Compressive Measurement Bounds for Wireless Sensor Networks in Structural Health Monitoring

Jae Young Park, Anna C. Gilbert, and Michael B. Wakin

Abstract

Structural Health Monitoring (SHM) systems are critical for monitoring aging infrastructure (such as buildings or bridges) in a cost-effective manner. Wireless sensor networks are particularly appealing for SHM applications due to their flexibility and low cost. However, in order to extend the battery life of wireless sensor nodes, it is essential to minimize the amount of data these sensors must collect and transmit. In this paper, we study two frameworks for Compressive Sensing (CS) in SHM systems. In particular, we suppose that individual sensor nodes sample vibration data over time; our two CS frameworks differ only in how the sample times are chosen (uniformly vs. randomly). We then suppose that the data is transmitted to a central node, where modal analysis can be used to detect damage in the structure or to populate a model for simulation. In particular, we examine the performance of a very simple technique for estimating the structure's mode shapes without requiring a traditional CS reconstruction of the vibration signals; all that is needed is to compute a single Singular Value Decomposition (SVD). We provide new finite sample analysis characterizing conditions under which this simple technique—also known as the Proper Orthogonal Decomposition (POD) in structural dynamics—can correctly estimate the mode shapes of the structure.

I. Introduction

A. Overview

Modal analysis plays an important role in Structural Health Monitoring (SHM) systems. Modal parameters of a structure (such as its modal frequencies, mode shapes, and modal damping ratios) describe the vibrational characteristics when external forces such as wind, earthquakes, or vehicle loadings are applied to the structure. Many damage detection algorithms make use of modal parameters to detect, localize, and assess the severity of damage [1], [2].

We consider the problem of determining the modes shapes of a structure from a collection of vibrational time series recordings as might be collected by a network of battery operated wireless sensor nodes deployed across the structure. Our primary concerns are: (i) guaranteeing the accuracy of the estimated mode shapes, and (ii) minimizing the amount of data that must be collected and then transmitted to a central node for processing, in order to extend the life of the sensor network. As a technique for estimating the mode shapes from the matrix of collected data, we study a very simple method known as the Singular Value Decomposition (SVD) in signal processing and as the Proper Orthogonal Decomposition (POD) in structural dynamics. It has previously been shown that, under certain assumptions, the singular vectors (also known as proper orthogonal modes) returned by such a technique converge to the true mode shapes asymptotically as the number of observations goes to infinity [3], [4]. We believe that our work is new in two respects:

- When time samples are collected *uniformly* in time from the sensors, we quantify the accuracy of the estimated mode shapes in terms of the sampling rate, total observation time, and spacing of the modal frequencies. This provides finite sample bounds for the SVD/POD technique that, to the best of our knowledge, have not appeared in the literature.
- When time samples are collected *randomly* in time from the sensors, we quantify the accuracy of the estimated mode shapes in terms of the number of samples, total observation time, and spacing of the modal frequencies. In some cases, we show that mode shape estimates obtained using a given number of random samples can be more accurate than the mode shape estimates obtained using an equivalent number of uniform time samples.

We support our analysis with representative experiments. Although the SVD/POD technique that we study is a simple one, and although we restrict our analysis to idealized N-degree-of-freedom systems under free vibration and with no damping, our goal is to extend our new understanding of this problem to more sophisticated techniques for modal analysis. Moreover, even with these naive assumptions, our experiments demonstrate that under certain conditions, this method can accurately recover the modal parameters on both synthetic data (see Section IV) and real data (see [5]).

This paper is closely related to our recently published work [5], where we also studied the two types of time sampling above. As we explain in Section III, however, our accuracy bounds in [5] were sensitive to cases where two or more modes were oscillating with the same (or nearly the same) amplitude. In this paper we extend our theorems from [5] to account for such cases. As we show, while the simple SVD-based estimation technique fails to accurately recover *individual* mode shapes in such instances, this technique can accurately recover the relevant *subspaces* spanned by the mode shapes of interest.

This paper is also related to [6], [7], which have explored the application of Compressive Sensing (CS) [8], [9], [10] to SHM systems. CS is a new framework for capturing high-dimensional signals using small numbers of randomized time

*JYP and ACG are with the Department of Electrical Engineering and Computer Science at the University of Michigan. Email: jaeypark,annacg@umich.edu. MBW is with the Department of Electrical Engineering and Computer Science at the Colorado School of Mines. Email: mwakin@mines.edu. This work was partially supported by NSF grants CCF-1161233 and CIF-0910765, AFOSR grant FA9550-09-1-0465, and NSF CAREER grant CCF-1149225.

samples or other compressive measurement protocols. CS is typically applied to one signal at a time, where a receiver uses a regularizing model (such as an assumption of sparsity in some basis) in order to recover each high-dimensional signal from its low-dimensional compressive measurements. Unfortunately, these recent works [6], [7] have found that a relatively high number of CS measurements can be required for signal-by-signal reconstruction to be accurate enough to permit adequate assessment of the structural health. Although our acquisition framework can be viewed as being CS-like in its motivation to minimize the requisite number of samples for modal analysis, our algorithmic framework for processing the samples is very different from traditional CS papers. In particular, we do not attempt to reconstruct the individual vibration signals and then perform modal analysis. Rather, we apply a single SVD to a matrix formed by concatenating the (low-dimensional) measured samples, and this immediately provides our estimates of the mode shapes.

B. Paper Organization

Section II lays the groundwork for our analysis, beginning with the modal superposition equation for the idealized N-degree-of-freedom systems that we consider. This section also details our uniform and random sampling strategies and explains how the SVD/POD can be used to estimate mode shapes from the vibrational time sample data. Section III presents and discusses our new theoretical results concerning the accuracy of this SVD-based technique for modal analysis. Section IV uses a series of simulations to demonstrate the relative tradeoffs of mode shape accuracy with the number of samples, the duration of sampling, and the closeness of the amplitudes of the oscillating modes. These simulations confirm that, while the SVD-based estimation technique may fail to accurately recover individual mode shapes in situations where two modes oscillating with nearly the same amplitude, this technique can accurately recover the relevant subspace spanned by the mode shapes of interest.

II. Preliminaries

A. Modal Superposition Equation

An unforced N-degree multiple-degree-of-freedom (MDOF) system[1] can be formulated as

$$[M]\{\ddot{u}(t)\} + [C]\{\dot{u}(t)\} + [K]\{u(t)\} = \{0(t)\}, \tag{1}$$

where $[M]$ is an $N \times N$ diagonal mass matrix, $[C]$ is a symmetric $N \times N$ damping matrix, $[K]$ is an $N \times N$ symmetric stiffness matrix, and $\{u(t)\}$ is an $N \times 1$ vector of displacement signals. Note that

$$\{u(t)\} = \left\{ \begin{array}{c} u_1(t) \\ u_2(t) \\ \vdots \\ u_N(t) \end{array} \right\},$$

and each $\{u(t)\}(l) = u_l(t)$, $l = 1, \ldots, N$, is a displacement signal. One can view $u_l(t)$ as the signal being observed at the lth sensor node.

We consider an undamped system and set $[C] = [0]$. We also assume that the mass matrix $[M]$ is proportional to the identity $[I]$ (i.e., that all N masses are equal). If this is not the case, each displacement signal can be normalized by the square root of the mass at that node (see [3], [4] for details); this yields an equivalent set of dynamic equations having a mass matrix equal to the identity.

Under these assumptions, the general solution to (1) takes the form

$$\{u(t)\} = \sum_{n=1}^{N} \{\psi_n\} \rho_n \cos(\omega_n t + \theta_n), \tag{2}$$

where $\{\psi_1\}, \ldots, \{\psi_N\}$ are a collection of real-valued $N \times 1$ spatial vectors known as the *mode shape vectors*. These vectors are normalized to have unit energy (i.e., $\|\{\psi_n\}\|_2 = 1$) and—importantly—they are orthogonal to one another. In (2), $\omega_1, \ldots, \omega_N$ are real-valued, positive numbers known as the *modal frequencies*. For simplicity in this paper, we assume the modal frequencies are distinct. The parameters ρ_n and θ_n denote the magnitude and phase, respectively, of the oscillating nth mode shape; these parameters may be affected by an input that induces vibrations in the structure prior to the observation period (during which we assume the structure is unforced). In the structural dynamics community, (2) is known as the *modal superposition equation*.

[1] Theoretically, a structure will have infinitely many degrees of freedom. However, the number of mode shapes that one can detect is equal to the number of sensors placed on the structure. In the following, whenever we deal with an N-degree MDOF system, we are implicitly assuming that we have N sensor nodes deployed on the structure.

B. The Analytic Signal of $\{u(t)\}$

Our framework and analysis will involve sampling what is known as the analytic signal of $\{u(t)\}$. A signal $v(t)$ is said to be *analytic* iff $V(f) = 0$ for $f < 0$, where $V(f)$ is the Fourier transform of $v(t)$ [Definition 1.2.1, [11]]. An analytic signal can be obtained by removing all negative frequencies in a given signal. The analytic signal is a frequently used representation in mathematics, signal processing, and communications; in some problems (such as ours) it can simplify the mathematical manipulations.

To discuss the analytic signal of $\{u(t)\}$, let us examine each entry in $\{u(t)\}$, i.e., $u_l(t)$. Extending the derivation in Section II-A, each $u_l(t)$ can be written as

$$\begin{aligned} u_l(t) &= \sum_{n=1}^{N} \{\psi_n\}(l)\rho_n \cos(\omega_n t + \theta_n) \\ &= \sum_{n=1}^{N} \{\psi_n\}(l)\frac{\rho_n}{2}\left(e^{i(\omega_n t+\theta_n)} + e^{-i(\omega_n t+\theta_n)}\right) \\ &= \sum_{n=1}^{N} \{\psi_n\}(l)(A_n e^{i\omega_n t} + A_n^* e^{-i\omega_n t}), \end{aligned}$$

where $A_n = \frac{\rho_n}{2}e^{i\theta_n}$ is a complex amplitude encoding the magnitude and phase of the oscillating nth mode shape. Thus, the analytic signal of $u_l(t)$, represented as $v_l(t)$, is simply

$$v_l(t) = \sum_{n=1}^{N} \{\psi_n\}(l)A_n e^{i\omega_n t},$$

and the analytic signal of the entire vector $\{u(t)\}$, denoted as $\{v(t)\}$, can be written as

$$\{v(t)\} = \sum_{n=1}^{N} \{\psi_n\}A_n e^{i\omega_n t}. \tag{3}$$

Note that $\{v(t)\}$ is no longer real but complex.

Obtaining an analytic signal in practice involves the application of a Hilbert transform. However, detailed discussion of this matter is out of scope of this paper and we will refer interested readers to [12] for more information. For the remainder of this paper, we will assume that we have successfully extracted the analytic signal from each $u_l(t)$. Thus, all derivations from here onwards will be in terms of $\{v(t)\}$.

C. Time Sampling and Matrix Formation

If certain modal frequencies have not been excited, it is possible that some of the complex amplitudes A_n may be zero. We let K denote the number of nonzero amplitudes. While in general K may be as large as N, in some cases K may be smaller than N and it will not be possible to estimate the modal parameters corresponding to the zero amplitudes. Without loss of generality, we also assume the complex amplitudes are ordered from large to small, and thus we will have $|A_1| \geq \cdots \geq |A_K| > A_{K+1} = \cdots = A_N = 0$. (We do not assume any particular ordering to the modal frequencies.)

Let us suppose that we sample each component of the analytic displacement signal vector $\{v(t)\}$ at M distinct points in time $t_1, \ldots, t_M$. We assume $M \geq K$ and denote the resulting $N \times M$ data matrix as

$$[V] = \begin{bmatrix} v_1(t_1) & v_1(t_2) & \ldots & v_1(t_M) \\ v_2(t_1) & v_2(t_2) & \ldots & v_2(t_M) \\ \vdots & \vdots & \ddots & \vdots \\ v_N(t_1) & v_N(t_2) & \ldots & v_N(t_M) \end{bmatrix} \in \mathbb{C}^{N\times M}. \tag{4}$$

We can write the analytic modal superposition equation (3) in matrix-vector multiplication format as

$$\{v(t)\} = [\Psi][\Gamma]\{s(t)\},$$

where $[\Psi] = [\{\psi_1\}, \ldots, \{\psi_K\}]$ denotes an $N \times K$ matrix formed from the mode shape vectors corresponding to the nonzero amplitudes (note that the columns of $[\Psi]$ are orthonormal due to our assumption that the mass matrix $[M]$ is proportional to the identity),

$$[\Gamma] = \sqrt{M} \cdot \begin{bmatrix} A_1 & 0 & \ldots & 0 \\ 0 & A_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & A_K \end{bmatrix}$$

Algorithm 1: Pseudo-code for mode shape estimation

Input: Data matrix $[V]$ as defined in (4)
Output: $[\widehat{\Psi}]$ (left singular vectors of $[V]$)
$\mathrm{SVD}([\mathrm{V}]) = [\widehat{\Psi}][\widehat{\Gamma}][\widehat{S}]$

denotes a $K \times K$ diagonal matrix, and

$$\{s(t)\} = \frac{1}{\sqrt{M}} \cdot \begin{Bmatrix} e^{i\omega_1 t} \\ e^{i\omega_2 t} \\ \vdots \\ e^{i\omega_K t} \end{Bmatrix}$$

denotes a $K \times 1$ modal coordinate vector. The sampling of $\{v(t)\}$ at $t_1, \ldots, t_M$ implies the sampling of $\{s(t)\}$ at the exact same time points which leads us to define

$$[S] = \frac{1}{\sqrt{M}} \cdot \begin{bmatrix} e^{i\omega_1 t_1} & e^{i\omega_1 t_2} & \ldots & e^{i\omega_1 t_M} \\ e^{i\omega_2 t_1} & e^{i\omega_2 t_2} & \ldots & e^{i\omega_2 t_M} \\ \vdots & \vdots & \ddots & \vdots \\ e^{i\omega_K t_1} & e^{i\omega_K t_2} & \ldots & e^{i\omega_K t_M} \end{bmatrix} \in \mathbb{C}^{K \times M}$$

and allows us to write the matrix of samples as

$$[V] = [\Psi][\Gamma][S]. \tag{5}$$

In the results that follow, we use $\delta_{\min}$ and $\delta_{\max}$ to denote lower and upper bounds on the minimum and maximum separation of the modal frequencies. In other words, we assume that

$$\delta_{\min} \leq \min |\omega_l - \omega_n|,$$

where the minimum is taken over all $l, n \in \{1, \ldots, K\}$ with $l \neq n$, and that

$$\delta_{\max} \geq \max |\omega_l - \omega_n|,$$

where the maximum is also taken over all $l, n \in \{1, \ldots, K\}$. Furthermore, we use $t_{\max}$ to denote the total sampling time span. Finally for $a, b \in [0, 1]$, the quantity

$$D(a||b) := a(\log(a) - \log(b)) + (1 - a)(\log(1 - a) - \log(1 - b))$$

is known as the binary information divergence, or the Kullback-Leibler divergence [13].

D. Estimating Modal Frequencies via the SVD/POD

The method we consider for recovering the mode shape vectors from uniform time samples or from random time samples is very simple and is described in Algorithm 1. In particular, one simply computes the SVD of $[V]$ and returns the truncated matrix of left singular vectors $[\widehat{\Psi}] = [\{\widehat{\psi_1}\}, \ldots, \{\widehat{\psi_K}\}]$ as estimates of the true mode shape matrix $[\Psi] = [\{\psi_1\}, \ldots, \{\psi_K\}]$. (Although the full-size SVD of the $N \times M$ matrix $[V]$ will return N left singular vectors, we use only the first K of these as estimates of the mode shape vectors.) Informally, the intuition behind this technique is that (5) provides one way of decomposing $[V]$ into a product of three matrices: one with orthonormal columns, one that is diagonal, and one that will have *nearly* orthonormal rows under certain conditions on the sample times $t_1, \ldots, t_M$ (see Section III). Thus, since (5) is itself nearly an SVD of $[V]$, the true left singular vectors of $[V]$ will nearly match the first matrix in the decomposition appearing in (5), namely the matrix $[\Psi]$ containing the true mode shape vectors.

Although it has not been considered in the context of CS, the use of the SVD for estimating mode shapes has also appeared under the name of the Proper Orthogonal Decomposition (POD) in the structural dynamics literature. In fact, it has previously been shown that, under certain assumptions, the singular vectors (also known as proper orthogonal modes) returned by such a technique converge to the true mode shapes asymptotically as the number of observations goes to infinity [3], [4]. As mentioned in Section I-A, we believe that our work is new in two respects:

- When time samples are collected *uniformly* in time from the sensors, we quantify the accuracy of the estimated mode shapes in terms of the sampling rate, total observation time, and spacing of the modal frequencies. This provides finite (non-asymptotic) sample bounds for the SVD/POD technique that, to the best of our knowledge, have not appeared in the literature.
- When time samples are collected *randomly* in time from the sensors, we quantify the accuracy of the estimated mode shapes in terms of the number of samples, total observation time, and spacing of the modal frequencies. In some cases,

we show that mode shape estimates obtained using a given number of random samples can be more accurate than the mode shape estimates obtained using an equivalent number of uniform time samples.

III. Analysis

A. Uniform Time Sampling

Let us now suppose that we sample $\{v(t)\}$ uniformly in time with a *uniform sampling interval* denoted by T_s. The sampling times are given by $t_m = (m-1)T_s$, $m = 1, \ldots, M$. We are therefore sampling within the time span $[0, t_{\max}]$, where $t_{\max} := (M-1)T_s$.

In [5], we proved the following bound for recovering the mode shape vectors using the SVD on uniform time samples.

Theorem 1: Let $[V] = [\Psi][\Gamma][S]$ be as given in (5) (describing an N-degree-of-freedom with no damping, equal masses, and K nonzero amplitudes in free decay) sampled according to the uniform sampling scheme described above. For $0 < \epsilon < 1$, suppose we sample for a total time span of at least

$$t_{\max} \geq \frac{2\pi(\log\lfloor K/2 \rfloor + 1.01)}{\epsilon \delta_{\min}} \tag{6}$$

with sampling interval $T_s \leq \frac{\pi}{\delta_{\max}}$ and ensure that $M \geq K$. Then, the mode shape estimates $[\widehat{\Psi}]$ obtained via Algorithm 1 satisfy

$$|\langle \{\psi_n\}, \{\widehat{\psi}_n\} \rangle|^2 \geq 1 - \frac{\epsilon^2(1+\epsilon)}{1-\epsilon} \cdot \max_{l \neq n} \frac{|A_l|^2 |A_n|^2}{\min\limits_{c \in [-1,1]} ||A_l|^2 - |A_n|^2(1+c\epsilon)|^2} \tag{7}$$

for $n = 1, \ldots, K$.

In (7), we see that the error bound for the nth estimated mode shape vector mainly depends on ϵ and what is essentially the minimum separation between $|A_n|$ and all other $|A_l|$. Because the mode shape vectors $\{\psi_n\}$ and their estimates $\{\widehat{\psi}_n\}$ are normalized to have unit energy (i.e., $\|\{\psi_n\}\|_2 = \|\{\widehat{\psi}_n\}\|_2 = 1$), the correlation $|\langle \{\psi_n\}, \{\widehat{\psi}_n\} \rangle|^2$ cannot exceed 1. When the nth largest amplitude $|A_n|$ is well separated from all other amplitudes $|A_l|$ (for $l \neq n$), (7) states that the correlation $|\langle \{\psi_n\}, \{\widehat{\psi}_n\} \rangle|^2$ will be close to 1. The variable ϵ controls how close the rows of $[S]$ are to being orthogonal; a small choice of ϵ implies more orthogonal rows and leads to a better preservation of the mode shapes but requires more samples. Note that the amplitudes $|A_n|$ are dependent on the underlying structure and on the input that excited the structure prior to free vibration observation period, and thus these parameters may be beyond our control. In order to guarantee a small error in the nth mode shape when $|A_n|$ is close to some other $|A_l|$, one would need to make ϵ smaller.

In terms of the roles played by $\delta_{\min}$ and $\delta_{\max}$ (the minimum and maximum separation of the modal frequencies), Theorem 1 essentially tells us that we need to sample for a time span that is inversely proportional to the minimum spacing between the modal frequencies. Thus, the smaller the minimum spacing between the modal frequencies, the longer we must sample to get an accurate estimate. Also, since the theorem requires that $T_s \leq \frac{\pi}{\delta_{\max}}$, the maximum spacing between modal frequencies determines how fast we need to sample.

Theorem 1 does not guarantee that the nth estimated mode shape vector can be accurately recovered when $|A_n|$ is close to some other $|A_l|$ (unless one chooses ϵ to be very small, which would require taking more samples). However, it is possible to extend Theorem 1 to account for situations where $|A_n|$ is close to some other $|A_l|$. In particular, when there are a set of amplitudes that are similar in size to each other, it may still be possible to recover the *subspace* spanned by the corresponding mode shape vectors.

We can make this formal by using the concept of principal angles between two subspaces. For two P-dimensional subspaces $\mathcal{X}_1, \mathcal{X}_2$ of $\mathbb{C}^N$, the principal angles between $\mathcal{X}_1$ and $\mathcal{X}_2$ are a set of angles $\theta_1, \theta_2, \ldots, \theta_P \in [0, \pi/2]$. These angles are defined recursively, with the kth angle given by [14]

$$\cos(\theta_k) = \max_{\{x_1\} \in \mathcal{X}_1} \max_{\{x_2\} \in \mathcal{X}_2} |\langle \{x_1\}\{x_2\} \rangle| = |\langle \{x_{1,k}\}, \{x_{2,k}\} \rangle|,$$

subject to the constraints that $\|\{x_1\}\|_2 = \|\{x_2\}\|_2 = 1$, $\langle \{x_1\}, \{x_{1,i}\} \rangle = 0$, and $\langle \{x_2\}, \{x_{2,i}\} \rangle = 0$ for all $i = 1, 2, \ldots, k-1$. Our new theorem is stated below and proved in the Appendix.

Theorem 2: Let $[V]$, $t_{\max}$, T_s, and M be as given in Theorem 1, and let $[\widehat{\Psi}]$ denote the mode shape estimates obtained via Algorithm 1. For any index set $\mathcal{I} \subset \{1, 2, \ldots, K\}$, set $P = |\mathcal{I}|$, let $[\Psi_{\mathcal{I}}]$ denote the $N \times P$ matrix formed by restricting $[\Psi]$ to the columns indexed by $\mathcal{I}$, and let $[\widehat{\Psi}_{\mathcal{I}}]$ denote the $N \times P$ matrix formed by restricting $[\widehat{\Psi}]$ to the columns indexed by $\mathcal{I}$. Let $\theta_1, \theta_2, \ldots, \theta_P \in [0, \pi/2]$ denote the principal angles between $\mathrm{colspan}([\Psi_{\mathcal{I}}])$ and $\mathrm{colspan}([\widehat{\Psi}_{\mathcal{I}}])$. Then

$$\sum_{i=1}^{P} \sin^2(\theta_i) \leq \frac{\epsilon^2(1+\epsilon)}{1-\epsilon} \cdot \sum_{n \in \mathcal{I}} \max_{l \notin \mathcal{I}} \frac{|A_l|^2 |A_n|^2}{\min\limits_{c \in [-1,1]} ||A_l|^2 - |A_n|^2(1+c\epsilon)|^2}. \tag{8}$$

This bound will be favorable (i.e., all θ_i will be guaranteed to be small) in situations where all $|A_n|$ for $n \in \mathcal{I}$ are well separated from all $|A_l|$ for $l \notin \mathcal{I}$. This bound does not depend on the proximity of amplitudes within the index set $\mathcal{I}$, and in fact all $|A_n|$ for $n \in \mathcal{I}$ could even be equal.

B. Random Time Sampling

Let us now suppose that we sample $\{v(t)\}$ at M time points $t_1, \ldots, t_M$ chosen *uniformly at random* in the time interval $[0, t_{\max}]$. We do not assume that these sample times are ordered (i.e., that $t_1 < t_2 < \cdots < t_M$), although in practice there is no harm in generating them randomly and then sorting them.

In [5], we proved the following bound for recovering the mode shape vectors using the SVD on random time samples.

Theorem 3: Let $[V] = [\Psi][\Gamma][S]$ be as given in (5) (describing an N-degree-of-freedom with no damping, equal masses, and K nonzero amplitudes in free decay) sampled according to the random sampling scheme described above. Suppose we sample for a total time span of at least

$$t_{\max} \geq \frac{40(\log\lfloor K/2 \rfloor + 1.01)}{\epsilon \delta_{\min}}, \tag{9}$$

and within this time span suppose we take a number of measurements satisfying

$$M > \max\left(\frac{\log(K) + \log(2/\tau)}{\min(D_1, D_2)},\ K \right) \tag{10}$$

where

$$D_1 = D((1+\epsilon)/K || (1+\epsilon/10)/K),$$
$$D_2 = D((1-\epsilon)/K || (1-\epsilon/10)/K).$$

Then with probability at least $1 - \tau$ all of the mode shape estimates $[\widehat{\Psi}]$ obtained via Algorithm 1 will satisfy (7).

This result for random time sampling looks somewhat similar to Theorem 1 for uniform time sampling. The recovery guarantee is the same and the required time span differs only by a constant.

However, Theorem 3 can differ from Theorem 1 in the number of samples M required for successful recovery of the mode shape vectors. For Theorem 1, recall that $t_{\max} = (M-1)T_s$. If we set $T_s = \frac{\pi}{\delta_{\max}}$ (the largest permitted by Theorem 1), then in order to satisfy the lower bound on $t_{\max}$ given in (6), we would need to collect at least

$$M \geq \max\left(\frac{2(\log\lfloor K/2 \rfloor + 1.01)}{\epsilon} \frac{\delta_{\max}}{\delta_{\min}} + 1,\ K \right) \tag{11}$$

total uniform time samples, and if we were to reduce the sampling interval T_s we would need to take more total samples M to cover the same time span $t_{\max}$. The necessary condition (11) on M is fairly satisfactory in its logarithmic dependence on K and its linear dependence on $\frac{1}{\epsilon}$ (this assumes the left hand term dominates in (11)). However, it also scales with the ratio $\frac{\delta_{\max}}{\delta_{\min}}$, which depends on the structure. For some structures this ratio could in fact be large, and in the absence of additional information about the structure, one may need to assume this ratio is large.

For Theorem 3, the requisite number of samples M in (10) no longer depends on the ratio $\frac{\delta_{\max}}{\delta_{\min}}$. However, when K is large and ϵ is small, the denominator in (10) will scale like $\frac{\epsilon^2}{K}$, and so the requisite number of samples will scale like $\frac{K \log(K)}{\epsilon^2}$. This represents a stronger dependence on K compared to what appears in (11), but only by a logarithmic factor (because the right hand term in (11) scales like K). It also represents a stronger dependence on ϵ compared to what appears in (11). Ultimately, we see that in some cases random time sampling could provide a significant reduction in the number of samples for systems where $\frac{\delta_{\max}}{\delta_{\min}}$ is large or unknown. For a given problem, the better choice between uniform and random time sampling may depend on the particular circumstances and the parameters of the system under study. This is reflected in our simulations in Section IV.

Like Theorem 1, Theorem 3 does not guarantee that the nth estimated mode shape vector can be accurately recovered when $|A_n|$ is close to some other $|A_l|$. However, using the concept of principal angles between two subspaces, it is possible to extend Theorem 3 to account for situations where $|A_n|$ is close to some other $|A_l|$. In particular, when there are a set of amplitudes that are similar in size to each other, it may still be possible to recover the *subspace* spanned by the corresponding mode shape vectors. We formalize this in the following new theorem.

Theorem 4: Let $[V]$, $t_{\max}$, M be as given in Theorem 3, and let $[\widehat{\Psi}]$ denote the mode shape estimates obtained via Algorithm 1. Then with probability at least $1-\tau$, the following holds for any index set $\mathcal{I} \subset \{1, 2, \ldots, K\}$: Set $P = |\mathcal{I}|$, let $[\Psi_{\mathcal{I}}]$ denote the $N \times P$ matrix formed by restricting $[\Psi]$ to the columns indexed by $\mathcal{I}$, and let $[\widehat{\Psi}_{\mathcal{I}}]$ denote the $N \times P$ matrix formed by restricting $[\widehat{\Psi}]$ to the columns indexed by $\mathcal{I}$. Let $\theta_1, \theta_2, \ldots, \theta_P \in [0, \pi/2]$ denote the principal angles between $\mathrm{colspan}([\Psi_{\mathcal{I}}])$ and $\mathrm{colspan}([\widehat{\Psi}_{\mathcal{I}}])$. Then (8) is satisfied.

Like Theorem 2, this bound will be favorable (i.e., all θ_i will be guaranteed to be small) in situations where all $|A_n|$ for $n \in \mathcal{I}$ are well separated from all $|A_l|$ for $l \notin \mathcal{I}$. This bound does not depend on the proximity of amplitudes within the index set $\mathcal{I}$, and in fact all $|A_n|$ for $n \in \mathcal{I}$ could even be equal.

The proof of Theorem 4 closely follows the proof of Theorem 2, which is provided in the Appendix. The only difference is that we must argue that $\|[\Delta]\|_2 \leq \epsilon$ with probability at least $1-\tau$ when the time samples are chosen randomly, rather than uniformly. Again, however, we have previously provided a proof of this fact in [5], and so we do not repeat the argument here.

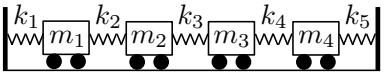
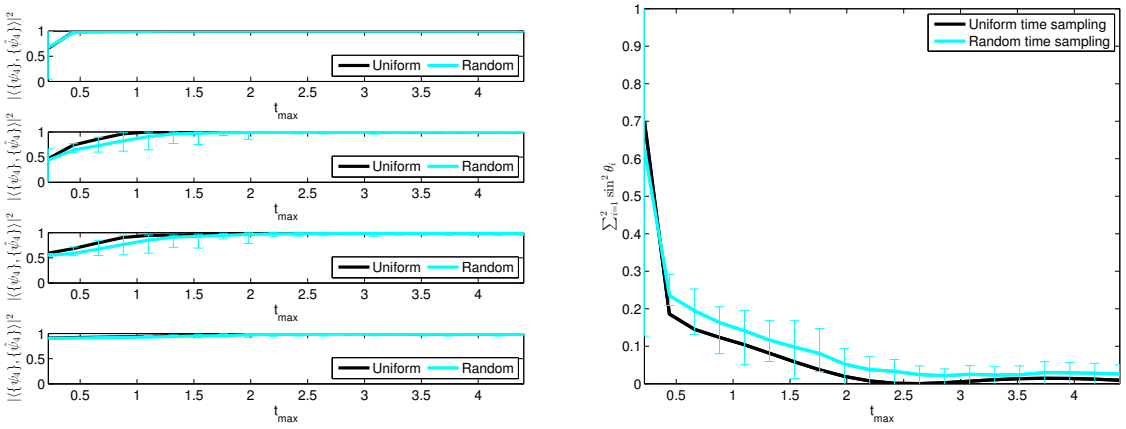


Fig. 1: *A simple 4-degree-of-freedom system made up of 4 box cars each of mass m_1, m_2, m_3, m_4. The box cars are connected via springs of stiffness $k_1, \ldots, k_5$. In the synthetic simulations, we simulate this particular system with $m_1 = m_2 = m_3 = m_4 = 1$kg, and $k_1 = 100$, $k_2 = 150$, $k_3 = 100$, $k_4 = 100$, $k_5 = 200$N/m.*



(a) (b)

Fig. 2: *Simulation results making use of the $[\Gamma]$ matrix as given in (12) with a set of modal frequencies with $\delta_{\min} = 0.715\pi$ rad/s. (a) Correlation $|\langle\{\psi_n\}, \{\widehat{\psi}_n\}\rangle|^2$ of ground truth mode shapes and mode shape estimates using uniform (black) time sampling scheme. Average correlation $|\langle\{\psi_n\}, \{\widehat{\psi}_n\}\rangle|^2$ of ground truth mode shapes and mode shape estimates using random (blue) time sampling scheme over 100 trials for each value of $t_{\max}$. The four subfigures correspond to the four mode shapes ($n = 1, 2, 3, 4$), and the results are plotted as a function of the total sampling time $t_{\max}$. (b) Plot of sum of $\sin^2$ of principal angles between colspan($[\{\widehat{\psi}_2\}, \{\widehat{\psi}_3\}]$) and colspan($[\{\psi_2\}, \{\psi_3\}]$) against $t_{\max}$. The plot of the random (blue) time sampling scheme has been averaged over 100 trials. The lowest and the highest point of the error bars in the two plots correspond to the minimum and maximum value of correlations that occurred within each of the 100 trials, respectively. Both plots illustrate that the mode shape estimates via uniform and random sampling schemes perform comparably.*

IV. Illustrative Simulations

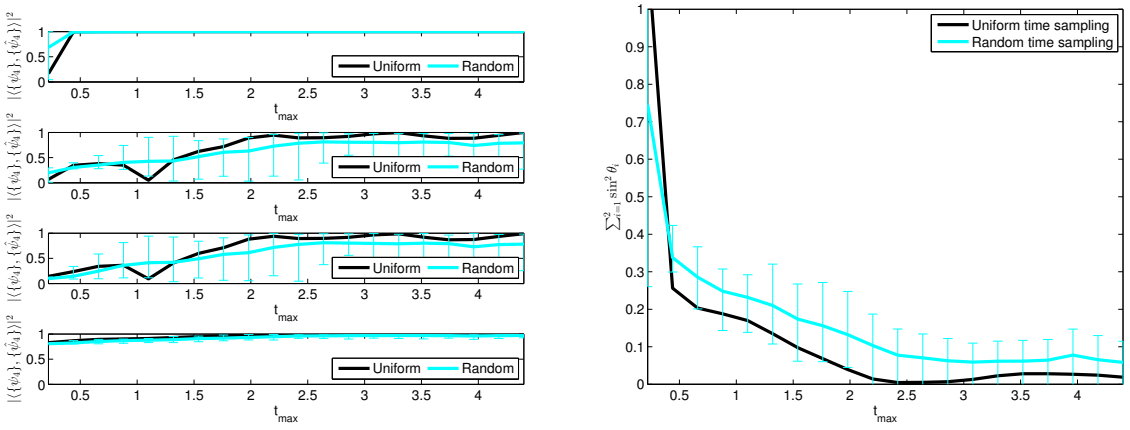
We demonstrate the effectiveness of the various methods using an idealized synthetic dataset. The system that we consider is illustrated in Figure 1. It is a simple 4-degree-of-freedom system consisting of four objects of masses $m_1, \ldots, m_4$ arranged in a line. Springs connect each adjacent pair of masses, where the springs have stiffness values of $k_1, \ldots, k_5$. We do not consider damping in this system; thus $[C] = [0]$. For the following simulation, we simulate the above system under free vibration with $m_1 = \cdots = m_4 = 1$kg, i.e., $[M] = I$, and with $k_1 = 100$, $k_2 = 150$, $k_3 = 100$, $k_4 = 100$, and $k_5 = 200$N/m. The stiffness matrix corresponding to the above system can be written as

$$[K] = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 & 0 \\ -k_2 & k_2 + k_3 & -k_3 & 0 \\ 0 & -k_3 & k_3 + k_4 & -k_4 \\ 0 & 0 & -k_4 & k_4 + k_5 \end{bmatrix} = \begin{bmatrix} 250 & -150 & 0 & 0 \\ -150 & 250 & -100 & 0 \\ 0 & -100 & 200 & -100 \\ 0 & 0 & -100 & 300 \end{bmatrix}.$$

Since $[M] = [I]$, the mode shape vectors $[\Psi]$ are simply the eigenvectors of $[K]$ and we use these mode shapes throughout this subsection. The modal frequencies are simply the square roots of the eigenvalues of $[K]$, which are $\omega_1 = 6.635\pi$, $\omega_2 = 5.920\pi$, $\omega_3 = 4.163\pi$, $\omega_4 = 2.220\pi$ rad/s. We also denote $[\Gamma]$ as the matrix of amplitudes throughout this subsection. Then, the modal superposition equation can be written as

$$\{v(t)\} = [\Psi][\Gamma] \underbrace{\begin{Bmatrix} e^{i\omega_1 t} \\ e^{i\omega_2 t} \\ e^{i\omega_3 t} \\ e^{i\omega_4 t} \end{Bmatrix}}_{\{s(t)\}}.$$

In our first experiment, we demonstrate the uniform and random time sampling methods by plotting the correlations $|\langle\{\psi_n\}, \{\widehat{\psi}_n\}\rangle|^2$ of the true and the estimated mode shapes obtained from the SVD of the respective sampled matrix $[V]$.



(a) (b)

Fig. 3: *Simulation results making use of the* $[\Gamma]$ *matrix as given in* (13) *with a set of modal frequencies with* $\delta_{\min} = 0.715\pi$ *rad/s. (a) Correlation* $|\langle\{\psi_n\}, \{\widehat{\psi}_n\}\rangle|^2$ *of ground truth mode shapes and mode shape estimates using uniform (black) time sampling scheme. Average correlation* $|\langle\{\psi_n\}, \{\widehat{\psi}_n\}\rangle|^2$ *of ground truth mode shapes and mode shape estimates using random (blue) time sampling scheme over 100 trials for each value of* $t_{\max}$*. The four subfigures correspond to the four mode shapes* ($n = 1, 2, 3, 4$)*, and the results are plotted as a function of the total sampling time* $t_{\max}$*. (b) Plot of sum of* $\sin^2$ *of principal angles between colspan*$([\{\widehat{\psi}_2\}, \{\widehat{\psi}_3\}])$ *and colspan*$([\{\psi_2\}, \{\psi_3\}])$ *against* $t_{\max}$*. The plot of the random (blue) time sampling scheme has been averaged over 100 trials. The lowest and the highest point of the error bars in the two plots correspond to the minimum and maximum value of correlations that occurred within each of the 100 trials, respectively.*

We use the following $[\Gamma]$ matrix

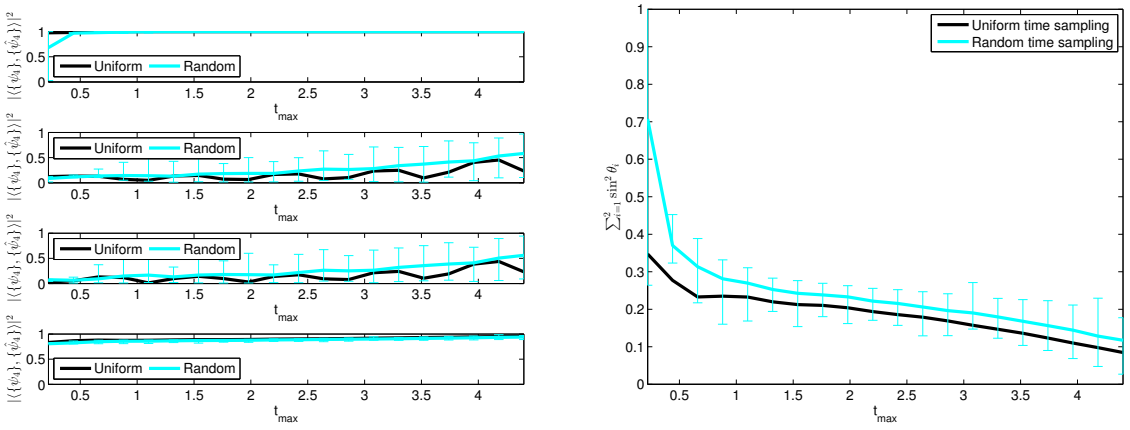
$$[\Gamma] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.45 & 0 & 0 \\ 0 & 0 & 0.15 & 0 \\ 0 & 0 & 0 & 0.01 \end{bmatrix}, \tag{12}$$

which has a well-separated set of $|A_n|$. For the uniform time sampling scheme, we use a sampling interval of $T_s = 0.22$s. This is just slightly faster than $\frac{\pi}{\delta_{\max}}$, which is the upper bound of what Theorem 1 prescribes. Using this fixed rate, we collect samples over a total time span of duration $t_{\max}$, and we repeat the experiment for values of $t_{\max} = [0 : T_s : 4.4]$s (the value of M therefore increases with $t_{\max}$). Figure 2(a) plots in black curves the correlation between the true and the estimated mode shape vectors against $t_{\max}$. For the random time sampling scheme, we use the same values of $t_{\max}$ and sample $\{v(t)\}$ uniformly at random within the interval $[0,\ t_{\max}]$. For each value of $t_{\max}$, the total number of samples M we collect is chosen to equal the corresponding value of M used for uniform time sampling above. For the random sampling scheme we plot in blue curves the average correlations between the true and the estimated mode shapes over 100 trials. The range of the error bars in the plot indicate the minimum and maximum value of correlations that occurred within the 100 trials. We see that overall, the performance of the two sampling schemes is comparable. This is in agreement with Theorems 1 and 3, as they both suggest the same reconstruction guarantees given that we satisfy the sampling conditions.

In the second experiment, we use the following $[\Gamma]$ matrix:

$$[\Gamma] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.55 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.01 \end{bmatrix}. \tag{13}$$

Note that this matrix contains closely spaced amplitude values of $|A_2| = 0.55$ and $|A_3| = 0.5$. Thus, from our theorems above we anticipate $\{\widehat{\psi}_2\}$ and $\{\widehat{\psi}_3\}$ will be less accurately preserved than in the previous experiment given the same values of $t_{\max}$. However, we do expect the estimates to get more accurate as we increase $t_{\max}$ (as increasing $t_{\max}$ would allow for a decrease of ϵ). Furthermore, while we expect the estimates $\{\widehat{\psi}_2\}$ and $\{\widehat{\psi}_3\}$ to degrade in accuracy, we expect the accuracy of the estimates $\{\widehat{\psi}_1\}$ and $\{\widehat{\psi}_4\}$ to stay more or less the same. All of this is in agreement with the simulation results plotted in Figure 3(a), which shows the results obtained by repeating the previous simulation but with the $[\Gamma]$ matrix given in (13). As we can see from the second and third panes, the proposed method struggles to obtain accurate estimates of $\{\psi_2\}$ and $\{\psi_3\}$. When compared with the second and third panes of Figure 2(a) the difference is especially pronounced when $t_{\max} \in [0.5,\ 1.5]$. The accuracy, however, does generally increase with increasing $t_{\max}$. The first and the fourth panes illustrate that the estimates $\{\widehat{\psi}_1\}$ and $\{\widehat{\psi}_4\}$ remain accurate despite the change in $[\Gamma]$.



(a) (b)

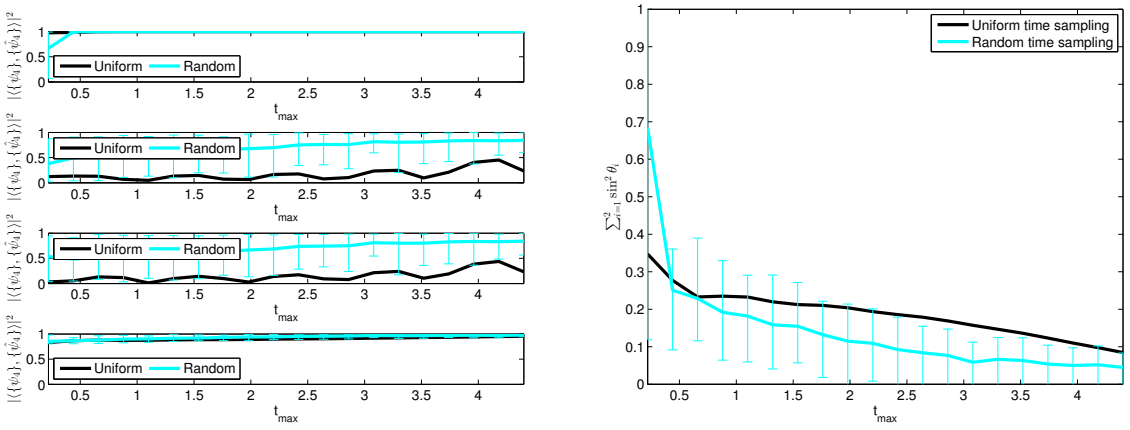
Fig. 4: *Simulation results making use of the* $[\Gamma]$ *matrix as given in* (13) *with a set of modal frequencies with* $\delta_{\min} = 0.286\pi$ *rad/s. (a) Correlation* $|\langle\{\psi_n\}, \{\widehat{\psi}_n\}\rangle|^2$ *of ground truth mode shapes and mode shape estimates using uniform (black) time sampling scheme. Average correlation* $|\langle\{\psi_n\}, \{\widehat{\psi}_n\}\rangle|^2$ *of ground truth mode shapes and mode shape estimates using random (blue) time sampling scheme over 100 trials for each value of* $t_{\max}$*. The four subfigures correspond to the four mode shapes* ($n = 1, 2, 3, 4$)*, and the results are plotted as a function of the total sampling time* $t_{\max}$*. (b) Plot of sum of* $\sin^2$ *of principal angles between colspan*$([\{\widehat{\psi}_2\}, \{\widehat{\psi}_3\}])$ *and colspan*$([\{\psi_2\}, \{\psi_3\}])$ *against* $t_{\max}$*. The plot of the random (blue) time sampling scheme has been averaged over 100 trials. The lowest and the highest point of the error bars in the two plots correspond to the minimum and maximum value of correlations that occurred within each of the 100 trials, respectively.*

Figure 3(a) confirms that the quality of the individual estimates $\{\widehat{\psi}_2\}$ and $\{\widehat{\psi}_3\}$ suffer when the corresponding amplitudes $|A_2|$ and $|A_3|$ are closely spaced. As we have discussed in Section III, however, the perturbations of these vectors are not completely arbitrary. More specifically, in Theorems 2 and 4 we have showed that the principal angles between the space spanned by $\{\widehat{\psi}_2\}$ and $\{\widehat{\psi}_3\}$ and the space spanned by $\{\psi_2\}$ and $\{\psi_3\}$ should be small if $|A_2|$ and $|A_3|$ are well separated from $|A_1|$ and $|A_4|$. Figure 3(b) confirms this by plotting $\sum_{i=1}^{2} \sin^2 \theta_i$ versus $t_{\max}$, where θ_1 and θ_2 are the two principal angles between the two-dimensional subspaces colspan$([\{\widehat{\psi}_2\}, \{\widehat{\psi}_3\}])$ and colspan$([\{\psi_2\}, \{\psi_3\}])$. Again, the plot of the random time sampling scheme (in blue) has been averaged over 100 trials and the range of the error bars indicate the minimum and maximum value of correlations that occurred within the 100 trials. As we can see, for both uniform and random sampling schemes the angles are reasonably small even for small $t_{\max}$ values, e.g., $t_{\max} \in [0.5, 1.5]$, even though the individual estimates $\{\widehat{\psi}_2\}$ and $\{\widehat{\psi}_3\}$ were highly inaccurate in this range of $t_{\max}$. For the sake of comparison, Figure 2(b) plots $\sum_{i=1}^{2} \sin^2 \theta_i$ versus $t_{\max}$ for the simulation setup used in Figure 2(a).

Our third experiment highlights the role played by $\delta_{\min}$ (the minimum separation of the modal frequencies). We consider two distinct systems. The first system is the same as above with $[\Gamma]$ as given in (13). We simulate a second system that uses the exact same mode shape vectors and $[\Gamma]$ but has a slightly different set of modal frequencies, such that $\omega_1 = 6.635\pi$, $\omega_2 = 6.349\pi$, $\omega_3 = 4.163\pi$, $\omega_4 = 2.220\pi$ rad/s.[2] Note that $\delta_{\min} = 0.715\pi$ in the first system, and $\delta_{\min} = 0.286\pi$ in the second system. Based on Theorems 1 and 3, we anticipate the need to sample for a longer time span (larger $t_{\max}$) when $\delta_{\min}$ is smaller. The correlations of the uniform sampling scheme and the average correlations (over 100 trials) of the random sampling scheme using the second set of modal frequencies are plotted in Figure 4(a). Comparing to the results in Figure 3(a), we see that a longer sampling duration $t_{\max}$ is indeed needed to achieve comparable accuracy in mode shapes. As in the previous simulation, in Figure 4(b) we again plot $\sum_{i=1}^{2} \sin^2 \theta_i$ versus $t_{\max}$, where θ_1 and θ_2 are the two principal angles between the two-dimensional subspaces colspan$([\{\widehat{\psi}_2\}, \{\widehat{\psi}_3\}])$ and colspan$([\{\psi_2\}, \{\psi_3\}])$. Again, for both uniform and random sampling schemes the subspace of colspan$([\{\psi_2\}, \{\psi_3\}])$ is reasonably well-preserved even for small $t_{\max}$ values.

For uniform time sampling, when T_s is fixed, increasing $t_{\max}$ will automatically increase the number of samples M to be acquired. For random time sampling, however, $t_{\max}$ and M can be chosen independently of one another. While in Figure 4 we have always used the same value of $t_{\max}$ and the corresponding M for random sampling as we used for uniform sampling, this is not actually necessary. In fact, our Theorems 3 and 4 suggest that, for random sampling when $\delta_{\min}$ is small, we can increase $t_{\max}$ without increasing the number of samples M. To demonstrate this, we conduct a fourth experiment, and for this we use the second (more closely spaced) set of modal frequencies above. For several values of M, we collect M samples using both uniform time sampling (for which $t_{\max}$ will be determined by M) and random time sampling. For

[2]Note that this system may not have the same setup as the one illustrated in Figure 1. However, in order to focus on the effect $\delta_{\min}$ has on the mode shape recovery we have developed this hypothetical scenario where we keep all parameters fixed except ω_2 in order to achieve a smaller $\delta_{\min}$. Similarly, in the simulations that follow, where necessary, we will change the modal frequencies but always employ the same mode shapes and $[\Gamma]$ matrix of (13).



(a) (b)

Fig. 5: *Simulation results making use of the* $[\Gamma]$ *matrix as given in* (13) *with a set of modal frequencies with* $\delta_{\min} = 0.286\pi$ *rad/s. For each value of* M*, the* $t_{\max}$ *for random sampling is chosen to be* 3 *seconds longer than the* $t_{\max}$ *for uniform sampling. (a) Correlation* $|\langle\{\psi_n\},\{\widehat{\psi}_n\}\rangle|^2$ *of ground truth mode shapes and mode shape estimates using uniform (black) time sampling scheme. Average correlation* $|\langle\{\psi_n\},\{\widehat{\psi}_n\}\rangle|^2$ *of ground truth mode shapes and mode shape estimates using random (blue) time sampling scheme over 100 trials for each value of* $t_{\max}$*. The four subfigures correspond to the four mode shapes* ($n = 1, 2, 3, 4$)*, and the results are plotted as a function of the number of measurements* M*. (b) Plot of sum of* $\sin^2$ *of principal angles between colspan*$([\{\widehat{\psi}_2\},\{\widehat{\psi}_3\}])$ *and colspan*$([\{\psi_2\},\{\psi_3\}])$ *against* M*. The plot of the random (blue) time sampling scheme has been averaged over 100 trials. The lowest and the highest point of the error bars in the two plots correspond to the minimum and maximum value of correlations that occurred within each of the 100 trials, respectively.*

each value of M with random time sampling, however, we choose $t_{\max}$ to be 3s longer than the value of $t_{\max}$ used for uniform sampling with the same value of M. The correlation result of the uniform time sampling and the averaged result (over 100 trials) of the random time sampling are shown in Figure 5. Both Figure 5(a) and (b) demonstrate that by simply increasing $t_{\max}$ without affecting M, the random time sampling scheme can accommodate the decreased value of $\delta_{\min}$.

APPENDIX

The proof of Theorem 2 closely follows the proof of Theorem 1 from [5]. Here we summarize the essential arguments.

To prove Theorem 2, we take a perturbation theoretic viewpoint. We start with the equation $[V] = [\Psi][\Gamma][S]$. Let us note that $[V][V]^* = [\Psi][\Gamma][S][S]^*[\Gamma]^*[\Psi]^*$, where $[\Psi]$ is an $N \times K$ matrix containing the true mode shape vectors as orthonormal columns, and $[S][S]^*$ is a $K \times K$ matrix with entries

$$([S][S]^*)_{l,n} = \begin{cases} 1, & l = n, \\ \frac{1}{M}\sum_{m=1}^{M} e^{i(\omega_l - \omega_n)t_m}, & l \neq n. \end{cases}$$

Thus, we can decompose this product as $[S][S]^* = [I] + [\Delta]$, where $[\Delta]$ contains the off-diagonal entries of $[S][S]^*$. Then

$$[V][V]^* = [\Psi][\Gamma][\Gamma]^*[\Psi]^* + [\Psi][\Gamma][\Delta][\Gamma]^*[\Psi]^*. \tag{14}$$

The above expression allows us to view $[V][V]^*$ as the summation of an "original" matrix $[\Psi][\Gamma][\Gamma]^*[\Psi]^*$ and a "perturbation" matrix $[\Psi][\Gamma][\Delta][\Gamma]^*[\Psi]^*$. Noting that the eigenvectors of the original matrix $[\Psi][\Gamma][\Gamma]^*[\Psi]^*$ are given by $[\Psi]$, our goal is to show that the eigenvectors of $[V][V]^*$ (which equal the left singular vectors of $[V]$) are close to those of this original matrix. At this point, it is helpful to define

$$\begin{aligned} [G] &:= [\Psi]^*[V][V]^*[\Psi] \\ &= [\Psi]^*[\Psi][\Gamma][\Gamma]^*[\Psi]^*[\Psi] + [\Psi]^*[\Psi][\Gamma][\Delta][\Gamma]^*[\Psi]^*[\Psi] \\ &= \underbrace{[\Gamma][\Gamma]^*}_{[H]} + \underbrace{[\Gamma][\Delta][\Gamma]^*}_{[\delta H]}. \end{aligned} \tag{15}$$

In this equation, we can view $[G]$ as the sum of an original matrix $[H]$ and a perturbation $[\delta H]$.

Recall that we write the SVD of $[V]$ as $[\widehat{\Psi}][\widehat{\Gamma}][\widehat{S}]$. Since the rank of $[V]$ must be K or less—and in fact we explain below conditions under which it is guaranteed to be equal to K—we can consider the truncated SVD of $[V]$ where $[\widehat{\Psi}]$ has size $N \times K$. Therefore, note that

$$[G] = [\Psi]^*[V][V]^*[\Psi] = [\Psi]^*[\widehat{\Psi}][\widehat{\Gamma}][\widehat{\Gamma}]^*[\widehat{\Psi}]^*[\Psi].$$

Since the column spans of $[\Psi]$ and $[\widehat{\Psi}]$ are the same, $[\Psi]^*[\widehat{\Psi}]$ is a $K \times K$ orthonormal matrix that contains the eigenvectors of $[G]$. The following theorem allows us to relate the eigenvectors of $[H]$ to those of $[G]$.

Theorem 5 (Theorem 1, [15]): Let $[H] = [R][\Upsilon][R]^*$ and $[G] = [H] + [\delta H] = [\widehat{R}][\widehat{\Upsilon}][\widehat{R}]^*$ be $K \times K$ positive definite matrices. Assume that $[R]$ and $[\widehat{R}]$ are unitary and that $[\Upsilon] = \text{diag}(\upsilon_1, \ldots, \upsilon_K)$ and $[\widehat{\Upsilon}] = \text{diag}(\widehat{\upsilon}_1, \ldots, \widehat{\upsilon}_K)$ are diagonal. Let $[B] = [R]^*[\widehat{R}]$, and assume

$$\eta = \|[H]^{-\frac{1}{2}}[\delta H][H]^{-\frac{1}{2}}\| < 1, \tag{16}$$

where $[H]^{-\frac{1}{2}} = [R][\Upsilon]^{-1/2}[R]^*$. Then for any n and for *any* set $\mathscr{T} \subseteq \{1, \ldots, K\}$ not containing n we have,

$$\sum_{l \in \mathscr{T}} |b_{l,n}|^2 \leq \max_{l \in \mathscr{T}} \frac{\upsilon_l \widehat{\upsilon}_n}{|\upsilon_l - \widehat{\upsilon}_n|^2} \frac{\eta^2}{1-\eta}. \tag{17}$$

We apply Theorem 5 with $[H]$ and $[\delta H]$ as defined in (15), allowing us to take $[R] = [I]$, $[\widehat{R}] = [\Psi]^*[\widehat{\Psi}]$, and $|b_{l,n}| = |\langle\{\widehat{\psi}_n\}, \{\psi_l\}\rangle| = |\langle\{\psi_l\}, \{\widehat{\psi}_n\}\rangle|$.

Note that the columns of $[\Psi]$ are orthonormal, the columns of $[\widehat{\Psi}]$ are orthonormal, and the column spans of $[\Psi]$ and $[\widehat{\Psi}]$ are equal. Without loss of generality, less us assume that the index set $\mathcal{I} = \{1, 2, \ldots, P\}$ (e.g., so that $[\Psi_{\mathcal{I}}]$ merely corresponds to the first P columns of $[\Psi]$). This allows us to partition $[\Psi]$ and $[\widehat{\Psi}]$ as follows:

$$[\Psi] = [[\Psi_{\mathcal{I}}]\ [\Psi_{\mathcal{I}^c}]]$$

and

$$[\widehat{\Psi}] = \left[[\widehat{\Psi}_{\mathcal{I}}]\ [\widehat{\Psi}_{\mathcal{I}^c}]\right]$$

We also let $[\Psi_\perp]$ denote an $N \times (N-K)$ matrix with orthonormal columns spanning the orthogonal complement of $\text{colspan}([\Psi])$. Our analysis will use the following theorem.

Theorem 6 ([14], [16]): Let $[X]$ and $[Z]$ be $N \times P$ matrices with orthonormal columns spanning the subspaces $\mathcal{X}$ and $\mathcal{Z}$, respectively. Let $\theta_1, \theta_2, \ldots, \theta_P \in [0, \pi/2]$ denote the principal angles between $\mathcal{X}$ and $\mathcal{Z}$. Let $[Z_\perp]$ denote an $N \times (N-P)$ matrix with orthonormal columns spanning the orthogonal complement of $\mathcal{Z}$. Let $\sigma_1([Z_\perp]^*[X]), \sigma_2([Z_\perp]^*[X]), \ldots, \sigma_P([Z_\perp]^*[X])$ denote the first P singular values of $[Z_\perp]^*[X]$. Then

$$\sigma_i([Z_\perp]^*[X]) = \sin(\theta_{P+1-i})$$

for $i = 1, 2, \ldots, P$.

To apply Theorem 6 we can take $[X] = [\widehat{\Psi}_{\mathcal{I}}]$, $[Z] = [\Psi_{\mathcal{I}}]$, and

$$[Z_\perp] = [[\Psi_{\mathcal{I}^c}]\ [\Psi_\perp]]\,.$$

Note that by summing up the squared singular values of $[Z_\perp]^*[X]$, we obtain the Frobenius norm of $[Z_\perp]^*[X]$:

$$\|[Z_\perp]^*[X]\|_F^2 = \sum_{i=1}^{P} \sigma_i^2([Z_\perp]^*[X]) = \sum_{i=1}^{P} \sin^2(\theta_{P+1-i}) = \sum_{i=1}^{P} \sin^2(\theta_i).$$

We also note that, since the column spans of $[\Psi]$ and $[\widehat{\Psi}]$ are equal,

$$[Z_\perp]^*[X] = \begin{bmatrix} [\Psi_{\mathcal{I}^c}]^*[\widehat{\Psi}_{\mathcal{I}}] \\ [\Psi_\perp]^*[\widehat{\Psi}_{\mathcal{I}}] \end{bmatrix} = \begin{bmatrix} [\Psi_{\mathcal{I}^c}]^*[\widehat{\Psi}_{\mathcal{I}}] \\ 0 \end{bmatrix},$$

and so

$$\|[Z_\perp]^*[X]\|_F^2 = \|[\Psi_{\mathcal{I}^c}]^*[\widehat{\Psi}_{\mathcal{I}}]\|_F^2 = \sum_{n \in \mathcal{I}} \sum_{l \notin \mathcal{I}} |\langle\{\psi_l\}, \{\widehat{\psi}_n\}\rangle|^2.$$

Therefore, assuming (17) holds, we will have

$$\sum_{i=1}^{P} \sin^2(\theta_i) = \sum_{n \in \mathcal{I}} \sum_{l \notin \mathcal{I}} |\langle\{\psi_l\}, \{\widehat{\psi}_n\}\rangle|^2 \leq \sum_{n \in \mathcal{I}} \max_{l \notin \mathcal{I}} \frac{\upsilon_l \widehat{\upsilon}_n}{|\upsilon_l - \widehat{\upsilon}_n|^2} \frac{\eta^2}{1-\eta}. \tag{18}$$

Let us examine the term η which appears in (18) and is defined in (16). Note that

$$\begin{aligned}
\eta &= \| \left([\Gamma][\Gamma]^*\right)^{-\frac{1}{2}} [\Gamma][\Delta][\Gamma]^* \left([\Gamma][\Gamma]^*\right)^{-\frac{1}{2}} \|_2 \\
&= \max_{\{x\} \neq 0} \sqrt{\frac{\{x\}^* \left([\Gamma][\Gamma]^*\right)^{-\frac{1}{2}} [\Gamma][\Delta]^*[\Delta][\Gamma]^* \left([\Gamma][\Gamma]^*\right)^{-\frac{1}{2}} \{x\}}{\{x\}^*\{x\}}} \\
&= \max_{\{y\} \neq 0} \sqrt{\frac{\{y\}^*[\Delta]^*[\Delta]\{y\}}{\{y\}^*\{y\}}} \\
&= \|[\Delta]\|_2.
\end{aligned}$$

In the second to last line we let $\{y\} = [\Gamma]^* \left([\Gamma][\Gamma]^*\right)^{-\frac{1}{2}} \{x\}$ and note that $\{y\}^*\{y\} = \{x\}^*\{x\}$.

Our Theorem 2 relies on proving that $\|[\Delta]\|_2 \leq \epsilon$. For uniform samples collected as prescribed in the theorem statement, we prove that $\|[\Delta]\|_2 \leq \epsilon$ in [5], and so we do not repeat the argument here.

Once we have established that $\eta = \|[\Delta]\|_2 \leq \epsilon$, we can plug this inequality into (18) and obtain the bound

$$\sum_{i=1}^{P} \sin^2(\theta_i) \leq \sum_{n \in \mathcal{I}} \max_{l \notin \mathcal{I}} \frac{\upsilon_l \widehat{\upsilon}_n}{|\upsilon_l - \widehat{\upsilon}_n|^2} \frac{\epsilon^2}{1-\epsilon}. \tag{19}$$

We note that (19) depends both on the original eigenvalues of $[H]$ (appearing as $\upsilon_1, \ldots, \upsilon_K$) and on the eigenvalues of the perturbed matrix $[G]$ (appearing as $\widehat{\upsilon}_1, \ldots, \widehat{\upsilon}_K$). We can relate these two collections of eigenvalues as follows. First, note that the eigenvalues of $[G]$ are the same as those of $[V][V]^*$. Furthermore, the eigenvalues of $[H]$ are the same as those of $[\Gamma][\Gamma]^*$. We focus on relating the eigenvalues of $[G]$ to those of $[\Gamma][\Gamma]^*$. Recall that we can write $[G] = [\Gamma][\Gamma]^* + [\Gamma][\Delta][\Gamma]^*$. Thus, for any $\{x\} \neq 0$,

$$\begin{aligned} \frac{\{x\}^*[G]\{x\}}{\{x\}^*[\Gamma][\Gamma]^*\{x\}} &= \frac{\{x\}^*[\Gamma][\Gamma]^*\{x\} + \{x\}^*[\Gamma][\Delta][\Gamma]^*\{x\}}{\{x\}^*[\Gamma][\Gamma]^*\{x\}} \\ &= 1 + \frac{\{x\}^*[\Gamma][\Delta][\Gamma]^*\{x\}}{\{x\}^*[\Gamma][\Gamma]^*\{x\}}. \end{aligned}$$

Let us denote $\{y\} = [\Gamma]^*\{x\}$ and use this to write

$$\frac{\{x\}^*[G]\{x\}}{\{x\}^*[\Gamma][\Gamma]^*\{x\}} = 1 + \frac{\{y\}^*[\Delta]\{y\}}{\{y\}^*\{y\}}.$$

If we can show that $\|[\Delta]\|_2 \leq \epsilon$, this means $-\epsilon \leq \frac{\{y\}^*[\Delta]\{y\}}{\{y\}^*\{y\}} \leq \epsilon$, giving us

$$1 - \epsilon \leq \frac{\{x\}^*[G]\{x\}}{\{x\}^*[\Gamma][\Gamma]^*\{x\}} \leq 1 + \epsilon.$$

Applying Lemma 1 in [17] allows us to conclude that

$$1 - \epsilon \leq \frac{\widehat{\upsilon}_l}{\upsilon_l} \leq 1 + \epsilon \tag{20}$$

for $l = 1, \ldots, K$. We note that, in order to apply Theorem 5 above, it is necessary that $[H]$ and $[G]$ be positive definite. The requirement that $[H]$ be positive definite will be satisfied by our assumption that the amplitudes $A_1, \ldots, A_K$ are nonzero. The requirement that $[G]$ be positive definite will then be satisfied as a consequence of (20). Note that this latter requirement can only be satisfied if $M \geq K$.

Supposing that (20) is satisfied, let us examine the term $\frac{\upsilon_l \widehat{\upsilon}_n}{|\upsilon_l - \widehat{\upsilon}_n|^2}$ appearing in (19). An upper bound on this expression is given by

$$\frac{\upsilon_l \upsilon_n (1+\epsilon)}{\min_{c \in [-1,1]} |\upsilon_l - \upsilon_n(1+c\epsilon)|^2}.$$

Plugging this bound into (19) and using the fact that $\upsilon_l = M|A_l|^2$ for $l = 1, \ldots, K$, we obtain the bound appearing in (8).

REFERENCES

[1] S. W. Doebling, C. R. Farrar, M. B. Prime, and D. W. Shevitz, "Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review," Los Alamos National Lab., NM, Tech. Rep., 1996.

[2] H. Sohn, C. R. Farrar, F. Hemez, and J. Czarnecki, "A review of structural health monitoring literature 1996–2001," Los Alamos National Lab., NM, Tech. Rep., 2001.

[3] B. F. Feeny and R. Kappagantu, "On the physical interpretation of proper orthogonal modes in vibrations," *Journal of Sound and Vibration*, vol. 211, no. 4, pp. 607–616, 1998.

[4] G. Kerschen and J.-C. Golinval, "Physical interpretation of the proper orthogonal modes using the singular value decomposition," *Journal of Sound and Vibration*, vol. 249, no. 5, pp. 849–865, 2002.

[5] J. Y. Park, M. B. Wakin, and A. C. Gilbert, "Modal analysis with compressive measurements," *IEEE Trans. Signal Process.*, vol. 62, no. 7, April 1 2014.

[6] D. Mascarenas, D. Hush, J. Theiler, and C. Farrar, "The application of compressed sensing to detecting damage in structures," in *Proc. 8th Int. Workshop on Structural Health Monitoring*, 2011.

[7] B. Yuequan, J. L. Beck, and L. Hui, "Compressive sampling for accelerometer signals in structural health monitoring," *Structural Health Monitoring*, vol. 10, no. 3, pp. 235–246, 2011.

[8] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, Apr. 2006.

[9] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[10] M. Wakin, S. Becker, E. Nakamura, M. Grant, E. Sovero, D. Ching, J. Yoo, J. Romberg, A. Emami-Neyestanak, and E. Candés, "A non-uniform sampler for wideband spectrally-sparse environments," *IEEE J. Emerging Sel. Topics Circuits Syst.*, vol. 2, no. 3, pp. 516–529, Sep. 2012.

[11] B. Boashash, *Time frequency signal analysis and processing: A comprehensive reference*. Elsevier, 2003.

[12] F. W. King, *Hilbert Transforms*. Cambridge University Press, 2009.

[13] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Found. Comput. Math.*, vol. 12, no. 4, pp. 389–434, 2012.
[14] A. Knyazev and P. Zhu, "Principal angles between subspaces and their tangents," *Arxiv preprint arXiv:1209.0523*, 2012.
[15] R. Mathias and K. Veselić, "A relative perturbation bound for positive definite matrices," *Linear Algebra and its Applications*, vol. 270, pp. 315–321, 1998.
[16] J. Sun, "Perturbation of angles between linear subspaces," *J. Comput. Math.*, vol. 5, pp. 58–61, 1987.
[17] J. Barlow and J. Demmel, "Computing accurate eigensystems of scaled diagonally dominant matrices," Tech. Rep., 1980.