

Uniform Recovery Bounds for Structured Random Matrices in Corrupted Compressed Sensing

Peng Zhang, Lu Gan, Cong Ling and Sumei Sun

Abstract—We study the problem of recovering an s -sparse signal $\mathbf{x}^* \in \mathbb{C}^n$ from corrupted measurements $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{z}^* + \mathbf{w}$, where $\mathbf{z}^* \in \mathbb{C}^m$ is a k -sparse corruption vector whose nonzero entries may be arbitrarily large and $\mathbf{w} \in \mathbb{C}^m$ is a dense noise with bounded energy. The aim is to exactly and stably recover the sparse signal with tractable optimization programs. In this paper, we prove the *uniform recovery guarantee* of this problem for two classes of structured sensing matrices. The first class can be expressed as the product of a unit-norm tight frame (UTF), a random diagonal matrix and a bounded column-wise orthonormal matrix (e.g., partial random circulant matrix). When the UTF is bounded (i.e. $\mu(\mathbf{U}) \sim 1/\sqrt{m}$), we prove that with high probability, one can recover an s -sparse signal exactly and stably by l_1 minimization programs even if the measurements are corrupted by a sparse vector, provided $m = \mathcal{O}(s \log^2 s \log^2 n)$ and the sparsity level k of the corruption is a constant fraction of the total number of measurements. The second class considers randomly sub-sampled orthonormal matrix (e.g., random Fourier matrix). We prove the uniform recovery guarantee provided that the corruption is sparse on certain sparsifying domain. Numerous simulation results are also presented to verify and complement the theoretical results.

Index Terms—Compressed sensing, corruption, dense noise, unit-norm tight frames.

I. INTRODUCTION

The theory of compressed sensing has been widely studied and applied in various promising applications over the recent years [1]–[5]. It provides an efficient way to recover a sparse signal from a relatively small number of measurements. Specifically, an s -sparse signal $\mathbf{x}^*$ is measured through

$$\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{w}, \quad (1)$$

where $\mathbf{A} \in \mathbb{C}^{m \times n}$ is referred to as the sensing matrix, $\mathbf{y} \in \mathbb{C}^m$ is the measurement vector and $\mathbf{w} \in \mathbb{C}^m$ represents the noise vector with the noise level $\|\mathbf{w}\|_2 \leq \varepsilon$. It has been shown that if $\mathbf{A}$ satisfies the restricted isometry property (RIP) and ε is small, the recovered signal $\hat{\mathbf{x}}$ obtained by l_1 norm minimization is close to the true $\mathbf{x}^*$, i.e. $\|\hat{\mathbf{x}} - \mathbf{x}^*\| \leq C\varepsilon$ with C being a small numerical constant. Many types of sensing matrices have been proven to satisfy the RIP condition. For

example, random Gaussian/Bernoulli matrices satisfy the RIP with high probability if $m \geq \mathcal{O}(s \log(n/s))$ [1], [3], whereas structured sensing matrices consisting of either randomly subsampled orthonormal matrix [6] or modulated unit-norm tight frames [7] have the RIP with high probability when m is about $\mathcal{O}(s \log^4 n)^1$.

This standard compressed sensing problem has been generalized to cope with the recovery of sparse signals when some unknown entries of the measurement vector $\mathbf{y}$ are severely corrupted. Mathematically, we have

$$\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{z}^* + \mathbf{w}, \quad (2)$$

where $\mathbf{z}^* \in \mathbb{C}^m$ is an unknown sparse vector. To reconstruct $\mathbf{x}^*$ from the measurement vector $\mathbf{y}$, the following penalized l_1 norm minimization has been proposed:

$$\min_{\mathbf{x}, \mathbf{z}} \|\mathbf{x}\|_1 + \lambda \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - (\mathbf{A}\mathbf{x} + \mathbf{z})\|_2 \leq \varepsilon. \quad (3)$$

In [10], it was shown that random Gaussian matrices can provide uniform recovery guarantees to this problem (3). In other words, a single random draw of a Gaussian matrix $\mathbf{A}$ is able to stably recover all s -sparse signals $\mathbf{x}^*$ and all k -sparse corruptions $\mathbf{z}^*$ simultaneously with high probability. On the other hand, for structured sensing matrices, the *nonuniform recovery guarantees*² can be proved for randomly subsampled orthonormal matrix [11] and its generalized model - bounded orthonormal systems³ [10]. Very recently, the uniform recovery guarantee for bounded orthonormal systems is shown in [13].

In this paper, we prove the uniform recovery guarantee for two different corrupted sensing models. In the first model, the measurement matrix is based on randomly modulated unit-norm frames [7] and the corruption is sparse on the identity basis. It is noted that the measurement matrix in the first model does not consist of a random subsampling operator, e.g., the partial random circulant matrix [14]. For the second model, we consider

$$\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{H}\mathbf{z}^* + \mathbf{w}, \quad (4)$$

where $\mathbf{A}$ represents a randomly subsampled orthonormal matrix, and the corruption $\mathbf{H}\mathbf{z}^*$ is assumed to be sparse on

¹Recent works [8] [9] for subsampled Fourier matrices show that the factor $\log^4 n$ can be reduced to $\log^3 n$.

²A nonuniform recovery result only states that a fixed pair of sparse signal and sparse corruption can be recovered with high probability using a random draw of the matrix. Sometimes, the signs of the non-zero coefficients of the sparse vector (and corruption) can be chosen at random to further simplify arguments. Uniform recovery is stronger than nonuniform recovery. (see [5, Chapter 9.2] [6, Section 3.1] for more details.)

³See [12] for the construction of the generalized model.

P. Zhang was with the Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, UK (e-mail: p.zhang12@imperial.ac.uk). He is now with the Institute for Infocomm Research, A*STAR, Singapore, 138632, Singapore (e-mail: zhangp@i2r.a-star.edu.sg).

C. Ling is with the Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, UK (e-mail: cling@ieee.org).

L. Gan is with the College of Engineering, Design and Physical Science, Brunel University, London, UB8 3PH, UK (e-mail: lu.gan@brunel.ac.uk).

S. Sun is with the Institute for Infocomm Research, A*STAR, Singapore, 138632, Singapore (e-mail: sunsm@i2r.a-star.edu.sg).

certain bounded domain (e.g., a discrete Fourier transform (DFT) matrix). Our results imply that many structured sensing matrices can be employed in the corrupted sensing model to ensure the exact and stable recovery of both $\mathbf{x}^*$ and $\mathbf{z}^*$, even when the sparsity of the corruption is up to a constant fraction of the total number of measurements. Thanks to the uniform recovery guarantee, our results can address the adversarial setting, which means that exact and stable recovery is still guaranteed even when $\mathbf{x}^*$, $\mathbf{z}^*$ and $\mathbf{w}$ are selected given knowledge of the sensing matrix $\mathbf{A}$. In addition, our analysis results are also applicable to demonstrate the recovery guarantee when the corrupted sensing problem is solved via nonconvex optimization.

A. Potential Applications

The problem of recovering sparse signal $\mathbf{x}^*$ and sparse corruption $\mathbf{z}^*$ from the measurement vector $\mathbf{y}$ arise from many applications, where the compressed measurements may be corrupted by impulse noise.

For example, in a sensor network, each sensor node measures the same signal $\mathbf{x}^*$ independently before sending the outcome to the center hub for analysis. In this setting, each sensor makes the measurement $\langle \mathbf{a}_i, \mathbf{x}^* \rangle$, and the resultant measurement vector is $\mathbf{A}\mathbf{x}^*$ by arranging each $\mathbf{a}_i$ as the rows of $\mathbf{A}$ [11], [15]. However, in practice, some sensor readings can be anomalous from the rest. These outliers could be caused by individually malfunctioned sensors, or due to some unusual phenomena or event happening in certain areas of the network [16] [17]. This anomaly effect can be modeled by a sparse vector $\mathbf{z}^*$. Mathematically, we have $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{z}^* + \mathbf{w}$, where $\mathbf{z}^*$ represents the outlier regions and $\mathbf{w}$ stands for possible small noise in the data transmission. Our results make it possible to recover both the underlying signal and detect the outlier regions simultaneously, which could be very useful for network monitoring.

Another application of sparse signal recovery from sparsely corrupted measurements is error correction in joint source-channel coding. In [11], [18], [19], compressed sensing has been exploited as a joint source-channel coding strategy for its efficient encoding and robust error correcting performance. For a signal $\mathbf{f}$ that is sparse in the domain Ψ , i.e., $\mathbf{f} = \Psi\mathbf{x}^*$, it can be encoded by a linear projection $\mathbf{y} = \Phi\mathbf{f} = \mathbf{A}\mathbf{x}^*$ with $\mathbf{A} = \Phi\Psi$. Existing works have investigated the situations where the encoded signal $\mathbf{y}$ is sent through either an erasure channel [18] or a gross error channel [11], [19]. Our results can not only be applied in these scenarios, but also provide a new design on the encoding matrix with uniform recovery guarantee.

In some scenarios, the measurement noise may be sparse or compressible in some sparsifying basis. One example is the recovery of video or audio signal that are corrupted by narrow-band interference (NBI) due to improperly designed equipment [20], [21]. Electric hum as a typical impairment is sparse in the Fourier basis. Another example is the application of compressed sensing to reduce the number of samples in convolution systems with deterministic sequences (e.g., m-sequence, Golay sequence). Such convolution systems are widely used in

communications, ultrasound and radar [22], [23]. In practice, the measurements may be affected by frequency domain interference or multi-tone jamming [24]. For instance, in CS-based OFDM channel estimation [25]–[28], suppose $\mathbf{x}$ is the channel response and that the pilot sequence $\mathbf{g}$ is constructed from Golay sequences, the time-domain received signal can be represented as [28], $\mathbf{y} = \sqrt{\frac{n}{m}}\mathbf{R}_{\Omega'}\mathbf{F}^*\text{diag}(\mathbf{g})\mathbf{F}\mathbf{x} + \mathbf{w}$, where $\mathbf{R}_{\Omega'}$ is a random subsampling operator and $\mathbf{F}$ denotes the DFT matrix. The recovery performance can be guaranteed by noticing that the sensing model is a subsampled version of the orthonormal matrix $\mathbf{F}^*\text{diag}(\mathbf{g})\mathbf{F}$. However, in OFDM-based powerline communications, the NBI due to intended or unintended narrow-band signals can severely contaminate the transmitted OFDM signal. The time-domain NBI vector is sparse in the Fourier basis [29], [30]. Our results cover these settings, and therefore, provide a CS-based method to jointly estimate the signal of interest and the NBI.

B. Notations and Organization of the paper

For an n -element vector $\mathbf{a}$, we denote by a_i , ($i \in [n] = \{0, \dots, n-1\}$), the i -th element of this vector. We represent a sequence of vectors by $\mathbf{a}_0, \dots, \mathbf{a}_{n-1}$ and a column vector with q ones by $\mathbf{1}_q$. The sparsity of a vector can be measured by its best s -term approximation error,

$$\sigma_s(\mathbf{a})_p = \inf_{\|\tilde{\mathbf{a}}\|_0 \leq s} \|\mathbf{a} - \tilde{\mathbf{a}}\|_p,$$

where $\|\cdot\|_p$ is the standard l_p norm on vectors. For a matrix $\mathbf{A}$, A_{jk} denotes the element on its j -th row and k -th column. The vector obtained by taking the j -th row (k -th column) of $\mathbf{A}$ is represented by $\mathbf{A}_{(j,:)}$ ($\mathbf{A}_{(:,k)}$). We denote by $\mathbf{A}_0, \dots, \mathbf{A}_{n-1}$ a sequence of matrices. $\mathbf{A}^{-1}$ and $\mathbf{A}^*$ represent the inverse and the conjugate transpose of $\mathbf{A}$. The Frobenius norm and the operator norm of matrix $\mathbf{A}$ are denoted by $\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}^*\mathbf{A})}$ and $\|\mathbf{A}\|_{2 \rightarrow 2} = \sup_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2$ respectively. We write $A \lesssim B$ if there is an absolute constant c such that $A \leq cB$. We denote $A \sim B$ if $c_1A \leq B \leq c_2A$ for absolute constants c_1 and c_2 .

The coherence $\mu(\mathbf{A})$ of an $\tilde{n} \times n$ matrix $\mathbf{A}$ describes the maximum magnitude of the elements of $\mathbf{A}$, i.e., $\mu(\mathbf{A}) = \max_{1 \leq j \leq \tilde{n}, 1 \leq k \leq n} |A_{jk}|$. For a unitary matrix $\Psi \in \mathbb{C}^{n \times n}$, we have $\frac{1}{\sqrt{n}} \leq \mu(\Psi) \leq 1$.

The rest of the paper is organized as follows. We start by reviewing some key notions and results in compressed sensing in Section II. In Section III, we prove the uniform recovery guarantee for two classes of structured random matrices. In Section IV, we conduct a series of simulations to reinforce our theoretical results. Conclusion is given in Section V. We defer most of the proofs to the Appendices.

II. PRELIMINARIES

A. RIP and structured sensing matrices

The restricted isometry property (RIP) is a sufficient condition that guarantees uniform and stable recovery of all s -sparse vectors via nonlinear optimization (e.g. l_1 -minimization). For a

matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $s < n$, the restricted isometry constant δ_s is defined as the smallest number such that

$$(1 - \delta_s) \|\mathbf{x}\|_2^2 \leq \|\mathbf{Ax}\|_2^2 \leq (1 + \delta_s) \|\mathbf{x}\|_2^2,$$

holds for all s -sparse vectors $\mathbf{x}$. Alternatively, the restricted isometry constant of $\mathbf{A}$ can be written as

$$\delta_s = \sup_{\mathbf{x} \in \mathcal{D}_{s,n}} \left| \|\mathbf{Ax}\|_2^2 - \|\mathbf{x}\|_2^2 \right|, \quad (5)$$

where $\mathcal{D}_{s,n} = \{\mathbf{x} \in \mathbb{C}^n : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_0 \leq s\}$.

Among the many structured sensing matrices that satisfy the RIP, two classes have been found to be applicable in various scenarios. One is the randomly subsampled orthonormal systems [6], which encompass structured sensing matrices like partial random Fourier [2], convolutional CS [28], [31] and spread spectrum [32]. The other is the UDB framework which consists of a unit-norm tight frame (UTF), a random diagonal matrix and a bounded column-wise orthonormal matrix [7]. Popular sensing matrices under this framework include partial random circulant matrices [14], random demodulation [33], random probing [34] and compressive multiplexing [35].

B. Recovery Condition

We review the definition of generalized RIP, which is useful to establish robustness and stability of the optimization algorithm.

Definition II.1. [10, Definition 2.1] For any matrix $\Theta \in \mathbb{C}^{r \times (n+m)}$, it has the (s, k) -RIP with constant $\delta_{s,k}$ if $\delta_{s,k}$ is the smallest value of δ such that

$$(1 - \delta)(\|\mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2) \leq \left\| \Theta \begin{bmatrix} \mathbf{x} \\ \mathbf{z} \end{bmatrix} \right\|_2^2 \leq (1 + \delta)(\|\mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2) \quad (6)$$

holds for any $\mathbf{x} \in \mathbb{C}^n$ with $\|\mathbf{x}\|_0 \leq s$ and any $\mathbf{z} \in \mathbb{C}^m$ with $\|\mathbf{z}\|_0 \leq k$.

Here, the generalized RIP is termed as the (s, k) -RIP for convenience. We note that the (s, k) -RIP is more stringent than the conventional RIP. In other words, the fact that a sensing matrix $\mathbf{A}$ satisfies the RIP does not mean that the associated matrix $\Theta = [\mathbf{A}, \mathbf{I}]$ would satisfy the (s, k) -RIP. The recovery performance of the penalized optimization (3) can be guaranteed by the following result.

Theorem II.2. [13, Theorem 3.7] Suppose $\mathbf{y} = \mathbf{Ax}^* + \mathbf{z}^* + \mathbf{w}$ and $\Theta = [\mathbf{A}, \mathbf{I}] \in \mathbb{C}^{m \times (n+m)}$ has the $(2s, 2k)$ -RIP constant $\delta_{2s,2k}$ satisfying

$$\delta_{2s,2k} < \frac{1}{\sqrt{1 + \left(\frac{1}{2\sqrt{2}} + \sqrt{\eta} \right)^2}}$$

with $\eta = \frac{s + \lambda^2 k}{\min\{s, \lambda^2 k\}}$. Then for $\mathbf{x}^* \in \mathbb{C}^n$, $\mathbf{z}^* \in \mathbb{C}^m$, and $\mathbf{w} \in \mathbb{C}^m$ with $\|\mathbf{w}\|_2 \leq \varepsilon$, the solution $(\hat{\mathbf{x}}, \hat{\mathbf{z}})$ to the penalized

optimization problem (3) satisfies

$$\begin{aligned} \|\hat{\mathbf{x}} - \mathbf{x}^*\|_1 + \|\hat{\mathbf{z}} - \mathbf{z}^*\|_1 &\leq c_1(\sigma_s(\mathbf{x})_1 + \lambda \sigma_k(\mathbf{z})) \\ &\quad + c_2 \sqrt{s + \lambda^2 k} \varepsilon \\ \|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 + \|\hat{\mathbf{z}} - \mathbf{z}^*\|_2 &\leq c_3 \left(1 + \eta^{1/4}\right) \left(\frac{\sigma_s(\mathbf{x})_1}{\sqrt{s}} + \frac{\sigma_k(\mathbf{z})_1}{\sqrt{k}} \right) \\ &\quad + c_4 \left(1 + \eta^{1/4}\right) \varepsilon, \end{aligned}$$

where the constants c_1, c_2, c_3, c_4 depend on $\delta_{2s,2k}$ only.

We note that similar theorem has been proven in [10] when both the signal and corruption are vectors with exact sparsity. The above result not only relaxes the requirement on the $(2s, 2k)$ -RIP constant, but also guarantees stable recovery of inexact sparse signals and corruptions. Therefore, for either sparse or compressible signals and corruptions, the key to establish the recovery guarantee for a sensing matrix is to prove the (s, k) -RIP.

III. MAIN RESULTS

In this section, we prove the (s, k) -RIP for two classes of structured sensing matrices. This result can then be combined with Theorem II.2 to prove the recovery guarantee. In addition, the extension to the recovery via nonconvex optimization is presented. Last but not least, we compare the main theorems to existing literature where relevant.

A. Randomly modulated unit-norm tight frames

We prove the uniform recovery guarantees for the class of structured sensing matrices that can be written as $\mathbf{A} = \mathbf{UDB}$, where $\mathbf{U} \in \mathbb{C}^{m \times \tilde{n}}$ is a UTF with $\mu(\mathbf{U}) \sim 1/\sqrt{m}$, $\mathbf{D} = \text{diag}(\boldsymbol{\xi})$ is a diagonal matrix with $\boldsymbol{\xi}$ being a length- $\tilde{n}$ random vector with independent, zero-mean, unit-variance, and L -subgaussian entries, and $\tilde{\mathbf{B}} \in \mathbb{C}^{\tilde{n} \times n}$, $\tilde{n} \geq n$, represents a column-wise orthonormal matrix, i.e. $\tilde{\mathbf{B}}^* \tilde{\mathbf{B}} = \mathbf{I}$.

The following result presents a bound on the required number of measurements m such that the corresponding matrix Θ has the (s, k) -RIP constant satisfying $\delta_{s,k} \leq \delta$ for any $\delta \in (0, 1)$.

Theorem III.1. Suppose $\mathbf{y} = \mathbf{Ax}^* + \mathbf{z}^* + \mathbf{w}$ with $\Theta = [\mathbf{A}, \mathbf{I}] \in \mathbb{C}^{m \times (n+m)}$, $\mathbf{A} = \mathbf{UDB}$ and $\mu(\mathbf{U}) \sim 1/\sqrt{m}$. If, for $\delta \in (0, 1)$,

$$\begin{aligned} m &\geq c_5 \delta^{-2} s \tilde{n} \mu^2(\tilde{\mathbf{B}}) \log^2 s \log^2 \tilde{n}, \\ m &\geq c_6 \delta^{-2} k \log^2 k \log^2 \tilde{n}, \end{aligned}$$

where c_5 and c_6 are some absolute constants, then with probability at least $1 - 2\tilde{n}^{-\log^2 s \log \tilde{n}}$, the (s, k) -RIP constant of Θ satisfies $\delta_{s,k} \leq \delta$.

Proof. The (s, k) -RIP constant $\delta_{s,k}$ can be equivalently expressed as

$$\delta_{s,k} = \sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} \left| \left\| \Theta \begin{bmatrix} \mathbf{x} \\ \mathbf{z} \end{bmatrix} \right\|_2^2 - \|\mathbf{x}\|_2^2 - \|\mathbf{z}\|_2^2 \right|, \quad (7)$$

where $\mathcal{T} := \{(\mathbf{x}, \mathbf{z}) : \|\mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2 = 1, \|\mathbf{x}\|_0 \leq s, \|\mathbf{z}\|_0 \leq k, \mathbf{x} \in \mathbb{C}^n, \mathbf{z} \in \mathbb{C}^m\}$. With $\Theta = [\mathbf{A}, \mathbf{I}]$, the RIP-constant can be further reduced to

$$\begin{aligned} \delta_{s,k} &= \sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} \left| \|\mathbf{Ax}\|_2^2 + \|\mathbf{z}\|_2^2 + 2\langle \mathbf{Ax}, \mathbf{z} \rangle - \|\mathbf{x}\|_2^2 - \|\mathbf{z}\|_2^2 \right| \\ &\leq \underbrace{\sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} \left| \|\mathbf{Ax}\|_2^2 - \|\mathbf{x}\|_2^2 \right|}_{\delta_1} + 2 \underbrace{\sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} |\langle \mathbf{Ax}, \mathbf{z} \rangle|}_{\delta_2} \end{aligned} \quad (8)$$

Our aim is to derive bounds on the number of measurements m such that for any $\delta \in (0, 1)$ the RIP-constant $\delta_{s,k}$ is upper bounded by δ . We have

$$\delta_1 \leq \sup_{\mathbf{x} \in \mathcal{D}_{s,n}} \left| \|\mathbf{Ax}\|_2^2 - \|\mathbf{x}\|_2^2 \right| \quad (9)$$

with $\sup_{\mathbf{x} \in \mathcal{D}_{s,n}} \left| \|\mathbf{Ax}\|_2^2 - \|\mathbf{x}\|_2^2 \right|$ being the restricted isometry constant in the standard RIP definition (5). Then, by [7, Theorem III.2], we reach the following result.

Suppose, for any $\delta \in (0, 1)$,

$$m \geq 4c_1\delta^{-2}s\tilde{n}\mu^2(\tilde{\mathbf{B}})(\log^2 s \log^2 \tilde{n}),$$

then $\delta_1 \leq \delta/2$ holds with probability at least $1 - \tilde{n}^{-(\log \tilde{n})(\log s)^2}$.

Therefore, proof of the (s, k) -RIP is reduced to bounding the inner product term δ_2 .

$$\begin{aligned} \delta_2 &= 2 \sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} |\langle \mathbf{UDB}\tilde{\mathbf{x}}, \mathbf{z} \rangle| = 2 \sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} |\mathbf{z}^* \mathbf{UDB}\tilde{\mathbf{x}}| \\ &= 2 \sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} |\mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\mathbf{x}) \boldsymbol{\xi}| = 2 \sup_{\mathbf{v} \in \mathcal{A}_v} |\langle \mathbf{v}, \boldsymbol{\xi} \rangle|, \end{aligned} \quad (10)$$

where $\mathbf{v} = (\mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\mathbf{x}))^*$, and

$$\mathcal{A}_v := \{\mathbf{v} : \|\mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2 = 1, \|\mathbf{x}\|_0 \leq s, \|\mathbf{z}\|_0 \leq k\}. \quad (11)$$

The following lemma is proved in Appendix A.

Lemma III.2. Suppose $\boldsymbol{\xi}$ is a length- $\tilde{n}$ random vector with independent, zero-mean, unit-variance, and L -subgaussian entries. For any $\delta \in (0, 1)$, if

$$\begin{aligned} m &\geq c_5\delta^{-2}s\tilde{n}\mu^2(\tilde{\mathbf{B}})\log^2 s \log^2 \tilde{n} \\ m &\geq c_6\delta^{-2}k\log^2 k \log^2 \tilde{n}, \end{aligned}$$

then $\sup_{\mathbf{v} \in \mathcal{A}_v} |\langle \mathbf{v}, \boldsymbol{\xi} \rangle| \leq \delta/2$ holds with probability exceeding $1 - \exp(-\log^2 s \log^2 \tilde{n})$, where c_5 and c_6 are some constants depending only on L .

Combining (10) with Lemma III.2, we have, for any $\tau > 0$, $\delta_2 \leq c\delta$ holds with probability exceeding $1 - \exp(-\log^2 s \log^2 \tilde{n})$ for some constant c .

Finally, Theorem III.1 can be obtained by combining the above results. Suppose, for any $\delta \in (0, 1)$, $m \geq c_5\delta^{-2}s\tilde{n}\mu^2(\tilde{\mathbf{B}})(\log^2 s \log^2 \tilde{n})$, $m \geq c_6\delta^{-2}k\log^2 k \log^2 \tilde{n}$ and $\mu(U) \sim 1/\sqrt{m}$, then we have $\delta_{s,k} \leq \delta_1 + \delta_2 \leq \delta$ with probability exceeding

$$\begin{aligned} &1 - \tilde{n}^{-(\log \tilde{n} \log^2 s)} - \exp(-c \log^2 s \log^2 \tilde{n}) \\ &= 1 - \tilde{n}^{-(\log \tilde{n} \log^2 s)} - \tilde{n}^{-c \log^2 s \log \tilde{n}} \\ &= 1 - 2\tilde{n}^{-\log^2 s \log \tilde{n}} \end{aligned}$$

The uniform recovery guarantee can be obtained by combining Theorem II.2 and III.1.

A few remarks are in order. First, when $\tilde{\mathbf{B}}$ is a bounded column-wise orthonormal matrix, i.e., $\mu(\tilde{\mathbf{B}}) \sim 1/\sqrt{\tilde{n}}$, the bound on the sparsity of $\mathbf{x}^*$ can be relaxed to $\|\mathbf{x}^*\|_0 \leq Cm/(\log^2 \tilde{n} \log^2 m)$. The sparsity $\|\mathbf{z}^*\|_0$ is always a constant fraction of the total number of measurements m regardless the magnitude of the coherence $\mu(\tilde{\mathbf{B}})$. When $\|\mathbf{w}\|_2 = 0$, Theorem III.1 implies that a sparse signal can be exactly recovered by tractable l_1 minimization even if some parts of the measurements are arbitrarily corrupted.

Second, the proposed class of structured sensing matrices is equivalent to the UDB framework [7] but with an additional requirement of $\mu(\mathbf{U}) \sim 1/\sqrt{m}$. The UDB framework has been proved to support uniform recovery guarantees for conventional CS problem, while with the extra condition it is now shown to provide uniform recovery guarantees for the CS with sparse corruptions problem. Theorem III.1 holds for many existing and new structured sensing matrices as long as they can be decomposed into $\mathbf{A} = \mathbf{UDB}$.

One application of the UDB framework is to simplify the mask design in double random phase encoding (DRPE) for optical image encryption. Consider an image $\mathbf{f}$ that is sparse in the domain Ψ , i.e., $\mathbf{f} = \Psi\mathbf{x}^*$, DRPE is based on random masks placed in the input and Fourier planes of the optical system [36], [37]. Mathematically, the measurements can be written as $\mathbf{y} = \sqrt{\frac{n}{m}}\mathbf{R}_\Omega\mathbf{F}^*\mathbf{A}_1\mathbf{F}\mathbf{A}_2\mathbf{f} + \mathbf{w}$, where $\mathbf{R}_\Omega : \mathbb{C}^n \rightarrow \mathbb{C}^m$ represents an arbitrary/deterministic subsampling operator with Ω being the set of selected row indices, $\mathbf{A}_1$ and $\mathbf{A}_2$ are random diagonal matrices. By the UDB framework, the random diagonal matrix $\mathbf{A}_2$ can be replaced by a deterministic diagonal matrix constructed from a Golay sequence $\mathbf{g}$. The reason is that the measurement model $\sqrt{\frac{n}{m}}\mathbf{R}_\Omega\mathbf{F}^*\mathbf{A}_1\mathbf{F}\text{diag}(\mathbf{g})\Psi\mathbf{x}^*$ can be decomposed into a UTF $\sqrt{\frac{n}{m}}\mathbf{R}_\Omega\mathbf{F}^*$, a random diagonal matrix $\mathbf{A}_1$, and an orthonormal matrix $\mathbf{F}\text{diag}(\mathbf{g})\Psi$ whose coherence is proven to be bounded for many orthonormal transforms Ψ , e.g., DCT, Haar wavelet [7, Lemma IV.2]. When the measurements are corrupted by impulse noise due to detector plane impairment, our theorem above provides a recovery guarantee on the image.

Furthermore, the UDB framework encompasses some popular structured sensing matrices, e.g., partial random circulant matrices [14] and random probing [34]. To elaborate, consider the partial random circulant matrices

$$\mathbf{A} = \frac{1}{\sqrt{m}}\mathbf{R}_\Omega\mathbf{C}_\epsilon$$

where $\mathbf{C}_\epsilon$ denotes the circulant matrix generated from ϵ . Suppose $\epsilon = \mathbf{F}^*\boldsymbol{\xi}$, where $\mathbf{F}$ is a normalized DFT matrix and $\boldsymbol{\xi}$ is a length- n random vector with independent, zero-mean, unit-variance, and sub-Gaussian entries. Let $\mathbf{D} = \text{diag}(\boldsymbol{\xi})$, we have $\mathbf{A} = \sqrt{\frac{n}{m}}\mathbf{R}_\Omega\mathbf{F}^*\mathbf{D}\mathbf{F}$. It can be observed that $\mathbf{U} = \sqrt{\frac{n}{m}}\mathbf{R}_\Omega\mathbf{F}^*$ is a UTF and $\mathbf{B} = \mathbf{F}$ is a unitary matrix. Hence, Theorem III.1 implies that any sparse signal $\mathbf{x}$ and sparse corruption $\mathbf{z}$ can be faithfully recovered from the measurement model $\mathbf{y} = \frac{1}{\sqrt{m}}\mathbf{R}_\Omega\mathbf{C}_\epsilon\mathbf{x}^* + \mathbf{z}^* + \mathbf{w}$ by the penalized recovery

algorithm. The sparse recovery from partial random circulant measurements can be applied in many common deconvolution tasks, such as radar [38] and coded aperture imaging [39]. In practice, where the measurements can be corrupted by impulsive noise due to bit errors in transmission, faulty memory locations, and buffer overflow [40], our theorem guarantees the recovery of both the signal of interest and the corruption.

In some situations, the proposed framework can still provide reliable recovery guarantee even if the corruption is sparse in some basis. Suppose the corruption is sparse under some fixed and known orthonormal transformation $\mathbf{H}$, i.e. $\mathbf{H}^*\mathbf{H} = \mathbf{I}$. We consider the measurement model

$$\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{H}\mathbf{z}^* + \mathbf{w}. \quad (12)$$

It is clear that this setting can be reduced to

$$\mathbf{H}^*\mathbf{y} = \mathbf{H}^*\mathbf{A}\mathbf{x}^* + \mathbf{z}^* + \mathbf{H}^*\mathbf{w}. \quad (13)$$

Notice that $\mathbf{H}^*\mathbf{A} = \mathbf{H}^*\mathbf{U}\mathbf{D}\mathbf{\tilde{B}} := \mathbf{\hat{U}}\mathbf{D}\mathbf{\tilde{B}}$, where $\mathbf{\hat{U}} = \mathbf{H}^*\mathbf{U}$ is still a UTF due to the orthogonality of $\mathbf{H}$. Therefore, if $\mu(\mathbf{\hat{U}}) \sim 1/\sqrt{m}$, Theorem III.1 still holds in this measurement model.

B. Randomly sub-sampled orthonormal system

Next, we consider the corrupted sensing measurement model for randomly sub-sampled orthonormal system. We prove the uniform recovery guarantee for such matrices provided that the corruption is sparse on certain sparsifying domain. Suppose $\boldsymbol{\lambda} \in \mathbb{R}^n$ is a random Bernoulli vector with i.i.d. entries such that $\mathbb{P}(\lambda_i = 1) = \frac{m}{n} \forall i \in [n]$ and $\Omega' = \{i : \lambda_i = 1\}$ with $|\Omega'| = M$, the random sampling operator $\mathbf{R}_{\Omega'} \in \mathbb{R}^{M \times n}$ is a collection of the i -th row of an n -dimensional identity matrix for all $i \in \Omega'$. Here, M is random with mean value m . The observation model is

$$\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{H}\mathbf{z}^* + \mathbf{w}, \quad (14)$$

where $\mathbf{A} = \sqrt{\frac{n}{M}}\mathbf{R}_{\Omega'}\mathbf{G}$, $\mathbf{G} \in \mathbb{C}^{n \times n}$ is an orthonormal basis and $\mathbf{H} \in \mathbb{C}^{M \times M}$ is a unitary matrix with $\mu(\mathbf{H}) \sim 1/\sqrt{M}$.

From our analysis in previous subsection, the uniform recovery performance can be guaranteed as long as the associated matrix $\boldsymbol{\Theta}$ satisfies the (s, k) -RIP. Since the matrix $\mathbf{A}$ satisfies the standard RIP, the problem of proving the (s, k) -RIP is again reduced to bounding the inner product term $\sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} |\langle \mathbf{A}\mathbf{x}, \mathbf{H}\mathbf{z} \rangle|$. Detail proof of the following result is given in Appendix B.

Theorem III.3. Suppose $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{H}\mathbf{z}^* + \mathbf{w}$ with $\boldsymbol{\Theta} = [\mathbf{A}, \mathbf{H}] \in \mathbb{C}^{M \times (n+M)}$, $\mathbf{A} = \sqrt{\frac{n}{M}}\mathbf{R}_{\Omega'}\mathbf{G}$ and $\mu(\mathbf{H}) \sim 1/\sqrt{M}$. If, for $\delta \in (0, 1)$,

$$\begin{aligned} m &\geq \max(c_7\delta^{-2}sn\mu^2(\mathbf{G})\log^2 s \log^2 n, c_8\delta^2 s \log^4 n, 2c_9 \log n), \\ m &\geq c_{10}\delta^{-2}kn\mu^2(\mathbf{G})\log^2 k \log^2 n, \\ m &\leq c_{11}\delta^2 n, \end{aligned}$$

where $\{c_i\}_{i=7,\dots,11}$ are constants, then with probability at least $1 - 2\tilde{n}^{-\log^2 s \log n} - n^{-c_9}$, the (s, k) -RIP constant of $\boldsymbol{\Theta}$ satisfies $\delta_{s,k} \leq \delta$.

When $\mathbf{G}$ is a bounded orthonormal basis, i.e., $\mu(\mathbf{G}) \sim 1/\sqrt{n}$, the bound on the sparsity of $\mathbf{x}^*$ can be relaxed to $m \geq \mathcal{O}(\delta^{-2}s \log^2 s \log^2 n, \delta^{-2}k \log^2 k \log^2 n)$, which implies that a sparse signal can be exactly recovered by tractable l_1 minimization even if the measurements are affected by corruption sparse on some bounded domain. A bounded orthonormal basis can include the Fourier transform or the Hadamard transform. In addition, in CS-based OFDM where the pilot is generated from a Golay sequence and a random subsampler is employed at the receiver (Section I-A), the effective orthonormal basis is also bounded, i.e., $\mu(\mathbf{F}^*\text{diag}(\mathbf{g})\mathbf{F}) \sim 1/\sqrt{n}$ [28].

C. Nonconvex optimization

We have shown the (s, k) -RIP for two popular classes of structured sensing matrices, and proven the performance guarantee for the recovery of the sparse signal and corruption via the l_1 -norm minimization algorithm (3). However, our (s, k) -RIP analysis on the structured sensing matrices is also applicable to proving the recovery guarantee for nonconvex optimization. Consider the following formulation of the problem

$$\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{H}\mathbf{z}^*, \quad (15)$$

It was demonstrated in [41] that the unique minimizer of the l_p minimization problem ($0 < p < 1$)

$$\min_{\mathbf{x}, \mathbf{z}} \|\mathbf{x}\|_p^p + \nu \|\mathbf{z}\|_p^p \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} + \mathbf{H}\mathbf{z} = \mathbf{y}, \quad (16)$$

is exactly the pair $(\mathbf{x}^*, \mathbf{z}^*)$ if the combined matrix $[\mathbf{A}, \mathbf{H}]$ satisfies the (s, k) -RIP, where ν is the regularization parameter. In addition, the l_p minimization approach still provides stable recovery even when there is additional dense noise as long as the (s, k) -RIP holds [41], [42]. The l_p minimization problem can be numerically solved via an iteratively reweighted least squares (IRLS) method [43]. However, [41] only considers the sensing model with $\mathbf{A}$ being random Gaussian matrices and $\mathbf{H}$ being an identity matrix. With our (s, k) -RIP analysis, many structured sensing matrices can be employed to provide exact/stable recovery performance for this corrupted sensing problem via l_p minimization.

D. Comparison with related literature

In this part, we compare our main results with related literature.

1) *Sparse signal, sparse corruption:* [10] proved that sensing matrices with independent Gaussian entries provide uniform recovery guarantee for corrupted CS by solving (3) for all vectors $\mathbf{x}^*$ and $\mathbf{z}^*$ satisfying $\|\mathbf{x}\|_0 \leq \alpha m / (\log(n/m) + 1)$ and $\|\mathbf{z}^*\|_0 \leq \alpha m$. The difference is that our theorems come with a tighter requirement on the sparsity of $\mathbf{x}^*$ and the sparsity of $\mathbf{z}^*$, which is a compensation on the employment of structured measurements.

[10] also proved the recovery guarantee for structured sensing matrices that belong to the framework proposed in [12]. Here, faithful recovery is guaranteed provided that $\|\mathbf{x}\|_0 \leq \alpha m / (\mu^2 \log^2 n)$ and $\|\mathbf{z}^*\|_0 \leq \beta m / \mu^2$, where μ is the coherence of the sensing matrix. [11] considered the corrupted

CS with sensing matrices that are randomly subsampled orthonormal matrix, and proved similar results. It is noted that the requirements on the sparsity of $\mathbf{x}^*$ in these works seem less strict than that in our results. However, in both [10] and [11], performance guarantees of their structured sensing matrices rely on the assumption that the support set of $\mathbf{x}^*$ or $\mathbf{z}^*$ is fixed and the signs of the signal are independently and equally likely to be 1 or -1 [10, Section 1.3.2] [11, Section II.B] (i.e. a nonuniform recovery guarantee). While in our theorem, two classes of structured sensing matrices (including randomly subsampled orthogonal matrix) are shown to provide uniform recovery guarantee for corrupted CS.

We note that recently the uniform recovery guarantee for bounded orthonormal systems is proven in [13]. The bounded orthonormal systems is more general than the random subsampled orthonormal matrix considered in our second class. However, the corruption models are different: the corruption vector in [13] is sparse in time domain, whereas our theorem considers corruption in sparsifying domain with $\mu(\mathbf{H}) \sim 1/\sqrt{M}$. Due to this difference in the corruption model, the techniques used to prove the (s, k) -RIP (specifically, bound the inner product $\sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} |\langle \mathbf{A}\mathbf{x}, \mathbf{H}\mathbf{z} \rangle|$) are essentially different.

2) *Structured signal, structured corruption*: In a recent work [44], sensing with random Gaussian measurements for general structured signals and corruptions (including sparse vectors, low rank matrix, sign vectors and etc) has been proven. However, our study departs from it in the following aspects: [44] proved a nonuniform recovery guarantee for the recovering of sparse signals from sparse corruptions and dense noise. In our paper, we established a uniform recovery guarantee for the corresponding problem. Moreover, [44] considered random Gaussian matrices, while we propose structured sensing matrices.

We have shown that a large class of structured sensing matrices can provide faithful recovery for the sparse sensing with sparse corruption. Whether such structured measurements can be applied in a general corrupted sensing problem (e.g. structured signal with structured corruption) is still open. Extension of our measurement framework to the general corrupted sensing problem is interesting for further study.

Other works related to the recovery of signals from corrupted measurements include [20], [45]–[51]. However, their models are different from the one in our paper.

Remark III.4. We note that the value of the regularization parameter can be chosen as $\lambda = \sqrt{s/k}$. In practice, when no a priori knowledge on the sparsity levels of the signal and the corruption is available, λ can usually be taken by cross validation. On the other hand, if it is known a priori that the corruption (the signal) is very sparse, one can increase (decrease) the value of λ to improve the overall recovery performance. Similar discussion on the theoretical and practical settings of the regularization parameter has also been noted in [10, Section 1.3.3], [11, Section II.E, Section VII], [44, Section III.B]. In addition, an iteratively reweighted l_1 minimization method can be used to adaptively improve the setting of λ in practice [13].

IV. NUMERICAL SIMULATIONS

In this section, we verify and reinforce the theoretical results of Section III with a series of simulations. We present experiments to test the recovery performance of the penalized recovery algorithm for the proposed structured sensing matrices. In each experiment, we used the CVX Matlab package [52], [53] to specify and solve the convex recovery programs.

Two different ways of generating sparse vectors are considered:

- Gaussian setting: the nonzero entries are drawn from a Gaussian distribution and their locations are chosen uniformly at random,
- Flat setting: the magnitudes of nonzero entries are 1 and their locations are chosen uniformly at random.

A. Penalized Recovery

This experiment is to investigate the empirical recovery performance of the penalized recovery algorithm (3) when the dense noise is zero. Here, the sensing matrix (Mtx-I) $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{B}$ of size $m \times n$ with $m = 256$ and $n = 512$ is constructed as below.

- 1) Arbitrarily select $m = 256$ rows from a 512×512 Hadamard matrix to form a new matrix, which is then normalized by $1/\sqrt{m}$ to form the UTF $\mathbf{U}$.
- 2) The diagonal entries of the diagonal matrix $\mathbf{D}$ are i.i.d. Bernoulli random variables.
- 3) $\mathbf{B}$ is a normalized Hadamard matrix.

We vary the signal sparsity and the corruption sparsity with $s \in [1, 100]$ and $k \in \{10, 20, 30\}$. For each pair of (s, k) , we draw a sensing matrix as described above and perform the following experiment 100 times:

- 1) Generate $\mathbf{x}^*$ with sparsity s by the Gaussian setting
- 2) Generate $\mathbf{z}^*$ with sparsity k by the Gaussian setting
- 3) Solve (3) by setting $\lambda = 1$
- 4) Declare success if⁴

$$\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 / \|\mathbf{x}^*\|_2 + \|\hat{\mathbf{z}} - \mathbf{z}^*\|_2 / \|\mathbf{z}^*\|_2 < 10^{-3}$$

The fraction of successful recovery averaged over the 100 iterations is presented in Fig. 1a. To demonstrate the performance for signals and corruptions that do not have i.i.d. signs, the experiment is repeated by generating the sparse vectors $\mathbf{x}^*$ and $\mathbf{z}^*$ based on the Flat setting as shown in Fig. 1b. It can be seen that in both scenarios the performance improves as the sparsity of the corruption decreases.

Next, we demonstrate the performance of the penalized recovery algorithm when the sensing matrix is from a randomly subsampled orthonormal matrix. The sensing matrix (Mtx-II) $\mathbf{A}$ is a collection of randomly selected $M = 256$ rows from a 512×512 Hadamard matrix, and normalized by $\sqrt{n/M}$. The corruption is $\mathbf{H}\mathbf{z}$, where $\mathbf{H}$ is an $M \times M$ normalized Hadamard matrix. For each pair of (s, k) , we repeat the above steps 100 times to obtain the probability of success (see Fig. 2). It is noted that the recovery performance of Mtx-I is better than that of Mtx-II. This seems consistent with our theoretical analysis as the random subsampled orthonormal matrix shows

⁴This criterion indicates that both $\mathbf{x}^*$ and $\mathbf{z}^*$ have been faithfully recovered.

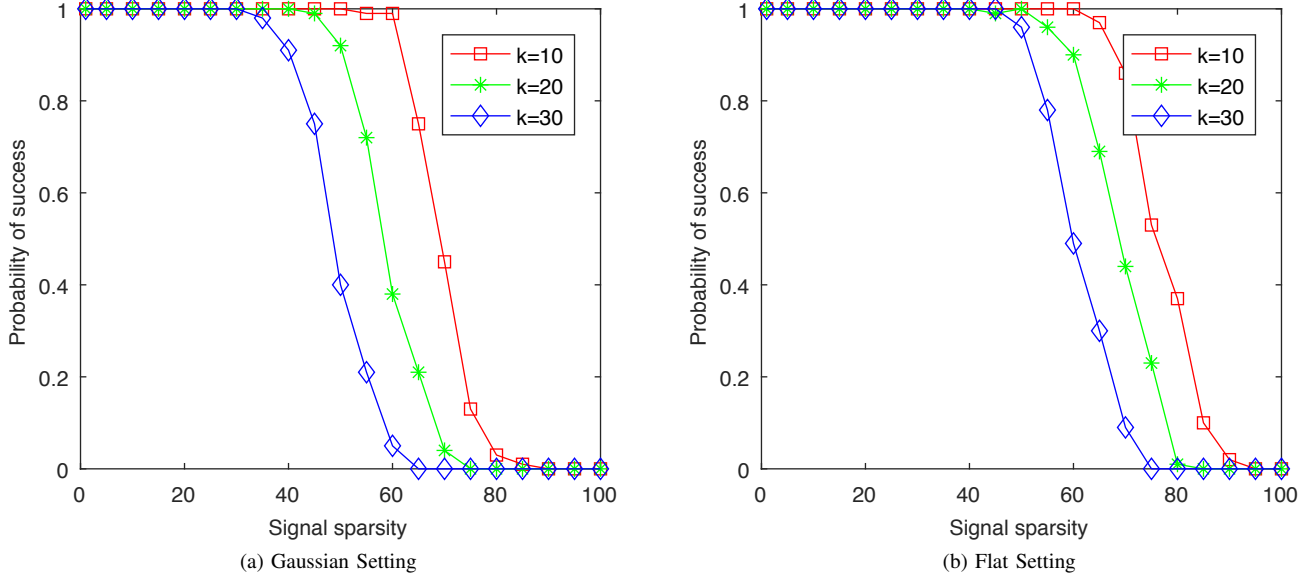


Fig. 1. Probability of success as a function of the signal sparsity s using penalized recovery with signal dimension $n = 512$, number of measurements $m = 256$, and the corruption sparsity $k = \{10, 20, 30\}$ for Mtx-I.

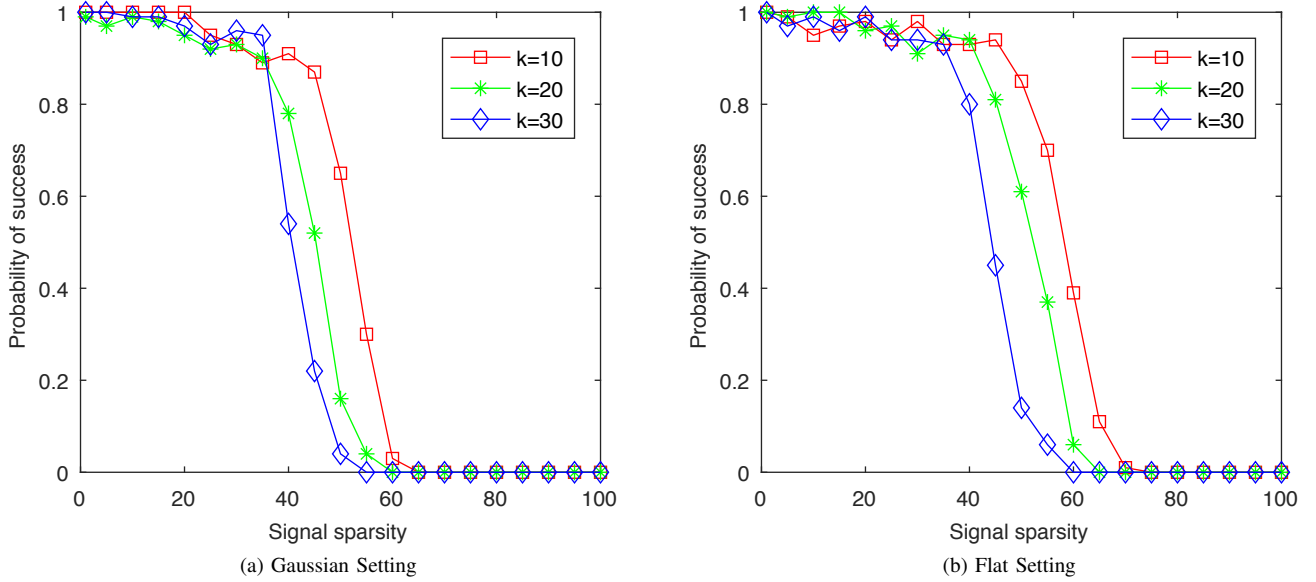


Fig. 2. Probability of success as a function of the signal sparsity s using penalized recovery with signal dimension $n = 512$, number of measurements $m = 256$, and the corruption sparsity $k = \{10, 20, 30\}$ for Mtx-II.

more stringent recovery condition than the UDB framework (see Theorem III.1 and III.3). However, since the (s, k) -RIP is a sufficient condition for the recovery guarantee, it may not fully reflect the performance gap between the two classes of structured sensing matrices. Further investigation based on a necessary and sufficient condition for the recovery guarantee of the corrupted CS problem is a difficult, but interesting open question.

B. Stable recovery

We study the stability of the penalized recovery algorithms when the dense noise is nonzero, i.e., $\varepsilon \neq 0$, and compare the performance of structured sensing matrix (Mtx-I) with random Gaussian sensing matrix. In this experiment, the 256-by-512 sensing matrix (Mtx-I) is constructed as in previous subsection. We fix the signal and corruption sparsity levels at $s = 10$ and $k = 10$ respectively. The dense noise $\mathbf{w}$ consists of i.i.d. Bernoulli entries with amplitude ε . We vary the noise level with $\varepsilon \in [0, 0.1]$, and perform the following experiment 100 times for each ε :

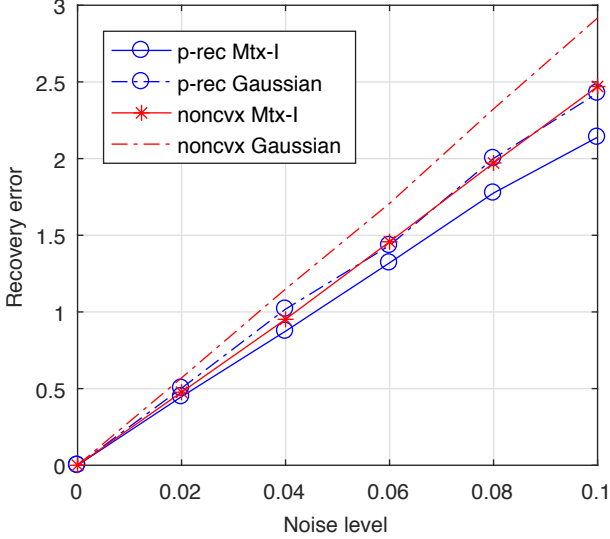


Fig. 3. Empirical recovery error versus the noise level ε .

- 1) Generate $\mathbf{x}^*$ with $s = 10$ by the Gaussian setting
- 2) Generate $\mathbf{z}^*$ with $k = 10$ by the Gaussian setting
- 3) Solve penalized recovery (p-rec) algorithm (3) by setting $\lambda = 1$
- 4) Record the empirical recovery error $\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 + \|\hat{\mathbf{z}} - \mathbf{z}^*\|_2$

An average recovery error is then obtained for each ε . Fig. 3 depicts the average error with varying noise levels. The results in Theorems II.2 and III.1 imply that the recovery errors are bounded by the noise level ε up to some constants. Fig. 3 clearly shows this linear relationship. In addition, we repeat the above experiments with an iteratively reweighted least squares approach [43] using $p = 0.5$. As shown in Fig. 3, the structured sensing matrix is still able to exhibit stable performance by the nonconvex optimization algorithm.

V. CONCLUSION

We have studied a generalized CS problem where the measurement vector is corrupted by both sparse noise and dense noise. We have proven that structured random matrices encompassed in the UDB framework or the randomly subsampled orthonormal system can satisfy the sufficient condition, i.e., the (s, k) -RIP. These structured matrices can therefore be applied to provide faithful recovery of both the sparse signal and the corruption by the penalized optimization algorithm as well as the nonconvex optimization algorithm. Our simulations have clearly illustrated and reinforced our theoretical results.

APPENDIX A PROOF OF LEMMA III.2

Throughout the proof in this and the following sections, C and c denote an absolute constant whose values may change from occurrence to occurrence.

A metric space is denoted by (T, d) , where T is a set and d is the notion of distance (metric) between elements

of the set. For a metric space (T, d) , the covering number $N(T, d, u)$ is the minimal number of open balls of radius u needed to cover (T, d) . A subset $\bar{T}$ of T is called a u -net of T if every point $\mathbf{x} \in T$ can be approximated to within u by some point $\bar{\mathbf{x}} \in \bar{T}$, i.e., $d(\mathbf{x}, \bar{\mathbf{x}}) \leq u$. The minimal cardinality of $\bar{T}$ is equivalent to the covering number $N(T, d, u)$. The p -th moment (or the L_p -norm) of a random variable is denoted by $\|X\|_{L_p} = (\mathbb{E}|X|^p)^{1/p}$.

We aim to upper bound the variable $\Delta := \sup_{\mathbf{v} \in \mathcal{A}_v} |\langle \mathbf{v}, \boldsymbol{\xi} \rangle|$ which is the supremum of a stochastic process with the index set $\mathcal{A}_v$. To complete the proof, we require the following important result due to Krahmer et al.:

Theorem A.1. [14, Theorem 3.5 (a)] *Let $\mathcal{A}$ be a set of matrices, and let $\boldsymbol{\xi}$ be a random vector whose entries ξ_j are independent, mean 0, variance 1, and L -subgaussian random variables. Set*

$$d_F(\mathcal{A}) = \sup_{\mathbf{S} \in \mathcal{A}} \|\mathbf{S}\|_F, \quad d_{2 \rightarrow 2}(\mathcal{A}) = \sup_{\mathbf{S} \in \mathcal{A}} \|\mathbf{S}\|_{2 \rightarrow 2},$$

$$N_{\mathcal{A}}(\boldsymbol{\xi}) := \sup_{\mathbf{S} \in \mathcal{A}} \|\mathbf{S}\boldsymbol{\xi}\|_2, \quad E = \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A}).$$

Then, for every $p \geq 1$,

$$\|N_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} \leq C(E + \sqrt{p}d_{2 \rightarrow 2}(\mathcal{A})), \quad (17)$$

where C is a constant depends only on L .

Here, $N_{\mathcal{A}}(\boldsymbol{\xi})$ represents the supremum of certain stochastic processes indexed by a set of matrices $\mathcal{A}$. The above Proposition implies that $N_{\mathcal{A}}(\boldsymbol{\xi})$ can be bounded by three parameters: the suprema of Frobenius norms $d_F(\mathcal{A})$, the suprema of operator norms $d_{2 \rightarrow 2}(\mathcal{A})$ and a γ_2 -functional $\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2})$, which can be bounded in terms of the covering numbers $N(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u)$ as below.

$$\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \leq c \int_0^{d_{2 \rightarrow 2}(\mathcal{A})} \sqrt{\log N(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u)} du,$$

where the integral is known as Dudley integral or entropy integral [54].

We can transfer the estimates on the moment (17) to a tail bound by the standard estimate due to Markov's inequality (see [5, Proposition 7.15]).

Proposition A.2. *Following the definitions in Theorem A.1, for $t \geq 1$,*

$$\mathbb{P}(N_{\mathcal{A}}(\boldsymbol{\xi}) \geq CE + Cd_{2 \rightarrow 2}(\mathcal{A})t) \leq \exp(-t^2). \quad (18)$$

It can be observed that Δ can be expressed in the form of $N_{\mathcal{A}}(\boldsymbol{\xi})$, where $\mathbf{S}$ and $\mathcal{A}$ are replaced with $\mathbf{v}$ and $\mathcal{A}_v$, respectively. Now, we only need to estimate the parameters $d_F(\mathcal{A}_v)$, $d_{2 \rightarrow 2}(\mathcal{A}_v)$ and $\gamma_2(\mathcal{A}_v, \|\cdot\|_{2 \rightarrow 2})$ before bounding Δ by using Theorem A.2. Since $\mathcal{A}_v$ is a set of vectors, we have $d_F(\mathcal{A}_v) = d_{2 \rightarrow 2}(\mathcal{A}_v)$ and $\gamma_2(\mathcal{A}_v, \|\cdot\|_{2 \rightarrow 2}) = \gamma_2(\mathcal{A}_v, \|\cdot\|_2)$.

For any vector $\mathbf{x} \in \mathcal{D}_{s,n}$, we denote by $\mathbf{x}^s$ the length- s vector that retains only the non-zero elements in $\mathbf{x}$. And correspondingly for any vector $\mathbf{b} \in \mathbb{C}^n$, we denote by $\mathbf{b}^s$ the length- s vector that retains only the elements that have the same indexes as those of the non-zero elements in $\mathbf{x}$. We

have, for any $\mathbf{v} \in \mathcal{A}_{\mathbf{v}}$,

$$\begin{aligned} \|\mathbf{v}\|_2 &= \|\mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\mathbf{x})\|_2 \leq \|\mathbf{z}^* \mathbf{U}\|_2 \|\tilde{\mathbf{B}}\mathbf{x}\|_\infty \\ &= \sqrt{\frac{\tilde{n}}{m}} \|\mathbf{z}\|_2 \max_{j \in [\tilde{n}]} \{|\langle \tilde{\mathbf{B}}_{(j,:)}^s, \mathbf{x}^s \rangle|\} \\ &= \sqrt{\frac{\tilde{n}}{m}} \|\mathbf{z}\|_2 \max_{j \in [\tilde{n}]} \{|\langle \tilde{\mathbf{B}}_{(j,:)}^s, \mathbf{x}^s \rangle|\} \\ &\leq \sqrt{\frac{\tilde{n}}{m}} \mu(\tilde{\mathbf{B}}) \sqrt{s} \|\mathbf{x}\|_2 \|\mathbf{z}\|_2 \leq \frac{1}{2} \sqrt{\frac{\tilde{n}}{m}} \mu(\tilde{\mathbf{B}}) \sqrt{s}, \end{aligned}$$

where the last inequality is due to $\|\mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2 = 1$. Therefore,

$$d_F(\mathcal{A}_{\mathbf{v}}) = d_{2 \rightarrow 2}(\mathcal{A}_{\mathbf{v}}) \leq \frac{1}{2} \sqrt{\frac{\tilde{n}}{m}} \mu(\tilde{\mathbf{B}}) \sqrt{s}. \quad (19)$$

Following the same steps, we can alternatively obtain, for any $\mathbf{v} \in \mathcal{A}_{\mathbf{v}}$,

$$\begin{aligned} \|\mathbf{v}\|_2 &= \|\mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\mathbf{x})\|_2 \leq \|\mathbf{z}^* \mathbf{U}\|_\infty \|\tilde{\mathbf{B}}\mathbf{x}\|_2 \\ &\leq \frac{1}{2} \mu(\mathbf{U}) \sqrt{k}. \end{aligned}$$

This provides another upper bound

$$d_F(\mathcal{A}_{\mathbf{v}}) = d_{2 \rightarrow 2}(\mathcal{A}_{\mathbf{v}}) \leq \frac{1}{2} \mu(\mathbf{U}) \sqrt{k}. \quad (20)$$

We note that both (19) and (20) are valid bounds, and they are not comparable to each other since the relationship between s and k is unknown. It will be clear later that both bounds are useful for computing the entropy integrals. In particular, (19) and (20) are used for computing I_1 and I_2 respectively (as in (23)).

Next, we bound γ_2 -functional $\gamma_2(\mathcal{A}_{\mathbf{v}}, \|\cdot\|_2)$ by estimating the covering numbers $N(\mathcal{A}_{\mathbf{v}}, \|\cdot\|_2, u)$. The derivation is divided into two steps.

Step 1. Decompose $N(\mathcal{A}_{\mathbf{v}}, \|\cdot\|_2, u)$. Let $\mathcal{D}_1 = \{\mathbf{x} \in \mathbb{C}^n : \|\mathbf{x}\|_2^2 \leq 1, \|\mathbf{x}\|_0 \leq s\}$ and define the semi-norm $\|\cdot\|_{K_1}$ as

$$\|\mathbf{x}\|_{K_1} = \|\mathbf{U} \text{diag}(\tilde{\mathbf{B}}\mathbf{x})\|_{2 \rightarrow 2} \quad \forall \mathbf{x} \in \mathbb{C}^n. \quad (21)$$

For the metric space $(\mathcal{D}_1, \|\cdot\|_{K_1})$, we take $\overline{\mathcal{D}_1}$ to be a $\frac{u}{2}$ -net of $\mathcal{D}_1$ with $|\overline{\mathcal{D}_1}| = N(\mathcal{D}_1, \|\cdot\|_{K_1}, \frac{u}{2})$. Let $\mathcal{D}_2 = \{\mathbf{z} \in \mathbb{C}^m : \|\mathbf{z}\|_2^2 \leq 1, \|\mathbf{z}\|_0 \leq k\}$ and define the semi-norm $\|\cdot\|_{K_2}$ as

$$\|\mathbf{z}\|_{K_2} = \|\tilde{\mathbf{B}}^* \text{diag}(\mathbf{U}^* \mathbf{z})\|_{2 \rightarrow 2} \quad \forall \mathbf{z} \in \mathbb{C}^m. \quad (22)$$

For the metric space $(\mathcal{D}_2, \|\cdot\|_{K_2})$, we take $\overline{\mathcal{D}_2}$ to be a $\frac{u}{2}$ -net of $\mathcal{D}_2$ with $|\overline{\mathcal{D}_2}| = N(\mathcal{D}_2, \|\cdot\|_{K_2}, \frac{u}{2})$.

Now, let $\overline{\mathcal{A}_{\mathbf{v}}} = \{(\tilde{\mathbf{z}}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\mathbf{x}))^* : \tilde{\mathbf{x}} \in \overline{\mathcal{D}_1}, \tilde{\mathbf{z}} \in \overline{\mathcal{D}_2}\}$ and remark that $|\overline{\mathcal{A}_{\mathbf{v}}}| \leq |\overline{\mathcal{D}_1}| |\overline{\mathcal{D}_2}|$. It remains to show that for all $\mathbf{v} \in \mathcal{A}_{\mathbf{v}}$, there exists $\tilde{\mathbf{v}} \in \overline{\mathcal{A}_{\mathbf{v}}}$ with $\|\mathbf{v} - \tilde{\mathbf{v}}\|_2 \leq u$.

For any $\mathbf{v} = (\mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\mathbf{x}))^* \in \mathcal{A}_{\mathbf{v}}$, there exist $\tilde{\mathbf{v}} = (\tilde{\mathbf{z}}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\tilde{\mathbf{x}}))^* \in \overline{\mathcal{A}_{\mathbf{v}}}$ with $\tilde{\mathbf{x}} \in \overline{\mathcal{D}_1}$ and $\tilde{\mathbf{z}} \in \overline{\mathcal{D}_2}$ obeying

$\|\mathbf{x} - \tilde{\mathbf{x}}\|_{K_1} \leq \frac{u}{2}$ and $\|\mathbf{z} - \tilde{\mathbf{z}}\|_{K_2} \leq \frac{u}{2}$. This gives

$$\begin{aligned} \|\mathbf{v} - \tilde{\mathbf{v}}\|_2 &= \|\mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\mathbf{x}) - \tilde{\mathbf{z}}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\tilde{\mathbf{x}})\|_2 \\ &= \|\mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\mathbf{x}) - \mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\tilde{\mathbf{x}}) \\ &\quad + \mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\tilde{\mathbf{x}}) - \tilde{\mathbf{z}}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\tilde{\mathbf{x}})\|_2 \\ &\leq \|\mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}(\mathbf{x} - \tilde{\mathbf{x}}))\|_2 + \|(\mathbf{z} - \tilde{\mathbf{z}})^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}\tilde{\mathbf{x}})\|_2 \\ &= \|\mathbf{z}^* \mathbf{U} \text{diag}(\tilde{\mathbf{B}}(\mathbf{x} - \tilde{\mathbf{x}}))\|_2 + \|\tilde{\mathbf{x}}^* \tilde{\mathbf{B}}^* \text{diag}(\mathbf{U}^*(\mathbf{z} - \tilde{\mathbf{z}}))\|_2 \\ &\leq \|\mathbf{z}\|_2 \|\mathbf{U} \text{diag}(\tilde{\mathbf{B}}(\mathbf{x} - \tilde{\mathbf{x}}))\|_{2 \rightarrow 2} \\ &\quad + \|\tilde{\mathbf{x}}\|_2 \|\tilde{\mathbf{B}}^* \text{diag}(\mathbf{U}^*(\mathbf{z} - \tilde{\mathbf{z}}))\|_{2 \rightarrow 2} \\ &\stackrel{(a)}{\leq} \|\mathbf{x} - \tilde{\mathbf{x}}\|_{K_1} + \|\mathbf{z} - \tilde{\mathbf{z}}\|_{K_2} \leq u, \end{aligned}$$

where (a) is due to the fact that $\|\mathbf{z}\|_2 \leq 1$ and $\|\tilde{\mathbf{x}}\|_2 \leq 1$.

Hence,

$$\begin{aligned} N(\mathcal{A}_{\mathbf{v}}, \|\cdot\|_2, u) &\leq |\overline{\mathcal{A}_{\mathbf{v}}}| \\ &\leq N(\mathcal{D}_1, \|\cdot\|_{K_1}, u/2) N(\mathcal{D}_2, \|\cdot\|_{K_2}, u/2). \end{aligned}$$

The γ_2 -functional $\gamma_2(\mathcal{A}_{\mathbf{v}}, \|\cdot\|_2)$ can now be estimated by

$$\begin{aligned} \gamma_2(\mathcal{A}_{\mathbf{v}}, \|\cdot\|_2) &\leq c \int_0^{d_{2 \rightarrow 2}(\mathcal{A})} \sqrt{\log N(\mathcal{A}_{\mathbf{v}}, \|\cdot\|_2, u)} du \\ &\lesssim \underbrace{\int_0^{d_{2 \rightarrow 2}(\mathcal{A})} \sqrt{\log N(\mathcal{D}_1, \|\cdot\|_{K_1}, u/2)} du}_{I_1} \\ &\quad + \underbrace{\int_0^{d_{2 \rightarrow 2}(\mathcal{A})} \sqrt{\log N(\mathcal{D}_2, \|\cdot\|_{K_2}, u/2)} du}_{I_2}. \end{aligned} \quad (23)$$

Step 2. Estimate the covering numbers $N(\mathcal{D}_1, \|\cdot\|_{K_1}, u/2)$ and $N(\mathcal{D}_2, \|\cdot\|_{K_2}, u/2)$ and the entropy integrals. We estimate each covering number in two different ways. For small value of u , we use a volumetric argument. For large value of u , we use the Maurey method ([14, Lemma 4.2], or [5, Problem 12.9]). Then, the resultant covering number estimates can be used to compute the entropy integrals I_1 and I_2 . Similar techniques on the covering number estimation and the entropy integral computation have been used in the CS literature, i.e., [6], [7], [14], [55].

From [7, Equation (28)] and (19), we have

$$I_1 \lesssim \sqrt{\frac{s\tilde{n}}{m}} \mu(\tilde{\mathbf{B}}) (\log s) (\log \tilde{n}). \quad (24)$$

It remains to estimate $N(\mathcal{D}_2, \|\cdot\|_{K_2}, u/2)$ and compute I_2 . 1) *small u .* We observe that $\mathcal{D}_2$ is a subset of the union of $\binom{m}{k}$ unit Euclidean balls $\mathcal{B}_2^k$,

$$\mathcal{B}_2^k := \{\mathbf{z} \in \mathbb{C}^m : \|\mathbf{z}\|_2 \leq 1, |\text{supp}(\mathbf{z})| \leq k\}. \quad (25)$$

For any $\mathbf{z} \in \mathcal{D}_2$,

$$\begin{aligned} \|\mathbf{z}\|_{K_2} &= \|\tilde{\mathbf{B}}^* \text{diag}(\mathbf{U}^* \mathbf{z})\|_{2 \rightarrow 2} \leq \|\mathbf{U}^* \mathbf{z}\|_\infty \leq \max_{i \in [n]} |\langle \mathbf{U}_{(:,i)}, \mathbf{z} \rangle| \\ &\leq \mu(\mathbf{U}) \|\mathbf{z}\|_1 \leq \mu(\mathbf{U}) \sqrt{k} \|\mathbf{z}\|_2 \leq \sqrt{\frac{k}{m}} \|\mathbf{z}\|_2, \end{aligned} \quad (26)$$

where the last step is due to the assumption that $\mu(\mathbf{U}) \sim \frac{1}{\sqrt{m}}$. Therefore,

$$\begin{aligned} N(\mathcal{D}_2, \|\cdot\|_{K_2}, u/2) &\leq \binom{m}{k} N(\mathcal{B}_2^k, \|\cdot\|_{K_2}, u/2) \\ &\leq \binom{m}{k} N(\mathcal{B}_2^k, \sqrt{\frac{k}{m}} \|\cdot\|_2, u/2) \\ &\leq \left(\frac{em}{k}\right)^k (1 + 4\sqrt{\frac{k}{m}} \frac{1}{u})^k, \end{aligned} \quad (27)$$

where the last inequality is an application of [6, Proposition 10.1] and [5, Lemma C.5].

2) *large u*. For any $\mathbf{z} \in \mathcal{D}_2$, we have $\|\mathbf{z}\|_1 \leq \sqrt{k}\|\mathbf{z}\|_2 \leq \sqrt{k}$, which gives

$$\mathcal{D}_2 \subset \sqrt{k}\mathcal{B}_1^m := \{\mathbf{z} \in \mathbb{C}^m : \|\mathbf{z}\| \leq \sqrt{k}\}.$$

Then,

$$\begin{aligned} N(\mathcal{D}_2, \|\cdot\|_{K_2}, u/2) &\leq N(\sqrt{k}\mathcal{B}_1^m, \|\cdot\|_{K_2}, u/2) \\ &= N(\mathcal{B}_1^m, \|\cdot\|_{K_2}, u/(2\sqrt{k})). \end{aligned}$$

Based on the Maurey method, for $0 < u < \frac{1}{2}\mu(\mathbf{U})\sqrt{k}$, the covering number can be estimated by [6, Lemma 8.3]

$$\begin{aligned} \sqrt{\log N(\mathcal{D}_2, \|\cdot\|_{K_2}, u/2)} &\lesssim \sqrt{k}\mu(\mathbf{U})\sqrt{\log \tilde{n} \log mu^{-1}} \\ &\leq \sqrt{\frac{k}{m}}\sqrt{\log \tilde{n} \log mu^{-1}}. \end{aligned} \quad (28)$$

We note that the estimation based on Maurey method depends on the range of the parameter u (see [6, Lemma 8.3]), which is the reason why we employ different bounds ((19) and (20)) when computing the entropy integrals I_1 and I_2 .

We now combine the results (27) and (28) to estimate the entropy integral I_2 : we apply the first bound for $0 < u \leq \frac{1}{10}\sqrt{\frac{1}{m}}$, and the second bound for $\frac{1}{10}\sqrt{\frac{1}{m}} < u \leq d_{2 \rightarrow 2}(\mathcal{A}_v) = \frac{1}{2}\sqrt{\frac{k}{m}}$. It reveals that

$$I_2 \lesssim \sqrt{\frac{k}{m}} \log \tilde{n} \log k. \quad (29)$$

Combine (23), (24) and (29)

$$\begin{aligned} \gamma_2(\mathcal{A}_v, \|\cdot\|_2) &\lesssim \sqrt{\frac{s\tilde{n}}{m}} \mu(\tilde{\mathbf{B}}) (\log s) (\log \tilde{n}) \\ &\quad + \sqrt{\frac{k}{m}} \log \tilde{n} \log k. \end{aligned} \quad (30)$$

Finally, we are ready to complete the proof by applying Proposition A.2. For the assumption on m and p , $\delta \in (0, 1)$,

$$\begin{aligned} m &\geq c_1 \delta^{-2} s \tilde{n} \mu^2(\tilde{\mathbf{B}}) \log^2 s \log^2 \tilde{n} \\ m &\geq c_2 \delta^{-2} k \log^2 k \log^2 \tilde{n}, \end{aligned}$$

we have, by (19),

$$d_F(\mathcal{A}_v) = d_{2 \rightarrow 2}(\mathcal{A}_v) \lesssim \frac{\delta}{\log s \log \tilde{n}}, \quad \gamma_2(\mathcal{A}_v, \|\cdot\|_2) \lesssim \delta.$$

By substituting the above results into Proposition A.2 (let $t = \log s \log \tilde{n}$), one obtains

$$\mathbb{P}(\sup_{\mathbf{v} \in \mathcal{A}_v} |\langle \mathbf{v}, \boldsymbol{\xi} \rangle| \leq c\delta) \geq 1 - \exp(-\log^2 s \log^2 \tilde{n}). \quad (31)$$

The proof is completed by incorporating the constant c into c_1, c_2 .

APPENDIX B PROOF OF THEOREM III.3

Recall that in the measurement model $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{H}\mathbf{z}^* + \mathbf{w}$, $\mathbf{A} = \sqrt{\frac{n}{M}}\mathbf{R}_{\Omega'}\mathbf{G}$ is a randomly sub-sampled unitary matrix and $\mathbf{H} \in \mathbb{C}^{M \times M}$ is a unitary matrix with $\mu(\mathbf{H}) \sim 1/\sqrt{M}$.

The following Lemma from [56] is needed.

Lemma B.1 (Theorem 3.3 [56]). *For the matrix $\mathbf{A} = \sqrt{\frac{n}{m}}\mathbf{R}_{\Omega'}\mathbf{G}$, if for $\delta \in (0, 1)$,*

$$m \geq c\delta^{-2}s \log^4 n, \quad (32)$$

then with probability at least $1 - n^{-\log^3 n}$ the restricted isometry constant δ_s of $\mathbf{A}$ satisfies $\delta_s \leq \delta$.

The (s, k) -RIP associated with $\Theta = [\mathbf{A}, \mathbf{H}]$ can be bounded by

$$\delta_{s,k} \leq \underbrace{\sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} \left| \|\mathbf{A}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right|}_{\delta_1} + 2 \underbrace{\sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} |\langle \mathbf{A}\mathbf{x}, \mathbf{H}\mathbf{z} \rangle|}_{\delta_2}, \quad (33)$$

where $\mathcal{T} := \{(\mathbf{x}, \mathbf{z}) : \|\mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2 = 1, \|\mathbf{x}\|_0 \leq s, \|\mathbf{z}\|_0 \leq k, \mathbf{x} \in \mathbb{C}^n, \mathbf{z} \in \mathbb{C}^m\}$.

By Lemma B.1, we have $\delta_1 \leq \delta/2$ holds with probability $1 - n^{-\log^3 n}$ for any $\delta \in (0, 1)$ if $m \geq c\delta^{-2}s \log^4 n$.

Define a random vector $\mathbf{d} \in \mathbb{C}^n$ with i.i.d. entries satisfying $\lambda_i = \sqrt{\frac{m(n-m)}{n^2}}d_i + \frac{m}{n}$. Assume $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ and $\mathbf{H}^*\mathbf{R}_{\Omega'} = \mathbf{U}'$. We have,

$$\begin{aligned} \delta_2 &= 2 \sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} \left| \left\langle \sqrt{\frac{n}{M}}\mathbf{R}_{\Omega'}\mathbf{G}\mathbf{x}, \mathbf{H}\mathbf{z} \right\rangle \right| \\ &= 2 \sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} \left| \sqrt{\frac{n}{M}}\mathbf{z}^*\mathbf{U}'\mathbf{\Lambda}\mathbf{G}\mathbf{x} \right| \\ &= 2 \sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} \left| \sqrt{\frac{n}{M}}\mathbf{z}^*\mathbf{U}'\text{diag}(\mathbf{G}\mathbf{x})\boldsymbol{\lambda} \right| \\ &\leq 2 \underbrace{\sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} \left| \frac{1}{2}\sqrt{\frac{n}{M}}\mathbf{z}^*\mathbf{U}'\text{diag}(\mathbf{G}\mathbf{x})\mathbf{d} \right|}_{t_1} \\ &\quad + 2 \underbrace{\sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} \left| \sqrt{\frac{m^2}{Mn}}\mathbf{z}^*\mathbf{U}'\mathbf{G}\mathbf{x} \right|}_{t_2}, \end{aligned}$$

where the last inequality is due to the fact that $\sqrt{\frac{m(n-m)}{n^2}} \leq \frac{1}{2}$ for any $m \leq n$.

Since $\boldsymbol{\lambda}$ is a random Bernoulli vector with i.i.d. entries, by construction $\mathbf{d}$ is a length- n random vector with independent, zero-mean, unit-variance, and L -subgaussian entries. Hence, the bound for t_1 can be formulated as the supremum of a stochastic process with the index $\mathcal{A}_r$, where $\mathbf{r} = \sqrt{\frac{n}{M}}\mathbf{z}^*\mathbf{U}'\text{diag}(\mathbf{G}\mathbf{x})$ and $\mathcal{A}_r := \{\mathbf{r} : \|\mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2 =$

1, $\|\mathbf{x}\|_0 \leq s, \|\mathbf{z}\|_0 \leq k\}$. For any $\mathbf{r} \in \mathcal{A}_{\mathbf{r}}$,

$$\begin{aligned} \|\mathbf{r}\|_2 &= \sqrt{\frac{n}{M}} \|\mathbf{z}^* \mathbf{U}' \text{diag}(\mathbf{G}\mathbf{x})\|_2 \\ &\leq \sqrt{\frac{n}{M}} \|\mathbf{z}^* \mathbf{H}^* \mathbf{R}_{\Omega'}\|_2 \|\mathbf{G}\mathbf{x}\|_\infty \\ &= \sqrt{\frac{n}{M}} \|\mathbf{z}\|_2 \max_{j \in [n]} \{|\langle \mathbf{G}_{(j,:)}, \mathbf{x} \rangle|\} \\ &\leq \frac{1}{2} \sqrt{\frac{n}{M}} \mu(\mathbf{G}) \sqrt{s}, \\ \|\mathbf{r}\|_2 &= \sqrt{\frac{n}{M}} \|\mathbf{z}^* \mathbf{U}' \text{diag}(\mathbf{G}\mathbf{x})\|_2 \\ &\leq \sqrt{\frac{n}{M}} \|\mathbf{z}^* \mathbf{H}^*\|_\infty \|\mathbf{R}_{\Omega'} \mathbf{G}\mathbf{x}\|_2 \\ &\leq \sqrt{\frac{n}{M}} \mu(\mathbf{H}) \sqrt{k} \|\mathbf{G}_{(\Omega',:)} \mathbf{x}\|_2 \\ &\leq \frac{1}{2} \sqrt{n} \mu(\mathbf{G}) \mu(\mathbf{H}) \sqrt{k}. \end{aligned}$$

Therefore,

$$\begin{aligned} d_F(\mathcal{A}_{\mathbf{r}}) &\leq \frac{1}{2} \sqrt{\frac{n}{M}} \mu(\mathbf{G}) \sqrt{s}, \\ d_F(\mathcal{A}_{\mathbf{r}}) &\leq \frac{1}{2} \sqrt{n} \mu(\mathbf{G}) \mu(\mathbf{H}) \sqrt{k}. \end{aligned}$$

By following the same proof steps as in Appendix A, we have

$$\mathbb{P}(\sup_{\mathbf{r} \in \mathcal{A}_{\mathbf{r}}} |\langle \mathbf{r}, \mathbf{d} \rangle| \leq c\delta) \geq 1 - \exp(-\log^2 s \log^2 n) \quad (34)$$

provided that

$$\begin{aligned} M &\geq c\delta^{-2} s n \mu^2(\mathbf{G}) \log^2 s \log^2 n \\ M &\geq c\delta^{-2} k n \mu^2(\mathbf{G}) \log^2 k \log^2 n. \end{aligned}$$

Bernstein's inequality [57, Theorem A.1.13] gives, for any $\nu > 0$,

$$\mathbb{P}(M > (1 - \nu)m) \geq 1 - \exp\left(-\frac{m\nu^2}{2}\right). \quad (35)$$

Hence, if

$$\begin{aligned} m &\geq \frac{1}{1 - \nu} c\delta^{-2} s n \mu^2(\mathbf{G}) \log^2 s \log^2 n \\ m &\geq \frac{1}{1 - \nu} c\delta^{-2} k n \mu^2(\mathbf{G}) \log^2 k \log^2 n, \end{aligned}$$

then

$$\begin{aligned} \mathbb{P}(\sup_{\mathbf{r} \in \mathcal{A}_{\mathbf{r}}} |\langle \mathbf{r}, \mathbf{d} \rangle| \leq c\delta) &\geq 1 - \exp(-\log^2 s \log^2 n) \\ &\quad - \exp\left(-\frac{m\nu^2}{2}\right). \end{aligned}$$

By assuming $m \geq 2c' \log n$ and $\nu = \sqrt{\frac{2c' \log n}{m}}$, the above probability of success can be written as

$$\mathbb{P}(\sup_{\mathbf{r} \in \mathcal{A}_{\mathbf{r}}} |\langle \mathbf{r}, \mathbf{d} \rangle| \leq c\delta) \geq 1 - n^{-\log^2 s \log n} - n^{-c'}. \quad (36)$$

For the second term, we have

$$\begin{aligned} t_2 &= \sup_{(\mathbf{x}, \mathbf{z}) \in \mathcal{T}} \left| \frac{m^2}{Mn} \mathbf{z}^* \mathbf{U}' \mathbf{G}\mathbf{x} \right| \\ &\leq \frac{m^2}{Mn} \|\mathbf{z}\|_2 \|\mathbf{x}\|_2 \\ &\leq \frac{1}{2} \frac{m^2}{Mn}, \end{aligned}$$

where the last inequality is due to $\|\mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2 = 1$. Therefore, $t_2 \leq \delta/2$ for any $\delta \in (0, 1)$ if $M \sim m$ and $m \leq \delta n$. By Bernstein's inequality, this condition can be satisfied with probability exceeding $1 - n^{-c'}$ as long as $m \leq c\delta^2 n$. Theorem III.3 is proved by combining the above results.

REFERENCES

- [1] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006.
- [2] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [4] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge University Press, 2012.
- [5] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.
- [6] H. Rauhut, "Compressive sensing and structured random matrices," *Theoretical foundations and numerical methods for sparse recovery*, vol. 9, pp. 1–92, 2010.
- [7] P. Zhang, L. Gan, S. Sun, and C. Ling, "Modulated unit-norm tight frames for compressed sensing," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3974–3985, Aug 2015.
- [8] J. Bourgain, "An improved estimate in the restricted isometry problem," in *Geometric Aspects of Functional Analysis*. Springer, 2014, pp. 65–70.
- [9] I. Haviv and O. Regev, "The restricted isometry property of subsampled fourier matrices," in *Geometric Aspects of Functional Analysis*. Springer, 2017, pp. 163–179.
- [10] X. Li, "Compressed sensing and matrix completion with constant proportion of corruptions," *Constructive Approximation*, vol. 37, no. 1, pp. 73–99, 2013.
- [11] N. Nguyen and T. Tran, "Exact recoverability from dense corrupted observations via l_1 -minimization," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2017–2035, April 2013.
- [12] E. J. Candes and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.
- [13] B. Adcock, A. Bao, A. Narayan *et al.*, "Compressed sensing with sparse corruptions: Fault-tolerant sparse collocation approximations," *arXiv preprint arXiv:1703.00135*, 2017.
- [14] F. Krahmer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property," *Communications on Pure and Applied Mathematics*, 2014.
- [15] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 92–101, 2008.
- [16] C. Franke and M. Gertz, "ORDEN: Outlier region detection and exploration in sensor networks," in *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. ACM, 2009, pp. 1075–1078.
- [17] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *Commun. Surveys Tuts.*, vol. 12, no. 2, pp. 159–170, 2010.
- [18] Z. Charbiwala, S. Chakraborty, S. Zahedi, Y. Kim, M. Srivastava, T. He, and C. Bisdikian, "Compressive oversampling for robust data transmission in sensor networks," in *INFOCOM, 2010 Proceedings IEEE*, March 2010, pp. 1–9.
- [19] Z. Li, F. Wu, and J. Wright, "On the systematic measurement matrix for compressed sensing in the presence of gross errors," in *Data Compression Conference (DCC), 2010*. IEEE, 2010, pp. 356–365.

- [20] J. N. Laska, M. Davenport, R. G. Baraniuk *et al.*, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*. IEEE, 2009, pp. 1556–1560.
- [21] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, 2012.
- [22] B. M. Popovic, "Synthesis of power efficient multitone signals with flat amplitude spectrum," *IEEE Trans. Commun.*, vol. 39, no. 7, pp. 1031–1033, 1991.
- [23] J. A. Davis and J. Jedwab, "Peak-to-mean power control in OFDM, Golay complementary sequences, and Reed-Muller codes," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2397–2417, 1999.
- [24] B. Levitt, "FH/MFSK performance in multitone jamming," *IEEE J. Sel. Areas Commun.*, vol. 3, no. 5, pp. 627–643, 1985.
- [25] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708–1721, 2010.
- [26] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, 2010.
- [27] J. Meng, W. Yin, Y. Li, N. T. Nguyen, and Z. Han, "Compressive sensing based high-resolution channel estimation for OFDM system," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 1, pp. 15–25, 2012.
- [28] K. Li, L. Gan, and C. Ling, "Convolutional compressed sensing using deterministic sequences," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 740–752, 2013.
- [29] D. Umehara, H. Nishiyori, and Y. Morihira, "Performance evaluation of CMFB transmultiplexer for broadband power line communications under narrowband interference," in *Power Line Communications and Its Applications, 2006 IEEE International Symposium on*. IEEE, 2006, pp. 50–55.
- [30] A. Gomaa and N. Al-Dhahir, "A sparsity-aware approach for NBI estimation in MIMO-OFDM," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1854–1862, 2011.
- [31] J. Romberg, "Compressive sensing by random convolution," *SIAM J. Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, 2009.
- [32] G. Puy, P. Vanderghynst, R. Gribonval, and Y. Wiaux, "Universal and efficient compressed sensing by spread spectrum and application to realistic Fourier imaging techniques," *EURASIP J. Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–13, 2012.
- [33] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, 2010.
- [34] J. Romberg and R. Neelamani, "Sparse channel separation using random probes," *Inverse Problems*, vol. 26, no. 11, p. 115015, 2010.
- [35] J. P. Slavinsky, J. N. Laska, M. A. Davenport, and R. G. Baraniuk, "The compressive multiplexer for multi-channel compressive sensing," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 3980–3983.
- [36] Y. Rivenson, A. Stern, and B. Javidi, "Single exposure super-resolution compressive imaging by double phase encoding," *Optics express*, vol. 18, no. 14, pp. 15 094–15 103, 2010.
- [37] J. Li, J. S. Li, Y. Y. Pan, and R. Li, "Compressive optical image encryption," *Scientific reports*, vol. 5, 2015.
- [38] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, 2009.
- [39] R. F. Marcia, C. Kim, J. Kim, D. J. Brady, and R. M. Willett, "Fast disambiguation of superimposed images for increased field of view," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 2620–2623.
- [40] D.-S. Pham and S. Venkatesh, "Improved image recovery from compressed data contaminated with impulsive noise," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 397–405, 2012.
- [41] M. Filipović, "Reconstruction of sparse signals from highly corrupted measurements by nonconvex minimization," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.
- [42] R. Saab, R. Chartrand, and O. Yilmaz, "Stable sparse approximations via nonconvex optimization," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 3885–3888.
- [43] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2008, pp. 3869–3872.
- [44] R. Foygel and L. Mackey, "Corrupted sensing: Novel guarantees for separating structured signals," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1223–1247, Feb 2014.
- [45] Y. Chen, C. Caramanis, and S. Mannor, "Robust sparse regression under adversarial corruption," in *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 2013, pp. 774–782.
- [46] C. Studer and R. G. Baraniuk, "Stable restoration and separation of approximately sparse signals," *Applied and Computational Harmonic Analysis*, vol. 37, no. 1, pp. 12–35, 2014.
- [47] G. Pope, A. Bracher, and C. Studer, "Probabilistic recovery guarantees for sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3104–3116, May 2013.
- [48] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.
- [49] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 429–443, 2011.
- [50] Y. Chi, "Convex relaxations of spectral sparsity for robust super-resolution and line spectrum estimation," in *Wavelets and Sparsity XVII*, vol. 10394. International Society for Optics and Photonics, 2017, p. 103941G.
- [51] C. Fernandez-Granda, G. Tang, X. Wang, and L. Zheng, "Demixing sines and spikes: Robust spectral super-resolution in the presence of outliers," *arXiv preprint arXiv:1609.02247*, 2016.
- [52] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, Mar. 2014.
- [53] —, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.
- [54] M. Talagrand, *The generic chaining*. Springer, 2005, vol. 154.
- [55] A. Eftekhari, H. L. Yap, C. J. Rozell, and M. B. Wakin, "The restricted isometry property for random block diagonal matrices," *arXiv preprint arXiv:1210.3395*, 2012.
- [56] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, 2008.
- [57] N. Alon and J. H. Spencer, *The probabilistic method*. John Wiley & Sons, 2004.