\mathbf{e}} \tau_3(\mathbf{w},\mathbf{e}) = \max_{\mathbf{w},\mathbf{e}} \left[\gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 - \frac{1}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} \right], \quad \text{subject to } \mathbf{w}^{\mathrm{T}} \mathbf{x}_i = e_i, \tag{10}$$

where γ is an unknown positive constant that is determined by solving this optimization problem. Since e_i represent the within-class variance, the first term of an objective function in Eq. (10) is equivalent to the variance of the projected variables in Eq. (3) except an additional coefficient of $\gamma/2$. Moreover, Eq. (10) explicitly shows that the regularization term of $\mathbf{w}^{\mathrm{T}}\mathbf{w}/2$ is included to avoid over-fitting. Note that maximizing negative $\mathbf{w}^{\mathrm{T}}\mathbf{w}/2$ in Eq. (10) is equivalent to minimizing $\mathbf{w}^{\mathrm{T}}\mathbf{w}/2$ in Eq. (9). Compared with Eq. (9), the constraint in Eq. (10) represents the single target value is zero and a bias, i.e., b in Eq. (9), is also zero since data are centered [30]. By solving this optimization problem, one can find $\mathbf{w}$, i.e., one of the eigenvectors in input space that maximizes the variance of data covariance, and the corresponding principal component e_i that is projected on $\mathbf{w}$.

3.5. Nonlinear principal component analysis

The mathematical formulation for linear principal component analysis can be easily extended to nonlinear principal component analysis by using kernel method and a nonlinear mapping $\boldsymbol{\varphi}(\cdot)$: $\mathbb{R}^m \rightarrow \mathbb{R}^h$ where $m<h$. The nonlinear mapping transforms the features in the input space into a higher, possibly infinite, dimensional space so that linear principal component analysis can be conducted in the extended dimensional space as shown in Fig. 2(b).

Suppose linear principal component analysis is performed in the transformed space. Then, an objective function similar to Eq. (10) can be constructed by using the transformed data, $\boldsymbol{\varphi}(\mathbf{x}_i)$:

$$\max_{\mathbf{w},\mathbf{e}} \tau_3(\mathbf{w},\mathbf{e}) = \max_{\mathbf{w},\mathbf{e}} \left[\gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 - \frac{1}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} \right], \quad \text{subject to } \mathbf{w}^{\mathrm{T}} \boldsymbol{\varphi}(\mathbf{x}_i) = e_i, \tag{11}$$

From Eq. (11), Lagrangian becomes:

$$L_2(\mathbf{w},\mathbf{e},\boldsymbol{\alpha}) = \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 - \frac{1}{2} \mathbf{w}^{\mathrm{T}} \mathbf{w} - \sum_{i=1}^{N} \alpha_i (e_i - \mathbf{w}^{\mathrm{T}} \boldsymbol{\varphi}(\mathbf{x}_i)), \tag{12}$$

where α_i is a Lagrange multiplier and $\boldsymbol{\alpha} = [\alpha_1 \dots \alpha_N]^{\mathrm{T}} \in \mathbb{R}^{N\times 1}$. The optimality can be obtained by:

$$\frac{\partial L_2}{\partial \mathbf{w}} = \mathbf{0} \rightarrow \mathbf{w} - \sum_{i=1}^{N} \alpha_i \boldsymbol{\varphi}(\mathbf{x}_i) = \mathbf{0}, \tag{13}$$

$$\frac{\partial L_2}{\partial e_i} = 0 \rightarrow \gamma e_i - \alpha_i = 0, \tag{14}$$

$$\frac{\partial L_2}{\partial \alpha_i} = 0 \rightarrow e_i - \mathbf{w}^{\mathrm{T}} \boldsymbol{\varphi}(\mathbf{x}_i) = 0, \tag{15}$$

Substituting Eqs. (13) and (14) into Eq. (15) leads to the following eigenvalue problem:

$$\mathbf{K}\boldsymbol{\alpha} = \lambda \boldsymbol{\alpha}, \tag{16}$$

where $\lambda(=1/\gamma)$ and $\boldsymbol{\alpha}$ are eigenvalues and eigenvectors, respectively. The ijth entity of $\mathbf{K}$ matrix, $\mathbf{K}_{ij}$, is defined as:

$$\mathbf{K}_{ij} \equiv \boldsymbol{\varphi}(\mathbf{x}_i)^{\mathrm{T}} \boldsymbol{\varphi}(\mathbf{x}_j) = k(\mathbf{x}_i, \mathbf{x}_j), \quad i,j = 1, \dots, N. \tag{17}$$

In Eq. (17), the kernel method is used based on the Mercer's theorem; for any symmetric, continuous, and positive-semidefinite function $k(\mathbf{x}_i,\mathbf{x}_j)$, there exists a Hillbert space, $\mathscr{H}$, a nonlinear map, $\boldsymbol{\varphi}$: $R^m \rightarrow \mathscr{H}$, and the

inner product of two nonlinearly transformed features, $\boldsymbol{\varphi}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}_j)$, can be expressed as a kernel function [30]. This kernel method allows the inner products of two transformed feature vectors to be calculated in the original input space without explicit computation of the nonlinear transformation, $\boldsymbol{\varphi}(\cdot)$. Furthermore, nonlinear principal component analysis is achieved by performing linear principal component analysis in the transformed space [28]. Kernel functions frequently used in practice are polynomial, Gaussian, and sigmoid kernels. In this study, a Gaussian kernel, $k(\mathbf{x}_i,\mathbf{x}_j)=\exp(-||\mathbf{x}_i-\mathbf{x}_j||^2/\rho^2)$, is selected, and ρ is called a width. The selection scheme for ρ is described later. Further discussion on kernels can be found in Refs. [25,27,28].

The kth nonlinear principal component for a feature vector $\mathbf{x}_j$ is computed:

$$\mathrm{NLPC}_k(\mathbf{x}_j)=(\mathbf{w}_k)^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}_j)=\sum_{i=1}^{N}(\boldsymbol{\alpha}_k)_i\boldsymbol{\varphi}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}_j)=\sum_{i=1}^{N}(\boldsymbol{\alpha}_k)_i k(\mathbf{x}_i,\mathbf{x}_j), \tag{18}$$

where $\mathrm{NLPC}_k(\mathbf{x}_j)$ and $(\boldsymbol{\alpha}_k)_i$ represent the kth nonlinear principal component for a feature vector $\mathbf{x}_j$ and the ith component of the kth eigenvector $\boldsymbol{\alpha}_k$, respectively.

In nonlinear principal component analysis, centering data, i.e., setting the mean of transformed features to be zero, is not a trivial task, since the nonlinear mapping, $\boldsymbol{\varphi}(\cdot)$, is not explicitly computed. To overcome this difficulty, a centered kernel matrix, $\mathbf{K}^*$, is employed instead of $\mathbf{K}$ in Eq. (16). $\mathbf{K}^*$ is derived by subtracting the mean of the transformed feature vectors, $\boldsymbol{\mu}(\boldsymbol{\varphi}(\mathbf{x}))=(1/N)\Sigma_j\boldsymbol{\varphi}(\mathbf{x}_j)$ from the transformed features. The ijth entry of the centered $\mathbf{K}^*$ matrix, $\mathbf{K}^*_{ij}$, is related to the original kernel matrix, $\mathbf{K}$, in Eq. (17):

$$\begin{aligned}\mathbf{K}^*_{ij} &\equiv \left(\boldsymbol{\varphi}(\mathbf{x}_i)-\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\varphi}(\mathbf{x}_k)\right)^{\mathrm{T}}\left(\boldsymbol{\varphi}(\mathbf{x}_j)-\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\varphi}(\mathbf{x}_k)\right)\\ &=\boldsymbol{\varphi}(\mathbf{x}_i)^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}_j)-\boldsymbol{\varphi}(\mathbf{x}_i)\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\varphi}(\mathbf{x}_k)-\boldsymbol{\varphi}(\mathbf{x}_j)\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\varphi}(\mathbf{x}_k)+\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\boldsymbol{\varphi}(\mathbf{x}_k)^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}_l)\\ &=k(\mathbf{x}_i,\mathbf{x}_j)-\frac{1}{N}\sum_{k=1}^{N}k(\mathbf{x}_i,\mathbf{x}_k)-\frac{1}{N}\sum_{k=1}^{N}k(\mathbf{x}_j,\mathbf{x}_k)+\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}k(\mathbf{x}_k,\mathbf{x}_l).\end{aligned} \tag{19}$$

The kth nonlinear principal components for a feature vector $\mathbf{x}_j$ using the centered kernel matrix can be computed similarly.

4. Sequential probability ratio test

In the previous section, nonlinear principal component analysis is formulated based on unsupervised least-squares support vector machine. For baseline data, $\mathbf{x}^u_j(t)$, obtained from an undamaged structure, AR-ARX coefficients denoted as $\{(b^u_i,c^u_j):\ i,j=1,\ldots,M\}$ are estimated by Eq. (2) and corresponding nonlinear principal components, $\mathrm{NLPC}_k(b^u_i,c^u_j)$, can be extracted.

When a new data set, $\mathbf{x}^d_t$, is measured from a possibly damaged structure, the computation of the AR-ARX coefficients denoted as $\{(b^d_i,c^d_j):\ i,j=1,\ldots,M\}$ and the corresponding nonlinear principal components, $\mathrm{NLPC}_k(b^d_i,c^d_j)$, is repeated. Then, the coefficients, $\{(b^*_i,c^*_j):i,j=1,\ldots,M\}$, whose principal components have the minimum Euclidean distance from those of new data, are selected among (b^u_i,c^u_j) in order to compute the residual errors, $\varepsilon^d_x(t)$ and $\varepsilon^d_y(t)$:

$$\varepsilon^d_x(t)=x^d(t)-\sum_{i=1}^{p}b^d_i x^d(t-i)-\sum_{j=1}^{q}c^d_j e^d_x(t-j),\quad \varepsilon^d_y(t)=x^d(t)-\sum_{i=1}^{p}b^*_i x^d(t-i)-\sum_{j=1}^{q}c^*_j e^d_x(t-j), \tag{20}$$

where $p<M$ and $q<M$. The residual errors, $\varepsilon^d_y(t)$ and $\varepsilon^d_x(t)$, are employed for subsequent damage diagnosis.

The first damage indicator is defined as the standard deviation ratio of two residual errors, i.e., $\sigma(\varepsilon^d_y(t))/\sigma(\varepsilon^d_x(t))$. If a new data set is obtained from a damaged system, the ratio of two standard deviations, i.e.,

$\sigma(\varepsilon_y^d(t))/\sigma(\varepsilon_x^d(t))$, is expected to increase at the damage location as the discrepancy between (b_i^d, c_j^d) and (b_i^*, c_j^*) rises.

Secondly, a simple hypothesis test is performed for damage diagnosis using two standard deviations of the residual errors, i.e., $\sigma(\varepsilon_x^d(t))$ and $\sigma(\varepsilon_y^d(t))$. This test is also based on the premise that damage, if exist, will increase a standard deviation of the residual error $\varepsilon_y^d(t)$ beyond a specified threshold σ_1:

$$H_0 : \sigma(\varepsilon_y^d(t)) \leqslant \sigma_0, \quad H_1 : \sigma(\varepsilon_y^d(t)) \geqslant \sigma_1, \quad 0 < \sigma_0 < \sigma_1 < \infty, \tag{21}$$

where σ_0 and σ_1 are user-specific thresholds that are related with $\sigma(\varepsilon_x^d(t))$. For example, $\sigma_0 = 1.3\sigma(\varepsilon_x^d(t))$ and $\sigma_1 = 1.4\sigma(\varepsilon_x^d(t))$ in this study as identical as those employed in Sohn et al. [17]. Generally speaking, σ_0 and σ_1 are determined by using signals from undamaged cases as well as a few damaged cases [17].

This hypothesis test diagnoses the structure being monitored as undamaged, i.e., accepts the null hypothesis H_0, if $\sigma(\varepsilon_y^d(t))$ is not larger than a user-specified value, σ_0. On the other hand, if $\sigma(\varepsilon_y^d(t))$ becomes equal or larger than σ_1, then the null hypothesis H_0 is rejected and the system is diagnosed as damaged.

A sequential probability ratio test adopted here is one of the simplest statistical inference methods. Compared to the conventional fixed sample size test, the sequential probability ratio test is shown to perform a hypothesis test with a smaller number of observations [16]. Furthermore, the sequential probability ratio test is suitable for continuous monitoring. For a sequence of residual errors, $\{\varepsilon_y^d(t) : t = 1, \ldots, T\}$, a set of data $E = [\varepsilon_y^d(1), \ldots, \varepsilon_y^d(T)]$ can be constructed.

Using the sequential probability ratio test, the hypothesis test in Eq. (21) can be reformatted as follows. At stage $n(\leqslant T)$, the sequential probability ratio test makes three kinds of decisions based on the following Z_n statistics:

$$\text{Accept } H_0 \text{ if } Z_n \leqslant z^u, \quad \text{reject } H_0 \text{ if } Z_n \geqslant z^l, \quad \text{continue monitoring if } z^l \leqslant Z_n \leqslant z^u, \tag{22}$$

where z^l and z^u represent lower and upper bounds of Z_n, respectively. These lower and upper bounds are approximated by Wald [31]:

$$z^u \cong \ln\frac{\beta}{1-\kappa} \quad \text{and} \quad z^l \cong \ln\frac{1-\beta}{\kappa}, \tag{23}$$

where κ and β are user-specified upper bounds for type I and II errors, respectively [31]. Z_n is the transformed random variable and it can be computed as follows:

$$Z_n = \sum_{i=1}^{N} z_i, \quad \text{where } z_i = \frac{1}{2}(\sigma_0^{-2} - \sigma_1^{-2})(\varepsilon_y(i) - \mu)^2 - \ln\frac{\sigma_1}{\sigma_0}, \tag{24}$$

where μ is mean of a normal distribution that defines a conditional distribution for z_i [17].

5. An experimental damage diagnosis example considering data normalization

The proposed method is applied to experimental data sets in order to investigate its capability of damage diagnosis especially in the presence of an unmeasured operational variation. In Oh and Sohn [26], the proposed data normalization technique is applied to the numerical data simulated from a computer hard disk drive model. Data were simulated based on an assumption that dynamic properties of the system depend on temperature, and novelty detection incorporated with kernel principal component analysis was able to detect damage in the presence of temperature variation.

In this study, data normalization capability of the proposed method to an external operational variation is investigated using an eight-dof mass–spring system shown in Fig. 3. For meaningful comparison, the same data set used by Sohn et al. [17] is utilized.

The experimental system is composed of eight masses connected in series with springs. The eight masses are denoted as m_1, m_2,…,m_8 starting from the closest one to the shaker (see Fig. 3). The motions of eight masses are constrained in one direction, i.e., along the rod, and all masses, m_2–m_8, weigh 419.4 g except the first mass, m_1, that weights 559.3 g due to an extra component connected to the shaker. Stiffness of all springs is 56.7 k N m^{-1} and damping is caused primarily by Coulomb friction. The measured quantities are acceleration

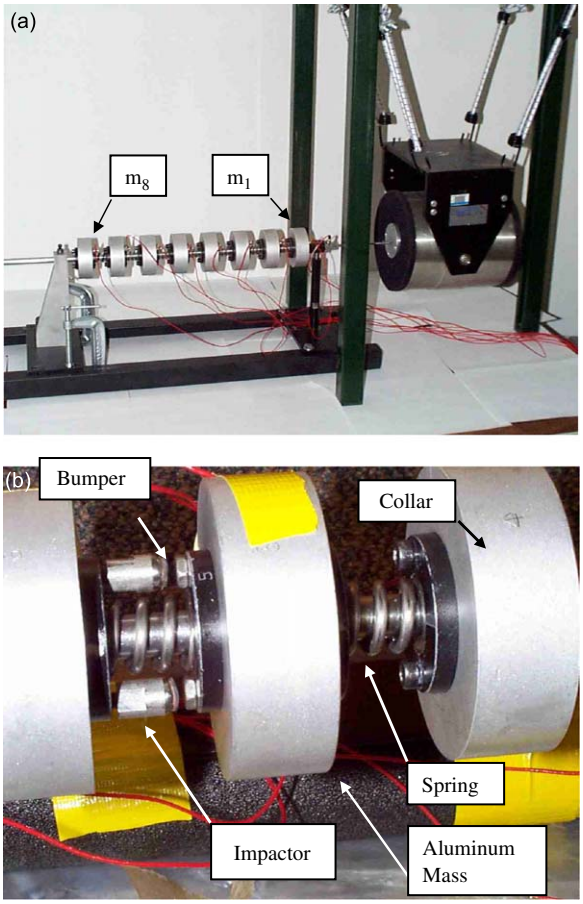

Fig. 3. Experimental setup of an eight-dof system: (a) an eight-dof system attached to a shaker (b) an installed bumper to simulate nonlinear damage.

Table 1
Damage scenarios investigated in this study.

Damage case	Damage location	Input levels (V)	No. of data
0	–[a]	3, 4, 5, 6, 7	75
1	m_1–m_2	3, 4, 5, 6, 7	25
2	m_5–m_6	3, 4, 5, 6, 7	25
3	m_7–m_8	4, 5, 6, 7	20

[a]"–" means no bumper is installed for this case.

responses of all masses to random excitations generated by a 215 N peak force electro-dynamic shaker. Additional details of the experimental setup can be found in Ref. [17].

Nonlinear damage is simulated by an impact of two adjacent masses, and the impact is caused by an installed bumper that limits the movement between those masses. The simulated nonlinear damage can be regarded as the closing of a crack in vibration and initial clearance is set to zero for all damage cases.

Table 1 shows four different damage scenarios selected in this study. Damage case 0 corresponds to an undamaged state. Damage located at m_1, for example, implies that a bumper is installed between m_1 and m_2, and all other damage cases are defined similarly. For each damage case, the root mean square input level varies from 3 to 7 V and the same experiment is repeated 5 times for each input level except damage cases 0 and 3. For damage case 0, the experiment is performed 15 times for each input level, and, for damage case 3,

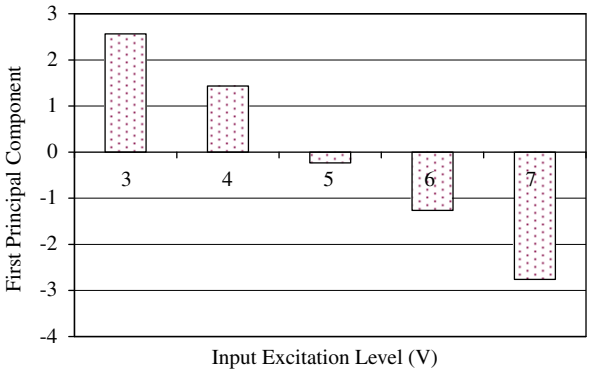

Fig. 4. Correlation between the input excitation (V) and the first principal component at location m_2.

the input level varies from 4 to 7 V instead of 3 to 7 V. Among 75 time signals, i.e., 15 time series from each input level, obtained for the undamaged case, 45 time series are used as baseline data. An operational variation of the eight-dof system is represented by the unmeasured changes of the input levels.

The coefficients (b_i^u, c_j^u) of AR-ARX models are estimated at each dof by using the acceleration responses measured from the undamaged structure: $\{(a_i^u) : i = 1, \ldots, 30\}$ and $\{(b_i^u, c_j^u) : i,j = 1, \ldots, 5\}$ are selected to be consistent with Ref. [17]. The correlation between the operational variation, i.e., the input excitation level, and the first principal component at mass m_2 is shown in Fig. 4. Note that only the first principal component is extracted, since the input excitation level is considered the only operational variability. Monotonically decreasing relationship reveals latent correlation between the first principal component and the input excitation level, which is seemingly undistinguishable but immanent inside of data. Similar monotonic relationships are found at the other mass locations as well. In reality, however, it may be difficult to decide the correct number of nonlinear principal components in advance, since it is not simple to identify all potential environmental and operational variations that affect the extracted features. Note that the proposed kernel principal component analysis is more flexible than auto-associative neural network because it can extract as many principal components as desired. On the other hand, auto-associative neural network needs to re-design the network, i.e., the number of neurons in mapping and de-mapping layers, every time the number of principal components needs to be changed.

For kernal principal component analysis, a width of Gaussian kernel, ρ, needs to be determined prior to the calculation of kernel functions, $k(\mathbf{x}_i,\mathbf{x}_j) = \exp(-||\mathbf{x}_i-\mathbf{x}_j||^2/\rho^2)$, $i,j = 1,\ldots N$. The ρ value is automatically determined to maximize the information (variance) of the first principal component since it is relevant to the operational variation. To accomplish this, the width value that maximizes the difference between the first and second eigenvalues is selected. For example, the difference reaches its maximum when $\rho = 1.95$ for the baseline data obtained from m_8. The widths for all the other locations are estimated in a similar fashion, and they are 0.70, 1.60, 1.00, 0.75, 1.70, 2.30, 2.85 and 1.95 at m_1 through m_8, respectively.

When a new data set is recorded from an unknown state, the first principal component at each dof is computed. Then, $\{(b_i^*, c_j^*) : i,j = 1, \ldots, 5\}$, whose principal component has the minimum Euclidean distance to that of the new data, are selected from the baseline data in order to compute $\varepsilon_y^d(t)$ in Eq. (20). This procedure is repeated for each dof. Note that it is not necessary to measure the input excitation level during this procedure.

The first diagnosis is performed using the standard deviation ratio of $\varepsilon_y^d(t)$ to $\varepsilon_x^d(t)$, i.e., $\sigma(\varepsilon_y^d(t))/\sigma(\varepsilon_x^d(t))$, for damage cases 0–3. When the AR-ARX coefficients are extracted from a possibly damaged structure, the residual error, $\varepsilon_y^d(t)$, is expected to increase with respect to $\varepsilon_x^d(t)$ as stated before. The ratios for all damage cases and mass locations are listed in Table 2. Each entity in Table 2 is the averaged value obtained from 75, 25, 25, and 20 time signals for damage cases 0, 1, 2, and 3, respectively.

The ratio is around 1 when damage is absent, while the ratio reaches its maximum at the damage location. As shown in Table 2, the maximum value of $\sigma(\varepsilon_y^d(t))/\sigma(\varepsilon_x^d(t))$ is observed at the location where the bumper is installed except damage case 1. Similar results have been reported by Sohn et al. [17] as shown in parentheses in Table 2. For damage case 1, it was speculated that responses were masked by the excitation signal, since m_1 was rigidly connected to the shaker.

Table 2
The ratios of standard deviations $\sigma(\varepsilon_y^d)/\sigma(\varepsilon_x^d)$.

Damage case	Degree of freedom							
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
0	1.0043[a]	1.0138	1.0526	1.0285	1.0350	1.0410	1.0294	1.0350
	(1.0021)	(1.0061)	(1.0185)	(1.0106)	(1.0213)	(1.0278)	(1.0226)	(1.0230)
1	1.0144	**1.8651**	1.1514	1.0768	1.0614	1.0391	1.0256	1.0724
	(1.0152)	**(1.5322)**	(1.1246)	(1.0695)	(1.0461)	(1.0373)	(1.0308)	(1.0287)
2	1.0033	1.0155	1.0645	1.0354	**1.7100**	1.2187	1.0571	1.0621
	(1.0024)	(1.0116)	(1.0290)	(1.0282)	**(1.7309)**	(1.2194)	(1.0510)	(1.0347)
3	1.0033	1.0205	1.0510	1.0614	1.0915	1.0672	**1.8225**	1.4707
	(1.0018)	(1.0141)	(1.0347)	(1.0186)	(1.0689)	(1.1765)	**(1.7158)**	(1.3566)

[a]The ratio presented in this table is the averaged value from all input levels. The values in the parenthesis are obtained from Sohn et al. [17] and provided here for comparison. Boldface with an underline indicates the identified damage.

Table 3
Results of a sequential probability ratio test using $H_0 : \sigma(\varepsilon_y^d) \leqslant 1.3\sigma(\varepsilon_x^d)$ and $H_1 : \sigma(\varepsilon_y^d) \geqslant 1.4\sigma(\varepsilon_x^d)$.

Damage case	Degree of freedom							
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
0	0/75	0/75	0/75	0/75	0/75	0/75	0/75	0/75
	(0/75)	(0/75)	(0/75)	(0/75)	(0/75)	(0/75)	(0/75)	(0/75)
1	0/25	**25/25**	0/25	0/25	0/25	0/25	2/25	0/25
	(0/25)	**(25/25)**	(0/25)	(0/25)	(0/25)	(0/25)	(0/25)	(0/25)
2	0/25	0/25	0/25	0/25	**22/25**[a]	2/25	0/25	0/25
	(0/25)	(0/25)	(0/25)	(0/25)	**(23/25)**	(1/25)	(0/25)	(0/25)
3	0/20	0/20	0/20	0/20	0/20	0/20	**20/20**	13/20
	(0/20)	(0/20)	(0/20)	(0/20)	(0/20)	(2/20)	**(20/20)**	(16/20)

[a]'22/25' means that the null hypothesis is rejected 22 times, i.e., damage is detected 22 times, out of all 25 time signals tested. The values in the parenthesis are obtained from Sohn et al. [17] and provided here for comparison. Boldface with an underline indicates the identified damage.

Table 3 presents the second damage diagnosis performed by the sequential probability ratio test. For this sequential probability ratio test, the necessary parameters are set identical to the ones in Ref. [17] ($\beta = \kappa = 0.001$, $\sigma_0 = 1.3\sigma(\varepsilon_x^d(t))$ and $\sigma_1 = 1.4\sigma(\varepsilon_x^d(t))$). As was explained in Ref. [17], these upper and lower bounds, i.e., σ_0 and σ_1, are determined by obtaining measurements from a wide range of an undamaged system as well as of a damaged system to make damage diagnosis more reliable. The entity of '0/75' in Table 3, i.e. the one in the 'Damage Case 0' row and in the 'm_1' column means that the null hypothesis, H_0, is accepted for all 75 time signals tested, and '22/25' means that the null hypothesis is rejected 22 times, i.e., damage is detected 22 times, out of all 25 time signals tested. The location where the maximum number of rejection occurs coincides with the actual damage location. In terms of misclassification, the proposed method is as effective as the previously employed auto-associative neural network.

The efficiency of data normalization is investigated by studying a false positive prediction rate of damage. Table 4 shows false prediction of damage by the sequential probability ratio test that utilizes only a fraction of 45 baseline data. For example, the entity of '5/75' represents that 5 false indications are observed out of 75 number of entire tests when only 9 time signals excited at 3 V input level are selected to comprise $\{(b_i^u, c_j^u) : i,j = 1,\ldots,5\}$. Note that, when the entire 45 baseline signals from a range of 3 to 7 V are used, no false indication of damage is observed among 75 test signals from an undamaged state as shown in the last row of Table 4. Table 4 demonstrates that using a limited data set as the baseline data can increase the false alarm rate. The entities in parentheses in Table 4 present the false positive study reported in Sohn et al. [17].

Table 4
Results of a sequential probability ratio test using $H_0 : \sigma(\varepsilon_y^d) \leqslant 1.3\sigma(\varepsilon_x^d)$ and $H_1 : \sigma(\varepsilon_y^d) \geqslant 1.4\sigma(\varepsilon_x^d)$.

Baseline data	Degree of freedom							
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
3 V	0/75	0/75	0/75	0/75	0/75	0/75	**5/75**[a]	0/75
	(0/75)	(0/75)	**(16/75)**	(0/75)	(0/75)	**(28/75)**	**(43/75)**	(0/75)
4 V	0/75	0/75	0/75	0/75	0/75	**1/75**	0/75	0/75
	(0/75)	**(6/75)**	(0/75)	(0/75)	(0/75)	**(4/75)**	(0/75)	**(4/75)**
5 V	0/75	0/75	0/75	0/75	**1/75**	0/75	0/75	0/75
	(0/75)	(0/75)	(0/75)	(0/75)	(0/75)	**(16/75)**	(0/75)	(0/75)
6 V	0/75	0/75	0/75	0/75	**2/75**	**1/75**	**1/75**	0/75
	(0/75)	(0/75)	(0/75)	(0/75)	(0/75)	(0/75)	**(5/75)**	**(11/75)**
7 V	0/75	0/75	0/75	0/75	**2/75**	**7/75**	0/75	0/75
	(0/75)	(0/75)	(0/75)	(0/75)	**(9/75)**	**(4/75)**	(0/75)	**(69/75)**
3–7 V	0/75	0/75	0/75	0/75	0/75	0/75	0/75	0/75

[a]Each value in this table shows the number of null hypothesis rejections obtained by sequential probability ratio test when a portion of baseline data is utilized to construct $\{(b_i^u, c_j^u), i,j = 1,\ldots,5\}$. For example, '5/75' indicates that the null hypothesis is rejected 5 times out of 75 entire signals when only time signals measured at 3 V input level are used as baseline data. The values in the parenthesis are obtained from Sohn et al. [17] and provided here for comparison. Boldface with an underline indicates the identified false positive prediction.

Table 5
The effect of limited baseline data on the sequential probability ratio test using $H_0 : \sigma(\varepsilon_y^d) \leqslant 1.3\sigma(\varepsilon_x^d)$ and $H_1 : \sigma(\varepsilon_y^d) \geqslant 1.4\sigma(\varepsilon_x^d)$: only 27 time signals in the range of 3 V, 4 V and 5 V are utilized as baseline data.

Damage case	Degree of freedom							
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
0	0/75	0/75	0/75	0/75	0/75	0/75	0/75	0/75
1	0/25	**25/25**[a]	0/25	0/25	0/25	0/25	2/25	0/25
2	0/25	0/25	0/25	0/25	**23/25**	4/25	0/25	0/25
3	0/20	0/20	0/20	0/20	0/20	0/20	**20/20**	17/20

[a]Each value in this table shows the number of null hypothesis rejections obtained by sequential probability ratio test when a limited baseline data are utilized to construct $\{(b_i^u, c_j^u), i,j = 1,\ldots,5\}$. For example, '25/25' indicates that the null hypothesis is rejected 25 times out of 25 entire signals when only time signals measured at 3 V, 4 V, and 5 V input level are used as baseline data. Boldface with an underline indicates the identified damage.

Note that the same subset of the entire 45 data sets was employed to construct the baseline data. Comparison reveals that kernel principal component analysis produces a smaller number of false alarms than auto-associative neural network.

In Sohn et al. [17], it was claimed that baseline data need to be acquired from a wide range of environmental and operational conditions in order to make reliable damage prediction. In this study, generalization performance of kernel principal component analysis is examined to see if reliable damage detection can be achieved when only a limited range of baseline data is available. Table 5 shows data normalization capability of the proposed kernel principal component analysis by using a limited number of baseline data, i.e., 27 time signals consisting of 9 acceleration histories from each of 3, 4 and 5 V input excitation levels. Note that the previous sequential probability ratio test in Table 3 utilizes 45 baseline signals obtained from 3 to 7 V input excitation levels. Except the number of the baseline data, all the other parameters in Table 5 are kept identical to the ones in Table 3. Damage diagnosis is performed via the sequential probability ratio test for each damage case defined in Table 1.

Sequential probability ratio test results performed along with kernel principal component analysis in Table 5 in comparison with those along with auto-associative neural network in Table 3 indicate that the

proposed method may be applicable for data normalization even when all possible environmental and operational conditions are not fully covered during the construction of the baseline data; boldfaces with underlines such as 25/25, 23/25, and 20/20 for damage case 1, 2, and 3, respectively, indicate the number of identified damage, which is identical for both methods. However, this finding should not be generalized in haste, since it is concluded only from one experimental example. Note that a large amount of data spanning a wide range of operational conditions is typically required to achieve reliable data normalization. The results presented in Tables 4 and 5 indicate that the success of data normalization depends on not only the availability of baseline data but also the nonlinear principal component analysis method selected to characterize the latent relationship. In the next section, advantages of kernel principal component analysis over auto-associative neural network are addressed.

6. Comparison between kernel principal component analysis and auto-associative neural network

There are several ways to realize nonlinear principal component analysis including auto-associative neural network, kernel principal component analysis, principal curves, and so on. Here, advantages of kernel principal component analysis over auto-associative neural network are described. Refer to Diamantaras and Kung [32] for more details of other nonlinear principal component analysis algorithms.

The first advantage of kernel principal component analysis is its explicit use of regularization and better generalization ability due to regularization. Regularization is one of the techniques that control over-fitting by adding a penalty term, for example, $\mathbf{w}^T\mathbf{w}$ in an objective function [29]. Regularization is known not only to prevent over-fitting but also to enhance prediction capability of new data sets [29]. Table 5 demonstrates the generalization performance of kernel principal component analysis. Comparison between Tables 3 and 5 reveals that the proposed method using kernel principal component analysis is able to achieve damage diagnosis comparable to the one obtained by auto-associative neural network although a smaller set of acceleration signals are used for baseline data: the baseline data used for kernel principal component analysis consist of the signals excited by 3 to 5 V input levels while the entire baseline data excited by 3 to 7 V are employed for auto-associative neural network.

Secondly, it should be emphasized that the solution obtained by kernel principal component analysis is unique and global, since kernel principal component analysis solves a simple eigenvalue problem. On the other hand, auto-associative neural network requires solving a complex nonlinear optimization problem by starting the optimization from several different initial values to avoid local minima.

In addition, kernel principal component analysis is able to control the number of extracted principal components pertaining to environmental and operational parameters without re-designing the algorithm. For the eight-dof mass–spring system examined in this study, only the first principal component is extracted since there is only one operational variation: the excitation level. However, because the number of operational and environmental parameters is not *known a priori* in practice, the flexibility in extracting as many principal components as necessary without additional efforts is advantageous. Note that the architecture of auto-associative neural network needs to be re-designed every time when different numbers of principal components are extracted.

Finally, kernel principal component analysis is computationally efficient. Since kernel principal component analysis solves a simple eigenvalue problem, it is computationally more attractive than auto-associative neural network that needs to solve a complex nonlinear optimization problem. This advantage renders the proposed method to be suitable for on-line health monitoring.

7. Conclusions

This paper introduces a data normalization method that can be employed for damage diagnosis in the presence of environmental and operational variations. Kernel principal component analysis, i.e., nonlinear principal component analysis based on unsupervised support vector machine, is integrated with an AR-ARX discrete-time prediction model and a sequential probability ratio test in order to distinguish the effect of damage from those caused by environmental and operational variations on damage diagnosis.

The proposed damage diagnosis method can be summarized as follows: (1) the coefficients of an AR-ARX model are estimated by fitting the AR-ARX model to the time signals obtained from the pristine condition of the structure, and these coefficients are used as the baseline data for subsequent kernel principal component analysis; (2) kernel principal component analysis performs nonlinear principal component analysis and characterizes the nonlinear correlations between the extracted AR-ARX coefficients and unmeasured operational parameters for data normalization; (3) when new time signals are obtained from a potentially damaged structure, the computation of the AR-ARX coefficients and the corresponding nonlinear principal components is repeated as before; (4) the closest AR-ARX coefficients, whose principal components have the minimum Euclidean distance to those of new data, are selected from the baseline data; (5) if a new data set is obtained from a damaged condition, the discrepancy between the AR-ARX coefficients estimated from the potentially damaged condition and the closest AR-ARX coefficients obtained from the baseline data is expected to increase. Subsequently, the prediction errors obtained by fitting the closest baseline AR-ARX model to the new data set will increase with respect to the prediction errors obtained using the new AR-ARX model. Based on this premise, a damage classifier is developed using sequential probability ratio test.

The eight-dof mass–spring example presented in this study demonstrates that the proposed method is able to detect damage under a time-varying excitation condition without explicitly measuring the excitation level. Compared with auto-associative neural network previously reported in Ref. [17], kernel principal component analysis produces smaller number of false alarms and provides better damage diagnosis even when a limited number of baseline data is used.

In view of the presented results, this study demonstrates that the proposed method can be a promising on-line tool for data normalization. Several advantages that characterize the proposed method are:

(1) reliable damage diagnosis with smaller numbers of misclassification (see Tables 2 and 3),
(2) better generalization performance in damage diagnosis (see Tables 4 and 5),
(3) assurance of a unique and global solution,
(4) flexibility in adjusting the number operational and environmental parameters without re-designing the algorithm,
(5) rapid operation.

So far, the proposed method has been applied to a simulated data set from a computer hard disk drive model [26] and an experimental data set from the eight-dof system. Additional studies are underway to investigate the robustness of the proposed method in harsh field environments.

Acknowledgements

This research is supported by the Radiation Technology Program under Korea Science and Engineering Foundation (KOSEF) and the Ministry of Science and Technology (M20703000015-07N0300-01510) and Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2007-331-D00462).

References

[1] H. Sohn, Effects of environmental and operational variability on structural health monitoring, *A Special Issue of Philosophical Transactions of the Royal Society A on Structural Health Monitoring* 365 (2007) 539–560.
[2] M.G. Wood, Damage Analysis of Bridge Structures Using Vibrational Techniques, PhD Thesis, University of Aston, Birmingham, UK, 1992.
[3] S. Moorty, C.W. Roeder, Temperature-dependent bridge movements, *ASCE Journal of Structural Engineering* 118 (1992) 1090–1105.
[4] Y. Fujino, Y. Yoshida, Wind induced vibration and control of trans-Tokyo bay crossing bridge, *ASCE Journal of Structural Engineering* 128 (8) (2002) 1012–1025.
[5] C.Y. Kim, D.S. Jung, N.S. Kim, J.G. Yoon, Effect of vehicle mass on the measured dynamic characteristics of bridges from traffic-induced test, *Proceedings of the IMAC-XIX*, Kissimmee, FL, February 2001, pp. 1106–1111.
[6] Q.W. Zhang, L.C. Fan, W.C. Yuan, Traffic-induced variability in dynamic properties of cable-stayed bridge, *Earthquake Engineering and Structural Dynamics* 31 (2002) 2015–2021.

[7] H. Sohn, C.K. Oh, *Statistical Pattern Recognition in Structural Health Monitoring. Encyclopedia of Structural Health Monitoring*, Wiley, New York, 2009.
[8] B. Peeters, G. De Roeck, One year monitoring of the Z24-bridge: environmental influences versus damage events, *Proceedings of IMAC-XVIII*, San Antonio, TX, February, 2000, pp. 1570–1576.
[9] K. Worden, H. Sohn, C.R. Farrar, Novelty detection in a changing environmental: regression and interpolation approaches, *Journal of Sound and Vibration* 258 (4) (2002) 741–761.
[10] R.G. Rohrmann, M. Baessler, S. Said, W. Schmid, W.F. Ruecker, Structural causes of temperature affected modal data of civil structures obtained by long time monitoring, in: *Proceedings of the XVII International Modal Analysis Conference*, Ksskmmee, FL, February 1999.
[11] J. Kullaa, Is temperature measurement essential in structural health monitoring, *Proceedings of the Fourth International Workshop on Structural Health Monitoring*, Stanford University, CA, September 2003.
[12] R. Ruotolo, C. Surace, Damage detection using singular value decomposition, *Proceedings of DAMAS 97*: *Structural Damage Assessment using Advanced Signal Processing Procedures*, University of Sheffield, Sheffield, June–July 1997.
[13] D. Chelidze, M. Liu, Reconstructing slow-time dynamics from fast-time measurements, *Philosophical Transactions of the Royal Society A* 366 (2008) 729–745.
[14] S.B. Kim, H. Sohn, Instantaneous reference-free crack detection based on polarization characteristics of piezoelectric materials, *Smart Materials and Structures* 16 (2007) 2375–2387.
[15] H. Sohn, C.R. Farrar, Damage diagnosis using time series analysis of vibration signals, *Smart Materials and Structures* 10 (2001) 446–451.
[16] B.K. Ghosh, *Sequential Tests of Statistical Hypotheses*, Addison-Wesley, Menlo Park, CA, 1970.
[17] H. Sohn, K. Worden, C.R. Farrar, Statistical damage classification under changing environmental and operational conditions, *Journal of Intelligent Materials, Systems and Structures* 13 (9) (2002) 561–574.
[18] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis: Forecasting and Control*, Prentice Hall, Upper Saddle River, NJ, 1994.
[19] L. Ljung, *System Identification—Theory for the User*, Prentice Hall, Upper Saddle River, NJ, 1999.
[20] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, Wiley-Interscience, New York, 2000.
[21] I.T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.
[22] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, 1996.
[23] T. Hastie, W. Stuetzle, Principal curves, *Journal of the American Statistical Association* 84 (1989) 502–516.
[24] E. Oja, A simplified neuron model as a principal component analyzer, *Journal of Mathematical Biology* 15 (1982) 267–273.
[25] B. Schölkopf, A. Smola, K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation* 10 (1998) 1299–1319.
[26] C.K. Oh, H. Sohn, Unsupervised support vector machine based principal component analysis for structural health monitoring, *Proceedings of the 15th ICCES Conference*, Honolulu, Hawaii, March 2008.
[27] V.N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
[28] B. Schölkopf, A. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.
[29] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.
[30] J.A.K. Suykens, T.V. Gestel, J.D. Brabanter, B.D. Moor, J. Vandewalle, *Least-squares Support Vector Machines*, World Scientific Publishing Co. Pte. Ltd., Singapore, 2002.
[31] A. Wald, *Sequential Analysis*, Wiley, New York, 1947.
[32] K.I. Diamantaras, S.Y. Kung, *Principal Component Neural Networks*, Wiley, New York, 1996.