Atomic Norm Minimization for Modal Analysis
from Random and Compressed Samples

Shuang Li, Dehui Yang, Gongguo Tang, and Michael B. Wakin*

December 25, 2017

Abstract

Modal analysis is the process of estimating a system's modal parameters such as its natural frequencies and mode shapes. One application of modal analysis is in structural health monitoring (SHM), where a network of sensors may be used to collect vibration data from a physical structure such as a building or bridge. There is a growing interest in developing automated techniques for SHM based on data collected in a wireless sensor network. In order to conserve power and extend battery life, however, it is desirable to minimize the amount of data that must be collected and transmitted in such a sensor network. In this paper, we highlight the fact that modal analysis can be formulated as an atomic norm minimization (ANM) problem, which can be solved efficiently and in some cases recover perfectly a structure's mode shapes and frequencies. We survey a broad class of sampling and compression strategies that one might consider in a physical sensor network, and we provide bounds on the sample complexity of these compressive schemes in order to recover a structure's mode shapes and frequencies via ANM. A main contribution of our paper is to establish a bound on the sample complexity of modal analysis with random temporal compression, and in this scenario we prove that the required number of samples per sensor can actually decrease as the number of sensors increases. We also extend an atomic norm denoising problem to the multiple measurement vector (MMV) setting in the case of uniform sampling.

1 Introduction

Modal analysis is the process of estimating a system's modal parameters such as its natural frequencies, mode shapes, and damping factors. One application of modal analysis is in structural health monitoring (SHM), where a network of sensors may be used to collect vibration data from a physical structure such as a building or bridge. The vibration characteristics of a structure are captured in its modal parameters, which can be estimated from the recorded displacement data. Changes in these parameters over time may be indicative of damage to the structure. Modal analysis has been widely used in civil structures [1], space structures [2], acoustical instruments [3], and so on.

Due to the considerable time and expense required to perform manual inspections of physical structures, and the difficulty of repeating these inspections frequently, there is a growing interest in developing automated techniques for SHM based on data collected in a wireless sensor network. For example, one could envision a collection of battery-operated wireless sensors deployed across a structure that record vibrational displacements over time and then transmit this information to a central node for analysis. In order to conserve power and extend battery life, however, it is desirable to minimize the amount of data that must be collected and transmitted in such a sensor network [4].

In this paper, we highlight the fact that modal analysis can be formulated as an atomic norm minimization (ANM) problem, which can be solved efficiently and in some cases recover perfectly a structure's mode shapes and frequencies. We survey several possible protocols for data collection, compression, and transmission, and we review the sampling requirements in each case. ANM generalizes the widely used ℓ_1-minimization

---

*Department of Electrical Engineering, Colorado School of Mines. Email: {shuangli,dyang,gtang,mwakin}@mines.edu

framework for finding sparse solutions to underdetermined linear inverse problems [5]. It has recently been shown to be an efficient and powerful way for exactly recovering unobserved time-domain samples and identifying unknown frequencies in signals having sparse frequency spectra [6, 7, 8], in particular when the unknown frequencies are continuous-valued and do not belong to a discrete grid. Sampling guarantees have been established both in the single measurement vector (SMV) scenario [6, 9] and in the multiple measurement vector (MMV) scenario under a joint sparse model [10, 7]; these results characterize the number of uniform or random time-domain samples to achieve exact frequency localization as a function of the minimum separation between the unknown frequencies. In this sense, by achieving exact recovery, ANM can completely avoid the effects of basis mismatch [11, 12, 13] which can plague conventional grid-based compressive sensing techniques.

To provide context in this paper, we survey a broad class of sampling and compression strategies that one might consider in a physical sensor network, and we provide bounds on the sample complexity of these compressive schemes in order to recover a structure's mode shapes and frequencies via ANM. In total, we consider five measurement schemes:

- **uniform sampling**, where vibration signals at each sensor are sampled synchronously at or above the Nyquist rate and transmitted without compression to a central node,

- **synchronous random sampling**, where vibration signals at each sensor are sampled randomly in time (as a subset of the Nyquist grid), but at the same time instants at each sensor,

- **asynchronous random sampling**, where vibration signals at each sensor are sampled randomly in time (as a subset of the Nyquist grid), at different time instants at each sensor,

- **random temporal compression**, where Nyquist-rate samples are compressed at each sensor via random matrix multiplication, before transmission to a central node, and

- **random spatial compression**, where Nyquist-rate samples are compressed en route to the central node.

Note that all sensors share the same Nyquist grid in each of these measurement schemes.

Modal analysis is a particular instance of the joint sparse frequency estimation problem, which has also been commonly studied in the context of direction-of-arrival (DOA) estimation [14]. Some conventional methods for joint sparse frequency estimation, such as MUSIC [15], can identify frequencies using a sample covariance matrix as long as a sufficient number of snapshots is given. However, as noted in [10], these methods usually assume that the source signals are spatially uncorrelated and their performance would decrease with source correlations. ANM does not have this limitation. Moreover, joint sparse frequency estimation techniques such as MUSIC do not naturally accommodate the sort of randomized sampling and compression protocols that we consider in this paper.

In this work, we explain how the results from [7, 10] can be interpreted in the context of exactly recovering a structure's mode shapes and frequencies from uniform samples, synchronous random samples, and asynchronous random samples. These random sampling results have an unfortunate scaling, however, in that the number of samples per sensor actually increases as the number of sensors increases; intuition and simulations suggest that the opposite should be true. A main contribution of our paper, then, is to establish a bound on the sample complexity of modal analysis with random temporal compression, and in this scenario we prove that the required number of samples per sensor can actually *decrease* as the number of sensors increases. A similar phenomenon—that the estimation accuracy increases as the number of snapshots increases—occurs in covariance fitting methods for DOA estimation [14]. We also explain how our previous work [16] on super-resolution of complex exponentials from modulations with unknown waveforms can be used to establish a bound on the sample complexity of modal analysis with random spatial compression. For the noisy case, we extend the SMV atomic norm denoising problem in [17, 18] to an MMV atomic norm denoising problem in the case of uniform sampling. We derive theoretical guarantees for this MMV atomic norm denoising problem. Although we focus on modal analysis to put our analysis and simulations in a specific context, our results can apply to joint sparse frequency estimation more generally.
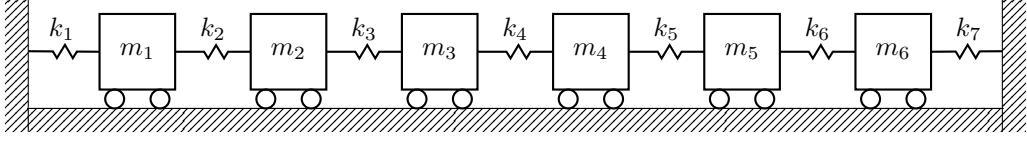
Figure 1: *Undamped 6-degree-of-freedom boxcar system with masses $m_1 = 1$, $m_2 = 2$, $m_3 = 3$, $m_4 = 4$, $m_5 = 5$, $m_6 = 6$ kg and stiffness values $k_1 = k_7 = 500$, $k_2 = k_6 = 150$, $k_3 = k_5 = 100$, $k_4 = 50$ N/m. See Section 4.1 for simulations related to this system.*

The remainder of this paper is organized as follows. In Section 2, we provide some background on modal analysis and on the atomic norm. In Section 3, we formulate our problem, and we characterize the ability of ANM to exactly recover a structure's mode shapes and frequencies under each of the above five measurement schemes. We also provide a bound on the performance of ANM in the case of noisy, uniform samples. In Section 4, we present simulation results to illustrate some of the essential trends in the theoretical results and to demonstrate the favorable performance of ANM. We prove our main theorems in Section 5, and we make some concluding remarks in Section 6. The Appendix provides supplementary theoretical results.

2 Background

2.1 Modal analysis

For an N degree-of-freedom linear time-invariant system [19], the second-order equations of motion which represent the dynamic behavior of the system can be formulated as

$$\mathbf{M}\ddot{\boldsymbol{x}}(t) + \mathbf{C}\dot{\boldsymbol{x}}(t) + \mathbf{K}\boldsymbol{x}(t) = \boldsymbol{f}(t) \tag{2.1}$$

where $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ denote the $N \times N$ mass, damping, and stiffness matrices, respectively. In this equation, $\boldsymbol{x}(t) \in \mathbb{R}^N$ represents a length-N vector of displacement values at time t, and $\boldsymbol{f}(t) \in \mathbb{R}^N$ represents the excitation force at the N nodes. We will assume that each of the N displacement values is associated with a wireless sensor node that can sample, record, and transmit that displacement value; as an example, one could equip each of the six boxcars shown in Fig. 1 with a displacement sensor.

As a simplification and as in [20], we consider systems in free vibration, where the forcing input $\boldsymbol{f}(t) = \mathbf{0}$ and the system vibrates freely in response to some set of initial conditions.[1] In this case, the general solution to (2.1) takes the form [25]

$$\boldsymbol{x}(t) = \sum_{k=1}^{K} \boldsymbol{\psi}_k q_k(t),$$

where $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_K \in \mathbb{R}^N$ are a collection of mode shape vectors whose span characterizes the set of possible displacement profiles of the structure, and $q_1(t), q_2(t), \ldots, q_K(t)$ are modal responses in the form of monotone exponentially-decaying sinusoids:

$$q_k(t) = u_k e^{-\zeta_k 2\pi f_k t} \cos(2\pi f_{d_k} t + \theta_k).$$

Here, u_k and θ_k are determined by the initial conditions, and the number of nonzero amplitudes $K \leq N$ corresponds to the number of active modes. The parameters ζ_k, f_k, and $f_{d_k} = \sqrt{1 - \zeta_k^2} f_k$ denote the kth damping ratio, the natural frequency, and the damped frequency, respectively, and these parameters along with the mode shapes are intrinsic properties of the system determined by the mass, damping, and stiffness

---

[1]One can also consider modal analysis with forced vibration [21, 22, 23, 24], when external forces such as an earthquake, wind, or vehicle loadings are applied.

matrices. Modal analysis refers to the identification of these parameters—particularly the mode shapes, damping ratios, and frequencies—from observations of the displacement vector $\boldsymbol{x}(t)$.

In recent years, blind source separation (BSS) based methods have become very popular in modal analysis. The authors in [26] and [27] propose a new modal identification algorithm based on sparse component analysis (SCA) to deal with even the underdetermined case where sensors may be highly limited compared to the number of active modes. In [28], a novel decentralized modal identification method termed PARAllel FACtor (PARAFAC) based sparse BSS (PSBSS) method is proposed. Independent component analysis (ICA) is a powerful method to solve the BSS problem. Yang et al. present an ICA based method to identify the modal parameters of lightly and highly damped systems, even in cases with heavy noise and nonstationarity [29]. Despite the favorable empirical performance of these methods, few of them have been supported by theoretical analysis. In this paper, inspired by promising recent work in line spectrum estimation, we provide theoretical support for ANM as a powerful technique for modal analysis.

Meanwhile, over the past decade, the development of compressive sensing (CS) has highlighted the possibility of capturing essential signal information with sampling rates much lower than the Nyquist rate [30, 31, 32]. Exploring the possibility of using CS in modal analysis, Park et al. provide a theoretical analysis of a singular value decomposition (SVD) based technique for estimating a structure's mode shapes in free vibration without damping [20]. The work in [20] builds upon a previous observation [33] that with a sufficiently large number of samples, the SVD technique can identify the mode shapes of systems with uniformly distributed mass (which leads to mutually orthogonal mode shapes) and light or no damping; as such, the results in [20] are limited to the assumption that the mode shapes are mutually orthogonal, which is not satisfied for general systems. In a more recent work, Yang et al. propose a BSS based method that can identify non-orthogonal mode shapes from video measurements [34]. The same group also develops another BSS based method to identify modal parameters from uniformly sub-Nyquist (temporally-aliased) video measurements [35]. Another recent work [36] also presents a new method based on a combination of CS and complexity pursuit (CP) to solve the modal identification problem.

In a very recent paper [9], Heckel and Soltanolkotabi consider solving a generalized line spectrum estimation problem with convex optimization. In particular, they recover a data vector from its compressed measurements using an ANM formulation that has been mentioned in [5]. Several different classes of random sensing matrices are considered for obtaining the measurements. The result for Gaussian random matrices in [9] can be viewed as a special case of our modal analysis result for random temporal compression (Theorem 3.4), where we have multiple measurement vectors and use a different sensing matrix for each measurement vector. Our analysis is inspired by their proof.

Finally, Lu et al. [37] propose a concatenated ANM approach for joint recovery of frequency sparse signals under a certain joint sparsity model. This problem differs from ours in the signal model, and the paper [37] does not establish theoretical bounds on the sample complexity.

2.2 The atomic norm

Frequency estimation from a mixture of complex sinusoids is a classical problem in signal processing. As mentioned in the introduction, ANM has recently been considered as a technique for solving this problem, both in the SMV ($N = 1$) and MMV ($N > 1$) scenarios. Under the right conditions, ANM can achieve exact frequency localization, avoiding the effects of basis mismatch which can plague grid-based techniques. In Section 3, we survey various formulations of the ANM problem that can be used with the various compressive measurement protocols for modal analysis. All of these formulations rely on the same core definition of an atomic norm, which is established in [7, 10] and repeated here.

Suppose each column of an $M \times N$ data matrix $\mathbf{X} = [\boldsymbol{x}_1, \ \boldsymbol{x}_2, \ \cdots, \ \boldsymbol{x}_N]$ is a spectrally sparse signal with K distinct frequency components and denoted as

$$\boldsymbol{x}_i = \sum_{k=1}^{K} c_{k,i} \boldsymbol{a}(f_k).$$

Here, the vector

$$\boldsymbol{a}(f) = [e^{j2\pi f 0}, e^{j2\pi f 1}, \cdots, e^{j2\pi f(M-1)}]^\top \tag{2.2}$$

corresponds to a collection of M samples of a complex exponential signal with frequency $f \in [0,1)$. Note that we use "$\top$" and "$*$" to denote transpose and conjugate transpose, respectively.

As shown in [7] and [10], one can define an atomic set to represent such a data matrix $\mathbf{X}$ with each *atom* defined as

$$\mathbf{A}(f, \boldsymbol{b}) = \boldsymbol{a}(f)\boldsymbol{b}^*, \tag{2.3}$$

where $f \in [0,1)$, $\boldsymbol{b} \in \mathbb{C}^N$ with $\|\boldsymbol{b}\|_2 = 1$. The corresponding *atomic set* can be defined as

$$\mathcal{A} = \{\mathbf{A}(f, \boldsymbol{b}) : \ f \in [0,1), \ \|\boldsymbol{b}\|_2 = 1\}.$$

The *atomic norm* of $\mathbf{X}$ is then defined as

$$\|\mathbf{X}\|_{\mathcal{A}} = \inf \{t > 0 : \mathbf{X} \in t \ \mathrm{conv}(\mathcal{A})\}$$

$$= \inf \left\{\sum_k c_k : \mathbf{X} = \sum_k c_k \mathbf{A}(f_k, \boldsymbol{b}_k), \ c_k \geq 0 \right\},$$

where $\mathrm{conv}(\mathcal{A})$ is the convex hull of $\mathcal{A}$. This atomic norm is equivalent to the solution of the following semidefinite program (SDP):

$$\|\mathbf{X}\|_{\mathcal{A}} = \inf_{\boldsymbol{u} \in \mathbb{C}^M, \ \boldsymbol{V} \in \mathbb{C}^{N \times N}} \left\{\frac{1}{2M} \mathrm{Tr}(\mathcal{T}(\boldsymbol{u})) + \frac{1}{2}\mathrm{Tr}(\mathbf{V}) : \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{X} \\ \mathbf{X}^* & \mathbf{V} \end{bmatrix} \succeq 0 \right\},$$

where $\mathcal{T}(\boldsymbol{u})$ is the Hermitian Toeplitz matrix with the vector $\boldsymbol{u}$ as its first column. $\mathrm{Tr}(\cdot)$ denotes the trace of a square matrix. The proof of this SDP form can be found in [6, 7].

The *dual norm* of $\|\cdot\|_{\mathcal{A}}$ is defined as

$$\begin{aligned}
\|\mathbf{Q}\|_{\mathcal{A}}^* &= \sup_{\|\mathbf{X}\|_{\mathcal{A}} \leq 1} \langle \mathbf{Q}, \mathbf{X} \rangle_{\mathbb{R}} \\
&= \sup_{f \in [0,1), \|\boldsymbol{b}\|_2 = 1} \langle \mathbf{Q}, \boldsymbol{a}(f)\boldsymbol{b}^* \rangle_{\mathbb{R}} \\
&= \sup_{f \in [0,1), \|\boldsymbol{b}\|_2 = 1} \langle \boldsymbol{b}, \mathbf{Q}^* \boldsymbol{a}(f) \rangle_{\mathbb{R}} \\
&= \sup_{f \in [0,1)} \|\mathbf{Q}^* \boldsymbol{a}(f)\|_2 \\
&= \sup_{f \in [0,1)} \|\mathcal{Q}(f)\|_2,
\end{aligned} \tag{2.4}$$

where $\mathcal{Q}(f) = \mathbf{Q}^* \boldsymbol{a}(f)$ is known as the dual polynomial. Above, $\langle \mathbf{Q}, \mathbf{X} \rangle_{\mathbb{R}} = \mathrm{Re}(\langle \mathbf{Q}, \mathbf{X} \rangle)$ corresponds to the real inner product between two matrices.

3 Main Result

3.1 Preliminaries

As in [20], we assume the structure has no damping ($\mathbf{C} = \mathbf{0}$) and that the real-valued displacement[2] signal has been converted to its complex analytic form, which can be accomplished by using the Hilbert transform

---

[2]Although we refer to displacement data in this work, the ANM method can also be applied to acceleration data. In particular, the ground truth acceleration signal vector will have the same form as (3.1) but with a different set of amplitudes A_k.

in practice [38, 39]. In this case, the ground truth displacement signal vector can be written as

$$\boldsymbol{x}^\star(t) = \sum_{k=1}^{K} A_k \boldsymbol{\psi}_k^\star e^{j2\pi F_k t}, \tag{3.1}$$

which is a superposition of K complex sinusoids. Here, A_k, F_k, and $\boldsymbol{\psi}_k^\star = [\psi_{1,k}^\star, \psi_{2,k}^\star, \cdots, \psi_{N,k}^\star]^\top \in \mathbb{C}^N$ are the complex amplitudes, frequencies, and mode shapes, respectively. We assume without loss of generality that the mode shapes are normalized[3] such that $\|\boldsymbol{\psi}_k^\star\|_2 = 1$. As in Section 2.1, N denotes the number of sensors. Thus, $\sum_{k=1}^{K} A_k \psi_{n,k}^\star e^{j2\pi F_k t}$ is the displacement signal at the nth sensor.

Because we assume that $\boldsymbol{x}^\star(t)$ is an analytic signal, all $F_k \geq 0$.[4] Define $F_c = \max_{1 \leq k \leq K} F_k$. If one were to take regularly spaced Nyquist samples of $\boldsymbol{x}^\star(t)$ at times

$$T = \{t_1, t_2, \ldots, t_M\} = \{0, T_s, \ldots, (M-1)T_s\}, \tag{3.2}$$

where $T_s < \frac{1}{F_c}$, and then stack the signal from the nth sensor as the nth column of a data matrix $\mathbf{X}^\star$, one would have

$$
\begin{aligned}
\mathbf{X}^\star &= \sum_{k=1}^{K} A_k \begin{bmatrix} \psi_{1,k}^\star e^{j2\pi F_k t_1} & \psi_{2,k}^\star e^{j2\pi F_k t_1} & \cdots & \psi_{N,k}^\star e^{j2\pi F_k t_1} \\ \psi_{1,k}^\star e^{j2\pi F_k t_2} & \psi_{2,k}^\star e^{j2\pi F_k t_2} & \cdots & \psi_{N,k}^\star e^{j2\pi F_k t_2} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{1,k}^\star e^{j2\pi F_k t_M} & \psi_{2,k}^\star e^{j2\pi F_k t_M} & \cdots & \psi_{N,k}^\star e^{j2\pi F_k t_M} \end{bmatrix} \\
&= \sum_{k=1}^{K} |A_k| \begin{bmatrix} \psi_{1,k} e^{j2\pi f_k 0} & \psi_{2,k} e^{j2\pi f_k 0} & \cdots & \psi_{N,k} e^{j2\pi f_k 0} \\ \psi_{1,k} e^{j2\pi f_k 1} & \psi_{2,k} e^{j2\pi f_k 1} & \cdots & \psi_{N,k} e^{j2\pi f_k 1} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{1,k} e^{j2\pi f_k (M-1)} & \psi_{2,k} e^{j2\pi f_k (M-1)} & \cdots & \psi_{N,k} e^{j2\pi f_k (M-1)} \end{bmatrix} \\
&= \sum_{k=1}^{K} |A_k| \boldsymbol{a}(f_k) \boldsymbol{\psi}_k^\top = \sum_{k=1}^{K} |A_k| \mathbf{A}(f_k, \boldsymbol{b}_k),
\end{aligned} \tag{3.3}
$$

where $\psi_{n,k} = \psi_{n,k}^\star e^{j\phi_{A_k}}$ with ϕ_{A_k} being the phase of the complex amplitude A_k, $f_k = F_k T_s \in [0, 1)$ are the discrete frequencies, $\boldsymbol{a}(f_k)$ are sampled complex exponentials as defined in (2.2), and $\mathbf{A}(f_k, \boldsymbol{b}_k)$ are the atoms defined in (2.3). The ability to express $\mathbf{X}^\star$ as a linear combination of K atoms from the atomic set $\mathcal{A}$ inspires the use of ANM to recover $\mathbf{X}^\star$ from partial information, as we discuss below.

3.2 Modal analysis for noiseless signals

In this section, we survey a broad class of sampling and compression strategies that one might consider in a physical sensor network, and we provide bounds on the sample complexity of these compressive schemes in order to recover a structure's mode shapes and frequencies via ANM.

The performance of ANM in these various scenarios will depend on the minimum separation of the discrete frequencies f_k's, which is defined to be

$$\Delta_f = \min_{k \neq j} |f_k - f_j|,$$

---

[3]Equation (3.1) will hold with any choice of normalization for the mode shapes, as any rescaling of $\boldsymbol{\psi}_k^\star$ can be simply absorbed into A_k. Any such renormalization can be applied after the Euclidean-normalized mode shapes are recovered and thus does not affect our results. A certain "mass normalization" is desired in some applications. Employing mass normalization requires either knowledge of the mass matrix or extra experiments [40].

[4]It would not be difficult to extend our analysis to the case where the frequencies F_k in (3.1) can be negative or positive: one would set $F_c = \max_{1 \leq k \leq K} |F_k|$ and $T_s < \frac{1}{2F_c}$, which yields discrete frequencies $f_k \in [-1/2, 1/2)$, which is equivalent to the interval $[0, 1)$.

where $|f_k - f_j|$ is understood as the wrap-around distance on the unit circle.

We also note that, because each complex mode shape vector $\boldsymbol{\psi}_k^\star$ is multiplied by a complex amplitude A_k in our model (3.1), recovery of these mode shapes and amplitudes is possible only up to a phase ambiguity. We denote our estimated mode shapes as $\hat{\boldsymbol{\psi}}_k$ and measure recovery performance using the absolute inner product $|\langle \boldsymbol{\psi}_k^\star, \hat{\boldsymbol{\psi}}_k \rangle|$. Because $\boldsymbol{\psi}_k^\star$ and $\hat{\boldsymbol{\psi}}_k$ are both normalized, achieving $|\langle \boldsymbol{\psi}_k^\star, \hat{\boldsymbol{\psi}}_k \rangle| = 1$ corresponds to exact recovery of the mode shape up to the unknown phase term.

3.2.1 Uniform sampling

As a baseline, we begin by considering a conventional data collection scheme, in which samples from each sensor are collected uniformly in time with sampling rate $F_s = \frac{1}{T_s} > F_c$. In this case, the data matrix $\mathbf{X}^\star$, defined in (3.3), is fully observed. In this section, we assume the samples are collected without noise; Section 3.3 revisits the uniform sampling scheme in the case of noisy samples.

To identify the mode shapes and frequencies in this scenario, it can be useful to consider the following ANM formulation:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{X}\|_{\mathcal{A}} \text{ s.t. } \mathbf{X} = \mathbf{X}^\star. \tag{3.4}$$

Although this problem has a trivial solution (namely $\hat{\mathbf{X}} = \mathbf{X}^\star$, which is already available), solving this problem in a certain way can reveal information about the mode shapes and frequencies. In particular, (3.4) is equivalent to the following SDP

$$\hat{\mathbf{X}}, \hat{\boldsymbol{u}}, \hat{\mathbf{V}} = \arg \min_{\mathbf{X}, \boldsymbol{u}, \mathbf{V}} \ \frac{1}{2M} \text{Tr}(\mathcal{T}(\boldsymbol{u})) + \frac{1}{2} \text{Tr}(\mathbf{V})$$

$$\text{s.t.} \quad \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{X} \\ \mathbf{X}^* & \mathbf{V} \end{bmatrix} \succeq 0, \ \ \mathbf{X} = \mathbf{X}^\star.$$

Certain approaches to solving this SDP[5] will return the dual solution $\mathbf{Q}$ directly. From this, following Proposition 1 in [7], one can formulate the dual polynomial $\mathcal{Q}(f) = \mathbf{Q}^* \boldsymbol{a}(f)$ to identify the frequencies and mode shapes. In particular, one can identify the frequencies by localizing where the dual polynomial achieves $\|\mathcal{Q}(f)\|_2 = 1$. Note that the data matrix $\mathbf{X}^\star$ can also be rewritten as

$$\mathbf{X}^\star = [\boldsymbol{a}(f_1) \ \cdots \ \boldsymbol{a}(f_K)][|A_1|\boldsymbol{\psi}_1 \ \cdots \ |A_K|\boldsymbol{\psi}_K]^\top.$$

Then, using the estimated data matrix $\hat{\mathbf{X}}$ and estimated frequencies $\hat{f}_k$, one can solve a least-squares problem to estimate the complex amplitudes and mode shapes:

$$\hat{\boldsymbol{\Psi}}_A \triangleq [|\hat{A}_1|\hat{\boldsymbol{\psi}}_1 \ \cdots \ |\hat{A}_K|\hat{\boldsymbol{\psi}}_K] = (\mathbf{A}_{\hat{f}}^\dagger \hat{\mathbf{X}})^\top,$$

where $\mathbf{A}_{\hat{f}}^\dagger$ denotes the pseudoinverse of the matrix $\mathbf{A}_{\hat{f}} = [\boldsymbol{a}(\hat{f}_1) \ \cdots \ \boldsymbol{a}(\hat{f}_K)]$. Because the true mode shapes are assumed to be normalized, we normalize the estimated mode shapes by setting $|\hat{A}_k| = \|\hat{\boldsymbol{\Psi}}_A(:,k)\|_2$ and $\hat{\boldsymbol{\psi}}_k = \frac{\hat{\boldsymbol{\Psi}}_A(:,k)}{\|\hat{\boldsymbol{\Psi}}_A(:,k)\|_2}$.

The following theorem is adapted from Theorem 4 in [10].

Theorem 3.1. *[10] Assume the data matrix $\mathbf{X}^\star$ shown in (3.3) is obtained by uniformly sampling the displacement signal vector $\boldsymbol{x}^\star(t)$ in time with a sampling interval $T_s < \frac{1}{F_c} = \frac{1}{\max_k F_k}$. If the number of samples M from each sensor satisfies*

$$M \geq \max \left\{ \frac{4}{\Delta_f} + 1, \ 257 \right\}, \tag{3.5}$$

then ANM perfectly recovers all frequencies f_k and all mode shapes up to a phase ambiguity, i.e., $|\langle \boldsymbol{\psi}_k^\star, \hat{\boldsymbol{\psi}}_k \rangle| = 1$.

---

[5]In practice, one can use the CVX software package [41].

This theorem indicates that more measurements are needed at each sensor for perfect recovery as the minimum separation Δ_f decreases, i.e., as the true frequencies become closer to each other.[6] As noted in Section 2.1, Park et al. [20] have provided a theoretical analysis of an SVD based technique for modal analysis. That work also considers uniform sampling and provides a bound on M that is similar to what is required in (3.5). However, the analysis in [20] is limited to the case where the true mode shapes are mutually orthogonal, a condition that we do not require here. Moreover, the SVD guarantees apply only to approximate recovery of the mode shapes; as confirmed in Theorem 3.1, ANM can offer exact recovery of both the mode shapes and frequencies.

3.2.2 Synchronous random sampling

As a first alternative to conventional uniform sampling, we now consider the case where vibration signals at each sensor are sampled randomly in time, but at the same time instants at each sensor. We refer to this data collection scheme as *synchronous random sampling*.

To be specific, we suppose that the random sample times are chosen as a subset of the Nyquist grid; thus, the random samples are merely a subset of a collection of uniform samples. Equivalently, we suppose that we observe $\mathbf{X}^\star$ (which is defined in (3.3)) on a set of indices $\Omega_S \times [N]$ with $[N] \triangleq \{1, 2, \ldots, N\}$ and $\Omega_S \subset T$, where T is defined in (3.2). Thus, each column of $\mathbf{X}^\star$ is observed at the same times, indexed by Ω_S.

Denoting the observed matrix as $\mathbf{X}^\star_{\Omega_S \times [N]}$, the ANM problem can be formulated as

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X}\|_{\mathcal{A}} \ \text{s.t.} \ \mathbf{X}_{\Omega_S \times [N]} = \mathbf{X}^\star_{\Omega_S \times [N]}, \tag{3.6}$$

which is equivalent to the following SDP

$$\hat{\mathbf{X}}, \hat{\boldsymbol{u}}, \hat{\mathbf{V}} = \arg\min_{\mathbf{X}, \boldsymbol{u}, \mathbf{V}} \ \frac{1}{2M} \text{Tr}(\mathcal{T}(\boldsymbol{u})) + \frac{1}{2} \text{Tr}(\mathbf{V})$$

$$\text{s.t.} \ \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{X} \\ \mathbf{X}^* & \mathbf{V} \end{bmatrix} \succeq 0, \ \mathbf{X}_{\Omega_S \times [N]} = \mathbf{X}^\star_{\Omega_S \times [N]}.$$

As in Section 3.2.1, one can solve the above SDP, obtain the estimated frequencies from the dual polynomial, and recover the mode shapes and amplitudes by solving a least-squares problem.

The following theorem from [10] shows that we can recover $\mathbf{X}^\star$ and estimate the frequencies accurately with high probability.

Theorem 3.2. *[10] Suppose the data matrix $\mathbf{X}^\star$ is observed on the index set $\Omega_S \times [N]$, with Ω_S selected uniformly at random as a subset of T. Assume that $\{\boldsymbol{\psi}^\star_k\}_{k=1}^K$ are independent random vectors with $\mathbb{E}\boldsymbol{\psi}^\star_k = \mathbf{0}$, chosen independently of the fixed amplitudes A_k. That is, each $\boldsymbol{\psi}^\star_k$ is sampled independently from a zero-mean distribution on the complex hyper-sphere; this distribution may vary with k. If $M \geq \frac{4}{\Delta_f} + 1$, then there exists a numerical constant C such that*

$$|\Omega_S| \geq C \max \left\{ \log^2 \frac{\sqrt{N}M}{\delta}, K \log \frac{K}{\delta} \log \frac{\sqrt{N}M}{\delta} \right\} \tag{3.7}$$

is sufficient to guarantee that we can exactly recover $\mathbf{X}^\star$ via (3.6) and perfectly recover the frequencies and mode shapes up to a phase ambiguity with probability at least $1 - \delta$.

The above theorem indicates that the number of measurements $|\Omega_S|$ needed from each sensor for perfect recovery scales almost linearly with the number of active modes K. The number of random samples per sensor $|\Omega_S|$ also increases logarithmically with M, the number of underlying uniform samples in T. The lower limit on M is similar to what appears in Theorem 3.1 for the uniform sampling case. Thus, the total

---

[6]The resolution limit $\frac{4}{M}$ is commonly encountered in atomic norm minimization. A recent work [42] proposes a reweighted atomic norm minimization algorithm that can improve upon this resolution limit.

duration over which the signals must be observed does not change. However, what is significant is that in many cases the lower bound on $|\Omega_S|$ will be smaller than M, which confirms that in general during this time span it is not necessary to fully sample the uniform data matrix $\mathbf{X}^\star$; sensing and communication costs can be reduced by randomly subsampling this matrix.

Because its derivation relies on concentration arguments such as Hoeffding's inequality, Theorem 3.2 does assume the mode shapes to be generated randomly, which is not physically plausible. Moreover, this result actually requires the number $|\Omega_S|$ of measurements per sensor to *increase* (albeit logarithmically) as the number of sensors N increases. However, intuition suggests that the opposite should be true: the N signals share a common structure (analogous to a joint sparse model [43] in distributed compressive sensing), and observing more signals gives more information about this common structure.

It is reasonable to expect that one will actually need *fewer* measurements per sensor as the number of sensors increases. This is supported by simulations (not shown in this paper), and it remains an open question to support this with theory for the case of synchronous random sampling.[7]

3.2.3   Asynchronous random sampling

We now consider the case where vibration signals at each sensor are sampled randomly in time, but at different time instants at each sensor. We refer to this data collection scheme as *asynchronous random sampling*. To be specific, suppose that we observe $\mathbf{X}^\star$ on the indices $\Omega_A \subset T \times [N]$. This observation model allows each column of $\mathbf{X}^\star$ to be observed at different times, all of which are still restricted to be drawn from the uniform sampling grid T, defined in (3.2).

Recovery of the full data matrix from asynchronous random samples is analogous to the conventional matrix completion problem. This problem can be solved using ANM with the same formulation as in (3.6) but with $\Omega_S \times [N]$ replaced by Ω_A. The following theorem from [7] shows that one can again recover $\mathbf{X}^\star$ and estimate the frequencies and mode shapes accurately with high probability.[8]

Theorem 3.3. *[7] Suppose the data matrix $\mathbf{X}^\star$ is observed on the index set $\Omega_A \subset T \times [N]$, which is selected uniformly at random. Assume the signs $\frac{\psi^\star_{n,k}}{|\psi^\star_{n,k}|}$ are drawn independently from the uniform distribution on the complex unit circle (and independently of the fixed amplitudes A_k) and that $M \geq \frac{4}{\Delta_f} + 1$. Then there exists a numerical constant C such that*

$$|\Omega_A| \geq CN \max\left\{ \log^2 \frac{MN}{\delta}, K \log \frac{KN}{\delta} \log \frac{MN}{\delta} \right\} \tag{3.8}$$

is sufficient to guarantee that we can exactly recover $\mathbf{X}^\star$ via (3.6) and exactly recover the frequencies and mode shapes up to a phase ambiguity with probability at least $1 - \delta$.

Up to small differences in the logarithmic factors, the total number $|\Omega_A|$ of samples required in Theorem 3.3 is comparable to the number of sensors N times the number of samples per sensor $|\Omega_S|$ required in Theorem 3.2. Many of the same comments on the theorem apply here, including the use of a (different) randomness assumption on the mode shapes, and the fact that the required average number of measurements per sensor $|\Omega_A|/N$ again increases logarithmically as the number of sensors N increases. Again, the lower limit on M is similar to what happens in Theorem 3.1, and so the total duration over which the signals must be observed does not change. However, during this time span the requisite number of asynchronous random samples may be far lower than the number of uniform samples demanded by the Nyquist rate.

Comparing the logarithmic terms in the right hand sides of (3.7) and (3.8), we do see that the measurement requirement is slightly stronger for asynchronous random sampling than for synchronous random sampling. However, simulations (e.g., Fig. 3(b)) indicate that the opposite may be true. It remains an open question to support this with theory.

---

[7]After the initial submission of this manuscript, [44] has appeared as a complement to [10] and improves upon the sample complexity in Theorem 3.2 under a different random assumption on the mode shapes.

[8]An explicit proof of Theorem 3.3, which concerns MMV ANM, is not included in [7]. We do note that essentially the same conclusion holds if one applies SMV ANM and uses a union bound over the N sensors.

3.2.4 Random temporal compression

Inspired by alternatives to random sampling that have been considered in CS, we now consider the case where the displacement signal at each sensor is compressed via random matrix multiplication before being transmitted to a central node. We refer to this as *random temporal compression*. This compression strategy was also considered in [20], which provided a theoretical analysis of an SVD based technique for modal analysis.

Let $\boldsymbol{x}_n^\star$ be the nth column of the data matrix $\mathbf{X}^\star$, and let $\boldsymbol{\Phi}_n \in \mathbb{R}^{M' \times M}$ be a compressive matrix generated randomly with independent and identically distributed (i.i.d.) Gaussian entries. At each sensor, assume that we compute the measurements

$$\boldsymbol{y}_n = \boldsymbol{\Phi}_n \boldsymbol{x}_n^\star, \quad n = 1, \ldots, N.$$

At the central node, the problem of recovering the original data matrix $\mathbf{X}^\star$ from the compressed measurements can be formulated as

$$
\begin{aligned}
\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \; & \|\mathbf{X}\|_{\mathcal{A}} \\
\text{s.t. } & \boldsymbol{y}_n = \boldsymbol{\Phi}_n \boldsymbol{x}_n, \quad n = 1, \ldots, N, \\
& \mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N],
\end{aligned}
\tag{3.9}
$$

which is equivalent to

$$
\begin{aligned}
\hat{\mathbf{X}}, \hat{\boldsymbol{u}}, \hat{\mathbf{V}} = \arg\min_{\mathbf{X}, \boldsymbol{u}, \mathbf{V}} \; & \frac{1}{2M} \operatorname{Tr}(\mathcal{T}(\boldsymbol{u})) + \frac{1}{2} \operatorname{Tr}(\mathbf{V}) \\
\text{s.t. } & \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{X} \\ \mathbf{X}^* & \mathbf{V} \end{bmatrix} \succeq 0, \quad \boldsymbol{y}_n = \boldsymbol{\Phi}_n \boldsymbol{x}_n, \quad n = 1, \ldots, N, \\
& \mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N]
\end{aligned}
$$

As in Section 3.2.1, one can solve the above SDP, obtain the estimated frequencies from the dual polynomial, and recover the mode shapes and amplitudes by solving a least-squares problem. We have the following result.

Theorem 3.4. *Suppose $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \ldots, \boldsymbol{\Phi}_N \in \mathbb{R}^{M' \times M}$ are independently drawn random matrices with i.i.d. Gaussian entries having mean zero and variance 1. Assume that the true frequencies satisfy the minimum separation condition*

$$\Delta_f \geq \frac{1}{\lfloor (M-1)/4 \rfloor}.
\tag{3.10}$$

Then there exists a numerical constant C such that if

$$M' \geq C K \log(M),
\tag{3.11}$$

$\mathbf{X}^\star$ *is the unique optimal solution of the ANM problem* (3.9) *and we can exactly recover the frequencies and mode shapes up to a phase ambiguity with probability at least $1 - e^{-\frac{N(M'-2)}{8}}$.*

We prove Theorem 3.4 in Section 5.2, extending analysis from Theorem 2 in [9], which concerned the SMV ($N = 1$) version of this problem.

From Theorem 3.4, we see that modal analysis is possible in the random temporal compression scenario using a number M' of measurements per sensor that scales linearly with the number of active modes K. One logarithmic term remains in this bound (3.11), while some of the other logarithmic terms appearing in (3.7) and (3.8) have disappeared. Expressing the failure probability as δ for easier comparison with Theorems 3.2 and 3.3, we see that when

$$M' \geq \max\left\{ \frac{8}{N} \log\left(\frac{1}{\delta}\right) + 2, C K \log(M) \right\},
\tag{3.12}$$

exact recovery of the mode shapes and frequencies is possible with probability at least $1 - \delta$. In the first term inside this expression, the number M' of measurements per sensor actually *decreases* as the number N of sensors increases, a contrast with Theorem 3.2 and 3.3 where the opposite trend holds. We note that the benefit of increasing the number of sensors N comes from the fact that in the setup of Theorem 3.4, each sensor uses a different measurement matrix Φ_n. When the second term in (3.12) dominates, the number M' of measurements per sensor still does not increase with N. Similar to the random sampling schemes, the total duration over which the signals must be observed does not change, but within this time span significant compression may be possible.

Also in contrast with Theorems 3.2 and 3.3, Theorem 3.4 requires no randomness assumption on the mode shapes. This allows for its use in practical scenarios, where in general mode shapes will not be generated randomly. It would be interesting to remove the randomness condition from Theorems 3.2 and 3.3; we leave this as a question for future work.

3.2.5 Random spatial compression

As a final strategy to compress the data matrix, we consider the scenario where each sensor modulates its sample value by a random number at each time sample, then the sensors transmit these values coherently to the central node, where the modulated values add to result in a single measurement vector. As discussed in [45], such randomized spatial aggregation of the measurements can be achieved as part of using phase-coherent analog transmissions to the base station. This strategy is sometimes referred to as *compressive wireless sensing*; we refer to it as *random spatial compression*.

In random spatial compression, the collected measurements can be expressed as

$$y_m = \langle \mathbf{X}^{\star\top}(:,m), \tilde{\boldsymbol{b}}_m \rangle = \langle \mathbf{X}^{\star\top}, \tilde{\boldsymbol{b}}_m \boldsymbol{e}_m^\top \rangle, \quad m = 1, \ldots, M$$

where $\tilde{\boldsymbol{b}}_m \in \mathbb{C}^{N \times 1}$ and $\boldsymbol{e}_m \in \mathbb{R}^{M \times 1}$ is the mth canonical basis vector. This allows us to formulate the modal analysis problem as an ANM problem:

$$
\begin{aligned}
\hat{\mathbf{X}} &= \arg\min_{\mathbf{X}} \|\mathbf{X}\|_{\mathcal{A}} \\
&\text{s.t. } y_m = \langle \mathbf{X}^\top, \tilde{\boldsymbol{b}}_m \boldsymbol{e}_m^\top \rangle, \quad 1 \le m \le M,
\end{aligned}
\tag{3.13}
$$

which is equivalent to

$$
\begin{aligned}
\hat{\mathbf{X}}, \hat{\boldsymbol{u}}, \hat{\mathbf{V}} &= \arg\min_{\mathbf{X}, \boldsymbol{u}, \mathbf{V}} \ \frac{1}{2M} \operatorname{Tr}(\mathcal{T}(\boldsymbol{u})) + \frac{1}{2} \operatorname{Tr}(\mathbf{V}) \\
&\text{s.t. } \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{X} \\ \mathbf{X}^* & \mathbf{V} \end{bmatrix} \succeq 0, \quad y_m = \langle \mathbf{X}^\top, \tilde{\boldsymbol{b}}_m \boldsymbol{e}_m^\top \rangle, \quad 1 \le m \le M.
\end{aligned}
$$

As in Section 3.2.1, one can solve the above SDP, obtain the estimated frequencies from the dual polynomial, and recover the mode shapes and amplitudes by solving a least-squares problem. The following theorem follows from our previous work [16] on super-resolution of complex exponentials from modulations with unknown waveforms.

Theorem 3.5. *[16] Suppose we observe the data matrix $\mathbf{X}^\star$ with the above random spatial compression scheme. Assume that the random vectors $\tilde{\boldsymbol{b}}_m$ are i.i.d. samples from an isotropic and $\mu-$incoherent distribution (see [16] for details). Also, suppose that $\boldsymbol{\psi}_k^\star$ are drawn i.i.d. from the uniform distribution on the complex unit sphere and that the minimum separation condition (3.10) is satisfied. Then there exists a numerical constant C such that[9]*

$$M \ge C \mu K N \log\left(\frac{MKN}{\delta}\right) \log^2\left(\frac{MN}{\delta}\right) \tag{3.14}$$

---

[9]Note that the bound in (3.14) contains M on both sides. One could remove M from the right hand side using the Lambert W-function as in [46, (64)–(66)]. Here, we prefer the form in (3.14) for simplicity and to highlight the relationship between M and K, N.

is sufficient to guarantee that we can exactly recover $\mathbf{X}^\star$ via (3.13) and exactly recover the frequencies and mode shapes up to a phase ambiguity with probability at least $1 - \delta$.

In random spatial compression, the central node receives exactly one compressed measurement at each time instant. Thus, the number of time samples M equals the total number of compressed measurements. Thus, Theorem 3.5 states that it is sufficient for the total number of compressed measurements to scale essentially linearly with KN (as long as (3.10) is also satisfied). Since K is the number of active modes and N is the number of sensors (and thus the length of each unknown mode shape), the number of unknown degrees of freedom in this problem scales with KN. In this sense, the result in Theorem 3.5 compares favorably with those in Theorems 3.2, 3.3, and 3.4. Theorem 3.5 does require a randomness assumption on the mode shapes.

We note that, in some applications, once (3.10) is satisfied (which imposes a lower bound on M that is comparable to what appears in Theorems 3.1–3.4), it could be the case that (3.14) is also satisfied. In this case, the same uniform data matrix $\mathbf{X}^\star$ that suffices for perfect recovery according to Theorem 3.1 can be completely compressed in the spatial dimension, reducing the total number of samples from MN to M. Thus, significant savings may be possible in structures where the number of sensors N is large. Indeed, up to the point where (3.14) becomes a stronger condition than (3.10), one could continue adding sensors without increasing the requisite number of compressed samples.

3.3 Modal analysis for noisy signals

In this section, we revisit the uniform sampling scenario where a data matrix is fully observed, but we now consider the case where the samples are corrupted by additive white Gaussian noise. While the analysis in this section may be of its own independent interest, it is also used in our proof of Theorem 3.4.

We consider observations of the form $\mathbf{Y} = \mathbf{X}^\star + \mathbf{W}$, where the entries of $\mathbf{W}$ satisfy $\mathcal{CN}(0, \sigma^2)$, and we consider the following atomic norm denoising problem:

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda\|\mathbf{X}\|_{\mathcal{A}}. \tag{3.15}$$

The theorem below provides an upper bound on the recovery error in Frobenius norm and is proved in Section 5.3.

Theorem 3.6. *Suppose the true data matrix $\mathbf{X}^\star$ is given as in (3.3) with the true frequencies satisfying the minimum separation condition (3.10). Given the noisy data $\mathbf{Y} = \mathbf{X}^\star + \mathbf{W}$, where the entries of $\mathbf{W}$ are i.i.d. complex Gaussian random variables which satisfy $\mathcal{CN}(0, \sigma^2)$, the estimate $\hat{\mathbf{X}}$ obtained by solving the atomic norm denoising problem (3.15) (with regularizing parameter $\lambda = \eta\sigma\sqrt{4MN\log(M)}$ and with $\eta \in (1, \infty)$ chosen sufficiently large) will satisfy*

$$\|\hat{\mathbf{X}} - \mathbf{X}^\star\|_F^2 \le C\sigma^2 KN\log(M) \tag{3.16}$$

with probability at least $1 - \frac{1}{M^2}$ for a numerical constant C.

Corollary 3.1. *Under the assumptions of Theorem 3.6, the estimate $\hat{\mathbf{X}}$ obtained by solving the atomic norm denoising problem (3.15) (with regularizing parameter $\lambda = \eta\sigma\sqrt{4MN\log(M)}$ and with $\eta \in (1, \infty)$ chosen sufficiently large) will satisfy*

$$\mathbb{E}\|\hat{\mathbf{X}} - \mathbf{X}^\star\|_F^2 \le C\sigma^2 KN\log(M). \tag{3.17}$$

Note that Theorem 3.6 provides a bound which extends the SMV case in [18] to the MMV case. Corollary 3.1 is a direct consequence of Theorem 3.6 and is proved in Section 5.4. The above MMV atomic norm denoising problem (3.15) is also considered in [7]. However, there the authors provide only an asymptotic bound on $\mathbb{E}\|\hat{\mathbf{X}} - \mathbf{X}^\star\|_F^2$.

4 Simulation Results

In this section, we present some experiments on synthetic data to exhibit the performance of ANM based modal analysis in the various sampling and compression scenarios.[10] We use the modal assurance criterion (MAC) to evaluate the quality of recovered mode shapes, which is defined as

$$\text{MAC}(\boldsymbol{\psi}_k^\star, \hat{\boldsymbol{\psi}}_k) = |\langle \boldsymbol{\psi}_k^\star, \hat{\boldsymbol{\psi}}_k \rangle|,$$

where $\hat{\boldsymbol{\psi}}_k$ is the kth estimated mode shape and $\boldsymbol{\psi}_k^\star$ is the kth true mode shape.[11] A value of MAC = 1 would indicate perfect recovery of the true mode shape. We consider mode shape recovery to be a success if $\text{MAC}(\boldsymbol{\psi}_k^\star, \hat{\boldsymbol{\psi}}_k) \geq 0.99$ for all k. We consider data matrix recovery to be a success if $\frac{\|\hat{\mathbf{X}} - \mathbf{X}^\star\|_F}{\|\mathbf{X}^\star\|_F} \leq 10^{-5}$.

4.1 Uniform sampling

As mentioned in Section 3.2.1, the full data, uniform sampling case has previously been considered in [10]. In this experiment, we compare the ANM based algorithm with the SVD based algorithm from [20] on a simple 6-degree-of-freedom boxcar system. The boxcar system is shown in Fig. 1. We consider this undamped system under the context of free vibration. The system parameters are set as follows: the masses are $m_1 = 1$, $m_2 = 2$, $m_3 = 3$, $m_4 = 4$, $m_5 = 5$, $m_6 = 6$ kg, and the stiffness values are $k_1 = k_7 = 500$, $k_2 = k_6 = 150$, $k_3 = k_5 = 100$, $k_4 = 50$ N/m. Thus, the mass, damping and stiffness matrices in (2.1) are given as $\mathbf{M} = \text{diag}([m_1, \ldots, m_6])$, $\mathbf{C} = \mathbf{0}$, and

$$\mathbf{K} = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 & 0 & 0 & 0 \\ -k_2 & k_2 + k_3 & -k_3 & 0 & 0 & 0 \\ 0 & -k_3 & k_3 + k_4 & -k_4 & 0 & 0 \\ 0 & 0 & -k_4 & k_4 + k_5 & -k_5 & 0 \\ 0 & 0 & 0 & -k_5 & k_5 + k_6 & -k_6 \\ 0 & 0 & 0 & 0 & -k_6 & k_6 + k_7 \end{bmatrix},$$

respectively. The true mode shapes and natural frequencies of this system can be obtained from the (normalized) generalized eigenvectors and square root of the generalized eigenvalues of the stiffness matrix $\mathbf{K}$ and mass matrix $\mathbf{M}$. In particular, the true frequencies are $F_1 = 0.5384$, $F_2 = 0.8964$, $F_3 = 1.2404$, $F_4 = 1.7434$, $F_5 = 1.7881$, and $F_6 = 4.1218$ Hz. We collect $M = 100$ uniform samples from this system with sampling interval $T_s = 0.9/F_c$, where $F_c = \max_{1 \leq k \leq 6} |F_k|$. The amplitudes are set as $A_1 = 1$, $A_2 = 0.85$, $A_3 = 0.7$, $A_4 = 0.5$, $A_5 = 0.25$ and $A_6 = 0.1$. We apply both ANM and SVD to the obtained data matrix to identify the modal parameters of this boxcar system. (Note that the SVD based algorithm only estimates the mode shapes and not the frequencies.) Figure 2(a) shows how the dual polynomial from ANM can be used to localize the frequencies. In particular, we identify the frequencies by identifying where the dual polynomial achieves $\|\mathcal{Q}(f)\|_2 = 1$. The true frequencies and estimated frequencies are presented in Fig. 2(b). The estimated mode shapes from the two algorithms are illustrated in Fig. 2(c). As mentioned in Section 2.1, the SVD based algorithm can only return mode shape estimates that are mutually orthogonal (in fact they are orthogonal singular vectors of the data matrix). The true mode shapes in this experiment are not orthogonal, which hampers the performance of the SVD. The ANM algorithm is not restricted to returning orthogonal mode shape estimates, and in this experiment it recovers the mode shapes perfectly. In particular, the MAC for AMN is $(1, 1, 1, 1, 1, 1)$, while the MAC for SVD is $(0.9176, 0.7231, 0.8060, 0.9749, 0.9659, 0.9883)$.

In this experiment, the minimum separation $\Delta_f = 0.0098$. Theorem 3.1 guarantees that perfect recovery is possible via ANM when $M \geq \max\{\frac{4}{\Delta_f} + 1, 257\} = \max\{410, 275\}$. We see perfect recovery in this simulation with $M = 100$ uniform samples.

In the following sections, for convenience we will use random mode shapes and/or discrete frequencies to test the ANM based algorithms.

---

[10]We use ADMM [7] to solve the atomic norm denoising problem in Section 4.5. All other simulations are implemented with CVX [41].

[11]To correctly pair the recovered mode shapes with the true ones, we assume the true frequencies f_k are in ascending order, and we adopt the same convention for the estimated frequencies.
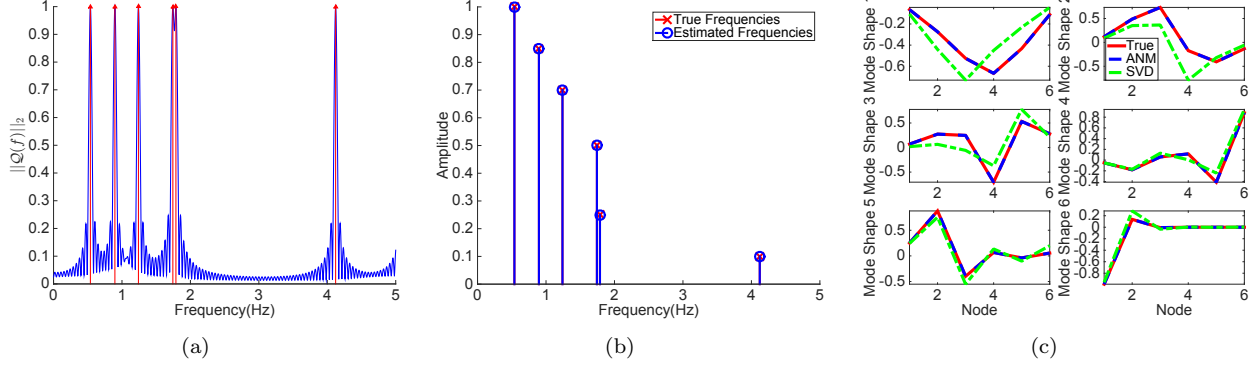
Figure 2: Uniform sampling in the boxcar system. (a) ANM based frequency localization from dual polynomial $\mathcal{Q}(f)$. The horizontal axis is shown in units of Hz, which corresponds to the digital frequency f times the sampling rate $\frac{1}{T_s}$. (b) Estimated frequencies by ANM. (c) Estimated mode shapes by ANM and SVD.

4.2 Asynchronous vs. synchronous random sampling

In this experiment, we compare the performance of asynchronous and synchronous random sampling in a case where the mode shapes are randomly generated but also correlated.[12] An example of such correlated mode shapes is shown in Fig. 3(a). (Only the first two mode shapes are correlated.) The true discrete frequencies are set to $f_1 = 0.1$, $f_2 = 0.15$, and $f_3 = 0.5$. We collect $M = 80$ uniform samples from each sensor. However, from these, on average, we keep only $M' < M$ random samples from each sensor, where the value for M' ranges from 2 to 20. In the case of synchronous random sampling, we keep exactly M' samples from each sensor at the same times. In the case of asynchronous random sampling, we generate Ω_A uniformly at random, with $|\Omega_A| = M'N$.

Other parameters are set the same as in Section 4.1. We perform 300 trials (each with a new set of mode shapes) for each value of M'. Figure 3(b) shows that when compared with synchronous sampling, asynchronous random sampling needs fewer observed measurements to achieve the same probability of successful recovery. This observation is reasonable (given the additional diversity in the asynchronous observations) but stands in contrast with the relative difference between the theoretical bounds in Theorems 3.2 and 3.3. More work may be needed to theoretically characterize the performance difference between asynchronous and synchronous random sampling.

4.3 Random temporal compression

In the next set of experiments, we generate a series of random matrices $\mathbf{\Phi}_n \in \mathbb{C}^{M' \times M}$, $n = 1, \ldots, N$ to compress a set of $M = 80$ uniform samples at each sensor. In the first experiment, we recover the data matrix and mode shapes both jointly (via the MMV approach from (3.9)) and separately (by solving N separate SMV problems) to show the advantage of joint recovery. We choose M' between 3 and 30 and perform 100 trials for each value of M'. We use random mode shapes, all generated with i.i.d. Gaussian entries and then normalized. The true discrete frequencies are set to $f_1 = 0.1$, $f_2 = 0.15$, and $f_3 = 0.5$, giving a separation of $\Delta_f = 0.05$, which is slightly smaller than the separation condition prescribed in (3.10). Figures 4(a), (b) show the probability of successful recovery for the data matrix and mode shapes, respectively, when we use separate ANM (dashed lines) and joint ANM (solid lines). The number on each line denotes the number of sensors N used in the experiments. It can be seen that joint recovery outperforms separate recovery significantly. Moreover, these results also indicate that the number of sensors has an important effect on the performance of joint recovery. In particular, for a given number of measurements

---

[12]We generate the first and third mode shapes randomly with i.i.d. Gaussian entries and then normalize. The second mode shape is generated by slightly perturbing the first mode shape and then normalizing.
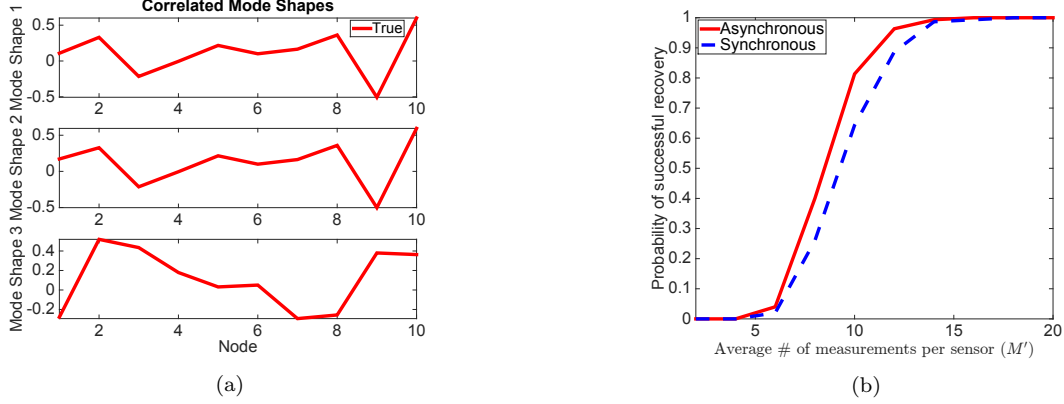
(a)

(b)

Figure 3: Asynchronous vs. synchronous random sampling. (a) Correlated random mode shapes used in experiments. (b) Probability of successful recovery of mode shapes with each sampling scheme.
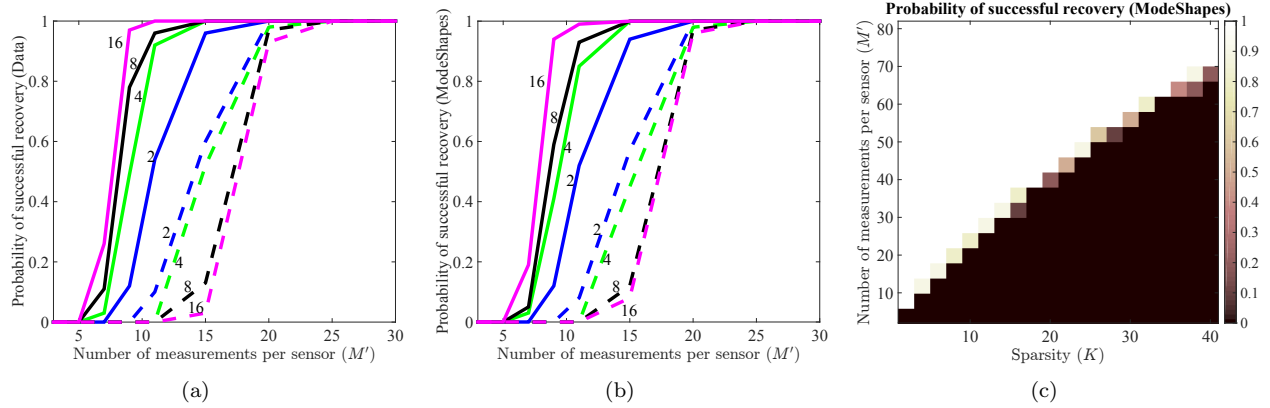

(a)

(b)

(c)

Figure 4: Random temporal compression. (a) Recovery of data matrix via joint ANM (solid lines) and separate ANM (dashed lines). (b) Recovery of mode shapes via joint ANM (solid lines) and separate ANM (dashed lines). (c) Probability of successful recovery for mode shapes with ANM when $N = 10$ is fixed.

M', the probability of successful joint recovery will increase as the number of sensors increases, which is consistent with our theoretical analysis. In contrast, the probability of successful separate recovery will decrease.

In the second experiment, we set the number of sensors to be $N = 10$ and investigate the minimal number of measurements per sensor M' needed for perfect joint recovery with various numbers of active modes K. The true mode shapes are generated randomly and we set $M = 100$. For each value of K, we randomly pick K discrete frequencies from a frequency set $\mathcal{F} = 0.03 : \text{Sep} : 0.99$, where $\text{Sep} = 2/M$. The amplitudes A_k, $k = 1, \ldots, K$ are chosen randomly from the uniform distribution between 0 and 1. It can be seen in Fig. 4(c) that the minimal number of measurements needed by each sensor for perfect recovery does scale roughly linearly with the number of active modes K, as indicated in Theorem 3.4.

4.4 Random spatial compression

Using the same parameters as in Section 4.2, we simulate the random spatial compression strategy with ANM based modal analysis and compare to the SVD algorithm studied in [20], which we apply to a full $M \times N$ data matrix. We generate mode shapes randomly, testing both orthogonal mode shapes and correlated mode shapes. It can be seen from Fig. 5(a), (b) that the SVD based method performs poorly when the mode
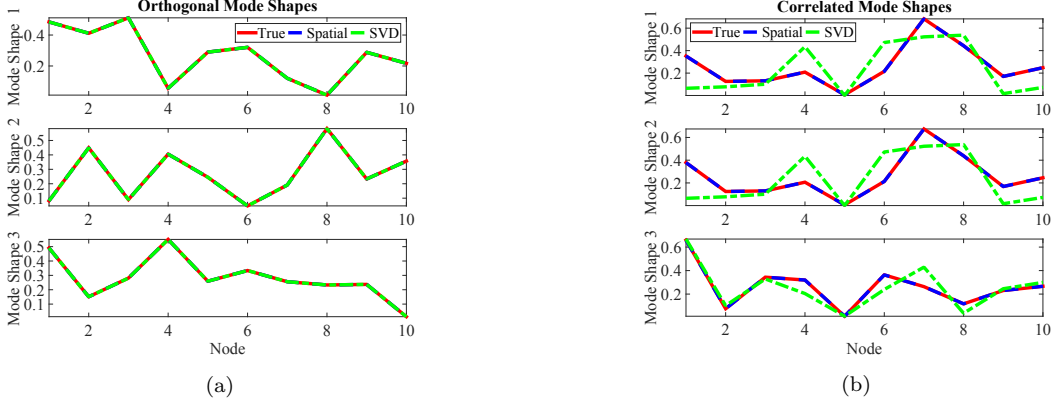
Figure 5: Random spatial compression. (a) Estimated mode shapes by ANM (on compressed data) and SVD (on uncompressed data) with orthogonal mode shapes. (b) Estimated mode shapes with correlated mode shapes.
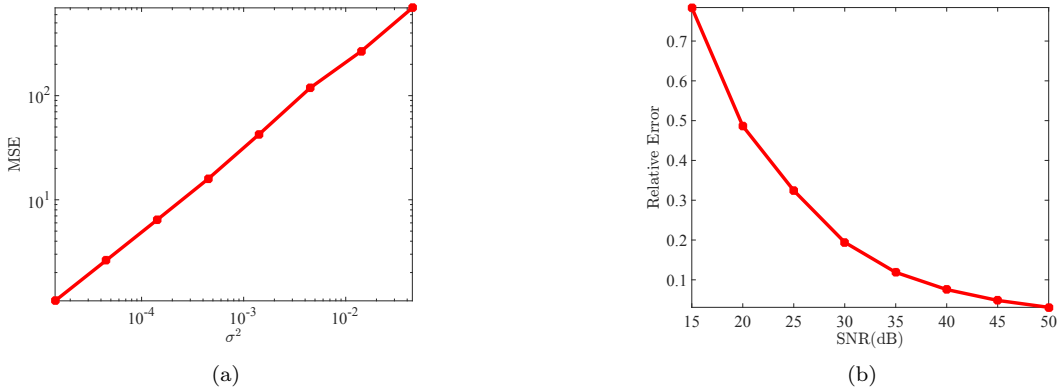

Figure 6: Atomic norm denoising. (a) Recovered MSE with respect to different noise levels. (b) Relative error with respect to different SNR levels.

shapes are correlated. However, the proposed ANM based algorithm performs well in both cases. Note that although SVD based method performs very well when the mode shapes are orthogonal, this is using $M \cdot N = 80 \cdot 10 = 800$ samples while the ANM based algorithm uses only $M = 80$ spatially compressed measurements to recover both the mode shapes and the frequencies.

4.5 Noisy data

Finally, we simulate the atomic norm denoising problem presented in Section 3.3. We assume the entries of $\mathbf{W}$ are i.i.d. random variables from the distribution $\mathcal{CN}(0, \sigma^2)$. We set the signal-to-noise ratio (SNR) between 15 and 50 dB, which corresponds to values of σ between 0.1754 and 0.0031. We perform 50 trials for each value of σ, with random mode shapes in each trial. The regularization parameter is set to $\lambda = \sigma \sqrt{4MN \log(M)}$ with $M = 80$ and $N = 10$. Other parameters are set the same as in Section 4.2. We define $\|\hat{\mathbf{X}} - \mathbf{X}^\star\|_F^2$ and $\frac{\|\hat{\mathbf{X}} - \mathbf{X}^\star\|_F}{\|\mathbf{X}^\star\|_F}$ as the mean square error (MSE) and relative error, respectively. It can be seen in Fig. 6(a) that the MSE is linearly correlated with σ^2, which is consistent with the theory in Section 3.3. We also present Fig. 6(b) to illustrate how the relative error behaves with different noise levels.

5 Proofs

In this section, we will prove Theorem 3.4, Theorem 3.6, and Corollary 3.1. Some proof techniques are inspired by the work in [5, 9, 17, 18].

5.1 Convex analysis

We first review some basic concepts from convex analysis [47]. A set Λ is a cone if

$$\theta \boldsymbol{x} \in \Lambda \ \text{ for any } \boldsymbol{x} \in \Lambda,$$

where θ is a nonnegative constant. Λ is a convex cone if

$$\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 \in \Lambda$$

holds for any $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in \Lambda$ and θ_1, $\theta_2 \geq 0$. The polar cone of a cone Λ is

$$\Lambda^\circ \triangleq \{\boldsymbol{z} : \langle \boldsymbol{x}, \boldsymbol{z} \rangle \leq 0 \ \forall \ \boldsymbol{x} \in \Lambda\}.$$

The tangent cone and normal cone at $\mathbf{X}$ with respect to the scaled unit ball $\|\mathbf{X}\|_{\mathcal{A}} \operatorname{conv}(\mathcal{A})$ are defined as

$$\mathcal{T}_{\mathcal{A}}(\mathbf{X}) \triangleq \operatorname{cone}\{\mathbf{D} : \|\mathbf{X} + \mathbf{D}\|_{\mathcal{A}} \leq \|\mathbf{X}\|_{\mathcal{A}}\},$$

and

$$\mathcal{N}_{\mathcal{A}}(\mathbf{X}) \triangleq \{\mathbf{S} : \langle \mathbf{S}, \mathbf{D} \rangle \leq 0, \forall \ \mathbf{D} \ \text{s.t.} \ \|\mathbf{X} + \mathbf{D}\|_{\mathcal{A}} \leq \|\mathbf{X}\|_{\mathcal{A}}\},$$

respectively. Note that the tangent cone $\mathcal{T}_{\mathcal{A}}(\mathbf{X})$ is the set of descent directions of the atomic norm $\| \cdot \|_{\mathcal{A}}$ at $\mathbf{X}$. The normal cone $\mathcal{N}_{\mathcal{A}}(\mathbf{X})$ is the polar cone of the tangent cone $\mathcal{T}_{\mathcal{A}}(\mathbf{X})$ and vice-versa.

Let Ω be a subset of $\mathbb{S}^{MN-1}$, where $\mathbb{S}^{MN-1} \triangleq \{\mathbf{Z} \in \mathbb{C}^{M \times N} : \|\mathbf{Z}\|_F = 1\}$ denotes the unit sphere. Then, the Gaussian width of Ω is

$$\omega(\Omega) \triangleq \mathbb{E}_{\mathbf{H}} \left[\max_{\mathbf{Z} \in \Omega} \langle \mathbf{H}, \mathbf{Z} \rangle \right], \tag{5.1}$$

where $\mathbf{H}$ is a Gaussian matrix with i.i.d. entries from the distribution $\mathcal{N}(0, 1)$.

5.2 Proof of Theorem 3.4

We start by showing that the number of measurements M' needed for perfect recovery can be lower bounded with a Gaussian width in Section 5.2.1. Then, we upper bound the Gaussian width with the expectation of the recovery error obtained from the atomic norm denoising problem (3.15) in Section 5.2.2.

5.2.1 Bounding M' with a Gaussian width

It can be shown that the ANM problem in (3.9) is equivalent to the following optimization problem

$$\min_{\mathbf{D}} \|\mathbf{X}^\star + \mathbf{D}\|_{\mathcal{A}}$$
$$\text{s.t. } \boldsymbol{\Phi}_n \boldsymbol{d}_n = \mathbf{0}, \ \ n = 1, \ldots, N,$$
$$\mathbf{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_N].$$

In particular, the above equivalent optimization problem can be obtained by eliminating the equality constraints in (3.9). Let $\mathbf{X} = \mathbf{X}^\star + \mathbf{D}$ with $\boldsymbol{\Phi}_n \boldsymbol{d}_n = \mathbf{0}$, $n = 1, \ldots, N$. It can be seen that the two optimization problems are equivalent.

Let $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_N]$. With a slight abuse of notation, we define

$$\text{null}(\boldsymbol{\Phi}) \triangleq \{\mathbf{D} \in \mathbb{C}^{M \times N} : \ \boldsymbol{\Phi}_n \boldsymbol{d}_n = \mathbf{0}, \ \ n = 1, \ldots, N\}$$

as a set of matrices with columns belonging to the null space of the corresponding sensing matrix. We also define a block diagonal matrix $\text{diag}(\mathbf{Z})$ corresponding to a matrix $\mathbf{Z} = [\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_N] \in \mathbb{C}^{M \times N}$ as

$$\text{diag}(\mathbf{Z}) \triangleq \begin{bmatrix} \boldsymbol{z}_1 & & & \\ & \boldsymbol{z}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{z}_N \end{bmatrix} \in \mathbb{C}^{MN \times N}. \tag{5.2}$$

Inspired by [5], we have the following proposition which gives us an optimal condition for exact recovery.

Proposition 1. $\hat{\mathbf{X}} = \mathbf{X}^\star$ *is the unique optimal solution to the ANM (3.9) if and only if*

$$\mathcal{T}_{\mathcal{A}}(\mathbf{X}^\star) \cap \text{null}(\boldsymbol{\Phi}) = \{\mathbf{0}\}.$$

Proof. On one hand, if $\mathbf{X}^\star$ is the unique optimal solution, then $\|\mathbf{X}^\star + \mathbf{D}\|_{\mathcal{A}} > \|\mathbf{X}^\star\|_{\mathcal{A}}$ holds for all $\mathbf{D} \in \text{null}(\boldsymbol{\Phi})/\{\mathbf{0}\}$. It follows that $\mathbf{D} \notin \mathcal{T}_{\mathcal{A}}(\mathbf{X}^\star)$. Thus, we can get $\mathcal{T}_{\mathcal{A}}(\mathbf{X}^\star) \cap \text{null}(\boldsymbol{\Phi}) = \{\mathbf{0}\}$.

On the other hand, if $\mathcal{T}_{\mathcal{A}}(\mathbf{X}^\star) \cap \text{null}(\boldsymbol{\Phi}) = \{\mathbf{0}\}$, then, for all $\mathbf{D} \in \text{null}(\boldsymbol{\Phi})/\{\mathbf{0}\}$, i.e., $\boldsymbol{\Phi}_n \boldsymbol{d}_n = \mathbf{0}, \ \ n = 1, \ldots, N$, we have $\|\mathbf{X}^\star + \mathbf{D}\|_{\mathcal{A}} > \|\mathbf{X}^\star\|_{\mathcal{A}}$. Thus, $\mathbf{X}^\star$ is the unique optimal solution. $\qquad\square$

Define $\Omega = \mathcal{T}_{\mathcal{A}}(\mathbf{X}^\star) \cap \mathbb{S}^{MN-1}$, where $\mathbb{S}^{MN-1} = \{\mathbf{Z} \in \mathbb{C}^{M \times N} : \|\mathbf{Z}\|_F = 1\}$. It can be seen that Ω is a subset of the unit sphere. According to Proposition 1, to show that $\mathbf{X}^\star$ is the unique optimal solution to (3.9), we hope to demonstrate that $\mathbf{Z} \notin \text{null}(\boldsymbol{\Phi})$ holds for any $\mathbf{Z} = [\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_N] \in \Omega$, i.e.,

$$\sum_{n=1}^{N} \|\boldsymbol{\Phi}_n \boldsymbol{z}_n\|_2 > 0,$$

which will hold if

$$\sum_{n=1}^{N} \|\boldsymbol{\Phi}_n \boldsymbol{z}_n\|_2^2 = \|[\boldsymbol{\Phi}_1 \boldsymbol{z}_1 \ \boldsymbol{\Phi}_2 \boldsymbol{z}_2 \ \cdots \ \boldsymbol{\Phi}_N \boldsymbol{z}_N]\|_F^2 = \|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F^2 > 0.$$

Therefore, we need to show

$$\|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F > 0,$$

where $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_N] \in \mathbb{R}^{M' \times MN}$, and $\hat{\mathbf{Z}} = \text{diag}(\mathbf{Z}) \in \mathbb{C}^{MN \times N}$ as is defined in (5.2). Note that we have $\|\hat{\mathbf{Z}}\|_F = \|\mathbf{Z}\|_F = 1$ since $\mathbf{Z}$ is in a subset of the unit sphere.

Next, we will show that $\|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F > 0$ holds for all $\mathbf{Z} \in \Omega = \mathcal{T}_{\mathcal{A}}(\mathbf{X}^\star) \cap \mathbb{S}^{MN-1}$ with high probability. It then follows that $\mathbf{X}^\star$ is the unique optimal solution of ANM (3.9).

Theorem 5.1. *Let $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_N] \in \mathbb{R}^{M' \times MN}$ be a random matrix with i.i.d Gaussian entries which satisfy $\mathcal{N}(0, 1)$. Let Ω be a subset of $\mathbb{S}^{MN-1}$. For all $\mathbf{Z} = [\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_N] \in \Omega$, define $\hat{\mathbf{Z}} = \text{diag}(\mathbf{Z})$ as (5.2). Then, we have*

$$\mathbb{E}\left[\min_{\mathbf{Z} \in \Omega} \|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F\right] \geq \lambda_{M'N} - \omega(\Omega),$$

where $\lambda_{M'N} = \mathbb{E}_{\mathbf{G} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \|\mathbf{G}\|_F \geq \frac{\sqrt{N}M'}{\sqrt{M'+1}}$ and $\omega(\Omega)$ is the Gaussian width of Ω defined in (5.1).

Remark 5.1. *The above theorem is based on Gordon's work [48], and provides us a lower bound for the minimum gain of the operator $\boldsymbol{\Phi}$ restricted to a set Ω. The proof of this theorem is given in Appendix A.1.*

18

Corollary 5.1. *Let* $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_N] \in \mathbb{R}^{M' \times MN}$ *be a random matrix with i.i.d. Gaussian entries which satisfy* $\mathcal{N}(0,1)$. *Define* $\Omega \triangleq \mathcal{T}_\mathcal{A}(\mathbf{X}^\star) \cap \mathbb{S}^{MN-1}$ *as a subset of the unit sphere. Then,* $\mathbf{X}^\star$ *is the unique optimal solution of ANM (3.9) with probability at least* $1 - e^{-\frac{N(M'-2)}{8}}$ *if*

$$M' \geq \frac{4}{N}\omega^2(\Omega). \tag{5.3}$$

Remark 5.2. *Corollary 5.1 is an immediate consequence of Theorem 5.1. It can be seen that the number of measurements* M' *needed for exact recovery can be lower bounded by a Gaussian width. The proof details (presented in Appendix A.2) are based on a Gaussian concentration inequality [49].*

5.2.2 Bounding the Gaussian width

This section is dedicated to finding an upper bound on $\omega(\mathcal{T}_\mathcal{A}(\mathbf{X}^\star) \cap \mathbb{S}^{MN-1})$. First, we define the mean-square distance of a set $\mathcal{C}$ as

$$\mathrm{dist}(\mathcal{C}) = \mathbb{E}_{\mathbf{G} \sim \mathcal{N}(\mathbf{0},\mathbf{I})}\left[\min_{\mathbf{D} \in \mathcal{C}} \|\mathbf{G} - \mathbf{D}\|_F^2\right].$$

Inspired by the work in [5] and [9], we have

$$\omega^2(\mathcal{T}_\mathcal{A}(\mathbf{X}^\star) \cap \mathbb{S}^{MN-1})$$
$$\leq \left[\mathbb{E}_\mathbf{G}\left[\min_{\mathbf{D} \in \mathcal{T}_\mathcal{A}^\circ(\mathbf{X}^\star)} \|\mathbf{G} - \mathbf{D}\|_F\right]\right]^2 \leq \mathbb{E}_\mathbf{G}\left[\min_{\mathbf{D} \in \mathcal{T}_\mathcal{A}^\circ(\mathbf{X}^\star)} \|\mathbf{G} - \mathbf{D}\|_F\right]^2$$
$$= \mathbb{E}_\mathbf{G}\left[\min_{\mathbf{D} \in \mathcal{T}_\mathcal{A}^\circ(\mathbf{X}^\star)} \|\mathbf{G} - \mathbf{D}\|_F^2\right] = \mathrm{dist}(\mathcal{T}_\mathcal{A}^\circ(\mathbf{X}^\star))$$
$$= \mathrm{dist}(\mathrm{cone}(\partial\|\mathbf{X}^\star\|_\mathcal{A})) \leq \mathrm{dist}(\lambda\partial\|\mathbf{X}^\star\|_\mathcal{A})$$
$$= \max_{\sigma > 0} \frac{\mathbb{E}_\mathbf{G}[\|\hat{\mathbf{X}}(\mathbf{X}^\star + \sigma\mathbf{G}, \sigma\lambda) - \mathbf{X}^\star\|_F^2]}{\sigma^2},$$

where the first two inequalities follow from Proposition 3.6 in [5] and Jensen's inequality, respectively. The third equality and last inequality can be found in equation (67) in [9] while the last equality comes from Theorem 1.1 in [50]. Note that $\mathcal{T}_\mathcal{A}^\circ$ is the polar cone of $\mathcal{T}_\mathcal{A}$, as defined at the beginning of Section 5.1. Here, $\hat{\mathbf{X}}(\mathbf{Y}, \lambda)$ is the solution to the MMV atomic norm denoising problem (3.15) with λ being the regularization parameter. With the inequality given in (3.17), we have

$$\frac{\mathbb{E}_\mathbf{G}[\|\hat{\mathbf{X}}(\mathbf{X}^\star + \sigma\mathbf{G}, \sigma\lambda) - \mathbf{X}^\star\|_F^2]}{\sigma^2} \leq CKN\log(M),$$

which implies

$$\omega^2(\mathcal{T}_\mathcal{A}(\mathbf{X}^\star) \cap \mathbb{S}^{MN-1}) \leq CKN\log(M). \tag{5.4}$$

Thus, plugging (5.4) into (5.3), we can get (3.11) and finish the proof of Theorem 3.4. $\qquad\square$

5.3 Proof of Theorem 3.6

In this section, we prove Theorem 3.6 by extending the results in [17] and [18] to the MMV case. For the SMV case, it is shown in [17] that a good choice of the regularization parameter can achieve accelerated convergence rates. Inspired by their choice of the regularization parameter, we use

$$\lambda \approx \eta\mathbb{E}\|\mathbf{W}\|_\mathcal{A}^*$$

in the MMV atomic norm denoising problem (3.15). Here, $\eta \in (1, \infty)$ is some constant which ultimately must be set large enough to enable the proof of Lemma 5.6, and $\mathbf{W}$ is a complex Gaussian matrix. To set λ, we need to find an upper bound for $\mathbb{E}\|\mathbf{W}\|_\mathcal{A}^*$.

5.3.1 Bounding $\mathbb{E}\|\mathbf{W}\|_{\mathcal{A}}^*$

Lemma 5.1. *Let $\mathbf{W} \in \mathbb{C}^{M \times N}$ be a random matrix with i.i.d. complex Gaussian entries from the distribution $\mathcal{CN}(0, \sigma^2)$. Then, there exists a numerical constant C such that*

$$\mathbb{E}\|\mathbf{W}\|_{\mathcal{A}}^* \leq C\sigma\sqrt{MN\log(M)}.$$

Proof. According to the definition of the dual atomic norm in (2.4), we have

$$(\|\mathbf{W}\|_{\mathcal{A}}^*)^2 = \sup_{f \in [0,1)} \|\mathbf{W}^*\boldsymbol{a}(f)\|_2^2 = \sup_{f \in [0,1)} \sum_{n=1}^{N} \left| \sum_{m=0}^{M-1} \mathbf{W}_{mn}^* e^{j2\pi fm} \right|^2$$

$$= \sup_{f \in [0,1)} \sum_{n=1}^{N} \sum_{m,p=0}^{M-1} \mathbf{W}_{mn}^* \mathbf{W}_{pn} e^{j2\pi f(m-p)}$$

$$= \sup_{f \in [0,1)} \mathcal{W}_M\left(e^{j2\pi f}\right),$$

where the polynomial $\mathcal{W}_M$ is defined as

$$\mathcal{W}_M\left(e^{j2\pi f}\right) \triangleq \sum_{n=1}^{N} \sum_{m,p=0}^{M-1} \mathbf{W}_{mn}^* \mathbf{W}_{pn} e^{j2\pi f(m-p)}.$$

For all $f_1, f_2 \in [0, 1)$ we have

$$\mathcal{W}_M\left(e^{j2\pi f_1}\right) - \mathcal{W}_M\left(e^{j2\pi f_2}\right) \leq \left| e^{j2\pi f_1} - e^{j2\pi f_2} \right| \sup_{f \in [0,1)} \mathcal{W}_M'\left(e^{j2\pi f}\right)$$

$$\leq 2\pi M |f_1 - f_2| \sup_{f \in [0,1)} \mathcal{W}_M\left(e^{j2\pi f}\right)$$

by using the mean value theorem and Bernstein's inequality for polynomials [51]. By letting f_2 take any of the L values $0, \frac{1}{L}, \ldots, \frac{L-1}{L}$, we have

$$\sup_{f \in [0,1)} \mathcal{W}_M\left(e^{j2\pi f}\right) \leq \max_{l=0,\ldots,L-1} \mathcal{W}_M\left(e^{j2\pi l/L}\right) + \frac{2\pi M}{L} \sup_{f \in [0,1)} \mathcal{W}_M\left(e^{j2\pi f}\right).$$

It follows that

$$(\|\mathbf{W}\|_{\mathcal{A}}^*)^2 = \sup_{f \in [0,1)} \mathcal{W}_M\left(e^{j2\pi f}\right) \leq \left(1 - \frac{2\pi M}{L}\right)^{-1} \max_{l=0,\ldots,L-1} \mathcal{W}_M\left(e^{j2\pi l/L}\right)$$

$$\leq \left(1 + \frac{4\pi M}{L}\right) \max_{l=0,\ldots,L-1} \mathcal{W}_M\left(e^{j2\pi l/L}\right),$$

where the last inequality holds if $L \geq 4\pi M$. Then, we get

$$\|\mathbf{W}\|_{\mathcal{A}}^* \leq \left(1 + \frac{4\pi M}{L}\right)^{\frac{1}{2}} \left[\max_{l=0,\ldots,L-1} \mathcal{W}_M\left(e^{j2\pi l/L}\right) \right]^{\frac{1}{2}} \tag{5.5}$$

and

$$\mathbb{E}\|\mathbf{W}\|_{\mathcal{A}}^* \leq \left(1 + \frac{4\pi M}{L}\right)^{\frac{1}{2}} \left[\mathbb{E} \max_{l=0,\ldots,L-1} \mathcal{W}_M\left(e^{j2\pi l/L}\right) \right]^{\frac{1}{2}}.$$

Define

$$x_l \triangleq \frac{1}{\sigma\sqrt{M}} \sum_{m=0}^{M-1} \mathbf{W}_{mn}^* e^{j2\pi lm/L}$$

for $l = 0, 1, \ldots L-1$. It can be seen that x_l is complex Gaussian variable satisfying $\mathcal{CN}(0, 1)$ since the entries of $\mathbf{W}$ satisfy $\mathcal{CN}(0, \sigma^2)$. Then, we have

$$
\begin{aligned}
\mathbb{E}\left[\max_{l=0,\ldots,L-1} \mathcal{W}_M\left(e^{j2\pi l/L}\right)\right] &= \mathbb{E}\left[\max_{l=0,\ldots,L-1} \sum_{n=1}^{N}\left|\sum_{m=0}^{M-1}\mathbf{W}_{mn}^* e^{j2\pi lm/L}\right|^2\right] \\
&\leq \sum_{n=1}^{N}\mathbb{E}\left[\max_{l=0,\ldots,L-1}\left|\sum_{m=0}^{M-1}\mathbf{W}_{mn}^* e^{j2\pi lm/L}\right|^2\right] \\
&= \frac{1}{2}\sigma^2 M\sum_{n=1}^{N}\mathbb{E}\left[\max_{l=0,\ldots,L-1} 2\left|x_l\right|^2\right] \\
&\leq \sigma^2 MN(\log L + 1),
\end{aligned}
$$

where the last inequality uses the result that $\mathbb{E}\left[\max_{l=0,\ldots,L-1} 2\left|x_l\right|^2\right] \leq 2\log(L) + 2$, see Lemma 5 in [17].

Choosing $L = 4\pi M \log(M)$ gives

$$
\begin{aligned}
\mathbb{E}\|\mathbf{W}\|_{\mathcal{A}}^* &\leq \sigma\left(1 + \frac{1}{\log(M)}\right)^{\frac{1}{2}}\sqrt{MN[\log(M) + \log(4\pi\log(M)) + 1]} \\
&\leq C\sigma\sqrt{MN\log(M)}.
\end{aligned}
$$

Note that C is a constant which belongs to $(1, 2)$ when M is large. $\qquad\square$

5.3.2 Bounding $\|\hat{\mathbf{X}} - \mathbf{X}^\star\|_F^2$

Now, we can set the regularizing parameter in the MMV atomic norm denoising problem (3.15) as $\lambda = \eta\sigma\sqrt{4MN\log(M)}$ for some $\eta \in (1, \infty)$. Then

$$
\|\mathbf{W}\|_{\mathcal{A}}^* \leq \frac{\lambda}{\eta} \tag{5.6}
$$

holds with high probability. In particular, we have the following lemma.

Lemma 5.2.

$$
\mathbb{P}\left[\|\mathbf{W}\|_{\mathcal{A}}^* \geq \frac{\lambda}{\eta}\right] \leq \frac{1}{M^2}.
$$

Proof. As is shown in (5.5), letting $L = M$, we have

$$
\begin{aligned}
\|\mathbf{W}\|_{\mathcal{A}}^* &\leq (1 + 4\pi)^{\frac{1}{2}}\left[\max_{l=0,\ldots,M-1}\mathcal{W}_M\left(e^{j2\pi l/M}\right)\right]^{\frac{1}{2}} \\
&= (1 + 4\pi)^{\frac{1}{2}}\left[\max_{l=0,\ldots,M-1}\sum_{n=1}^{N}\left|\sum_{m=0}^{M-1}\mathbf{W}_{mn}^* e^{j2\pi ml/M}\right|^2\right]^{\frac{1}{2}} \\
&\leq (1 + 4\pi)^{\frac{1}{2}}\left[\max_{l=0,\ldots,M-1}\left(\sum_{n=1}^{N}\left|\sum_{m=0}^{M-1}\mathbf{W}_{mn}^* e^{j2\pi ml/M}\right|\right)^2\right]^{\frac{1}{2}} \\
&= (1 + 4\pi)^{\frac{1}{2}}\max_{l=0,\ldots,M-1}\sum_{n=1}^{N}\left|\sum_{m=0}^{M-1}\mathbf{W}_{mn}^* e^{j2\pi ml/M}\right| \\
&\triangleq (1 + 4\pi)^{\frac{1}{2}}\max_{l=0,\ldots,M-1} Z_l.
\end{aligned}
$$

It can be seen that Z_l is stochastically upper bounded by a zero-mean Gaussian random variable with variance $\sigma^2 MN$ [9]. Then, for any $\beta > \frac{1}{\sqrt{2\pi}}$, we have

$$\mathbb{P}[|Z_m| \geq \sigma\sqrt{MN}\beta] \leq 2e^{-\beta^2}.$$

Let $\beta = \sqrt{4\log(M)}$, we can get

$$\mathbb{P}\left[\|\mathbf{W}\|_{\mathcal{A}}^* \geq \frac{\lambda}{\eta}\right] = \mathbb{P}\left[\|\mathbf{W}\|_{\mathcal{A}}^* \geq \sigma\sqrt{4MN\log M}\right] \leq 2Me^{-4\log(M)} \leq \frac{1}{M^2},$$

where the first inequality comes from the union bound. $\qquad\square$

The following lemma derived from convex analysis provides optimality conditions for $\hat{\mathbf{X}}$ to be the solution of (3.15).

Lemma 5.3. *(Optimality Conditions):* $\hat{\mathbf{X}}$ *is the solution of (3.15) if and only if*

1. $\|\mathbf{Y} - \hat{\mathbf{X}}\|_{\mathcal{A}}^* \leq \lambda$,

2. $\langle \mathbf{Y} - \hat{\mathbf{X}}, \hat{\mathbf{X}} \rangle = \lambda\|\hat{\mathbf{X}}\|_{\mathcal{A}}$.

Define an atomic measure as

$$\boldsymbol{\mu}(f) = \sum_{k=1}^{K} |A_k|\boldsymbol{b}_k\delta(f - f_k)$$

with $f \in [0, 1)$, $\|\boldsymbol{b}_k\|_2 = 1$. Then, we have

$$\mathbf{X}^\star = \sum_{k=1}^{K} |A_k|\boldsymbol{a}(f_k)\boldsymbol{b}_k^* = \int_0^1 \boldsymbol{a}(f)\boldsymbol{\mu}^*(f)df.$$

Similarly, the recovered data $\hat{\mathbf{X}}$ can be represented as

$$\hat{\mathbf{X}} = \int_0^1 \boldsymbol{a}(f)\hat{\boldsymbol{\mu}}^*(f)df$$

for some measure $\hat{\boldsymbol{\mu}}(f)$. Define the difference measure as $\boldsymbol{\nu} = \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$. Then, the error matrix is given as

$$\mathbf{E} \triangleq \hat{\mathbf{X}} - \mathbf{X}^\star = \int_0^1 \boldsymbol{a}(f)\boldsymbol{\nu}^*(f)df.$$

It follows that

$$\begin{aligned}
\|\mathbf{E}\|_F^2 = |\langle \mathbf{E}, \mathbf{E} \rangle| &= \left|\left\langle \mathbf{E}, \int_0^1 \boldsymbol{a}(f)\boldsymbol{\nu}^*(f)df \right\rangle\right| \\
&= \left|\int_0^1 \langle \mathbf{E}, \boldsymbol{a}(f)\boldsymbol{\nu}^*(f) \rangle df\right| \\
&= \left|\int_0^1 \boldsymbol{a}^*(f)\mathbf{E}\boldsymbol{\nu}(f)df\right| \qquad\qquad (5.7) \\
&= \left|\int_0^1 \boldsymbol{\xi}^*(f)\boldsymbol{\nu}(f)df\right| \\
&\leq \left|\int_F \boldsymbol{\xi}^*(f)\boldsymbol{\nu}(f)df\right| + \sum_{k=1}^{K}\left|\int_{N_k} \boldsymbol{\xi}^*(f)\boldsymbol{\nu}(f)df\right|
\end{aligned}$$

where $\boldsymbol{\xi}(f) \triangleq \mathbf{E}^* \boldsymbol{a}(f)$ is defined as a vector-valued error function. Here,

$$
\begin{aligned}
N_k &\triangleq \{f : d(f, f_k) \leq 0.16/M\}, \ k = 1, \dots, K \\
F &\triangleq [0, 1)/ \cup_{k=1}^{K} N_k
\end{aligned}
\tag{5.8}
$$

are defined as the kth near region corresponding to f_k and the far region, respectively.

With a little abuse of notation, we define

$$
\|\boldsymbol{\xi}(f)\|_{2,\infty} \triangleq \sup_{f \in [0,1)} \|\boldsymbol{\xi}(f)\|_2.
\tag{5.9}
$$

It turns out that

$$
\begin{aligned}
\|\boldsymbol{\xi}(f)\|_{2,\infty} &= \sup_{f \in [0,1)} \|\mathbf{E}^* \boldsymbol{a}(f)\|_2 \\
&= \sup_{f \in [0,1)} \|(\hat{\mathbf{X}} - \mathbf{Y})^* \boldsymbol{a}(f) + \mathbf{W}^* \boldsymbol{a}(f)\|_2 \\
&\leq \sup_{f \in [0,1)} \|(\mathbf{Y} - \hat{\mathbf{X}})^* \boldsymbol{a}(f)\|_2 + \sup_{f \in [0,1)} \|\mathbf{W}^* \boldsymbol{a}(f)\|_2 \\
&= \|\mathbf{Y} - \hat{\mathbf{X}}\|_{\mathcal{A}}^* + \|\mathbf{W}\|_{\mathcal{A}}^* \\
&\leq 2\lambda
\end{aligned}
$$

if the bound condition in (5.6) holds. The last inequality also follows from the first optimality condition in Lemma 5.3.

Next, we extend Lemmas 1, 2 and 3 in [18] to our MMV case and then bound the energy of the error matrix.

Lemma 5.4. *Since each entry of the vector-valued error function $\boldsymbol{\xi}(f) = \mathbf{E}^* \boldsymbol{a}(f)$ is an order-M trigonometric polynomial, we have*

$$
\|\mathbf{E}\|_F^2 \leq \|\boldsymbol{\xi}(f)\|_{2,\infty} \left[\int_F \|\boldsymbol{\nu}(f)\|_2 df + I_0 + I_1 + I_2 \right]
\tag{5.10}
$$

with

$$
\begin{aligned}
I_0^k &= \left\| \int_{N_k} \boldsymbol{\nu}(f) df \right\|_2, \\
I_1^k &= M \left\| \int_{N_k} (f - f_k) \boldsymbol{\nu}(f) df \right\|_2, \\
I_2^k &= \frac{M^2}{2} \int_{N_k} (f - f_k)^2 \|\nu(f)\|_2 df \\
I_l &= \sum_{k=1}^{K} I_l^k, \ \text{for} \ l = 0, 1, 2.
\end{aligned}
$$

The proof of Lemma 5.4 is given in Appendix A.3.1.

Lemma 5.5. *There exist some numerical constants C_0 and C_1 such that*

$$
\begin{aligned}
I_0 &\leq C_0 \left(\frac{K\lambda}{M} + I_2 + \int_F \|\boldsymbol{\nu}(f)\|_2 df \right), \\
I_1 &\leq C_1 \left(\frac{K\lambda}{M} + I_2 + \int_F \|\boldsymbol{\nu}(f)\|_2 df \right).
\end{aligned}
$$

The proof of Lemma 5.5 is given in Appendix A.3.2.

Lemma 5.6. *For some sufficiently large $\eta > 1$, set the regularizing parameter as $\lambda = \eta\sigma\sqrt{MN\log(M)}$. Then there exists a numerical constant C such that*

$$\int_F \|\boldsymbol{\nu}(f)\|_2 df + I_2 \leq \frac{CK\lambda}{M}$$

holds if $\|\mathbf{W}\|_{\mathcal{A}}^* \leq \frac{\lambda}{\eta}$.

The proof of Lemma 5.6 is given in Appendix A.3.3.

Using the above three lemmas, we have that

$$\|\hat{\mathbf{X}} - \mathbf{X}^\star\|_F^2 \leq \frac{CK\lambda^2}{M} = C\sigma^2 KN\log(M)$$

holds if η is sufficiently large and if $\|\mathbf{W}\|_{\mathcal{A}}^* \leq \frac{\lambda}{\eta}$. It follows from Lemma 5.2 that the above recovery error bound holds with probability at least $1 - \frac{1}{M^2}$. Thus, we finish the proof of Theorem 3.6. □

5.4   Proof of Corollary 3.1

As mentioned in the previous sections, once we set $\lambda = \eta\sigma\sqrt{4MN\log(M)}$, then

$$\|\mathbf{W}\|_{\mathcal{A}}^* \leq \frac{\lambda}{\eta}$$

holds with probability at least $1 - \frac{1}{M^2}$, which implies that the error bound in (3.16) holds with probability at least $1 - \frac{1}{M^2}$.

Note that

$$\mathbb{E}[\|\mathbf{E}\|_F^2] = \mathbb{E}\left[\|\mathbf{E}\|_F^2 \mathbb{1}\left\{\|\mathbf{W}\|_{\mathcal{A}}^* < \frac{\lambda}{\eta}\right\}\right] + \mathbb{E}\left[\|\mathbf{E}\|_F^2 \mathbb{1}\left\{\|\mathbf{W}\|_{\mathcal{A}}^* \geq \frac{\lambda}{\eta}\right\}\right], \tag{5.11}$$

where $\mathbb{1}$ is an indicator function. Theorem 3.6 implies that

$$\mathbb{E}\left[\|\mathbf{E}\|_F^2 \mathbb{1}\left\{\|\mathbf{W}\|_{\mathcal{A}}^* < \frac{\lambda}{\eta}\right\}\right] \leq C\sigma^2 KN\log(M), \tag{5.12}$$

which can be proved with the definition of expectation. Also, by the Cauchy-Schwarz inequality, we have

$$\mathbb{E}\left[\|\mathbf{E}\|_F^2 \mathbb{1}\left\{\|\mathbf{W}\|_{\mathcal{A}}^* \geq \frac{\lambda}{\eta}\right\}\right] \leq \sqrt{\mathbb{E}[\|\mathbf{E}\|_F^4]}\sqrt{\mathbb{P}\left[\|\mathbf{W}\|_{\mathcal{A}}^* \geq \frac{\lambda}{\eta}\right]} \leq C\sigma^2 N\log(M), \tag{5.13}$$

where the last inequality follows from the following lemma and Lemma 5.2.

Lemma 5.7.

$$\sqrt{\mathbb{E}\left[\|\mathbf{E}\|_F^4\right]} \leq C\sigma^2 MN\log(M). \tag{5.14}$$

Combining (5.11), (5.12), and (5.13) completes the proof of Corollary 3.1. Now, we are left with proving Lemma 5.7.

Proof. To prove Lemma 5.7, we first prove the following lemma.

Lemma 5.8. *There exists some numerical constant C such that the energy of the error matrix $\mathbf{E}$ can be upper bounded with*

$$\|\mathbf{E}\|_F^2 \leq 8\|\mathbf{W}\|_F^2 + C\lambda^2\frac{K}{M} = 8\|\mathbf{W}\|_F^2 + C\eta^2\sigma^2 KN\log(M).$$

24

Proof. The optimality of $\hat{\mathbf{X}}$ implies

$$\frac{1}{2}\|\mathbf{Y} - \hat{\mathbf{X}}\|_F^2 + \lambda\|\hat{\boldsymbol{\mu}}\|_{2,TV} \le \frac{1}{2}\|\mathbf{Y} - \mathbf{X}^\star\|_F^2 + \lambda\|\boldsymbol{\mu}\|_{2,TV},$$

which is equivalent to

$$\frac{1}{2}\|\mathbf{E}\|_F^2 + \lambda(\|\mathcal{P}_\mathcal{F}(\hat{\boldsymbol{\mu}})\|_{2,TV} + \|\mathcal{P}_{\mathcal{F}^c}(\hat{\boldsymbol{\mu}})\|_{2,TV}) \le \langle\mathbf{E}, \mathbf{W}\rangle + \lambda\|\boldsymbol{\mu}\|_{2,TV}. \tag{5.15}$$

Since $\hat{\boldsymbol{\mu}} = \boldsymbol{\nu} + \boldsymbol{\mu}$, we have

$$\begin{aligned}
\|\mathcal{P}_\mathcal{F}(\hat{\boldsymbol{\mu}})\|_{2,TV} &= \|\mathcal{P}_\mathcal{F}(\boldsymbol{\nu} + \boldsymbol{\mu})\|_{2,TV} \\
&= \|\boldsymbol{\mu}\|_{2,TV} + \|\mathcal{P}_\mathcal{F}(\boldsymbol{\nu})\|_{2,TV} \\
&= \|\boldsymbol{\mu}\|_{2,TV} + \|\boldsymbol{\nu}\|_{2,TV} - \|\mathcal{P}_{\mathcal{F}^c}(\boldsymbol{\nu})\|_{2,TV} \\
&= \|\boldsymbol{\mu}\|_{2,TV} + \|\boldsymbol{\nu}\|_{2,TV} - \|\mathcal{P}_{\mathcal{F}^c}(\hat{\boldsymbol{\mu}})\|_{2,TV}.
\end{aligned}$$

Then, the inequality (5.15) becomes

$$\frac{1}{2}\|\mathbf{E}\|_F^2 + \lambda\|\boldsymbol{\nu}\|_{2,TV} \le \langle\mathbf{E}, \mathbf{W}\rangle,$$

which implies

$$\begin{aligned}
\frac{1}{2}\|\mathbf{E}\|_F^2 &\le \langle\mathbf{E}, \mathbf{W}\rangle - \lambda\|\boldsymbol{\nu}\|_{2,TV} \\
&= \langle\mathbf{E}, \mathbf{W}\rangle - \lambda\int_0^1 \boldsymbol{\nu}^*(f)\mathcal{Q}(f)df \\
&= \langle\mathbf{E}, \mathbf{W}\rangle - \lambda\langle\mathbf{E}, \mathbf{Q}\rangle.
\end{aligned} \tag{5.16}$$

We have used Parseval's theorem in the last equality. Moreover,

$$\begin{aligned}
-\lambda\langle\mathbf{E}, \mathbf{Q}\rangle &\le \lambda\|\mathbf{E}\|_F\|\mathbf{Q}\|_F = \left(\frac{\|\mathbf{E}\|_F}{2}\right)(2\lambda\|\mathbf{Q}\|_F) \\
&\le \frac{1}{8}\|\mathbf{E}\|_F^2 + 2\lambda^2\|\mathbf{Q}\|_F^2.
\end{aligned} \tag{5.17}$$

Combining (5.16) and (5.17) and using the Cauchy-Schwarz inequality, we get

$$\begin{aligned}
\frac{3}{8}\|\mathbf{E}\|_F^2 &\le \left(\frac{\|\mathbf{E}\|_F}{2}\right)(2\|\mathbf{W}\|_F) + 2\lambda^2\|\mathbf{Q}\|_F^2 \\
&\le \frac{1}{8}\|\mathbf{E}\|_F^2 + 2\|\mathbf{W}\|_F^2 + 2\lambda^2\|\mathbf{Q}\|_F^2.
\end{aligned}$$

As a consequence, we have

$$\begin{aligned}
\|\mathbf{E}\|_F^2 &\le 8\|\mathbf{W}\|_F^2 + 8\lambda^2\|\mathbf{Q}\|_F^2 \\
&= 8\|\mathbf{W}\|_F^2 + 8\lambda^2\langle\mathcal{Q}(f), \mathcal{Q}(f)\rangle \\
&\le 8\|\mathbf{W}\|_F^2 + 8\lambda^2\|\mathcal{Q}(f)\|_1\|\mathcal{Q}(f)\|_\infty \\
&\le 8\|\mathbf{W}\|_F^2 + C\lambda^2\frac{K}{M},
\end{aligned}$$

where the last inequality follows from $\|\mathcal{Q}(f)\|_1 \le C\frac{K}{M}$ as is shown in (A.12). This completes the proof of Lemma 5.8. $\qquad\square$

25

Now, continuing the proof of Lemma 5.7, note that $\|\mathbf{W}/\sigma\|_F^2$ is a χ^2 distribution with MN degrees of freedom. It follows that

$$\mathbb{E}[\|\mathbf{W}\|_F^2] = \sigma^2 MN,$$
$$\mathbb{E}[\|\mathbf{W}\|_F^4] = \sigma^4 MN(MN+2) \leq C\sigma^4 M^2 N^2.$$

Then, with Lemma 5.8, we have

$$\mathbb{E}[\|\mathbf{E}\|_F^4] \leq 64\mathbb{E}[\|\mathbf{W}\|_F^4] + C\mathbb{E}[\|\mathbf{W}\|_F^2]\eta^2\sigma^2 KN\log(M) + C\eta^4\sigma^4 K^2 N^2 \log^2(M)$$
$$\leq C\sigma^4 M^2 N^2 + C\eta^2\sigma^4 MKN^2 \log(M) + C\eta^4\sigma^4 K^2 N^2 \log^2(M)$$
$$\leq C\sigma^4 M^2 N^2 \log^2(M).$$

By taking the square root, we obtain (5.14). This completes the proof of Lemma 5.7. $\qquad\square$

6 Conclusion

ANM can be used for modal analysis under a variety of different spatio-temporal sampling and compression schemes. In the noiseless case, theoretical analysis shows that perfect recovery of mode shapes and frequencies is possible under certain conditions. In particular, a minimum separation condition is required for all of the theorems in Section 3.2: to distinguish any closely spaced pair of frequencies, it is necessary to observe the signals over a long time span, regardless of whether or how the samples are compressed.

While compression does not allow the observation time span to be shortened, it does allow the number of samples to be reduced. Using synchronous random sampling, asynchronous random sampling, and random temporal compression, for example, exact recovery of mode shapes and frequencies is possible when the average number of samples per sensor is roughly proportional to the number of active modes. Using random spatial compression, exact recovery is possible when the total number of compressed measurements scales with the number of degrees of freedom.

Currently, the theoretical results for random sampling and for random spatial compression require randomness assumptions on the mode shapes and, for random sampling, indicate that the performance worsens as the number of sensors increases. Removing and improving these aspects of the results would be worthy of further study. Another open question is to theoretically characterize the performance improvements of asynchronous random sampling over synchronous random sampling.

Damping and external forces may be present in practical scenarios. In this work, we have focused on developing a theoretical foundation for the idealized model of free vibration without damping. Extending the ANM framework and analysis to accommodate damping and forced inputs are interesting questions for future work. We do note, however, that although our theory has focused on systems with no damping, ANM can empirically work well even on systems with slight proportional damping. To show this, we repeat the uniform sampling experiment on the 6-degree-of-freedom system but with damping. For simplicity, we only consider proportional damping of the form $\mathbf{C} = \alpha\mathbf{M} + \beta\mathbf{K}$ with $\alpha = 0.01$ and $\beta = 0.006$. Here, we change the stiffness values to $k_1 = k_7 = 500$, $k_2 = 300$, $k_3 = 100$, $k_4 = 50$, $k_5 = 400$, $k_6 = 200$ N/m in order to slightly increase the minimum separation of true frequencies. The other parameters are the same as in Section 4.1. Similarly, the true mode shapes and undamped natural frequencies of this system are obtained from the (normalized) generalized eigenvectors and square root of the generalized eigenvalues of $\mathbf{K}$ and $\mathbf{M}$. In particular, the true undamped natural frequencies are $F_1 = 0.6239$, $F_2 = 0.9820$, $F_3 = 1.7161$, $F_4 = 1.9365$, $F_5 = 2.3272$, $F_6 = 4.6879$ Hz. The damping ratios of this system are $\xi_1 = 0.0130$, $\xi_2 = 0.0193$, $\xi_3 = 0.0328$, $\xi_4 = 0.0369$, $\xi_5 = 0.0442$, $\xi_6 = 0.0885$, and the true damped natural frequencies are given as $F_{d_k} = F_k\sqrt{1 - \xi_k^2}$. In particular, $F_{d_1} = 0.6239$, $F_{d_2} = 0.9819$, $F_{d_3} = 1.7151$, $F_{d_4} = 1.9351$, $F_{d_5} = 2.3250$, $F_{d_6} = 4.6595$ Hz. The true amplitudes are set as $A_1 = 1$, $A_2 = 0.85$, $A_3 = 0.7$, $A_4 = 0.5$, $A_5 = 0.25$, $A_6 = 0.7$. In order to make the experiment as close to realistic as possible, we first collect 300 real-valued uniform samples from this system with sampling interval $T_s = 0.5/F_c$, where $F_c = \max_{1 \leq k \leq 6} |F_k|$. We then compute the Hilbert transform of
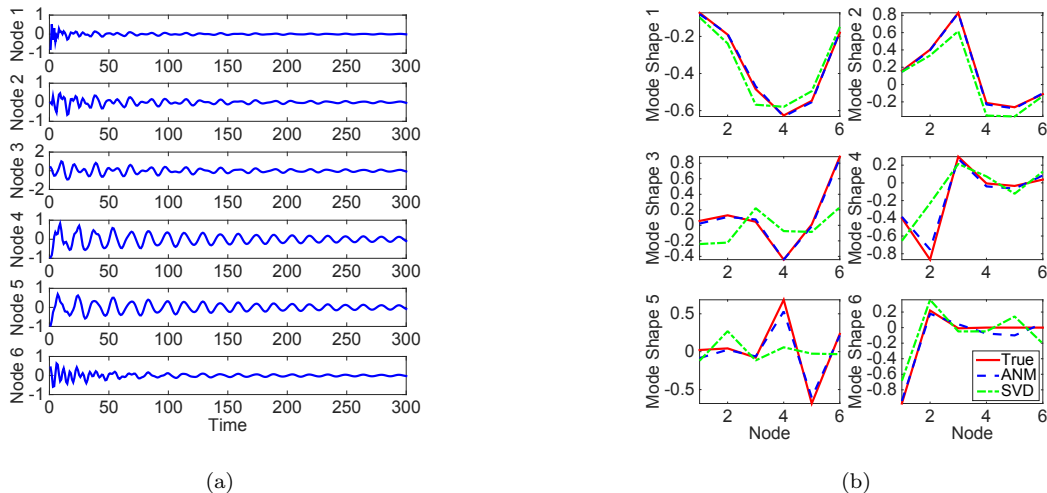
Figure 7: Uniform sampling in the boxcar system with damping. (a) Uniform samples from each node (sensor). (b) Estimated mode shapes by ANM and SVD.

these real-valued samples to obtain the analytic samples. To eliminate border effects, we perform ANM only on the first 100 analytical samples.

The damped real-valued uniform samples and the reconstructed mode shapes are presented in Figure 7. Although the input signals contain damping, we use the conventional undamped version of ANM to estimate the frequencies and the mode shapes. It can be seen that ANM still performs much better than SVD in general since the true mode shapes are not mutually orthogonal. In particular, the MAC for ANM is (0.9996, 0.9986, 0.9825, 0.9129, 0.8452, 0.9579), while the MAC for SVD is (0.9882, 0.9380, 0.8726, 0.7288, 0.9435, 0.7547). In addition, the estimated frequencies are $\hat{F}_1 = 0.6245$, $\hat{F}_2 = 0.9824$, $\hat{F}_3 = 1.7205$, $\hat{F}_4 = 1.9218$, $\hat{F}_5 = 2.3122$, $\hat{F}_6 = 4.5608$ Hz. Although the ANM algorithm returns reasonably accurate estimates, because it does not explicitly account for damping, it cannot perfectly recover the mode shapes and frequencies. We leave the development of such a damped ANM algorithm to future work.

Finally, in the noisy case, we have extended the SMV atomic norm denoising problem to an MMV atomic norm denoising problem and derived non-asymptotic theoretical bounds for the recovery error. While this analysis may be of its own independent interest, it is also used in our proof of Theorem 3.4.

7   Acknowledgements

References

[1] A. Cunha and E. Caetano, "Experimental modal analysis of civil engineering structures," *Sound and Vibration*, vol. 40, no. 6, pp. 12–20, 2006.

[2] D. C. Kammer, "Sensor placement for on-orbit modal identification and correlation of large space structures," *Journal of Guidance, Control, and Dynamics*, vol. 14, no. 2, pp. 251–259, 1991.

[3] K. D. Marshall, "Modal analysis of a violin," *The Journal of the Acoustical Society of America*, vol. 77, no. 2, pp. 695–709, 1985.

[4] S. O'Connor, J. Lynch, and A. Gilbert, "Compressed sensing embedded in an operational wireless sensor network to achieve energy efficiency in long-term monitoring applications," *Smart Materials and Structures*, vol. 23, no. 8, p. 085014, 2014.

[5] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[6] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[7] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1257–1269, 2016.

[8] Z. Yang and L. Xie, "Continuous compressed sensing with a single or multiple measurement vectors," in *IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 288–291, 2014.

[9] R. Heckel and M. Soltanolkotabi, "Generalized line spectral estimation via convex optimization," *arXiv preprint arXiv:1609.08198*, 2016.

[10] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5145–5157, 2014.

[11] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, 2013.

[12] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[13] M. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, 2010.

[14] Z. Yang, J. Li, P. Stoica, and L. Xie, "Sparse methods for direction-of-arrival estimation," *arXiv preprint arXiv:1609.09596*, 2016.

[15] R. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation.," Ph.D. dissertation, Stanford University, 1981.

[16] D. Yang, G. Tang, and M. B. Wakin, "Super-resolution of complex exponentials from modulations with unknown waveforms," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5809–5830, 2016.

[17] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.

[18] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 499–512, 2015.

[19] D. J. Ewins, *Modal testing: Theory and practice*, vol. 6. Research studies press Letchworth, 1995.

[20] J. Y. Park, M. B. Wakin, and A. C. Gilbert, "Modal analysis with compressive measurements," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1655–1670, 2014.

[21] D. Mead and S. Markus, "The forced vibration of a three-layer, damped sandwich beam with arbitrary boundary conditions," *Journal of Sound and Vibration*, vol. 10, no. 2, pp. 163–175, 1969.

[22] Y. Yamakoshi, J. Sato, and T. Sato, "Ultrasonic imaging of internal vibration of soft tissue under forced vibration," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 37, no. 2, pp. 45–53, 1990.

[23] O. S. Salawu and C. Williams, "Bridge assessment using forced-vibration testing," *Journal of Structural Engineering*, 1995.

[24] D. S. Cho, B. H. Kim, J.-H. Kim, N. Vladimir, and T. M. Choi, "Forced vibration analysis of arbitrarily constrained rectangular plates and stiffened panels using the assumed mode method," *Thin-Walled Structures*, vol. 90, pp. 182–190, 2015.

[25] D. Rade and V. Steffen Jr, "Structural dynamics and modal analysis," *EOLSS–Encyclopedia of Life Support*, 2008.

[26] Y. Yang and S. Nagarajaiah, "Output-only modal identification with limited sensors using sparse component analysis," *Journal of Sound and Vibration*, vol. 332, no. 19, pp. 4741–4765, 2013.

[27] K. Yu, K. Yang, and Y. Bai, "Estimation of modal parameters using the sparse component analysis based underdetermined blind source separation," *Mechanical Systems and Signal Processing*, vol. 45, no. 2, pp. 302–316, 2014.

[28] A. Sadhu, B. Hazra, and S. Narasimhan, "Decentralized modal identification of structures using parallel factor decomposition and sparse blind source separation," *Mechanical Systems and Signal Processing*, vol. 41, no. 1, pp. 396–419, 2013.

[29] Y. Yang and S. Nagarajaiah, "Time-frequency blind source separation using independent component analysis for output-only modal identification of highly damped structures," *Journal of Structural Engineering*, vol. 139, no. 10, pp. 1780–1793, 2012.

[30] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[31] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

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

[33] B. 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.

[34] Y. Yang, C. Dorn, T. Mancini, Z. Talken, G. Kenyon, C. Farrar, and D. Mascareñas, "Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification," *Mechanical Systems and Signal Processing*, vol. 85, pp. 567–590, 2017.

[35] Y. Yang, C. Dorn, T. Mancini, Z. Talken, S. Nagarajaiah, G. Kenyon, C. Farrar, and D. Mascareñas, "Blind identification of full-field vibration modes of output-only structures from uniformly-sampled, possibly temporally-aliased (sub-nyquist), video measurements," *Journal of Sound and Vibration*, vol. 390, pp. 232–256, 2017.

[36] Y. Yang and S. Nagarajaiah, "Output-only modal identification by compressed sensing: Non-uniform low-rate random sampling," *Mechanical Systems and Signal Processing*, vol. 56, pp. 15–34, 2015.

[37] Z. Lu, R. Ying, S. Jiang, P. Liu, and W. Yu, "Distributed compressed sensing off the grid," *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 105–109, 2015.

[38] B. Boashash, *Time-frequency signal analysis and processing: a comprehensive reference.* Academic Press, 2015.

[39] J. O. Smith III, *Spectral audio signal processing.* W3K publishing, 2011.

[40] M. López Aenlle, R. Brincker, and A. C. Fernández Canteli, "Some methods to determine scaled mode shapes in natural input modal analysis," in *Int. Modal Analysis Conference (IMAC)*, 2005.

[41] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," 2008.

[42] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 995–1006, 2016.

[43] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," *arXiv preprint arXiv:0901.3403*, 2009.

[44] Z. Yang, J. Tang, Y. C. Eldar, and L. Xie, "On the sample complexity of multichannel frequency estimation via convex optimization," *arXiv preprint arXiv:1712.05674*, 2017.

[45] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proceedings of the 5th International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 134–142, ACM, 2006.

[46] A. Eftekhari, H. L. Yap, M. B. Wakin, and C. J. Rozell, "Stabilizing embedology: Geometry-preserving delay-coordinate maps," *arXiv preprint arXiv:1609.06347*, 2016.

[47] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge University Press, 2004.

[48] Y. Gordon, "On Milman's inequality and random subspaces which escape through a mesh in R^n," in *Geometric Aspects of Functional Analysis*, pp. 84–106, Springer, 1988.

[49] G. Pisier, "Probabilistic methods in the geometry of Banach spaces," in *Probability and Analysis*, pp. 167–241, Springer, 1986.

[50] S. Oymak and B. Hassibi, "Sharp mse bounds for proximal denoising," *Foundations of Computational Mathematics*, vol. 16, no. 4, pp. 965–1029, 2016.

[51] A. Schaeffer, "Inequalities of A. Markoff and S. Bernstein for polynomials and related functions," *Bulletin of the American Mathematical Society*, vol. 47, no. 8, pp. 565–579, 1941.

[52] S. Li, D. Yang, and M. B. Wakin, "Atomic norm minimization for modal analysis," in *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2016.

[53] S. Li, D. Yang, and M. Wakin, "Atomic norm minimization for modal analysis with random spatial compression," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[54] J. Heinonen, *Lectures on Lipschitz analysis.* Jyväskylä Summer School, 2005.

[55] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[56] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.

A   Appendix

A.1   Proof of Theorem 5.1

This proof is modified from [48]. For $\mathbf{Z} \in \Omega \subset \mathbb{S}^{MN-1}$, $\tilde{\mathbf{Z}} \in \mathbb{S}^{M'N-1} \triangleq \{\tilde{\mathbf{Z}} \in \mathbb{C}^{M' \times N} : \|\tilde{\mathbf{Z}}\|_F = 1\}$, define two Gaussian processes

$$W_{\mathbf{Z},\tilde{\mathbf{Z}}} = \langle \mathbf{G}, \tilde{\mathbf{Z}} \rangle + \langle \mathbf{H}, \mathbf{Z} \rangle, \ \tilde{W}_{\mathbf{Z},\tilde{\mathbf{Z}}} = \langle \mathbf{\Phi}\hat{\mathbf{Z}}, \tilde{\mathbf{Z}} \rangle.$$

Here, $\mathbf{G}$ and $\mathbf{H}$ are random matrices with i.i.d. entries from the distribution $\mathcal{N}(0,1)$. For all $\mathbf{Z}$, $\mathbf{Z}' \in \Omega$, $\tilde{\mathbf{Z}}$, $\tilde{\mathbf{Z}}' \in \mathbb{S}^{M'N-1}$, it can be shown that

$$\mathbb{E}|W_{\mathbf{Z},\tilde{\mathbf{Z}}} - W_{\mathbf{Z}',\tilde{\mathbf{Z}}'}|^2 - \mathbb{E}|\tilde{W}_{\mathbf{Z},\tilde{\mathbf{Z}}} - \tilde{W}_{\mathbf{Z}',\tilde{\mathbf{Z}}'}|^2$$

$$=4 - \langle \tilde{\mathbf{Z}}', \tilde{\mathbf{Z}} \rangle - \langle \mathbf{Z}', \mathbf{Z} \rangle - \langle \tilde{\mathbf{Z}}, \tilde{\mathbf{Z}}' \rangle - \langle \mathbf{Z}, \mathbf{Z}' \rangle - \sum_{n=1}^{N} \langle z_n, z_n \rangle \langle \tilde{z}_n, \tilde{z}_n \rangle - \sum_{n=1}^{N} \langle z_n', z_n' \rangle \langle \tilde{z}_n', \tilde{z}_n' \rangle$$

$$+ \sum_{n=1}^{N} \langle z_n, z_n' \rangle \langle \tilde{z}_n', \tilde{z}_n \rangle + \sum_{n=1}^{N} \langle z_n', z_n \rangle \langle \tilde{z}_n, \tilde{z}_n' \rangle$$

$$=4 - 2\operatorname{Re}(\langle \tilde{\mathbf{Z}}', \tilde{\mathbf{Z}} \rangle) - 2\operatorname{Re}(\langle \mathbf{Z}, \mathbf{Z}' \rangle) - \sum_{n=1}^{N} \left(\|z_n\|^2 \|\tilde{z}_n\|^2 + \|z_n'\|^2 \|\tilde{z}_n'\|^2 - 2\operatorname{Re}(\langle z_n, z_n' \rangle \langle \tilde{z}_n', \tilde{z}_n \rangle) \right)$$

$$\geq \operatorname{Re} \left(4 - 2\langle \tilde{\mathbf{Z}}', \tilde{\mathbf{Z}} \rangle - 2\langle \mathbf{Z}, \mathbf{Z}' \rangle - \sum_{n=1}^{N} \sum_{m=1}^{N} \left(\|z_n\|^2 \|\tilde{z}_m\|^2 + \|z_n'\|^2 \|\tilde{z}_m'\|^2 - 2\langle z_n, z_n' \rangle \langle \tilde{z}_m', \tilde{z}_m \rangle \right) \right)$$

$$= \operatorname{Re} \left(2 - 2\langle \tilde{\mathbf{Z}}', \tilde{\mathbf{Z}} \rangle - 2\langle \mathbf{Z}, \mathbf{Z}' \rangle + 2\langle \mathbf{Z}, \mathbf{Z}' \rangle \langle \tilde{\mathbf{Z}}', \tilde{\mathbf{Z}} \rangle \right)$$

$$=2 \operatorname{Re} \left((1 - \langle \tilde{\mathbf{Z}}', \tilde{\mathbf{Z}} \rangle)(1 - \langle \mathbf{Z}, \mathbf{Z}' \rangle) \right)$$

$$\geq 0.$$

The first inequality holds since

$$\|z_n\|^2 \|\tilde{z}_m\|^2 + \|z_n'\|^2 \|\tilde{z}_m'\|^2 - 2\langle z_n, z_n' \rangle \langle \tilde{z}_m', \tilde{z}_m \rangle$$

$$\geq \|z_n\|^2 \|\tilde{z}_m\|^2 + \|z_n'\|^2 \|\tilde{z}_m'\|^2 - 2\|z_n\| \|\tilde{z}_m\| \|z_n'\| \|\tilde{z}_m'\| \geq 0.$$

Then, by [48] we have

$$\mathbb{E} \left[\min_{\mathbf{Z} \in \Omega} \max_{\tilde{\mathbf{Z}} \in \mathbb{S}^{M'N-1}} W_{\mathbf{Z},\tilde{\mathbf{Z}}} \right] \leq \mathbb{E} \left[\min_{\mathbf{Z} \in \Omega} \max_{\tilde{\mathbf{Z}} \in \mathbb{S}^{M'N-1}} \tilde{W}_{\mathbf{Z},\tilde{\mathbf{Z}}} \right],$$

which is equivalent to

$$\mathbb{E} \left[\min_{\mathbf{Z} \in \Omega} \max_{\tilde{\mathbf{Z}} \in \mathbb{S}^{M'N-1}} \langle \mathbf{G}, \tilde{\mathbf{Z}} \rangle + \langle \mathbf{H}, \mathbf{Z} \rangle \right] \leq \mathbb{E} \left[\min_{\mathbf{Z} \in \Omega} \max_{\tilde{\mathbf{Z}} \in \mathbb{S}^{M'N-1}} \langle \mathbf{\Phi}\hat{\mathbf{Z}}, \tilde{\mathbf{Z}} \rangle \right]. \tag{A.1}$$

Since $\|\tilde{\mathbf{Z}}\|_F = 1$, using Cauchy-Schwarz inequality, we have

$$\langle \mathbf{\Phi}\hat{\mathbf{Z}}, \tilde{\mathbf{Z}} \rangle \leq \|\mathbf{\Phi}\hat{\mathbf{Z}}\|_F. \tag{A.2}$$

Plugging (A.2) into (A.1) gives

$$
\mathbb{E}\left[\min_{\mathbf{Z}\in\Omega}\|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F\right] \geq \mathbb{E}\left[\min_{\mathbf{Z}\in\Omega}\max_{\tilde{\mathbf{Z}}\in\mathbb{S}^{M'N-1}}\langle\boldsymbol{\Phi}\hat{\mathbf{Z}},\tilde{\mathbf{Z}}\rangle\right]
$$

$$
\geq \mathbb{E}\left[\min_{\mathbf{Z}\in\Omega}\max_{\tilde{\mathbf{Z}}\in\mathbb{S}^{M'N-1}}\langle\mathbf{G},\tilde{\mathbf{Z}}\rangle + \langle\mathbf{H},\mathbf{Z}\rangle\right]
$$

$$
= \mathbb{E}_{\mathbf{G}}\|\mathbf{G}\|_F + \mathbb{E}_{\mathbf{H}}\left[\min_{\mathbf{Z}\in\Omega}\langle\mathbf{H},\mathbf{Z}\rangle\right]
$$

$$
\geq \mathbb{E}_{\mathbf{G}}\|\mathbf{G}\|_F - \mathbb{E}_{\mathbf{H}}\left[\max_{\mathbf{Z}\in\Omega}\langle\mathbf{H},\mathbf{Z}\rangle\right]
$$

$$
= \lambda_{M'N} - \omega(\Omega).
$$

Moreover, we have

$$
\lambda_{M'N} = \mathbb{E}_{\mathbf{G}}\|\mathbf{G}\|_F = \mathbb{E}\sqrt{\sum_{n=1}^{N}\|\boldsymbol{g}_n\|_2^2}
$$

$$
\geq \mathbb{E}\sqrt{N\min_{n\in[1,N]}\|\boldsymbol{g}_n\|_2^2} = \sqrt{N}\mathbb{E}\sqrt{\min_{n\in[1,N]}\|\boldsymbol{g}_n\|_2^2}
$$

$$
= \sqrt{N}\mathbb{E}\left[\min_{n\in[1,N]}\|\boldsymbol{g}_n\|_2\right]
$$

$$
\geq \frac{\sqrt{N}M'}{\sqrt{M'+1}},
$$

where the last inequality uses the result that the expected length of an M'-dimensional Gaussian random vector is lower bounded by $\frac{M'}{\sqrt{M'+1}}$ [5], [48]. $\qquad\square$

A.2 Proof of Corollary 5.1

First, we will show that the following function

$$
f : \boldsymbol{\Phi} \to \min_{\mathbf{Z}\in\Omega}\|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F
$$

is Lipschitz with respect to the Frobenius norm with constant 1.

Define a function $f_{\mathbf{Z}}(\boldsymbol{\Phi}) = \|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F$ with $\mathbf{Z} = [\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_N]$ and $\hat{\mathbf{Z}} = \mathrm{diag}(\mathbf{Z})$. The gradient of $f_{\mathbf{Z}}(\boldsymbol{\Phi})$ is

$$
\nabla_{\boldsymbol{\Phi}}f_{\mathbf{Z}}(\boldsymbol{\Phi}) = \frac{\boldsymbol{\Phi}\hat{\mathbf{Z}}\hat{\mathbf{Z}}^*}{\|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F}.
$$

Thus, using the mean value theorem, we can get

$$
|f_{\mathbf{Z}}(\boldsymbol{\Phi}_1) - f_{\mathbf{Z}}(\boldsymbol{\Phi}_2)| = |\langle\nabla_{\boldsymbol{\Phi}}f_{\mathbf{Z}}(\boldsymbol{\Phi}),\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\rangle| \leq \|\nabla_{\boldsymbol{\Phi}}f_{\mathbf{Z}}(\boldsymbol{\Phi})\|_F\|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\|_F
$$

$$
= \frac{1}{\|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F}\|\boldsymbol{\Phi}\hat{\mathbf{Z}}\hat{\mathbf{Z}}^*\|_F\|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\|_F
$$

$$
\leq \|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\|_F.
$$

So, the function $f_{\mathbf{Z}}(\boldsymbol{\Phi}) = \|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F$ is Lipschitz with respect to the Frobenius norm with constant 1. According to Lemma 2.1 in [54], we can conclude that $f : \boldsymbol{\Phi} \to \min_{\mathbf{Z}\in\Omega}\|\boldsymbol{\Phi}\hat{\mathbf{Z}}\|_F$ is also Lipschitz with respect to the Frobenius norm with constant 1.

The Gaussian concentration inequality for Lipschitz functions provided in [49] implies that

$$\mathbb{P}\left\{\min_{\mathbf{Z}\in\Omega}\|\mathbf{\Phi}\hat{\mathbf{Z}}\|_F \geq \mathbb{E}\left[\min_{\mathbf{Z}\in\Omega}\|\mathbf{\Phi}\hat{\mathbf{Z}}\|_F\right] - t\right\} \geq 1 - e^{-\frac{t^2}{2}}$$

holds for any $t > 0$. Using Theorem 5.1, we can get

$$\mathbb{P}\left\{\min_{\mathbf{Z}\in\Omega}\|\mathbf{\Phi}\hat{\mathbf{Z}}\|_F \geq \lambda_{M'N} - \omega(\Omega) - t\right\} \geq 1 - e^{-\frac{t^2}{2}}.$$

Thus, we can choose M' such that

$$\lambda_{M'N} - \omega(\Omega) - t > 0,$$

which is equivalent to

$$\lambda_{M'N}^2 > (\omega(\Omega) + t)^2 \triangleq \epsilon^2.$$

So, we can let

$$\frac{NM'^2}{M'+1} > \epsilon^2,$$

which means that we need

$$M' > \frac{\epsilon^2 + \sqrt{\epsilon^4 + 4N\epsilon^2}}{2N}. \tag{A.3}$$

Thus, we can choose

$$M' \geq \frac{\epsilon^2 + \sqrt{\epsilon^4 + 4N\epsilon^2 + 4N^2}}{2N} = \frac{1}{N}\epsilon^2 + 1 = \frac{1}{N}(\omega(\Omega) + t)^2 + 1$$

to satisfy the above condition (A.3). Moreover, if $M' \geq \frac{4}{N}\omega^2(\Omega)$, we have

$$M' = \frac{M'}{2} + \frac{M'-2}{2} + 1 \geq \frac{2}{N}\omega^2(\Omega) + 2\left(\frac{\sqrt{M'-2}}{2}\right)^2 + 1$$

$$\geq \left(\frac{\omega(\Omega)}{\sqrt{N}} + \frac{\sqrt{N(M'-2)}}{2\sqrt{N}}\right)^2 + 1$$

$$= \frac{1}{N}\left[\omega(\Omega) + \frac{\sqrt{N(M'-2)}}{2}\right]^2 + 1.$$

Taking $t = \frac{\sqrt{N(M'-2)}}{2}$, the proof for Corollary 5.1 is finished. □

A.3 Proof of key lemmas in Section 5.3.2

To prove these lemmas, we need the following theorem, which is an extension of Theorem 4 in [18].

Theorem A.1. *Let $\mathcal{B} = \{\boldsymbol{b} \in \mathbb{C}^N : \|\boldsymbol{b}\|_2 = 1\}$ be a set of vectors with unit ℓ_2 norm. For any $f_1, f_2, \ldots, f_K$ satisfying the minimum separation condition (3.10) , there exists a vector-valued trigonometric polynomial $\mathcal{Q}(f) = \mathbf{Q}^*\boldsymbol{a}(f)$ satisfying the following properties for some $\mathbf{Q} \in \mathbb{C}^{M \times N}$.*

  1. For each $k = 1, \ldots, K$, $\mathcal{Q}(f_k) = \boldsymbol{b}_k$ with $\boldsymbol{b}_k \in \mathcal{B}$.

2. In each near region $N_k = \{f : d(f, f_k) < 0.16/M\}$, there exist constants C_a and C'_a such that

$$\|\mathcal{Q}(f)\|_2 \leq 1 - \frac{C_a}{2} M^2 (f - f_k)^2 \tag{A.4}$$

$$\|\boldsymbol{b}_k - \mathcal{Q}(f)\|_2 \leq \frac{C'_a}{2} M^2 (f - f_k)^2. \tag{A.5}$$

3. In the far region, i.e., $f \in F = [0, 1)/ \cap_{k=1}^K N_j$, there exists a constant $C_b > 0$ such that

$$\|\mathcal{Q}(f)\|_2 \leq 1 - C_b. \tag{A.6}$$

Proof. Proposition 5.1 and inequality (47) in [10] guarantee the first and third statements. The second statement can be proved by following the proof strategies in Section 2.4 of [55] and Appendix A of [56]. $\square$

In the remainder of this section, we will prove the three key lemmas presented in Section 5.3.2.

A.3.1 Proof of Lemma 5.4

We define $\mathcal{F} \triangleq \{f_1, f_2, \ldots, f_K\}$ as a set containing the true frequencies. Recall that the near region N_k corresponding to f_k and the far region F are defined in (5.8). Using the Cauchy-Schwarz inequality, we have

$$\left| \int_F \boldsymbol{\xi}^*(f) \boldsymbol{\nu}(f) df \right| \leq \int_F |\boldsymbol{\xi}^*(f) \boldsymbol{\nu}(f)| df \leq \int_F \|\boldsymbol{\xi}(f)\|_2 \|\boldsymbol{\nu}(f)\|_2 df \leq \|\boldsymbol{\xi}(f)\|_{2,\infty} \int_F \|\boldsymbol{\nu}(f)\|_2 df, \tag{A.7}$$

where $\|\boldsymbol{\xi}(f)\|_{2,\infty}$ is defined in (5.9).

Let $\boldsymbol{u}$ be any vector with unit ℓ_2 norm. Define

$$\gamma(f) \triangleq \boldsymbol{u}^* \boldsymbol{\xi}(f),$$

which is a trigonometric polynomial with degree M. According to Bernstein's inequality for polynomials [51], we have

$$\sup_f |\gamma'(f)| \leq M \sup_f |\gamma(f)|, \tag{A.8}$$

$$\sup_f |\gamma''(f)| \leq M^2 \sup_f |\gamma(f)|. \tag{A.9}$$

Note that

$$\sup_f \|\boldsymbol{\xi}'(f)\|_2 = \sup_{f,\boldsymbol{u}} |\boldsymbol{u}^* \boldsymbol{\xi}'(f)| = \sup_{f,\boldsymbol{u}} |\gamma'(f)|$$

$$\leq M \sup_{f,\boldsymbol{u}} |\gamma(f)| = M \sup_{f,\boldsymbol{u}} |\boldsymbol{u}^* \boldsymbol{\xi}(f)| = M \sup_f \|\boldsymbol{\xi}(f)\|_2,$$

which implies

$$\|\boldsymbol{\xi}'(f)\|_{2,\infty} \leq M \|\boldsymbol{\xi}(f)\|_{2,\infty}.$$

Using a similar argument, we can get

$$\|\boldsymbol{\xi}''(f)\|_\infty \leq M^2 \|\boldsymbol{\xi}(f)\|_{2,\infty}.$$

Note that the Taylor expansion of $\gamma(f)$ at f_k is

$$\gamma(f) = \gamma(f_k) + (f - f_k)\gamma'(f_k) + \frac{1}{2}(f - f_k)^2 \gamma''(\tilde{f}_k)$$

for some $\tilde{f}_k \in N_k$. It follows that

$$\sup_{\boldsymbol{u}} |\gamma(f) - \gamma(f_k) - (f - f_k)\gamma'(f_k)|$$

$$= \frac{1}{2}(f - f_k)^2 \sup_{\boldsymbol{u}} |\gamma''(\tilde{f}_k)|$$

$$= \frac{1}{2}(f - f_k)^2 \sup_{\boldsymbol{u}} |\boldsymbol{u}^* \boldsymbol{\xi}''(\tilde{f}_k)|$$

$$\leq \frac{1}{2}(f - f_k)^2 \|\boldsymbol{\xi}''(f)\|_{2,\infty}.$$

Plugging in Bernstein's inequality (A.9), we have

$$\sup_{\boldsymbol{u}} |\gamma(f) - \gamma(f_k) - (f - f_k)\gamma'(f_k)| \leq \frac{M^2}{2}(f - f_k)^2 \|\boldsymbol{\xi}(f)\|_{2,\infty}.$$

Note that the left hand side of the above inequality is equal to

$$\sup_{\boldsymbol{u}} |\langle \boldsymbol{\xi}(f) - \boldsymbol{\xi}(f_k) - (f - f_k)\boldsymbol{\xi}'(f_k), \boldsymbol{u} \rangle|$$

$$= \|\boldsymbol{\xi}(f) - \boldsymbol{\xi}(f_k) - (f - f_k)\boldsymbol{\xi}'(f_k)\|_2.$$

Denoting $\boldsymbol{r}(f) = \boldsymbol{\xi}(f) - \boldsymbol{\xi}(f_k) - (f - f_k)\boldsymbol{\xi}'(f_k)$, we have

$$\|\boldsymbol{r}(f)\|_2 \leq \frac{M^2}{2}(f - f_k)^2 \|\boldsymbol{\xi}(f)\|_{2,\infty}.$$

Finally, we can obtain

$$\left\| \int_{N_k} \boldsymbol{\xi}^*(f)\boldsymbol{\nu}(f)df \right\|$$

$$\leq \|\boldsymbol{\xi}(f_k)\|_2 \left\| \int_{N_k} \boldsymbol{\nu}(f)df \right\|_2 + \|\boldsymbol{\xi}'(f_k)\|_2 \left\| \int_{N_k} (f - f_k)\boldsymbol{\nu}(f)df \right\|_2 + \int_{N_k} \|\boldsymbol{r}(f)\|_2 \|\boldsymbol{\nu}(f)\|_2 df \qquad (A.10)$$

$$\leq \|\boldsymbol{\xi}(f)\|_{2,\infty} \left(\left\| \int_{N_k} \boldsymbol{\nu}(f)df \right\|_2 + M \left\| \int_{N_k} (f - f_k)\boldsymbol{\nu}(f)df \right\|_2 + \frac{M^2}{2} \int_{N_k} (f - f_k)^2 \|\boldsymbol{\nu}_n(f)\|df \right).$$

We have plugged in $\boldsymbol{\xi}(f) = \boldsymbol{\xi}(f_k) + (f - f_k)\boldsymbol{\xi}'(f_k) + \boldsymbol{r}(f)$ and used the triangle inequality to get the first inequality. Substituting (A.7) and (A.10) into (5.7), we can obtain (5.10). $\qquad \square$

A.3.2   Proof of Lemma 5.5

Let $\mathcal{Q}(f) = \mathbf{Q}^* \boldsymbol{a}(f)$ be the dual polynomial as in [10]. Then, we have

$$|\langle \boldsymbol{\xi}(f), \mathcal{Q}(f) \rangle| \leq \int_0^1 |\mathcal{Q}^*(f)\boldsymbol{\xi}(f)|df \leq \int_0^1 \|\mathcal{Q}(f)\|_2 \|\boldsymbol{\xi}(f)\|_2 df \leq \|\boldsymbol{\xi}(f)\|_{2,\infty} \|\mathcal{Q}(f)\|_1, \qquad (A.11)$$

where $\|\mathcal{Q}(f)\|_1 \triangleq \int_0^1 \|\mathcal{Q}(f)\|_2 df$. The first and second inequalities follow from the triangle inequality and the Cauchy-Schwarz inequality, respectively. In this work, we use the dual polynomial $\mathcal{Q}(f)$ constructed in [10] of the form

$$\mathcal{Q}(f) = \sum_{f_k \in \mathcal{F}} \boldsymbol{\alpha}_k \mathcal{K}(f - f_k) + \sum_{f_k \in \mathcal{F}} \boldsymbol{\beta}_k \mathcal{K}'(f - f_k),$$

where $\mathcal{K}(f)$ is the squared Fejér kernel. All the $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ are $N \times 1$ vectors. Define $\|\mathbf{X}\|_{2,\infty} = \max_j \|\mathbf{X}_j\|_2$, with $\mathbf{X}_j$ being the jth column of $\mathbf{X}$. Denote $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_K]$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_K]$. It is shown in [10] that

$$\|\boldsymbol{\alpha}\|_{2,\infty} \leq C_\alpha, \quad \|\boldsymbol{\beta}\|_{2,\infty} \leq \frac{C_\beta}{M}$$

for some numerical constants C_α and C_β. Then, we have

$$
\begin{aligned}
\|\mathcal{Q}(f)\|_1 &\triangleq \int_0^1 \|\mathcal{Q}(f)\|_2 df \\
&\leq \int_0^1 \sum_{f_k \in \mathcal{F}} \|\boldsymbol{\alpha}_k\|_2 |\mathcal{K}(f - f_k)| df + \int_0^1 \sum_{f_k \in \mathcal{F}} \|\boldsymbol{\beta}_k\|_2 |\mathcal{K}'(f - f_k)| df \\
&\leq K C_\alpha \int_0^1 |\mathcal{K}(f)| df + \frac{C_\beta K}{M} \int_0^1 |\mathcal{K}'(f)| df \\
&\leq \frac{CK}{M}.
\end{aligned}
\tag{A.12}
$$

For the last inequality, we have used the results $\int_0^1 |\mathcal{K}(f)| df \leq \frac{C}{M}$ and $\int_0^1 |\mathcal{K}'(f)| df \leq C$, which can be found in Appendix C of [18].

Now, we can bound I_0 as

$$
\begin{aligned}
I_0 &= \sum_{k=1}^K \left\| \int_{N_k} \boldsymbol{\nu}(f) df \right\|_2 = \sum_{k=1}^K \int_{N_k} \boldsymbol{\nu}^*(f) df \frac{\int_{N_k} \boldsymbol{\nu}(\hat{f}) d\hat{f}}{\left\| \int_{N_k} \boldsymbol{\nu}(\hat{f}) d\hat{f} \right\|_2} \\
&= \sum_{k=1}^K \int_{N_k} \boldsymbol{\nu}^*(f) \frac{\int_{N_k} \boldsymbol{\nu}(\hat{f}) d\hat{f}}{\left\| \int_{N_k} \boldsymbol{\nu}(\hat{f}) d\hat{f} \right\|_2} df \\
&= \sum_{k=1}^K \int_{N_k} \boldsymbol{\nu}^*(f) \mathcal{Q}(f) df + \sum_{k=1}^K \int_{N_k} \boldsymbol{\nu}^*(f) \left[\frac{\int_{N_k} \boldsymbol{\nu}(\hat{f}) d\hat{f}}{\left\| \int_{N_k} \boldsymbol{\nu}(\hat{f}) d\hat{f} \right\|_2} - \mathcal{Q}(f) \right] df \\
&\leq \left| \int_0^1 \boldsymbol{\nu}^*(f) \mathcal{Q}(f) df \right| + \left| \int_F \boldsymbol{\nu}^*(f) \mathcal{Q}(f) df \right| + \sum_{k=1}^K \int_{N_k} \left| \boldsymbol{\nu}^*(f) \left[\frac{\int_{N_k} \boldsymbol{\nu}(\hat{f}) d\hat{f}}{\left\| \int_{N_k} \boldsymbol{\nu}(\hat{f}) d\hat{f} \right\|_2} - \mathcal{Q}(f) \right] \right| df \\
&\leq \left| \int_0^1 \boldsymbol{\nu}^*(f) \mathcal{Q}(f) df \right| + \int_F \|\boldsymbol{\nu}(f)\|_2 df + C_a' I_2.
\end{aligned}
$$

We have used the triangle inequality and the Cauchy-Schwarz inequality in the last two inequalities. We also use the fact that $\|\mathcal{Q}(f)\|_2 \leq 1$ and set $\boldsymbol{b}_k = \frac{\int_{N_k} \boldsymbol{\nu}(\hat{f}) d\hat{f}}{\left\| \int_{N_k} \boldsymbol{\nu}(\hat{f}) d\hat{f} \right\|_2}$, then use the boundary condition presented in (A.5) to get the last inequality. Note that the first term can be upper bounded with

$$
\begin{aligned}
\left| \int_0^1 \boldsymbol{\nu}^*(f) \mathcal{Q}(f) df \right| &= \left| \int_0^1 \boldsymbol{\nu}^*(f) \mathbf{Q}^* \boldsymbol{a}(f) df \right| = \left| \int_0^1 \langle \boldsymbol{a}(f) \boldsymbol{\nu}^*(f), \mathbf{Q} \rangle df \right| \\
&= \left| \left\langle \int_0^1 \boldsymbol{a}(f) \boldsymbol{\nu}^*(f) df, \mathbf{Q} \right\rangle \right| = |\langle \mathbf{E}, \mathbf{Q} \rangle| = |\langle \boldsymbol{\xi}(f), \mathcal{Q}(f) \rangle| \\
&\leq \frac{CK\lambda}{M},
\end{aligned}
$$

where the last equality follows from Parseval's theorem and the last inequality is a direct result from inequalities (A.11) and (A.12). As a consequence, we have that

$$I_0 \leq C_0 \left(\frac{K\lambda}{M} + I_2 + \int_F \|\boldsymbol{\nu}(f)\|_2 df \right)$$

holds for some numerical constant C_0. Similarly, we can bound I_1 in such a way. □

A.3.3 Proof of Lemma 5.6

Denote $\mathcal{P}_{\mathcal{F}}(\boldsymbol{\nu})$ as the projection of the difference measure $\boldsymbol{\nu}(f)$ on the support set $\mathcal{F} = \{f_1, f_2, \ldots, f_K\}$. Set $\mathcal{Q}(f)$ as the dual polynomial in Theorem A.1. To avoid confusion with the traditional definition of total variation (TV) norm, we use $\|\cdot\|_{2,TV}$ in this work to denote an extension of traditional TV norm. Then, we have

$$\begin{aligned}
\|\mathcal{P}_{\mathcal{F}}(\boldsymbol{\nu})\|_{2,TV} &= \int_0^1 \mathcal{P}_{\mathcal{F}}(\boldsymbol{\nu}^*)\mathcal{Q}(f)df = \int_{\mathcal{F}} \boldsymbol{\nu}^*(f)\mathcal{Q}(f)df \\
&\leq \left| \int_0^1 \boldsymbol{\nu}^*(f)\mathcal{Q}(f)df \right| + \left| \int_{\mathcal{F}^c} \boldsymbol{\nu}^*(f)\mathcal{Q}(f)df \right| \\
&\leq \frac{CK\lambda}{M} + \sum_{f_k \in \mathcal{F}} \left| \int_{N_k/\{f_k\}} \boldsymbol{\nu}^*(f)\mathcal{Q}(f)df \right| + \left| \int_F \boldsymbol{\nu}^*(f)\mathcal{Q}(f)df \right|,
\end{aligned}$$

where $\mathcal{F}^c$ is the complement set of $\mathcal{F}$ on $[0,1)$. Using (A.6) and (A.4), the integration over the far region F can be bounded by

$$\left| \int_F \boldsymbol{\nu}^*(f)\mathcal{Q}(f)df \right| \leq (1 - C_b) \int_F \|\boldsymbol{\nu}(f)\|_2 df,$$

and the integration over $N_k/\{f_k\}$ can be bounded by

$$\begin{aligned}
\left| \int_{N_k/\{f_k\}} \boldsymbol{\nu}^*(f)\mathcal{Q}(f)df \right| &\leq \int_{N_k/\{f_k\}} \|\boldsymbol{\nu}(f)\|_2 \|\mathcal{Q}(f)\|_2 df \\
&\leq \int_{N_k/\{f_k\}} \left(1 - \frac{1}{2}M^2 C_a(f - f_k)^2 \right) \|\boldsymbol{\nu}(f)\|_2 df \\
&\leq \int_{N_k/\{f_k\}} \|\boldsymbol{\nu}(f)\|_2 df - C_a I_2^k.
\end{aligned}$$

Therefore, we can get

$$\begin{aligned}
\|\mathcal{P}_{\mathcal{F}}(\boldsymbol{\nu})\|_{2,TV} &\leq \frac{CK\lambda}{M} + \sum_{f_k \in \mathcal{F}} \int_{N_k/\{f_k\}} \|\boldsymbol{\nu}(f)\|_2 df - C_a I_2 + (1 - C_b) \int_F \|\boldsymbol{\nu}(f)\|_2 df \\
&= \frac{CK\lambda}{M} + \|\mathcal{P}_{\mathcal{F}^c}(\boldsymbol{\nu})\|_{2,TV} - C_a I_2 - C_b \int_F \|\boldsymbol{\nu}(f)\|_2 df,
\end{aligned}$$

which implies

$$\|\mathcal{P}_{\mathcal{F}^c}(\boldsymbol{\nu})\|_{2,TV} - \|\mathcal{P}_{\mathcal{F}}(\boldsymbol{\nu})\|_{2,TV} \geq C_a I_2 + C_b \int_F \|\boldsymbol{\nu}(f)\|_2 df - \frac{CK\lambda}{M}. \tag{A.13}$$

Since $\hat{\mathbf{X}}$ is the solution of the MMV atomic norm denoising problem (3.15), we have

$$\frac{1}{2}\|\mathbf{Y} - \hat{\mathbf{X}}\|_F^2 + \lambda\|\hat{\mathbf{X}}\|_{\mathcal{A}} \leq \frac{1}{2}\|\mathbf{Y} - \mathbf{X}^\star\|_F^2 + \lambda\|\mathbf{X}^\star\|_{\mathcal{A}}.$$

As a consequence, we obtain

$$\lambda \|\hat{\mathbf{X}}\|_{\mathcal{A}} \leq \lambda \|\mathbf{X}^\star\|_{\mathcal{A}} + \langle \mathbf{W}, \hat{\mathbf{X}} - \mathbf{X}^\star \rangle$$

by elementary calculations. It follows that

$$\|\hat{\boldsymbol{\mu}}\|_{2,TV} \leq \|\boldsymbol{\mu}\|_{2,TV} + \frac{|\langle \mathbf{W}, \mathbf{E} \rangle|}{\lambda}.$$

Then, similar to Lemma 5.4, we have

$$
\begin{aligned}
|\langle \mathbf{W}, \mathbf{E} \rangle| &= \left| \left\langle \mathbf{W}, \int_0^1 \boldsymbol{a}(f)\boldsymbol{\nu}^*(f)df \right\rangle \right| = \left| \int_0^1 \langle \mathbf{W}, \boldsymbol{a}(f)\boldsymbol{\nu}^*(f) \rangle \, df \right| \\
&= \left| \int_0^1 \boldsymbol{a}^*(f) \mathbf{W} \boldsymbol{\nu}(f) df \right| \\
&\leq \|\mathbf{W}^* \boldsymbol{a}(f)\|_{2,\infty} \left(\frac{CK\lambda}{M} + I_0 + I_1 + I_2 \right) \\
&\leq C\frac{\lambda}{\eta} \left(\frac{K\lambda}{M} + I_2 + \int_F \|\boldsymbol{\nu}(f)\|_2 df \right).
\end{aligned}
$$

Here, the last inequality follows from $\|\mathbf{W}^* \boldsymbol{a}(f)\|_{2,\infty} = \|\mathbf{W}\|_{\mathcal{A}}^* \leq \frac{\lambda}{\eta}$ and Lemma 5.5. As a consequence, we have

$$
\begin{aligned}
\|\boldsymbol{\mu}\|_{2,TV} + C\eta^{-1} \left(\frac{K\lambda}{M} + I_2 + \int_F \|\boldsymbol{\nu}(f)\|_2 df \right) &\geq \|\hat{\boldsymbol{\mu}}\|_{2,TV} \\
&\geq \|\boldsymbol{\mu} + \boldsymbol{\nu}\|_{2,TV} \geq \|\boldsymbol{\mu}\|_{2,TV} - \|\mathcal{P}_{\mathcal{F}}(\boldsymbol{\nu})\|_{2,TV} + \|\mathcal{P}_{\mathcal{F}^c}(\boldsymbol{\nu})\|_{2,TV},
\end{aligned}
$$

which implies

$$\|\mathcal{P}_{\mathcal{F}^c}(\boldsymbol{\nu})\|_{2,TV} - \|\mathcal{P}_{\mathcal{F}}(\boldsymbol{\nu})\|_{2,TV} \leq C\eta^{-1} \left(\frac{K\lambda}{M} + I_2 + \int_F \|\boldsymbol{\nu}(f)\|_2 df \right). \tag{A.14}$$

Combining (A.13) and (A.14), we get

$$(C_a - C\eta^{-1})I_2 + (C_b - C\eta^{-1}) \int_F \|\boldsymbol{\nu}(f)\|_2 df \leq (1 + \eta^{-1}) \frac{CK\lambda}{M}. \tag{A.15}$$

Thus, Lemma 5.6 is proved under the assumption that η is large enough with respect to the constants appearing in (A.15). □