STRUCTURAL CONTROL AND HEALTH MONITORING
Struct. Control Health Monit. 2017; 24: e1851
Published online 23 February 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/stc.1851

Modeling and harnessing sparse and low-rank data structure: a new paradigm for structural dynamics, identification, damage detection, and health monitoring

Satish Nagarajaiah[1,2,*,†] and Yongchao Yang[3]

[1]*Department of Civil and Environmental Engineering, Rice University, Houston, TX 77005, USA*
[2]*Department of Mechanical Engineering, Rice University, Houston, TX 77005, USA*
[3]*Los Alamos National Laboratory, Los Alamos, NM 87545, USA; formerly Department of Civil and Environmental Engineering, Rice University, Houston, TX 77005, USA*

SUMMARY

This paper presents a new paradigm of explicitly modeling and harnessing the data structure to address the inverse problems in structural dynamics, identification, and data-driven health monitoring. In particular, it is shown that the structural dynamic features and damage information, intrinsic within the structural vibration response measurement data, possesses sparse and low-rank structure, which can be effectively modeled and processed by emerging mathematical tools such as sparse representation and compressed sensing, low-rank matrix decomposition and completion, as well as the unsupervised multivariate blind source separation. It is also discussed that explicitly modeling and harnessing the sparse and low-rank data structure could benefit future work in developing data-driven approaches toward rapid, unsupervised, and effective system identification, damage detection, as well as massive SHM data sensing and management. Copyright © 2016 John Wiley & Sons, Ltd.

Received 20 March 2015; Revised 20 November 2015; Accepted 24 January 2016

KEY WORDS: structural dynamics; system identification; structural health monitoring; sparse representation; low-rank representation; compressed sensing; machine learning; blind source separation

1. INTRODUCTION

During service, civil structures are subjected to operational loads and environmental effects, as well as various natural disasters (e.g., earthquakes and hurricanes) and man-made extreme events (e.g., blasts and impacts). Assessing health status and detecting damage of the structure as early as possible is essential to ensure structural integrity. It allows prompt maintenance and thus reduces the repair cost; in addition, timely damage information makes possible for informed decisions and immediate actions before catastrophic failure of structures occurs. To achieve this goal, structural health monitoring (SHM) systems with an array of networked sensors have been developed to continuously measure structural data for monitoring and assessing structural performance.

Vibration-based response measurements (e.g., strains, displacements, and accelerations) and analysis techniques such as modal analysis-based system identification and damage detection methods have been widely studied for SHM [1]. Traditional modal identification typically complies with the principle of system identification, which is based on the relationship of inputs and outputs [2]. This corresponds to an ideal situation where excitation to the system can be controlled or measured. For civil structures, typically large scale (e.g., bridges, buildings, dams, etc.), it is extremely difficult or expensive, if not

*Correspondence to: Satish Nagarajaiah, Department of Civil and Environmental Engineering and Department of Mechanical Engineering, Rice University, Houston, TX 77005, USA.
†E-mail: Satish.Nagarajaiah@rice.edu

Copyright © 2016 John Wiley & Sons, Ltd.

impossible, to apply controlled excitation to conduct input–output modal analysis. Accurate measurement of the ambient excitation (e.g., wind, traffic, etc) to structures is also challenging. Therefore in practical applications, it is often required to identify the structural dynamic properties and health status from only the available structural vibration response measurement data. This is essentially an ill-posed inverse problem, which hardly has analytical solutions. However, with some additional information and appropriate assumptions, one could hope to find solutions that may be sufficient in structural dynamics and health monitoring.

Solving the ill-posed inverse problem where only the structural vibration response measurements are available needs additional, prior, knowledge or assumption. If detailed knowledge of the structure is available, including material property, geometry, component connections and joints, boundary conditions, and so on, a common approach to solving the inverse problem is to build a physics-based or physical model of the structure, such as a finite element model, as the reference information of the initially healthy structure. Afterwards, the structural model is updated by fitting the model-predicted responses with the current structural responses (usually modal parameters) [3]. In the context of the need of performing output-only modal parameters identification from the current structural vibration response measurements, many established methods, such as Ibrahim time domain method [4], eigensystem realization algorithm [5], and stochastic subspace identification [6], (note: frequency domain decomposition (FDD) [7] is non-parametric) include a process of building a parametric dynamic model such as state space model, and then estimating the dynamic parameters of the dynamic model by fitting the structural response measurements. Finally, one obtains the system or dynamic parameters (e.g., by eigen analysis) from the updated structural model, and the discrepancy between the updated and reference models (physical or modal models) indicates structural damage.

While a reliable physics-based model containing comprehensive structural information is most desirable for SHM, such a parametric model-based approach requires an accurate model of the physical property of the structure (materials, geometry, etc.), which is often difficult and can be liable to model uncertainty and error [8]. In addition, the modal updating procedure is typically computationally demanding and requires expert attendance for parameter adjustments associated with the model fitting and parameters estimation process. For example, the model order problem [9] remains a challenge, for which using the stability chart demands expert interference and time-consuming computation: although effective for offline applications, they may not be suitable for real-time unsupervised processing of the large-scale data sets of civil structures.

An alternative approach is to directly exploit the available structural vibration response measurement data itself. Unlike parametric model-based methods, which are derived from the (mathematically) physical processes, data-driven approaches aim to extract the desirable information directly from the available data, without explicit knowledge of the physical or dynamic model of the underlying system. The non-parametric data-driven algorithms are efficient and have potential for real-time processing the massive SHM data.

Many signal processing-based system identification and damage detection algorithms that have been developed in the literature fall into this category, featuring efficient computation and adaptive implementation, such as those based on wavelet transform (WT) [10], Hilbert–Huang transform [11,12], and other time-frequency analysis techniques [13], to name a few. Successful implementations, however, require practitioners to wisely adjust the algorithm parameters; for example, the wavelet basis and the scales need to be carefully selected in the wavelet-based methods, and the prescription of the modal bandwidth as well as the sifting process also influence the abilities of such methods. In addition, the salient structural dynamic features and damage features in data have not been explicitly modeled or characterized with wide acceptance. For example, different signal processing methods give different features or damage indices such as wavelet coefficients and statistics-based indices, whose interpretations may not be an easy task.

While traditional research on SHM mainly exploits either the physical model or the use of different signal processing techniques, this paper contributes to present an alternative paradigm of explicitly modeling and harnessing the inherent data structure itself of the structural vibration response data to extract the desirable structural features and damage information, otherwise invisible. Particularly, the salient structural features and damage information intrinsic within the structural vibration response measurement data, usually large scale in SHM, possesses *sparsity nature and low-rank structure*,

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

which could be effectively modeled and processed by emerging mathematical tools such as sparse representation [14], compressed sensing (CS) [15], low-rank matrix decomposition and completion [16], as well as the unsupervised blind source separation (BSS) [17], toward rapid (close to real-time), automated, and effective system identification, damage detection, as well as massive SHM data management.

In this context, this paper attempts to provide physical interpretation and model of the data structure—sparsity and low-rank—to address the inverse problems of interest. It should be mentioned that the detailed mathematical theory of sparse representation and CS has been well documented in other fields and it is briefly reviewed in this paper. Also, it is not the attempt of this paper to provide comprehensive review of the large family of techniques in SHM; instead, the main focus of this paper is to present, based on the recent work by the authors, how this new approach—explicitly exploiting the data structure itself—could be taken advantage of to help address the structural dynamics, identification, monitoring, and data sensing and management problems in an innovative, efficient, and effective manner. It is finally discussed that a unified model of the data structure and characterization of the system dynamic and damage features could benefit some future work in structural dynamics, identification, and health monitoring. A framework with the presented new paradigm for the SHM process is shown in Figure 1.

The remainder of the paper is organized as follows. Section 2 gives mathematical definition and physical interpretations of sparse and low-rank representation. Section 3 presents the new insight and approach to address a few structural dynamics and identification, and data sensing and management problems in SHM. Conclusions and future work are discussed in Section 4.

2. DEFINITION AND MODELING OF SPARSITY AND LOW-RANK

2.1. Sparse representation

To mathematically express sparsity of a signal $x \in \mathbb{R}^N$, it is useful to define the ℓ_0-norm [14],

$$\|x\|_{\ell_0} = \#\{i : x_i \neq 0\} \tag{1}$$

simply counting the number of nonzeros in x. A signal x (vector) is K-sparse if it has at most K

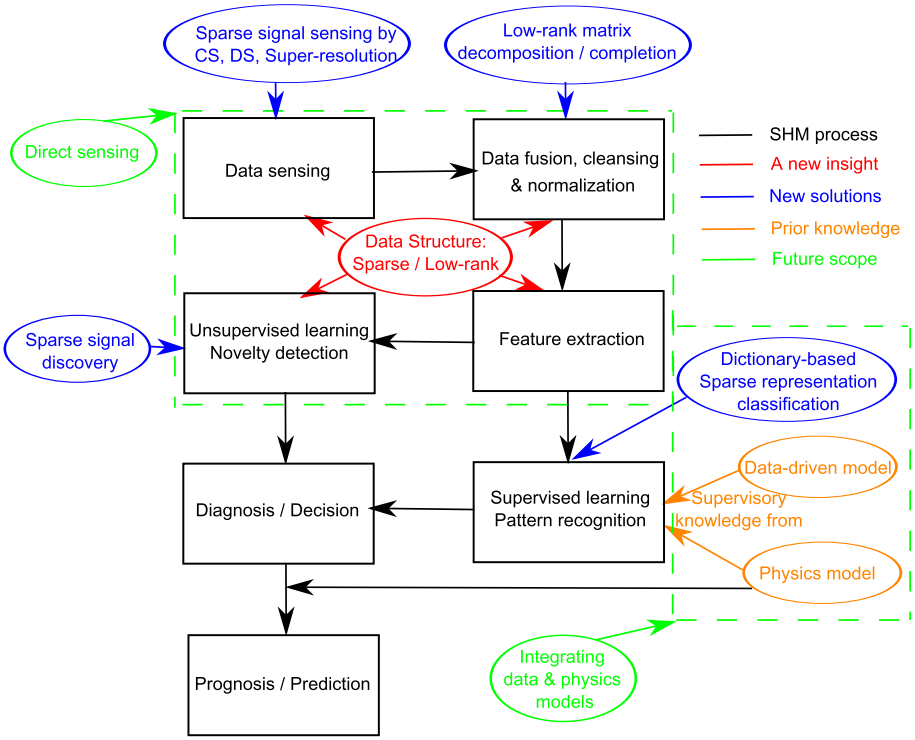

Figure 1. The framework of the new paradigm of explicitly modeling the sparse and low-rank data structure (presented in this paper) for structural health monitoring.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

nonzeros, that is, $\|x\|_{\ell_0} \leq K$. In analogy, a matrix $\mathbf{X}$ is also said to be sparse if most of its elements are zero.

In a more general perspective, x is said to be K-sparse (transform sparse) in a domain $\mathbf{\Psi}$ with a representation $\alpha \in \mathbb{R}^N$

$$x = \mathbf{\Psi}\alpha = \sum_{j=1}^{N} \alpha_j \psi_j \tag{2}$$

if $\|\alpha\|_{\ell_0} \leq K$. $\mathbf{\Psi} = [\psi_1, \ldots, \psi_N]^T \in \mathbb{R}^{N \times N}$ is an orthonormal basis (e.g., sinusoid, wavelet, etc), whose jth row is $\psi_j \in \mathbb{R}^N$ (or $\mathbb{C}^N$on Fourier basis). $\alpha \in \mathbb{R}^N$ is the coefficient sequence of $x \in \mathbb{R}^N$ on $\mathbf{\Psi}$, whose jth element $\alpha_j = \langle x, \psi_j \rangle$ (inner product). This generalization is particularly useful, because, in practice, x is typically sparse in an appropriate domain instead of its original domain. A simple example is the sinusoid, which is sparsest ($K = 1$) in the frequency domain. This actually underlies a sparse probability density function whose most elements are concentrated on the zero.

From a statistical view, a sparse distribution is easier to predict, while a uniform distribution provides little clue to trace. In fact, if of equal variance, a Gaussian-distributed variable ($\mathrm{p}(v) = 1/\sqrt{2\pi}e^{-v^2/2}$) is the most random or unstructured one [18]. Sparse distribution, such as Laplace distribution ($\mathrm{p}(v) = 1/\sqrt{2}e^{-\sqrt{2}|v|}$), has been extensively used in sparse models [14]. Figure 2 shows that the Laplace distribution is much more spiky than the Gaussian distribution (both normalized).

It turns out that the structural dynamic features and damage features of interest inherent in the structural vibration response measurement data are naturally sparse and can be readily revealed by the mathematical tools of sparse representation. In this paper, it has been a useful thread to explicitly exploit such data structure toward developing innovative data-driven system identification and damage detection approaches.

2.2. Low-rank structure

Structural vibration response measurements, from potentially hundreds of channels or sensors, can be represented as a data matrix. Analogous to the sparsity property of single-channel data (vector), the intrinsic low-dimensional data structure of multi-channel data matrix is also explicitly exploited and modeled, for example, by singular value decomposition (SVD) or principal component analysis (PCA) [19].

The data matrix $\mathbf{X} \in \mathbb{R}^{m \times N}$ with m sensors and N time history sampling points ($m < N$) has an SVD representation (Figure 3)

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \tag{3}$$

where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}$ is an orthonormal matrix associated with the channel (variable) dimension, called left-singular vectors or principal component directions; $\mathbf{\Sigma} \in \mathbb{R}^{m \times N}$ has m diagonal elements σ_i as the ith singular value ($\sigma_1 > \ldots > \sigma_r > \sigma_{r+1} = \ldots = \sigma_m = 0$), and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N] \in \mathbb{R}^{N \times N}$ is associated

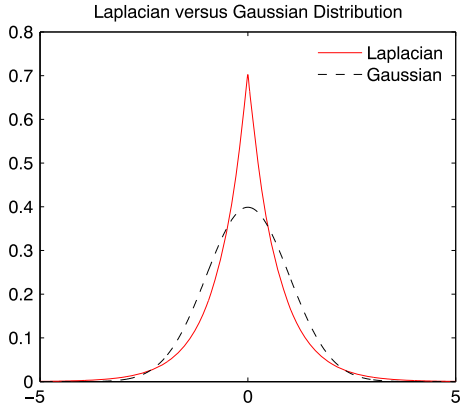

Figure 2. The probability density functions of standardized Laplacian and Gaussian distributions.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

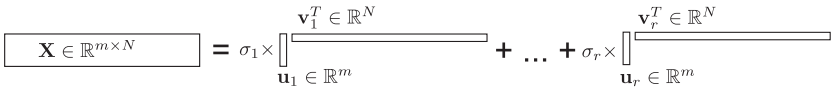

Figure 3. Interpretation of singular value decomposition. The singular value decomposition of the data matrix $\mathbf{X} \in \mathbb{R}^{m\times N}$ (e.g., m sensors and N time history points) as a linear combination of r active singular vector subspaces. If $r \ll \min(m, N) = m$, then explicitly $\mathbf{X}$ has a low-rank structure.

with the time history (measurement) dimension, called the right-singular vector matrix. SVD is closely related to the eigenvalue decomposition (EVD): the left-singular vector matrix $\mathbf{U}$ is obtained by the EVD of its covariance matrix

$$\mathbf{X}\mathbf{X}^T = \mathbf{U}\widehat{\mathbf{\Sigma}}^2\mathbf{U}^T \tag{4}$$

and similarly for the right-singular vector matrix $\mathbf{V}$,

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\widetilde{\mathbf{\Sigma}}^2\mathbf{V}^T \tag{5}$$

where $\widehat{\mathbf{\Sigma}} \in \mathbb{R}^{m\times m}$ and $\widetilde{\mathbf{\Sigma}} \in \mathbb{R}^{N\times N}$ are the zero-truncated and zero-padded versions of $\mathbf{\Sigma} \in \mathbb{R}^{m\times N}$, respectively. $\mathbf{X}$ is said to be low-rank if it has only few active (nonzero) singular values ($r \ll \min(m, N)$).

It is well understood that the ith singular value σ_i is related to the energy captured by the ith principal direction of $\mathbf{X}$. In structural dynamics, under some assumption, the principal directions would coincide with the mode directions [20] with the corresponding singular values indicating their participating energy in the structural responses $\mathbf{X}$, that is, the structural active modes are captured by r principal components under broadband excitation.

An empirical, but frequently sound, observation is that there are typically only few active modes in the structural vibration response measurements [21]; in other words, few of its singular values are active: r is typically quite small. If the sensor or channel number m is reasonably large, then $r \ll \min(m, N) = m$ and $\mathbf{X} \in \mathbb{R}^{m\times N}$ are said to be low rank. However, this is seldom so for large civil structures, because the sensor number m is not so much larger than (often times even less than) the number of the involved r modes; as a result, $r \ll m$ cannot be guaranteed for a low-rank representation.

A simple yet effective strategy—rank-invariant matrix reshape [22]—has been proposed to guarantee a low-rank representation of structural response data matrix, regardless of the original dimension of $\mathbf{X} \in \mathbb{R}^{m\times N}$. Essentially, mode information (few are active; hence, the rank of the structural response data matrix is small) remains approximately invariant (small) regardless of the reshape of the structural response data matrix. The low-rank structure can also model the highly correlated, slowing-changing background or irrelevant components, which will be explained in detail in Section 3.4.2.

2.3. Blind source separation

Real-time SHM requires continuous and efficient processing of the massive measured data with as little expert attendance as possible. BSS as a promising unsupervised multivariate machine learning technique is able to recover the hidden source signals and their characteristic factors using only the measured mixture signals, with high potential in unsupervised learning of the patterns and features hidden in the large-scale multi-channel SHM data set.

The linear instantaneous BSS model [17] is expressed as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) = \sum_{i=1}^{n} \mathbf{a}_i s_i(t) \tag{6}$$

where $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]^T$ is the observed mixture vector with m mixture signals and $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^T$ is the latten source vector with n sources; $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the unknown constant mixing matrix consisting of n columns with its ith column $\mathbf{a}_i \in \mathbb{R}^m$ associated with $s_i(t)$.

With only $\mathbf{x}(t)$ known, Eq. (6) may not be mathematically solved. To alleviate the problem, most BSS techniques, such as independent component analysis (ICA) [17], second-order blind identification (SOBI) [23], and complexity pursuit (CP) [24], exert a general assumption that the source signals $\mathbf{s}(t)$ are statistically independent (or as independent as possible) at each time instant t, and recover the components $\mathbf{y}(t) = [y_1(t), \ldots, y_n(t)]^T$ that are as mutually independent as possible

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \tag{7}$$

such that $\mathbf{y}(t) = \mathbf{s}(t)$ and $\mathbf{W} = \mathbf{A}^{-1}$.

In particular, ICA biases to recover sparse components that are of interest. The principle of ICA for estimation of the BSS model is based on the classic central limit theorem, which states that a sum of independent random variables distributes toward Gaussian; this implies for the generative BSS model Eq. (6) that a mixture x_i ($i = 1, \ldots, n$) is always more Gaussian than any of its individual one s_i. In the ICA de-mixing model Eq. (7), on the other hand, ICA finds $y_i = \mathbf{w}_i^T\mathbf{x} = \mathbf{w}_i^T\mathbf{A}\mathbf{s} = \mathbf{z}_i^T\mathbf{s}$, which is also a mixture of $\mathbf{s}$ and more Gaussian than any s_i but becomes least Gaussian when it equals s_i while rotating $\mathbf{w}_i^T$. Following such a principle, ICA's learning rule of maximization of non-Gaussianity would find the sources and the de-mixing matrix $\mathbf{W}$ (also $\mathbf{A}$), simultaneously. This learning rule is implemented by maximizing some contrast function of non-Gaussianity (independence), for example, negentropy

$$\mathrm{J}(y_i) = \mathrm{H}(y_{gau}) - \mathrm{H}(y_i) \tag{8}$$

where $\mathrm{H}(\cdot)$ is the entropy of a standardized random variable y_i and Gaussian random variable y_{gau}, measuring its uncertainty or randomness. A (standardized) Guassian random variable has largest entropy while one with a sparse probability density function possesses very small entropy. This is quite intuitively understood: Guassian is very random and unstructured, whereas a random variable with a sparse representation is much structured and easy to predict. This implies that maximization of negentropy propels ICA to recover sparse components. Figure 2 shows the probability density functions for a standardized Gaussian variable with flat shape, compared with the spiky Laplacian distribution, which is usually used for sparse models. In practice, negentropy is approximated by

$$\mathrm{J}(y_i) \propto \left[\mathrm{E}\{\mathrm{G}(y_i)\} - \mathrm{E}\left\{\mathrm{G}(y_{gau})\right\}\right]^2 \tag{9}$$

where $\mathrm{G}(\cdot)$ is some chosen function and $\mathrm{E}(\cdot)$ is the expectation operator. If taking $\mathrm{G}(y_i) = y_i^4$, then it becomes the well-known kurtosis (4th-order statistics) approximation. This family of functions is pointed out to possess the ability of promoting sparsity [14], which is the target structural dynamic and damage features of interest.

3. IMPLICATIONS OF THE SPARSE/LOW-RANK DATA STRUCTURE IN STRUCTURAL DYNAMICS AND SHM

3.1. Sparse representation and clustering of modal expansion

For an n-DOF linear time-invariant system, its equation of motion is

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = \mathbf{f}(t) \tag{10}$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are constant mass, diagonalizable damping, and stiffness matrices, respectively, and are real-valued and symmetric; $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]^T$ is the system response (displacement) vector, and $\mathbf{f}(t)$ is the external force vector. Under broadband excitation, the coupled $\mathbf{x}(t)$ may be expressed as linear combinations of the decoupled modal responses

$$\mathbf{x}(t) = \mathbf{\Phi}\mathbf{q}(t) = \sum_{i=1}^{n} \varphi_i q_i(t) \tag{11}$$

Unlike classic input–output system identification, output-only identification pursues to identify the modal parameters only from the knowledge of $\mathbf{x}(t)$ without the excitation or input information to the system, like identification of both $\mathbf{\Phi}$ and $\mathbf{q}(t)$ only from $\mathbf{x}(t)$ in Eq. (11). Such is an ill-posed problem and may not be solved mathematically. The challenges are that (1) existing time domain (stochastic subspace identification and eigensystem realization algorithm) output-only modal identification methods rely on parametric model (state-space model) fitting associated with the model order issue (e.g., spurious numerical modes); (2) the frequency domain method FDD usually requires users to judge the mode and is not well-suited for highly damped or complex modes. The emerging BSS-based methods [25] such as ICA is restricted to undamped and very lightly damped structures [26]; SOBI methods meet with difficulty in closely spaced modes, non-diagonalizable damping, and non-

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

stationary excitation [27–29]. Modified ICA and SOBI-based methods [28–33] have been developed to overcome some of these issues successfully.

3.1.2. Sparse clustering of modes. The spectral sparsity and spatially disjoint of the monotone modal responses was explicitly exploited by a new method, sparse component analysis (SCA) [34,35]. Transform Eq. (11) into the frequency domain f,

$$\mathbf{x}(f) = \mathbf{\Phi}\mathbf{q}(f) = \sum_{i=1}^{n} \varphi_i q_i(f) \tag{12}$$

Attributed to the spatially disjoint sparsity of q_j ($j=1,\ldots,n$), which is active only at f_k (the modal frequency of the jth mode) and elsewhere $f \neq f_k$, $q_j(f)=0$, Eq. (12) becomes

$$\mathbf{x}(f_k) = \varphi_j q_j(f_k) \tag{13}$$

which means that there is only a scale difference, $q_j(f_k)$, between $\mathbf{x}(f_k)$ and $\boldsymbol{\varphi}_j$ [35]. For the whole $f \in \Omega$, the scatter plot of $\mathbf{x}(f)$ (up to 3-dimension) then reveals all the n directions of the mode shape columns of $\mathbf{\Phi}$ (Figure 4). With a general value of the dimension m (may be larger than 3), the estimated vibration mode matrix $\mathbf{\Phi}$ can automatically be extracted by standard clustering algorithms such as fuzzy-C-means.

In determined case ($m=n$), time-domain modal responses $\mathbf{q}(t)$ are readily de-coupled by

$$\mathbf{q}(t) = \mathbf{\Phi}^{-1}\mathbf{x}(t) \tag{14}$$

thereby estimating the modal frequency and damping ratio from $\mathbf{q}(t)$. For underdetermined case ($m < n$) where the sensors are insufficient, $\mathbf{\Phi}$ is rectangular and recovery of $\mathbf{q}(f)$ from the underdetermined Eq. (12) is ill-posed. By looking to sparsity, the spatially sparsest (disjoint) representation of the modal responses $\mathbf{q}(f)$ can be recovered by the well-known sparsity optimization ℓ_0-minimization program (P_0) [14], at each $f \in \Omega$

$$(P_0): \quad \mathbf{q}^{\star}(f) = \arg\ \min \|\mathbf{q}(f)\|_{\ell_0} \quad \text{subject to} \quad \mathbf{\Phi}\mathbf{q}(f) = \mathbf{x}(f) \tag{15}$$

where $\|\mathbf{q}(f)\|_{\ell_0} = \#\{i : q_i(f) \neq 0\}$ is the ℓ_0-norm. (P_0) therefore finds a vector $\mathbf{q}^{\star}(f)$ with smallest ℓ_0-norm that explains the observation $\mathbf{x}(f)$. This ℓ_0-norm naturally guides (P_0) to seek the sparsest $\mathbf{q}^{\star}(f)$ with fewest nonzero entries among all feasible solutions. It has been proven, however, that solving (P_0) is in general NP-hard [14].

Fortunately, if the solution $\mathbf{q}^{\star}(f)$ is sufficiently sparse, then (P_0) can be safely replaced by a convex optimization program ℓ_1-minimization (P_1), known as basis pursuit [14],

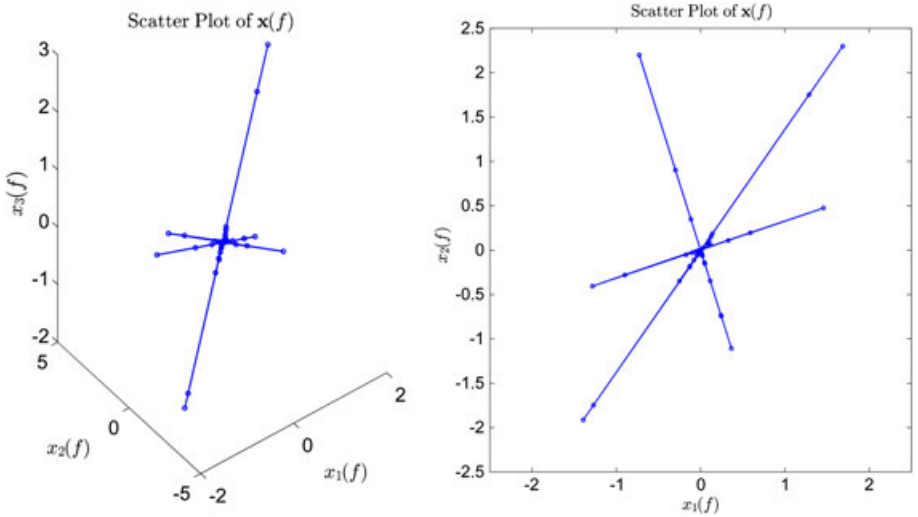

Figure 4. The scatter plot of the frequency–domain system responses in determined case with three sensors (left) and underdetermined case with two sensors (right).

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

$$(P_1): \quad \mathbf{q}^{\star}(f) = \arg\ \min \|\mathbf{q}\|_{\ell_1} \quad \text{subject to} \quad \mathbf{\Phi q}(f) = \mathbf{x}(f) \tag{16}$$

in which the ℓ_1-norm is defined by $\|\mathbf{q}(f)\|_{\ell_1} = \sum_{i=1}^{n} |q_i(f)|$. Because the underlying n-dimension $\mathbf{q}(f)$ is very sparse (theoretically $K = 1$) with only one nonzero entry, it is guaranteed to be accurately recovered by (P_1) from the incomplete m-dimension ($m < n$) observations $\mathbf{x}(f)$ and the retangular $\mathbf{\Phi}$. Using the inverse cosine transform, the time-domain modal responses $\mathbf{q}(t)$ can be readily recovered from $\mathbf{q}(f)$.

3.2. Independent component analysis for automated extraction of sparse modal responses

The similarity between the BSS model (Eq. (6)) and modal expansion (Eq. (11) naturally tempts one to solve the output-only modal identification problem in the BSS framework. Unfortunately, the direct application of time-domain ICA is only suitable for identification of very lightly damped structures [26]. The reason lies in that ICA only uses the statistical information of signals, while modal responses possess significant temporal structures [31]. Another view of the reason may lie in the concept of 'transform sparse': ICA recovers sparse components (Section 2.3), but modal responses are monotone only in frequency or time-frequency domain instead of their original time domain.

Therefore, transform the Eq. (11) to the sparse time-frequency domain using short-time-Fourier-transform (STFT), and concatenate all the windowed STFTs (to yield a 1-D formulation)

$$\mathbf{X}_{f\tau} = \mathbf{\Phi Q}_{f\tau} = \sum_{i=1}^{n} \varphi_i Q_{i,f\tau} \tag{17}$$

The available $\mathbf{X}_{f\tau}$ can then be blindly separated by ICA into $\mathbf{Q}_{f\tau} = [Q_{1,f\tau}, \ldots Q_{n,f\tau}]^T$ where $Q_{i,f\tau}\,(1, \ldots, n)$ is the sparse time-frequency representation of q_i,

$$\mathbf{Q}_{f\tau} = \mathbf{W X}_{f\tau} \tag{18}$$

$\mathbf{\Phi} = \mathbf{W}^{-1}$, and then the modal expansion can be de-coupled by Eq. (14). Furthermore, using a complex-ICA (cICA) algorithm, STFT-cICA is able to identify structures with complex modes [36], because Eq. (17) can be real valued using only their absolute values or complex values keeping the original complex STFTs. Similarly, the sparse time-frequency representation of modal responses has also been successfully exploited in output-only modal identification for wireless sensor network [37] and for structural seismic responses [38].

The SCA or STFT-ICA output-only modal identification to some extent yields resemblance to the well-known FDD method, but with some distinction. For FDD, it conducts SVDs of the cross spectra of the structural responses at some user-specified modal peaks, while SCA and ICA automatically separate all those modal peaks. Also, SCA and STFT-ICA can handle highly damped structures and STFT-cICA complex modes, whereas FDD assumes lightly damped ones and the performance of identification of complex ones is not studied. On the other hand, FDD using SVD incorporates underdetermined identification without modification, where SCA needs an additional sparse recovery step (ℓ_1-minimization).

3.3. Signal complexity pursuit

While modal responses possess sparsity in a transform (frequency) domain, the statistical property of their temporal structure can also be explicitly exploited. Specifically, the distinct temporal complexity between the modal responses and the system responses can lead to efficient separation of them, which is accomplished by another BSS technique CP learning rule [39].

The theoretically justified Stone's theorem [40] states that in the BSS model, the complexity of any mixture is always higher than the simplest constituent source (Figure 5). Using this basis to implement the CP learning rule, the simplest signal extracted from a set of mixtures would be a source signal. Relating to the modal expansion Eq. (11) where any system response $x_i(t)$ ($i = 1, \ldots, n$) is a linear mixture of the modal responses $\mathbf{s}(t)$, the complexity of any $x_i(t)$ will be higher than that of the simplest modal response. Specifically, the CP learning rule pursues to extract

$$y_i(t) = \mathbf{w}_i \mathbf{x}(t) \tag{19}$$

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

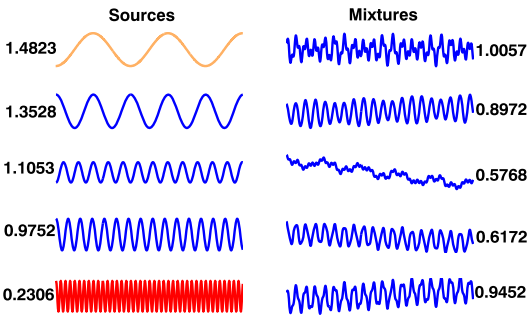

Figure 5. The temporal predictability (values) approximating signal complexity of the sources (left) and the mixtures (right).

such that $y_i(t)$ yields least complexity and corresponds to the simplest modal response. Theoretically, the complexity of y_i is rigorously measured by Kolmogorov complexity. Stone proposed a simple but robust approximation, termed temporal predictability [24], where the simplest signal is most predictable, and verse vice

$$\mathrm{F}(y_i) = \log \frac{\mathrm{V}(y_i)}{\mathrm{U}(y_i)} = \log \frac{\sum_{t=1}^{N} \left(\overline{y}_i(t) - y_i(t)\right)^2}{\sum_{t=1}^{N} \left(\widehat{y}_i(t) - y_i(t)\right)^2} \tag{20}$$

where the long-term predictor $\overline{y}_i(t)$ and short-term predictor $\hat{y}_i(t)$ are given, respectively, by

$$\begin{aligned} \overline{y}_i(t) &= \lambda_L \overline{y}_i(t-1) + (1-\lambda_L) y_i(t-1) \quad 0 \le \lambda_L \le 1 \\ \widehat{y}_i(t) &= \lambda_S \widehat{y}_i(t-1) + (1-\lambda_S) y_i(t-1) \quad 0 \le \lambda_S \le 1 \end{aligned} \tag{21}$$

The parameter λ is defined by the half-life parameter h as $\lambda = 2^{-1/h}$ where $h_S = 1$ and h_L is arbitrarily set (say, 900,000) as long as $h_L \gg h_S$.

While it is possible to extract the simplest source (modal response) one by one—after the simplest source is extracted as per Eq. (19) to Eq. (20), it is removed using the Gram–Schmidt orthonormalization—Stone proposed a more efficient algorithm based on a generalized eigenproblem, which simultaneously recovers all the sources (modal responses),

$$\mathbf{s}(t) = \mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \tag{22}$$

$\mathbf{\Phi} = \mathbf{W}^{-1}$. The CP method is found suitable for output-only modal identification of structures with closely spaced modes, complex highly damped modes, and in real-time identification of the time-varying cable tension time history; the details are referred to Ref. [39] and [41].

3.4. Data management via low-rank structure

Recently, many SHM systems, each with an array of networked sensors to continuously record structural data for monitoring and assessing structural performance, have raised the data-intensive issue. On the one hand, the continuously collected sensor data provides high-resolution and multi-dimensional information of the structure, which is vital for identifying and updating structural information, evaluating its health status, and detecting damage in real time. On the other hand, processing and managing the overwhelmingly voluminous data becomes a challenge. In this context, it is important to develop efficient and effective SHM data compression and cleansing algorithms. In the SHM community, spatial dependency [42,43] or spatio-temporal correlation [44] within the multi-channel structural vibration response measurements was taken advantage of for data compression. This section presents the approach of explicitly exploiting the sparse and low-rank data structure of the structural vibration response measurements to address this issue.

3.4.1. Low-rank structure and independent component analysis multivariate sparse representations for data compression. A relevant observation is that structural vibration responses are typically low-rank by SVD or PCA, that is, $\mathbf{X} \in \mathbb{R}^{m \times N}$ with small r, because in real world, only a few modes are

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

excited out and present in the structural vibration responses. Therefore, one strategy especially appealing to data compression is to drop those principal components with significantly small eigenvalues. As small eigenvalue indicates small energy of the corresponding principal component, it would cause little data loss by retaining those dominant components with larger eigenvalues. Meanwhile, it achieves higher compression by only encoding the retained components.

In Section 2.3, it has been derived that the learning rule of ICA leads itself into sparsifying the multivariate data set, which is useful for multi-channel data compression. There is more, on the other hand, to expect from it in a statistical learning viewpoint. ICA is essentially a multivariate analysis tool, pursuing to uncover the underlying structures of the given data set, say, $\mathbf{x}$, by transforming it into an independent representation space $\mathbf{y}$.

Looking back at the contrast function Eq. (8), entropy is closely related to the coding length of a signal: the larger entropy consumes more coding length. The Gaussian-distributed variable has the largest entropy among all variables with equal variance [18]. ICA's learning rule of maximization of negentropy can be perfectly interpreted as minimization of the mutual information of $\mathbf{x}$, such that the ICA-transformed components, $\mathbf{y}$, yield least mutual information, that is, their required coding length can no longer be reduced. Therefore, the statistical dependency within $\mathbf{x}$ is removed and the ICA-transformed components, $\mathbf{y}$, are most mutually independent. ICA thus naturally yields the optimal (linear) transformation adaptive to data itself for compression in statistical framework. The real-world examples presented in Ref. [21] show that the independent components indeed yield very sparse representations and thus are preferable to encoding.

3.4.2. Significant data compression by very low-rank representation of reshaped matrix. It is seen in the Section 3.4.1 that the dimension reduction for data compression is most effective when $r \ll m$ ($\mathbf{X} \in \mathbb{R}^{m \times N}$needs to be as low-rank as possible), that is, the channel (sensor) number needs to be *much* larger than that of the involved modes, which is in fact a common assumption where PCA is found effective such as in damage identification and feature extraction [45–47]. However, it is not satisfied in many situations: for civil engineering structures, typically large scale, the sensor number m is not so *much* more than the involved r modes; as a result, $r \ll m$ cannot be guaranteed for a low-rank representation. A matrix reshape scheme is proposed to remove this limitation for wider applicability of PCA, as detailed in the following.

Originally, $\mathbf{X} \in \mathbb{R}^{m \times N}$ is hardly low-rank; however, applying a simple matrix reshape scheme 'generates' a low-rank representation, by taking advantage of that, mode information (typically few are active, hence the rank of the structural response data matrix is small) remains *approximately invariant* regardless of reshaping the data matrix. First, divide the time history of each channel, say, $\mathbf{x}_i \in \mathbb{R}^N$ (ith channel), into l segments, yielding $(\mathbf{x}_i)_j \in \mathbb{R}^v$ as the jth segment of $\mathbf{x}_i$, where $v = N/l$. Then re-stack them into a new structural response matrix $\overline{\mathbf{X}} \in \mathbb{R}^{w \times v}$, where $w = m \times l$ and its ith row $\overline{\mathbf{x}}_i \in \mathbb{R}^v$ as a v-point segment of $\mathbf{x}_i \in \mathbb{R}^N$. Therefore, the SVD of $\overline{\mathbf{X}} \in \mathbb{R}^{w \times v}$ is

$$\overline{\mathbf{X}} = \overline{\mathbf{U}}\,\overline{\mathbf{\Sigma}}\,\overline{\mathbf{V}}^T = \sum_{i=1}^{r'} \sigma_i' \overline{\mathbf{u}}_i \overline{\mathbf{v}}_i^T \tag{23}$$

The key idea is that because there are still only r modes involved in the re-stacked matrix $\overline{\mathbf{X}} \in \mathbb{R}^{w \times v}$, so the rank of $\overline{\mathbf{X}} \in \mathbb{R}^{w \times v}$, r', will not have much difference with r, that is, $r' \approx r$ (usually r' is slightly larger than r); plus, $r \ll \min(w, v)$, so

$$\operatorname{rank}\left(\overline{\mathbf{X}}\right) = r' \approx r \ll \min(w, v) \tag{24}$$

that is, $\overline{\mathbf{X}}$ becomes a low-rank matrix. The matrix reshape scheme for low-rank representation is graphically illustrated in Figure 6, and theoretical justification is presented in Ref. [48].

To achieve a highest compression ratio $\rho' = r'/\min(w, v)$, it is best to *maximize* $\min(w, v)$, which suggests re-stacking $\overline{\mathbf{X}} \in \mathbb{R}^{w \times v}$ as *square* as possible, that is, making $w \approx v$ for most effective compression efficiency. The effectiveness of this simple strategy is detailed in Ref. [48]. It is also interesting to note that more sensors number or long time history (large dimension m or N, making $\min(w, v)$ larger), with $r' \approx r$ approximately invariant, is advantageous for a 'most low-rank' representation. Finally, effective data dimension reduction is realized by

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

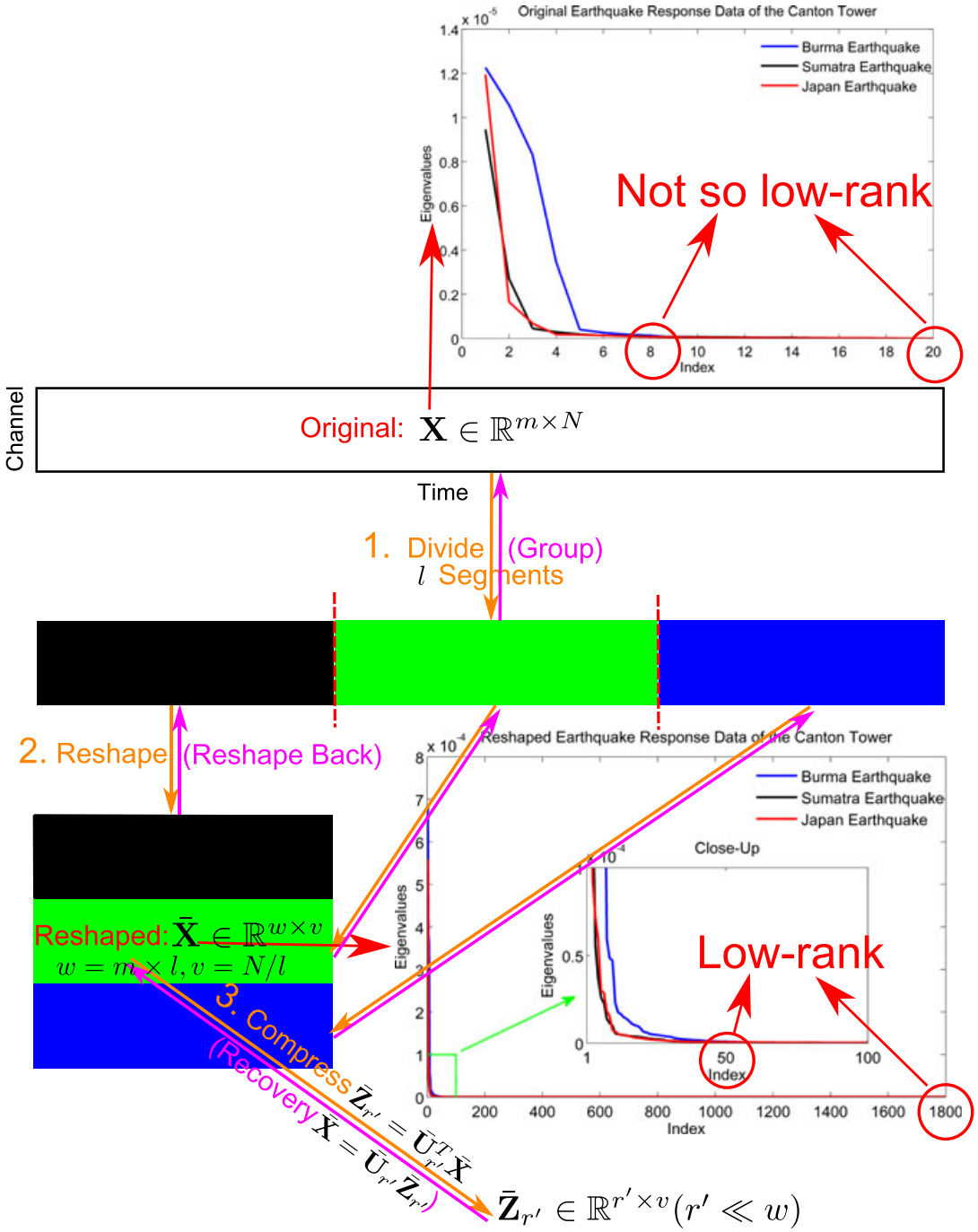

Figure 6. The data compression scheme with the matrix reshape strategy.

$$\overline{\mathbf{Z}}_{r'} = \overline{\mathbf{U}}_{r'}^T \overline{\mathbf{X}} \tag{25}$$

Because $r' \ll \min(w, v)$, a significant compression ratio $\rho' = r'/\min(w, v)$ can be achieved. Note that the re-stacking guaranteeing low-rank representation removes the constraint that the channel (sensor) number $m > r$ (to ensure redundancy) in original PCA-based or SVD-based methods. Recovery is performed by

$$\overline{\mathbf{X}} = \overline{\mathbf{U}}_{r'} \overline{\mathbf{Z}}_{r'} \tag{26}$$

and then restacking $\overline{\mathbf{X}} \in \mathbb{R}^{w \times v}$ back to $\mathbf{X} \in \mathbb{R}^{m \times N}$ (Figure 6).

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

3.4.3. Removing sparse outliers. Real-world measured structural response data typically contains considerable noise or errors. For example, the ambient vibration response data of the Canton Tower (Figure 7(a)), recorded by the SHM system, contains remarkable outliers (gross errors). Applications of traditional data processing methods can only deal with dense small noise. Taking advantage of the data structure of the multi-channel noisy structural vibration responses, robust PCA [16], termed PCP, is capable of effectively modeling the noisy data with outliers and thus simultaneously removing both the outliers and dense noise [22]. When the original data $\mathbf{X} \in \mathbb{R}^{m \times N}$ are addictively corrupted by both gross errors (outliers) and dense noise,

$$\widehat{\mathbf{X}} = \mathbf{X}_0 + \mathbf{N}_0 + \mathbf{Z}_0 \tag{27}$$

where $\mathbf{Z}_0 \in \mathbb{R}^{m \times N}$ has few (sparse) but gross outlier elements with arbitrarily large and located magnitudes, and $\mathbf{N}_0 \in \mathbb{R}^{m \times N}$ is entry-wise i.i.d. small dense noise. PCP aims to recover $\mathbf{X}_0$ by solving the following convex program

$$(P_*): \quad \min \|\mathbf{X}\|_* + \lambda \|\mathbf{Z}\|_{\ell_1} \quad \text{subject to} \quad \left\|\widehat{\mathbf{X}} - \mathbf{X} - \mathbf{Z}\right\|_{\mathrm{F}} \leq \delta \tag{28}$$

where $\|\mathbf{X}\|_* := \sum_i \sigma_i(\mathbf{X})$ is termed the nuclear norm of the matrix $\mathbf{X}$, which summates its singular values; $\|\mathbf{Z}\|_{\ell_1} := \sum_{ij} |z_{ij}|$ denotes the ℓ_1-norm of the matrix $\mathbf{Z}$, which is thought as a long vector; $\lambda = 1/\sqrt{N}$ is a trading parameter, $\|\mathbf{X}\|_{\mathrm{F}} := \sqrt{\sum_i \sigma_i^2}$ is the Frobenius norm of $\mathbf{X}$, and δ is some bounding parameter related to the small dense noise level. In analogy to the ℓ_1-norm of a vector, the nuclear norm is the tightest convex approximation to the rank of a matrix.

It is rigorously proved in Ref. [16] that if $\mathbf{X}_0$ is sufficiently low-rank and $\mathbf{Z}_0$ sparse, with overwhelmingly high probability, (P_*) accurately recovers the true low-rank $\mathbf{X}_0$ and sparse $\mathbf{Z}_0$. As mentioned, $\mathbf{X}_0 \in \mathbb{R}^{m \times N}$ in its original dimension is seldom very low-rank. The matrix reshape scheme is applied to make a low-rank reshaped matrix $\overline{\mathbf{X}}_0 \in \mathbb{R}^{w \times v}$: both the ℓ_1-norm and Frobenius norm of a matrix are summations of its entries and energy, respectively; as such, restacking will not essentially change the property that $\overline{\mathbf{Z}}_0 \in \mathbb{R}^{w \times v}$ remains sparse and $\overline{\mathbf{N}}_0 \in \mathbb{R}^{w \times v}$ bounded. With these assumptions satisfied, (P_*) accurately estimates the low-rank $\overline{\mathbf{X}}_0 \in \mathbb{R}^{w \times v}$ (and the outliers $\overline{\mathbf{Z}}_0 \in \mathbb{R}^{w \times v}$), which can then be readily re-stacked back to $\mathbf{X}_0^\star \in \mathbb{R}^{m \times N}$. (P_*) can be implemented using the Augmented Lagrange multiplier method. Inheriting from the virtue of convex program, the solution to (P_*) found by the Augmented Lagrange multiplier is always globally optimal. Figure 7(b) shows the structural vibration responses with the gross outliers removed, and more examples are presented in Ref. [22].

3.5. Damage detection by sparse signal discovery

Many existing signal-based techniques have shown significant promise in extracting damage features from measured structural response data; as opposed to structural model (physical or modal)-based methods, they feature efficient computation and make no prior assumption with respect to the structural

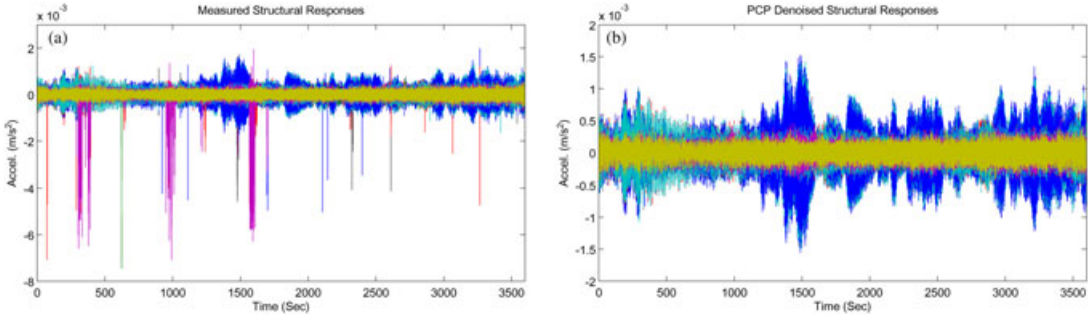

Figure 7. (a) The recorded ambient vibration accelerations with many outliers and (b) the PCP-denoised (reshape factor $l = 40$) of the Canton Tower from 12:00 am Jan. 20, 2010 to 1:00 pm Jan. 20, 2010. (20 channels of data are shown, available in Ref. [49].)

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

model, which makes them enjoy broad applicability in damage identification. However, several challenges hinder the effectiveness of the existing damage identification: damage features hidden in data have not been explicitly modeled or characterized; as a result, existing damage features proposed in the literatures are specific to the particular problem and algorithm without gaining wide acceptance and applicability, and there is no general clue and guidance on how to improve the effectiveness of existing damage detection algorithms.

Structural damage is typically observed as local phenomenon [1], which radiates pulse-like features hidden in the structural vibration response signals. Although not discussed in detail, the ability of revealing the hidden sparse pulse-like features has been the underlying success factor for the wavelet and Hilbert–Huang transform-based damage detection methods [10–13]. However, noise can easily destroy the effectiveness of damage detection method (Figure 8). The following in this section presents emerging mathematical tools for explicitly exploiting and modeling the sparse features for robust damage identification.

3.5.1. Independent component analysis for recovery of sparse damage feature. The derivation that ICA biases to recover sparse components, which is shown in Section 2.3, can immediately lead to a straightforward application in unsupervised damage identification. The observation lies in that even though damage information is invisible in the original structural response signals, transforming them to a proper domain (e.g., wavelet domain) radically reveals the spike-like damage features, which, however, are easily contaminated by noise (see the example in Figure 8). ICA, on the other hand, is able to recover the spike-like sparse components, which contains damage features, from the noisy response signals.

Consider transforming structural responses $\mathbf{x}(t)$ carrying the damage information into some wavelet scale l to expose the common spike-like feature (viewed as a latent sparse component $s_j(t)$) hidden within $\mathbf{x}(t)$ and incorporate it into the BSS model,

$$\mathbf{x}^l(t) = \mathbf{A}\mathbf{s}(t) \tag{29}$$

then ICA would blindly recover $y_j(t) = s_j(t)$ (the 'interesting' source) with outstanding spike features, directly indicating the damage instant(s). Furthermore, the simultaneously recovered $\mathbf{a}_j$ conveying the spatial signature of $s_j(t)$ also locates damage location(s) [50]. With WT properly exposing the signals to the sparse domain, WT-ICA then takes further advantage of ICA's ability of sparsity seeking.

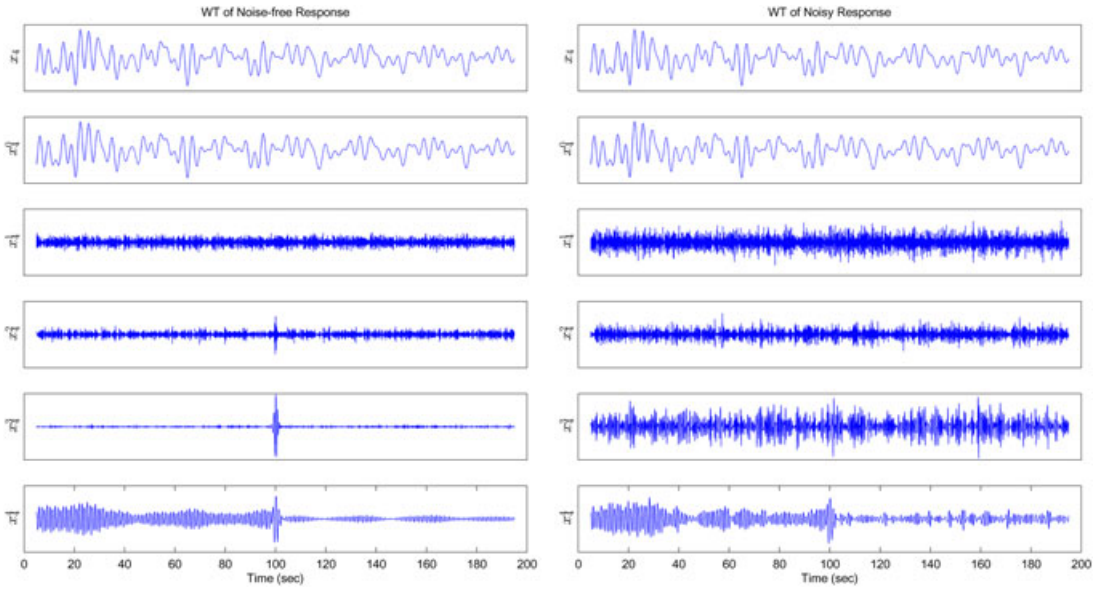

Figure 8. Wavelet transform of the simulated structural response measurements. The signals are decomposed into four levels by the db10 wavelet. In the noise-free case (left), the spikes at 2–4 levels clearly shows damage instant at the 100th second, but with small noise (about 1%) added to the response signal (right), the spikes are completely buried.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

The examples in Ref. [50] illustrate the effectiveness of explicit sparsity as a guiding principle in the identification of useful structural features.

3.5.2. Dynamic imaging for structural surveillance using low-rank plus sparse representation. Local structural assessment focuses on close-up inspection of structural health status and is meant to more accurately quantify structural damage (e.g., damage types and severity). Current practice of local structural assessment includes on-site visual inspection by experts and nondestructive testing (e.g., acoustic and ultrasonic). Although effective in many applications, they can be time-consuming and costly, and limited to areas that are accessible to experts, making them mostly suitable for offline evaluation. The video cameras—permanently mounted on appropriate positions—enable close-up imaging ('filming') of critical structural components, such as the anchorage of the stay cables and other critical connections, for continuous local structural damage assessment and damage diagnosis and alerts in real time. An unsupervised data-driven framework has been established to automate real-time detection of structural damage by explicitly modeling the fundamental spatiotemporal data structure of the multiple images (video stream) [51].

If restacking each of N temporal frame of the structure as a long column vector with a resolution of $M = M_1 \times M_2$ pixels, the multi-frame data matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ is obtained, whose ith ($i = 1, \ldots, N$) column $x_i \in \mathbb{R}^M$ represents the temporal frame at time T_i. PCP is able to blindly decompose $\mathbf{X} \in \mathbb{R}^{M \times N}$ into a superposition of a low-rank matrix $\mathbf{L} \in \mathbb{R}^{M \times N}$ and a sparse matrix $\mathbf{S} \in \mathbb{R}^{M \times N}$ as

$$\mathbf{X} = \mathbf{L} + \mathbf{S} \tag{30}$$

by solving (P_*). $\mathbf{S} \in \mathbb{R}^{M \times N}$ is said to be sparse if it has only few nonzero entries, and $\mathbf{L} \in \mathbb{R}^{M \times N}$ is low-rank in the sense that its SVD has few active singular values.

The $\mathbf{L} + \mathbf{S}$ representation has a novel insight into the data structure of the multiple temporal close-up frames of structures as a superposition of a background component and an innovation component: $\mathbf{L}$ represents the static or slowly changing correlated background component among the temporal frames, which is naturally low rank; $\mathbf{S}$ captures the innovation information in each frame induced by the evolutionary damage, which is naturally sparse standing out from the background. See the proposed dynamic imaging framework for continuous local structural assessment in Figure 9 and Ref. [51] for more details.

3.5.3. Damage identification via sparse classification. While extracting the sparse component from the structural vibration or image measurements could lead to efficient and effective identification of damage instants and locations, if incorporating a structural model or other structural reference information, one may perform supervised damage identification, in the pattern recognition framework [8], that can address even the problem of level 3, that is, the quantification of damage severity. Traditional supervised learning techniques, such as artificial neural networks [52], support vector machines [53,54], nearest neighbor [55], as well as Markov observers [56] and artificial optimization [57,58], typically involve three steps: *feature extraction*, *training (parametric classifier model)*, and *classification*. For damage identification, the extracted features from various predefined or reference damage classes, including different damage locations and damage extents, are used as inputs to train the classifiers (mostly parametric), which can then identify the damage class of the test feature representing the current state of the structure. The training process of the classifiers, however, could influence their performance.

Instead of building and training a parametric classifier model, Yang and Nagarajaiah [59] established a new damage identification method in the classification framework by exploiting the sparsity nature implied in the classification problem itself, via sparse representation classification of a test feature in terms of an adaptive reference dictionary (Figure 10); it is found to be intuitive and efficient.

In the damage identification problem, the features are chosen to be the mode shape columns and are blindly extracted by CP from the system responses of the model. For an n-DOF system, if simulating N different damage classes (with different damage locations and severities), then all the $w = N \times n$ (typically $n \ll w$) mode shape columns are concatenated to yield a reference matrix or dictionary $\mathbf{\Psi} \in \mathbb{R}^{n \times w}$

$$\mathbf{\Psi} = [\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_N] = \left[\varphi_{1,1}, \ldots, \varphi_{N,n}\right] \tag{31}$$

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

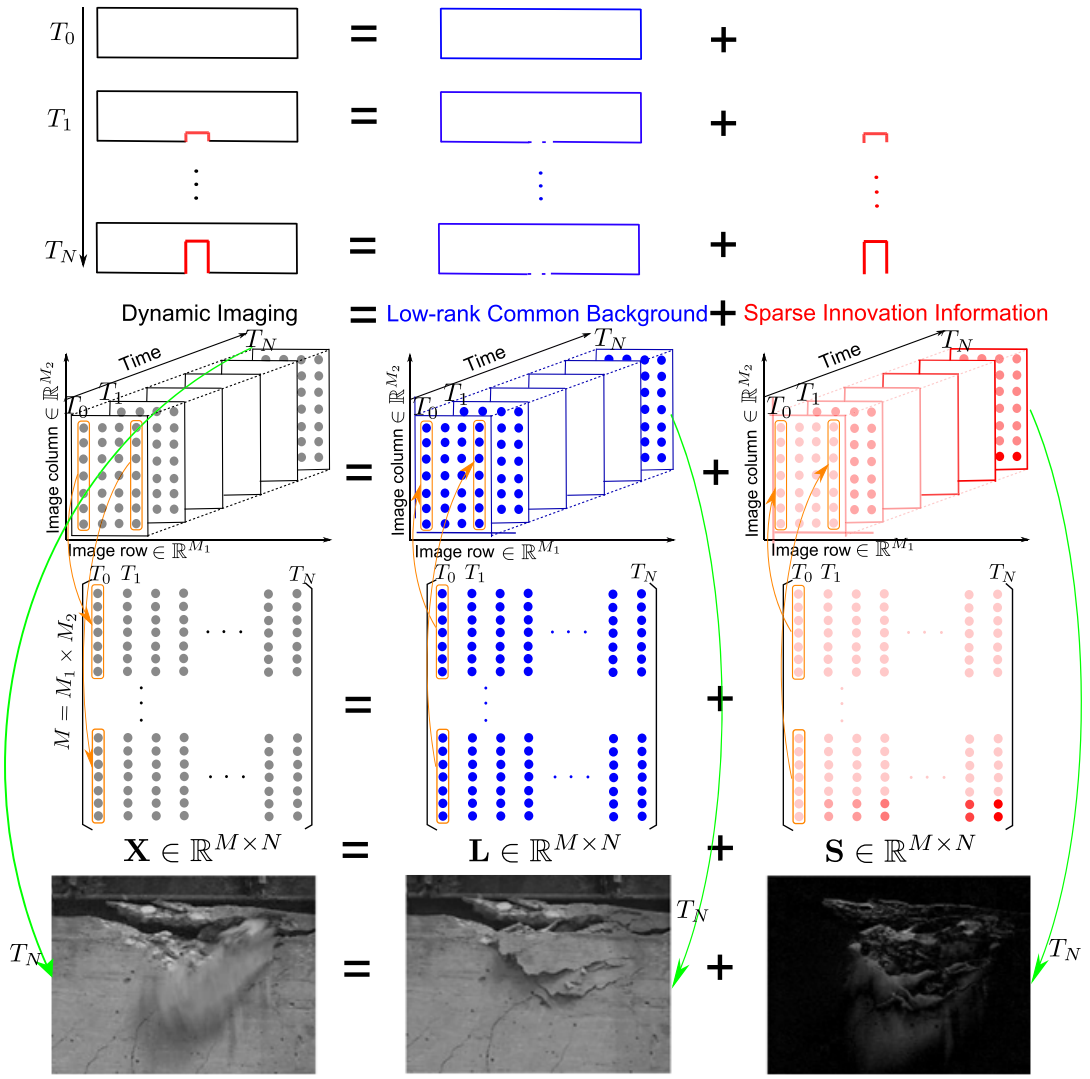

Figure 9. The dynamic imaging of structures framework based on low-rank plus sparse representation of the multiple temporal images of the structure.

Now suppose the test features $\widehat{\mathbf{\Phi}} = [\widehat{\varphi}_1, \ldots, \widehat{\varphi}_n] \in \mathbb{R}^{n \times n}$ are extracted from the current structural responses whose damage scenario coincides with one of the damage class of the reference matrix, say, the jth class (but it is unknown beforehand), then $\widehat{\varphi}_i$ ($i = 1, \ldots, n$) would equal $\varphi_{j,i}$ up to a scale difference. Expanding $\widehat{\varphi}_i$ in terms of the whole reference dictionary,

$$\widehat{\varphi}_i = \mathbf{\Psi}\boldsymbol{\alpha}_i = \sum_{k=1}^{N}\sum_{l=1}^{n} \alpha_{k,l}\varphi_{k,l} \tag{32}$$

where $\boldsymbol{\alpha}_i = [0, \ldots, 0, \alpha_{j,i}, 0, \ldots, 0]^T \in \mathbb{R}^w$ is its underlying sparse representation whose nonzero element $\alpha_{j,i}$ directly assigns the damage class the test feature belongs to. As introduced previously, finding the sparse solution $\boldsymbol{\alpha}_i$ to the (highly) underdetermined linear system of equations Eq. (32) from the knowledge of $\mathbf{\Psi} \in \mathbb{R}^{n \times w}$ and $\widehat{\varphi}_i$ can be efficiently accomplished by (P_1)

$$(P_1): \quad \boldsymbol{\alpha}_i^{\star} = \arg\ \min \|\boldsymbol{\alpha}_i\|_{\ell_1} \quad \text{subject to} \quad \mathbf{\Psi}\boldsymbol{\alpha}_i = \widehat{\varphi}_i \tag{33}$$

Sparse representation classification directly exploits the essence of the classification problem: the test feature can be most sparsely represented by the reference dictionary. It establishes an underdetermined linear system of equations whose underlying sparse solution can be efficiently recovered to dictate the damage class. Compared with other pattern recognition-based damage identification methods,

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

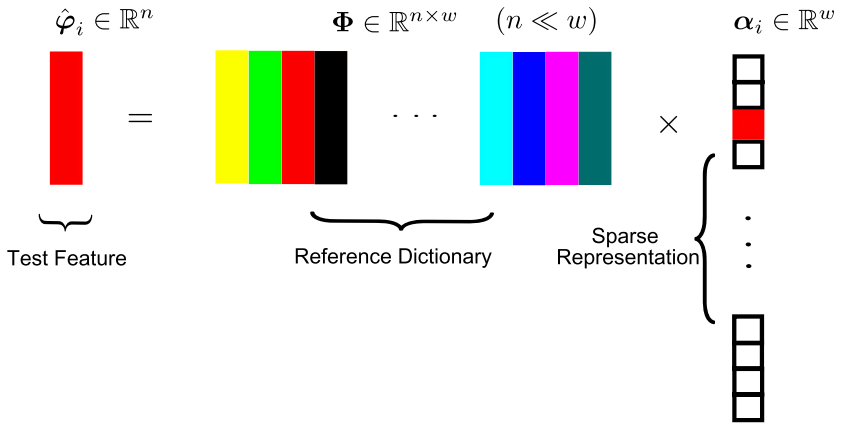

Figure 10. The sparse representation classification paradigm for damage identification. The test feature $\widehat{\boldsymbol{\phi}}_i \in \mathbb{R}^n$ (red column, e.g., mode shape column) only activates itself via its representation $\boldsymbol{\alpha}_i \in \mathbb{R}^w$ (read in its own location, white denotes unactivated zero) in terms of the large reference dictionary $\boldsymbol{\Psi} \in \mathbb{R}^{n\times w}$ $(n \ll w)$ (by concatenating all feature columns of all candidate reference damage classes), expressed as a highly underdetermined linear system of equations $\widehat{\boldsymbol{\phi}}_i = \boldsymbol{\Psi}\boldsymbol{\alpha}_i$. The unique nonzero element (red) in $\boldsymbol{\alpha}_i$ (recovered by ℓ_1-minimization) directly dictates which class the test feature belongs to, within the predefined reference dictionary.

the process of building and training the parametric classifier required by traditional pattern recognition-based methods is shifted to a model selection formulation Eq. (32), which is accomplished by (P_1).

3.6. Efficient sensing strategy enabled by sparse/low-rank data structure

It has been shown that explicitly modeling and harnessing the sparse and low-rank data structure of the structural vibration responses measurements can reveal the underlying structural dynamic and damage features for system identification and damage detection, as well as data management. This section illustrates that the implicit knowledge of the sparse and low-rank data structure enables efficient sensing of the structural vibration responses and strain measurements in the first place.

3.6.1. Output-only modal identification with compressed sensing. Using sparsity as the guiding principle, the established BSS-based methods are shown to be capable of efficiently performing output-only modal identification. One fundamental issue in practice remains, however, that it is usually difficult to identify higher (active) modes that are in many situations important, such as structures with high-speed excitation, for development of more accurate prediction model and so on. Capturing the active, potentially high-frequency, modes, in the traditional Shannon-Nyquist paradigm following the sampling theorem, requires extremely high uniform sampling rate (at least twice the maximum frequency of the highest active mode) in the data acquisition phase. Although feasible, it would result in sensing a large number of samples, which could significantly increase the computational demand in the data analysis (modal parameter estimation) phase. Especially, when the data acquisition is performed in the wireless sensor platform with only limited power and communication resources [60], transferring this resultant large number of samples to the data analysis base station would be a heavy burden. Inspired by the recent development of CS as well as its application in structural dynamics and SHM [35,61–63], a new modal analysis framework (Figure 11) in a non-uniform low-rate random sensing setting [64] was proposed based on CS and BSS.

Compressed sensing (CS) is largely motivated by transform coding but has significant distinction. Instead of transform coding's strategy of acquiring the full sets of a signal $x \in \mathbb{R}^N$, transforming $x \in \mathbb{R}^N$ to coefficients $\alpha \in \mathbb{R}^N$ but then keeping only a few most significant coefficients (discarding most of them), CS performs random under-sampling by correlating $x \in \mathbb{R}^N$ with a small subset of $M \ll N$ sensing waveforms $v_k \in \mathbb{R}^N$ $(k = 1, \ldots, M)$ *randomly* selected out of a full set of N waveforms, for example, spikes, sinusoids, and wavelets,

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

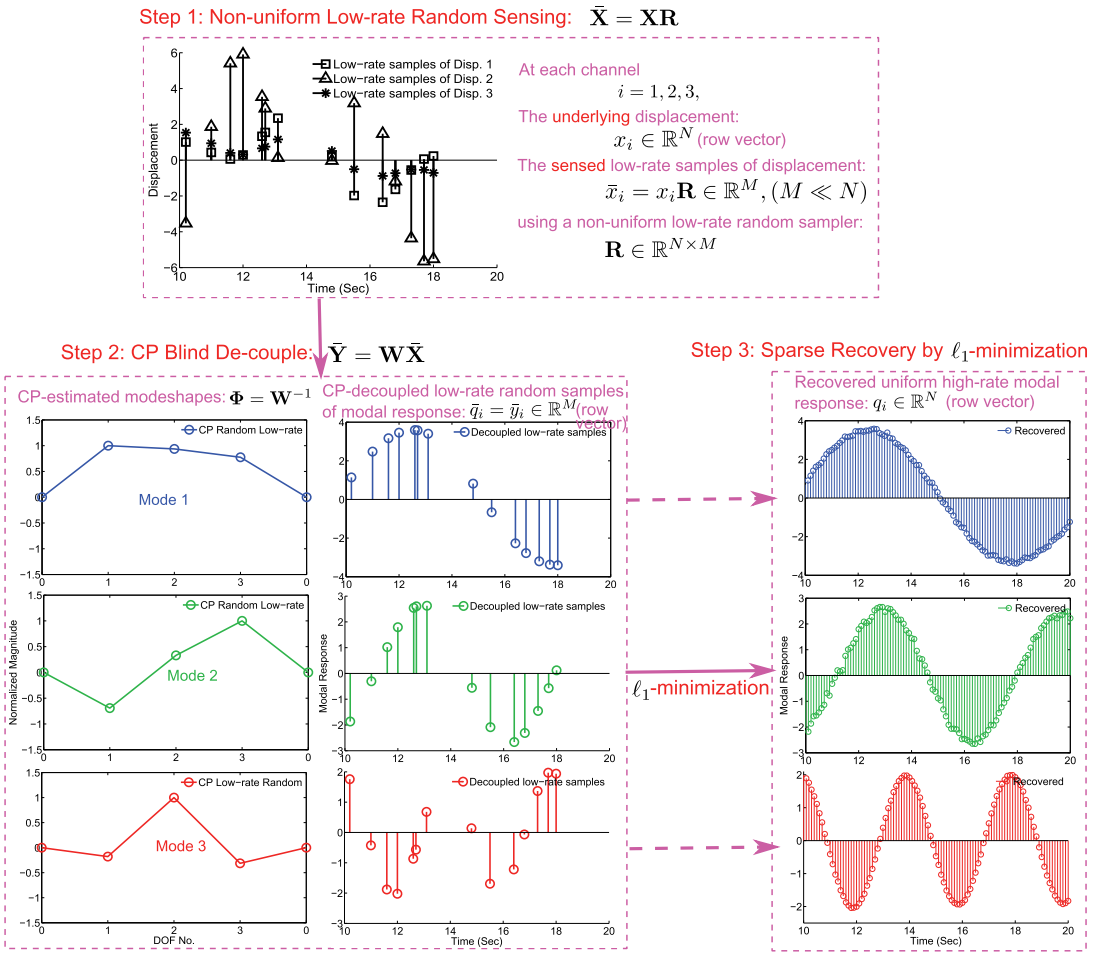

Figure 11. The non-uniform low-rate random sensing paradigm for output-only modal identification using blind source separation (BSS) technique complexity pursuit (CP) and compressed sensing (CS).

$$b_k = \langle x, v_k \rangle, \quad k = 1, \ldots, M \tag{34}$$

directly obtaining only $M \ll N$ (hence the term *compressed*) measurements $b \in \mathbb{R}^M$ ($b_k$is the kth measurement). For example, if $v_k \in \mathbb{R}^N$ ($k = 1, \ldots, M$) is sinusoid (adopted in this study), then $b \in \mathbb{R}^M$ ($M \ll N$) is a small subset of Fourier coefficients. In a matrix denotation, it amounts to

$$b = \mathbf{R}\Upsilon x = \mathbf{A}x \tag{35}$$

where $\Upsilon \in \mathbb{R}^{N\times N}$ is the orthonormal basis (Fourier in this study), $\mathbf{R} \in \mathbb{R}^{M\times N}$ randomly extracts M out of N, and $\mathbf{A} = \mathbf{R}\Upsilon \in \mathbb{R}^{M\times N}$ is the random undersensing matrix to obtain partial observation $b \in \mathbb{R}^M$ ($M \ll N$) about the underlying $x \in \mathbb{R}^N$. The CS theory states that if $x \in \mathbb{R}^N$ is sufficiently sparse (K is very small) in a transformed domain, then it can be exactly recovered from $b \in \mathbb{R}^M$ (M is proportional to K) in an incoherent domain by the ℓ_1-minimization program.

A first intuition of the applicability of CS on modal analysis is straightforward, because the target monotone modal responses are spectrally sparsest. Also, the spike-sensing paradigm has maximal incoherence with the frequency domain—these ingredients make the 'compressed' modal analysis feasible in a non-uniform low-rate random sensing framework.

Performing CS of the system responses $\mathbf{X} \in \mathbb{R}^{n\times N}$ ($x_i \in \mathbb{R}^N$, $i = 1, \ldots, n$) yields the non-uniform low-rate random measurements $\overline{\mathbf{X}} \in \mathbb{R}^{n\times M}$. At the ith channel, $\bar{x}_i \in \mathbb{R}^M$ is the under-sampled signal, where $\bar{x}_i = x_i\mathbf{R}$ and $\mathbf{R} \in \mathbb{R}^{N\times M}$ denotes the non-uniform low-rate sensing matrix, whose kth column $\mathbf{r}_k = \delta(t - \tau) \in \mathbb{R}^N$ ($k = 1, \ldots, M$) is a spike function and $\tau \in [1, N]$ is determined by randomly selecting the M sensing instants. It amounts to a 'compressed' modal expansion

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

$$\overline{\mathbf{X}} = \mathbf{XR} = \mathbf{\Phi QR} = \mathbf{\Phi}(\mathbf{QR}) = \sum_{i=1}^{n} \varphi_i (q_i \mathbf{R}) = \sum_{i=1}^{n} \varphi_i \overline{q}_i \tag{36}$$

This equation indicates that randomly sensing the system responses equals a linear combination of 'randomly sensing the modal response on each of the independent modal coordinates'. In the instantaneous model expression

$$\overline{\mathbf{x}}(k) = \mathbf{\Phi}\overline{\mathbf{q}}(k) = \sum_{i=1}^{n} \varphi_i \overline{q}_i(k) \tag{37}$$

If one can decouple $\overline{\mathbf{x}}(k)$ into a sparser $\overline{\mathbf{q}}(k)$, then recovery of $q_i \in \mathbb{R}^N$ $(i = 1, \ldots, n)$ from $\overline{q}_i \in \mathbb{R}^M$ would require least measurements (smallest M). Recall that BSS makes a very general assumption no more than the independence of sources at each time instant t. Using this basis, at any random sampling instant k $(k = 1, \ldots, M)$, the source independence holds for $\overline{\mathbf{q}}(k)$, then CP (or other BSS techniques) can still decouple the 'compressed' modal expansion

$$\overline{\mathbf{q}}(k) = \overline{\mathbf{y}}(k) = \mathbf{W}\overline{\mathbf{x}}(k) \tag{38}$$

such that $\mathbf{\Phi} = \mathbf{W}^{-1}$. According to CS, the original modal response $q_i \in \mathbb{R}^N$ can be recovered from the knowledge of the low-rate $\overline{q} = \overline{y}_i \in \mathbb{R}^M$ and $\mathbf{R} \in \mathbb{R}^{N \times M}$ and the fact that q_i has a $K = 1$-sparse representation $q_i = \boldsymbol{\alpha}_i \mathbf{C}$ in the discrete cosine transform domain $\mathbf{C}$, by solving (P_1)

$$(P_1): \quad \boldsymbol{\alpha}_i^{\star} = \arg\ \min \|\boldsymbol{\alpha}_i\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\alpha}_i \mathbf{CR} = \overline{y}_i \tag{39}$$

if the uniformly at random selected M $(\ll N)$ measurements exceeds $C \cdot \log N$ ($C$is a small positive constant). Then the time domain $q_i(t)$ is easily obtained by $q_i = \boldsymbol{\alpha}_i^{\star}\mathbf{C}$. The examples and the benefits are discussed in Ref. [64] of the proposed CP-CS application for output-only modal identification in the low-rate random sensing paradigm; the results match those in original Nyquist uniform sensing reasonably. Lately, similar successful examples are also seen in power spectral estimation of structural vibration responses from incomplete measurements [65,66] by utilizing the underlying sparse property of signals.

3.6.2. Compressed sensing of structural images. Exploiting the sparsity of the structural vibration responses, the CS technique can alleviate the data acquisition of modal analysis in a non-uniform low-rate random sampling framework, as introduced in Section 3.6.1. This sparsity principle with CS has also been used for sampling, transmission, and reconstruction of structural vibration measurements [67] For the computer vision component of the SHM system, the sparse data structure of the structural monitoring images, which are static measurements different from the dynamic vibration measurements, can also be exploited, leading to the CS-based central strategy: on some sparse domain, randomly encode large-scale image data into few relevant coefficients, which are then transferred (robust to data loss) and recovered (in base station) for subsequent structural health diagnosis. It turns out that by exploiting the sparse gradients or edges, which typically characterize the damage features (e.g., cracks), it suffices to recover the original image without losing the health diagnostic quality from highly undersampled (compressed) random encoded measurements.

After performing CS and obtaining the few random measurements $b \in \mathbb{R}^M$ about $x \in \mathbb{R}^N$, because $M \ll N$, Eq. (35) is highly underdetermined and ill-posed. The CS theory [14,15] states that if $x \in \mathbb{R}^N$ is sufficiently sparse (K is small in terms of $\mathbf{\Psi}$), then from the knowledge of $\mathbf{A} \in \mathbb{R}^{M \times N}$ and $b \in \mathbb{R}^M$, it can be accurately recovered with overwhelming probability of success by (P_1)

$$(P_1): \quad x^{\star} = \arg\ \min \|\mathbf{\Psi}^* x\|_{\ell_1} \quad \text{subject to} \quad \|b - \mathbf{A}x\|_{\ell_2} \leq \epsilon \tag{40}$$

provided that $M \propto K \log N$. ϵ is associated with the noise level. In practice, empirically M is approximately two to five times of the number of sparse coefficients K [15].

When $x \in \mathbb{R}^N$ is an image, it is often effective to use the finite differences as the sparsifying transform $\mathbf{\Psi}$ and (P_1) becomes

$$(P_{TV}): \quad x^{\star} = \arg\ \min \mathrm{TV}(x) \quad \text{subject to} \quad \|b - \mathbf{A}x\|_{\ell_2} \leq \epsilon \tag{41}$$

where TV(x) is the total variation of x, defined as the sum of the discrete gradient at each point (pixel). (P_{TV}) finds an $x^{\star} \in \mathbb{R}^N$ with fewest edges (nonzero gradients) that is consistent with the measurements $b \in \mathbb{R}^M$ when subjected to data fidelity constraint. This is useful for structural monitoring images

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

because the damage information such as a crack usually exhibits as the pronounced edge feature in the image. In addition, (P_{TV}) does not require any prior knowledge about the number, magnitudes, nor the distribution of the nonzero gradients; as long as there are sufficiently few nonzero gradients in x, which is usually true for structural monitoring images, then (P_{TV}) accurately recovers the true $x^\star \approx x$. The CS imaging framework is shown in Ref. [68].

3.6.3. Simultaneous multi-channel data recovery by low-rank matrix completion. To alleviate the wireless data sampling or communication burden, certain amounts of data are often discarded during sampling or before transmission. In addition, it is not uncommon that measurements are corrupted, such as shown in Section 3.4.3. In these cases, it is required to recover the original structural vibration responses from the available, incomplete, or corrupted data, which is an ill-posed inverse problem.

Explicitly harnessing the data structure itself of the structural vibration responses provides a solution to addressing this problem [69]. The observation is that typically there are only few modes active in the structural vibration responses; hence, a sparse representation (in frequency domain) of the single-channel data vector or a low-rank structure (by SVD) of the multi-channel data matrix. Exploiting such prior knowledge of data structure (intra-channel sparse or inter-channel low-rank), the new theories of ℓ_1-minimization sparse recovery (presented in Section 3.5.1 and not repeated here) and nuclear-norm-minimization low-rank matrix completion [70] (presented in the following) enable recovery of the missing or corrupted structural vibration response data.

If the incomplete data of each channel are fused into a matrix, which is then reshaped, then the available data $\mathbf{B}$ can be denoted as a random subset Ω of the complete $\overline{\mathbf{X}} \in \mathbb{R}^{w \times v}$,

$$\mathbf{B} = \mathcal{P}_\Omega\left(\overline{\mathbf{X}}\right) \tag{42}$$

where $\mathcal{P}_\Omega(\cdot)$ is a sampling operator that randomly extracts partial or a subset elements of $\overline{\mathbf{X}}$ such that $B_{ij} = \overline{X}_{ij}, (i,j) \in \Omega$. Using only the available knowledge of a randomly undersampled subset $\mathbf{B}$ with M' entries ($M' \ll w \times v = m \times N$), if the true, unknown, $\overline{\mathbf{X}} \in \mathbb{R}^{w \times v}$, is sufficiently low-rank, then it can be accurately recovered with overwhelming probability of success by solving the convex optimization program

$$(P_*): \quad \overline{\mathbf{X}}^\star = \arg\ \min \left\|\overline{\mathbf{X}}\right\|_* \quad \text{subject to} \quad \left\|\mathbf{B} - \mathcal{P}_\Omega\left(\overline{\mathbf{X}}\right)\right\|_{\mathrm{F}} \leq \delta \tag{43}$$

which can be interpreted as to find the $\overline{\mathbf{X}}^\star$ with minimal nuclear norm (smallest rank) that explains the available observation $B_{ij} = \overline{X}_{ij}, (i,j) \in \Omega$ within a bounded noise level δ. Finally, the found $\overline{\mathbf{X}}^\star \in \mathbb{R}^{w \times v}$ is reshaped back to $\mathbf{X}^\star \in \mathbb{R}^{m \times N}$ in its original dimension. The framework is shown in Figure 12 with more illustrations in Ref. [69].

4. FUTURE WORK

With the knowledge of the data structure/model of system/damage features—sparse and low-rank— it is useful to explore more advanced mathematical tools (e.g., redundant dictionary, overcomplete representation leading to sparser representation of signals), to explicitly target such data structure, which is more outstanding in high-dimensional data. In search of the sparse representation, recent development has shifted from the traditional orthonormal basis (e.g., wavelet and Fourier) to redundant or overcomplete bases [14,71]. This is motivated from the fact that redundant representation can achieve even higher sparsity of signals than orthonormal bases. While orthonormal bases are pre-selected, redundant bases or dictionary, as the term suggests, are not unique, and new methods have been developed for pursuit of the overcomplete bases or redundant dictionary learned from and optimally adaptive to the signal itself. Although usually more computationally extensive, it is advantageous, for future work, to seek appropriate redundant representations, which may reveal sparser components for more clear visualization and interpretation of the structural dynamic and damage features, especially when the noise level is high.

Current infrastructure monitoring and assessment systems follow such a canonical procedure: sensing (massive) data and then extracting (few) signal features (e.g., sparse spikes for structural dynamic and damage features) from them, most of which are redundant and end up being discarded. Explicitly modeling and

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

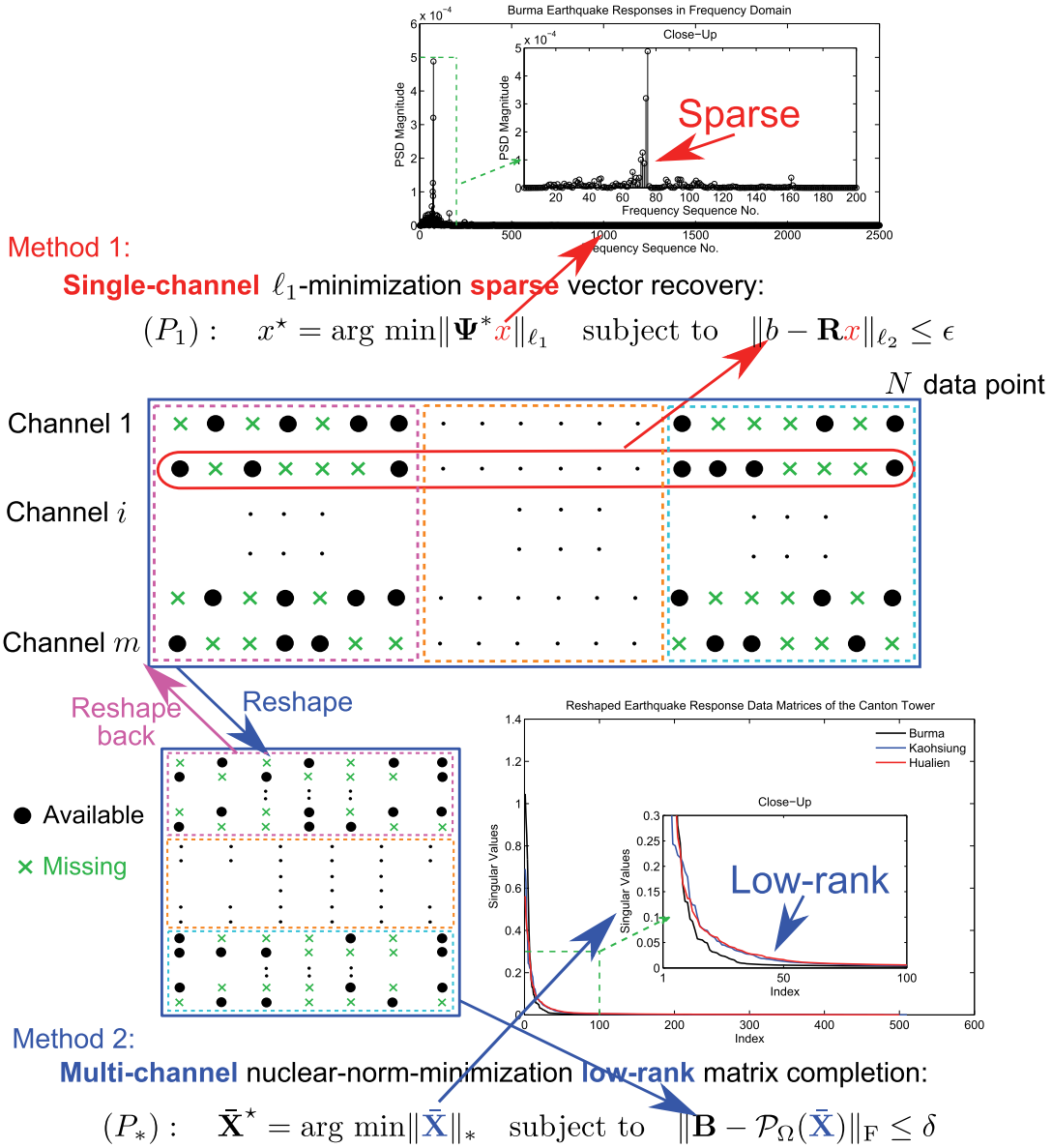

Figure 12. The data recovery using the single-channel sparse vector recovery method and the multi-channel matrix low-rank matrix completion method.

harnessing the sparse and low-rank data structure could inspire an innovative design of an intelligent sensing paradigm: direct sensing of the salient, relevant system/damage features, instead of sensing the naïve, inefficient full measurements and then discarding most of them. Such an intelligent sensing system could unify the big data sensing and processing procedures, which could significantly alleviate the data sampling, transferring, communication, and processing issues in wireless platform with limited resources. In addition, the knowledge of the data structure/model of the salient system and damage features could also inspire the development of *computational* super-resolution algorithms to achieve high-resolution structural information from the low-resolution structural measurement and imaging, as an economic alternative to the use of more expensive measurement system and imaging hardware of higher resolution.

It is also beneficial to integrate physics model-based and data based methods for a more effective and efficient diagnosis and prognosis of structures by taking advantage of both methods. Data-based methods are usually computationally efficient, which can be used for rapid system identification and damage detection in global SHM, for example, for preliminarily detecting the presence of damage

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

and finding the approximate damage region of the larger scale structures. Then, the established physics model of the (local) substructure—the found damage region, instead of the whole larger scale structure—can be used and updated for more accurate damage quantification and prognosis. In addition, although the reference dictionary depends on extracting the features by simulating different classes from an FEM model, it does not necessarily be so; reference features in the dictionary can learn from any available prior information, for example, features learned from experimental or historical field measurement data with labeled classes.

REFERENCES

1. Doebling SW, Farrar CR, Prime MB, Shevitz DW. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review, Report LA-13070-MS, UC-900, Los Alamos National Laboratory, New Mexico, 1996.
2. Ewin DJ. Modal Testing: Theory, Practice and Application (2nd edn). Research Studies Press: Hertfordshire, 2000.
3. Friswell MI, Mottershead JE. Finite Element Model Updating in Structural Dynamics. Kluwer Academic Publishers: London, 1995.
4. Ibrahim SR, Mikulcik EC. A time domain modal vibration test technique. *Shock and Vibration Bulletin* 1973; **43**(4):21–37.
5. Juang JN, Pappa RS. An eigensystem realization algorithm for modal parameter identification and model reduction. *Journal of Guidance, Control, and Dynamics* 1985; **8**:620–627.
6. VanOverschee P, De Moor B. Subspace Identification for Linear Systems: Theory, Implementation, Applications. Kluwer Academic Publishers: Boston, MA, 1996.
7. Brincker R, Zhang L, Andersen P. Modal identification of output-only systems using frequency domain decomposition. *Smart Material and Structures* 2001; **10**:441–445.
8. Farrar CR, Worden K. Structural Health Monitoring: A Machine Learning Perspective. Wiley, 2012.
9. Peeters B, Roeck GE. Stochastic system identification for operational modal analysis: a review. *Journal of Dynamic Systems, Measurement, and Control* 2001; **123**:659–667.
10. Kijewski T, Kareem A. Wavelet transforms for system identification in civil engineering. *Computer-aided Civil and Infrastructural Engineering* 2003; **18**(5):339–355.
11. Yang JN, Lei Y, Pan S, Huang N. System identification of linear structures based on Hilbert-Huang spectral analysis, part 1: normal modes. *Earthquake Engineering and Structural Dynamics* 2003; **32**:1443–1467.
12. Yang JN, Lei Y, Lin S, Huang N. Hilbert-Huang based approach for structural damage detection. *ASCE Journal of Engineering Mechanics* 2004; **130**:85–95.
13. Nagarajaiah S, Basu B. Output only identification and structural damage detection using time frequency and wavelet techniques. *Earthquake Engineering and Engineering Vibration* 2009; **8**:583–605.
14. Bruckstein AM, Donoho DL, Elad M. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review* 2008; **51**:34–81.
15. Candes E, Wakin M. An introduction to compressive sampling. *IEEE Signal Processing Magazine* 2008; **25**:21–30.
16. Candes E, Li X, Ma Y, Wright J. Robust principal component analysis. *Journal of ACM* 2009; **58**:1–37.
17. Hyvarinen A, Karhunen J, Oja E. Independent Component Analysis. Wiley: New York, NY, 2001.
18. Cover TM, Thomas JA. Elements of Information Theory. Wiley: New York, 1991.
19. Jolliffe I. Principal Component Analysis. Springer-Verlag: New York, 1986.
20. Feeny B, Kappagantu R. On the physical interpretation of proper orthogonal modes in vibration. *Journal of Sound and Vibration* 1998; **211**:607–616.
21. Yang Y, Nagarajaiah S. Data compression of structural seismic responses using principled independent component analysis. *ASCE Journal of Structural Engineering* 2014; **140**(7):0401–4032.
22. Yang Y, Nagarajaiah S. Blind denoising of structural responses with outliers via principal component pursuit. *Structural Control and Health Monitoring* 2014; **21**(6):962–978.
23. Belouchrani A, Abed-Meraim AK, Cardoso JF, Moulines E. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing* 1997; **45**:434–444.
24. Stone JV. Blind source separation using temporal predictability. *Neural Computation* 2001; **13**:1559–1574.
25. Antoni J, Braun S. Special issue: blind source separation. *Mechanical Systems and Signal Processing* 2005; **19**:1163–1380.
26. Kerschen G, Poncelet F, Golinval J-C. Physical interpretation of independent component analysis in structural dynamics. *Mechanical Systems and Signal Processing* 2007; **21**:1561–1575.
27. Poncelet F, Kerschen G, Golinval J-C, Verhelst D. Output-only modal analysis using blind source separation techniques. *Mechanical Systems and Signal Processing* 2007; **21**(2007):2335–2358.
28. McNeill SI, Zimmerman DC. A framework for blind identification using joint approximate diagonalization. *Mechanical Systems and Signal Processing* 2008; **22**:1526–1548.
29. Hazra B, Roffel AJ, Narasimhan S, Pandy MD. Modified cross-correlation method for the blind identification of structures. *ASCE Journal of Engineering Mechanics* 2010; **136**:889–897.
30. Abazarsa F, Ghahari SF, Nateghi F, Taciroglu E. Response-only modal identification of structures using limited sensors. *Structural Control and Health Monitoring* 2013; **20**:987–1006.
31. Yang Y, Nagarajaiah S. Time-frequency blind source separation using independent component analysis for output-only modal identification of highly-damped structures. *ASCE Journal of Structural Engineering* 2013; **139**:1780–1793.
32. Yang Y, Nagarajaiah S. Blind identification of damage in time-varying system using independent component analysis with wavelet transform. *Mechanical Systems and Signal Processing* 2014; **47**:3–20.
33. Antoni J, Chauhan S. A study and extension of second-order blind source separation to operational modal analysis. *Journal of Sound and Vibration* 2013; **332**:1079–1106.
34. Gribonval R, Lesage S. A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges. Proceedings of European Symposium on Artificial Neural Networks, Bruges, April 2006, pp. 323–330.

Copyright © 2016 John Wiley & Sons, Ltd.

DOI: 10.1002/stc

35. Yang Y, Nagarajaiah S. Output-only modal identification with limited sensors using sparse component analysis. *Journal of Sound and Vibration* 2013; **332**(19):4741–4765.
36. S. Nagarajaiah, Y. Yang. Blind modal identification of output-only non-proportionally-damped structures by time-frequency complex independent component analysis. *Smart Structures and Systems* 2015; **15**(1):81–97.
37. Sadhu A, Hazra B, Narasimhan S, Pandey MD. Decentralized modal identification using sparse blind source separation. *Smart Materials and Structures* 2011; **20**(12):125009.
38. Ghahari SF *et al.* Blind modal identification of structures from spatially sparse seismic response signals. *Structural Control and Health Monitoring* 2014; **21**(5):649–674.
39. Yang Y, Nagarajaiah S. Blind modal identification of output-only structures in time domain based on complexity pursuit. *Earthquake Engineering and Structural Dynamics* 2013; **42**:1885–1905.
40. Xie S, He Z, Fu Y. A note on Stone's conjecture of blind source separation. *Neural Computation* 2005; **17**:321–330.
41. Yang Y, Li S, Nagarajaiah S, Li H, Zhou P. Real-time output-only identification of time-varying cable tension via complexity pursuit. *ASCE Journal of Structural Engineering* 2015, in press, doi: 10.1061/(ASCE)ST.1943-541X.0001337, 04015083.
42. Zhang Y, Li J. Linear predictor-based lossless compression of vibration sensor data: systems approach. *Journal of Engineering Mechanics* 2007; **133**(4):431–441.
43. Yildirim U, Oguz O, Bogdanovic N. A prediction-error-based method for data transmission and damage detection in wireless sensor networks for structural health monitoring. *Journal of Vibration and Control* 2013; **19**(15):2244–2254.
44. Bogdanovic N, Ampeliotis D, Berberidis K, Casciati F, Plata-Chaves J. Spatio-temporal protocol for power-efficient acquisition wireless sensors based SHM. *Smart Structures and Systems* 2014; **14**(1):1–16.
45. De Boe P, Golinval J-C. Principal component analysis of a piezo-sensor array for damage localization. *International Journal of Structural Health Monitoring* 2003; **2**:137–144.
46. Friswell MI, Inman DJ. Sensor validation for smart structures. *Journal of Intelligent Material Systems and Structures* 1999; **10**:973–982.
47. Kerschen G, De Boe P, Golinval J, Worden K. Sensor validation using principal component analysis. *Smart Materials and Structures* 2005; **14**:36–42.
48. Yang Y, Nagarajaiah S. Data compression of very large-scale structural seismic and typhoon vibration responses by low-rank representation with matrix reshape. *Structural Control and Health Monitoring* 2015; **22**(8):1119–1131.
49. http://www.cse.polyu.edu.hk/benchmark/.
50. Yang Y, Nagarajaiah S. Blind identification of damage in time-varying systems using independent component analysis with wavelet transform. *Mechanical Systems and Signal Processing* 2014; **47**(1):3–20.
51. Yang Y, Nagarajaiah S. Dynamic imaging: real-time detection of local structural damage with blind separation of low-rank background and sparse innovation. *ASCE Journal of Structural Engineering,* 2015. doi:10.1061/(ASCE)ST.1943-541X.0001334.
52. Adeli H. Neural networks in civil engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering* 2001; **16**(2):126–142.
53. Widodo A, Yang B. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical System and Signal Processing* 2007; **21**:2560–2574.
54. Worden K, Lane AJ. Damage identification using support vector machines. *Smart Materials and Structures* 2001; **10**(3):540.
55. Trendafilova I, Cartmell MP, Ostachowicz W. Vibration-based damage detection in an aircraft wing scaled model using principal component analysis and pattern recognition. *Journal of Sound and Vibration* 2008; **313**:560–566.
56. Dharap P, Koh BH, Nagarajaiah S. Structural health monitoring using ARMarkov observers. *Journal of Intelligent Material Systems and Structures* 2006; **17**:469–481.
57. Sun H, Luş H, Betti R. Identification of structural models using a modified Artificial Bee Colony algorithm. *Computers & Structures* 2013; **116**:59–74.
58. Sun H, Waisman H, Betti R. Nondestructive identification of multiple flaws using XFEM and a topologically adapting artificial bee colony algorithm. *International Journal for Numerical Methods in Engineering* 2013; **95**(10):871–900.
59. Yang Y, Nagarajaiah S. Structural damage identification via a combination of blind feature extraction and sparse representation classification. *Mechanical Systems and Signal Processing* 2014; **45**(1):1–23.
60. Lynch JP, Loh KJ. A summary review of wireless sensors and sensor networks for structural health monitoring. *Shock and Vibration Digest* 2006; **38**:91–128.
61. Bao Y, Beck JL, Li H. Compressive sampling for accelerometer signals in structural health monitoring. *International Journal of Structural Health Monitoring* 2011; **10**:235–246.
62. Mascarenas D, Cattaneo A, Theiler J, Farrar C. Compressed sensing techniques for detecting damage in structures. *International Journal of Structural Health Monitoring* 2013; **12**(4):325–338.
63. O'Connor SM, Lynch JP, Gilbert AC. Compressed sensing embedded in an operational wireless sensor network to achieve energy efficiency in long-term monitoring applications. *Smart Materials and Structures* 2014; **23**(8):085014.
64. Yang Y, Nagarajaiah S. Output-only modal identification by compressed sensing: non-uniform low-rate random sampling. *Mechanical Systems and Signal Processing* 2015; **56**:15–34.
65. Tausiesakul B, Gkoktsi K, Giaralis A. Compressive sensing spectral estimation for output-only structural system identification. In *Proc. 7th Int. Conf. Computational Stochastic Mechanics*, 2014.
66. Comerford L, Kougioumtzoglou IA, Beer M. Compressive sensing based stochastic process power spectrum estimation subject to missing data. *Probabilistic Engineering Mechanics* 2015. doi:10.1016/j.probengmech.2015.09.015.
67. Bao Y, Li H, Sun X, Yu Y, Ou J. Compressive sampling based data loss recovery for wireless sensor networks used in civil structural health monitoring. *Structural Health Monitoring* 2012 1475921712462936.
68. Yang Y, Nagarajaiah S. Robust data transmission and recovery of images by compressed sensing for structural health diagnosis. *Structural Control and Health Monitoring* 2015 tentatively accepted.
69. Yang Y, Nagarajaiah S. Harnessing data structure for recovery of structural vibration responses from incomplete data: sparse representation versus low-rank structure. *Mechanical Systems and Signal Processing* 2015 tentatively accepted.
70. Candès E, Recht B. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics* 2009; **9**:717–772.
71. Mallat S. A wavelet tour of signal processing: the sparse way. Academic Press, 2009.

Copyright © 2016 John Wiley & Sons, Ltd.