Received April 3, 2019, accepted April 24, 2019, date of publication April 30, 2019, date of current version August 12, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2914116*

# Information Theoretic Bounds for Sparse Reconstruction in Random Noise

**JUNJIE CHEN, FANGQI ZHU, AND QILIAN LIANG, (Fellow, IEEE)**
Department of Electrical Engineering, The University of Texas at Arlington, Arlington, TX 76010, USA
Corresponding author: Junjie Chen (junjie.chen@mavs.uta.edu)

**ABSTRACT** Compressive sensing (CS) plays a pivotal role in the signal processing and we address on the issues, i.e., the information-theoretic analysis of CS under random noise in this paper. To distinguish from existing literature, we aim at providing a precise reconstruction of the source signal. From the analysis of the recovery performance, we calculate the lower band upper bound of the probability of error for CS. To be more specific, we provide more discussions for the case where both the source and the noise follow Gaussian distribution. It has been proved that perfect reconstruction of the signal vector is impossible if the corresponding conditions are not satisfied, which can be served as the theoretical reference of noisy CS. In terms of the necessary proofs, we leverage the results from information theory and estimation theory. The compression of real underwater acoustic sensor network (UWASN) data is applied to verify the theoretical bounds derived in this paper.

**INDEX TERMS** Compressive sensing, random noise, probability of error, theoretical performance bounds, underwater acoustic sensor networks.

## I. INTRODUCTION

Compressive sensing (CS) is one of the major tools [1]–[3], [12] that transforms an inherently sparse and high-dimensional signal vector in some space to a much lower dimensional representation. Compared to the classical Nyquist sampling requirement, the CS requires lower data to achieve similar performance, which demonstrates its tractability in real application.

To address the CS problem, let's consider the source signal vector $\underline{x} \in \mathbb{R}^N$ that is $K$-sparse in the basis $\Psi$ [1], where $\Psi \in \mathbb{R}^{N\times N}$ and also it is orthogonal. The compressed version of the original $\underline{x}$ can be represented as [1]–[3],

$$\underline{y} = \Phi\underline{x} + \underline{w} \tag{1}$$

where $\Phi \in \mathbb{R}^{M\times N}$ is the measurement matrix, and $\underline{w} \in \mathbb{R}^M$ is the additive noise and $\underline{y} \in \mathbb{R}^M$ is the observation vector. For the sparsity condition, we have $M \ll N$.

The purpose of reconstruction based on CS is to obtain the original source information $\underline{x}$ from the sparse observation $\underline{y}$. Since and $\underline{x}$ is sparse in terms of the basis $\Psi$, the reconstruction process is aim to find $\underline{x}$ such that $\Psi^{-1}\underline{x}$ has limited nonzero entries. The minimization of the $l_0$ norm of the $\underline{x}$ directly yields an intractable problem [13]. However, we can relax the above problem and recover the source by formulation the above problem as the following constrained optimization problem [2]:

$$\begin{aligned}\hat{\underline{\theta}} &= \text{argmin}\|\underline{\theta}\|_{l1}\ s.t.\ \|\underline{y} - \Phi\underline{x}\|_{l2} \\ &= \|\underline{y} - \Phi\Psi\underline{\theta}\|_{l2} \leq \epsilon \end{aligned} \tag{2}$$

For many scenarios, the data inherent in the model or the problem formulation lies on the fact the fact the signal is inherently sparse [3], the emerging CS provides a powerful tool to handle these problems in a concise and tractable way. A model-based CS theory with provable performance guarantees was introduced in [4] by investigating read-world and irregular signal models by dropping out simple assumptions in the CS theory. In [5], the robust PCA scheme permeated benefits from variable selection and compressive sampling. The algorithm related with CS can be also solved in a recursive way and an adaptive algorithms for recursive estimation and tracking of sparse signals is provided in [6]. Sparsity in the eigenspace of signal covariance matrix was discussed in [7]. The recovery of the unknown sparse vector of regression coefficients was studied in [8], and sparse TLS algorithms were provided to tackle the perturbation problem in nature. The work in [9] connects CS and linear regression to extend the range of sparsity-exploiting algorithms for model mismatch caused by unreliable sensors (outliers) that lead

The associate editor coordinating the review of this manuscript and approving it for publication was Jinhwan Koh.

to sparse residuals, even the non-sparse signal case is discussed. Another interesting case is about the CS recovery of the uniform quantization, which is studied in [10], and the authors show theoretically that the reconstruction error is bounded under certain assumptions. The application of CS to real SAR data was studied in [11], however in their work, they just provide limited cases which is not ubiquitous to other scenarios.

### A. RELATED WORK

One of the previous research focus is to analyze the performance of CS from information-theoretic point of view. The matrix completion problem under the information theoretic limit was given in [14]. Several theoretical results regarding the recovery of a low-rank matrix from just a few measurements were presented, and the error bounds regarding low-rank matrices were extended in [15]. One of the seminal manuscripts about the theory of CS was introduced in [22], in which the CS mechanism selects sensing vectors independently from a determined probability distribution without the requirement of the restricted isometry property and a given random model for the signal. Wainwright [16] studied the information-theoretic limits of sparsity recovery for a noisy linear observation model based on random measurement matrices drawn from general Gaussian measurement matrices. Both sufficient and necessary conditions were derived. The recovery of the non-asymptotic bounds, block-sparsity and low-rank matrix in the Gaussian ensemble were studied in [17]. Arias-Castro *et al.* [18] provided some MSE bounds of adaptive sensing estimation. In [19], The measurement matrix which is generated from the noisy projections scheme defined on different basis is proposed and the bounds of precise recovering of the support set of the CS signal is investigated by introducing the information theory. The concept of CS was extended to signals that are sparse in a redundant dictionary instead of the original orthonormal space [21]. Tang and Nehorai formulated the support recovery problems for jointly sparse signals as binary and multiple-hypothesis testings and derived corresponding bounds on the probability of error ($P_e$) in [20]. The metric for the quantity of measurements for recovering a compressed signal in the noise is analyzed in [23] and the average distortion introduced by quantization of compressive sensing (CS) measurements was studied in [24]. The reconstruction rate distortion performance was discussed in [28] and [29]. Reference [30] studied the upper and lower bounds of probability error for compressive sensing with Gaussian noise. A lower bound on the mean-squared error (MSE) achievable for sparse estimation with Gaussian noise was given in [31]. An approximation in terms of the minimum bit rate to achieve the best average performance from a CS system was developed in [27]. Rate distortion theoretic performance bounds to sensing and reconstruction of CS from noisy projections were provides in [32]. Reference [36] provided the lower bound of the $P_e$ for CS but the results are limited, and our work will extend to more cases about the theoretical limits for CS. The Cramer-Rao lower bound (CRLB) for CS was studied in [33] and [34] based on the assumptions that the location of the non-zero delayed taps in advance is known for the signal estimator.

### B. CONTRIBUTIONS IN THIS PAPER

By assuming that both the source and noise follow Gaussian distribution, we provide a information-theoretical analysis of noisy CS, and give the closed-form representation of the lower and upper bound in terms of the $P_e$. We have to mention that although the performance results are derived for the above special case those results still could be used as theoretical reference of real application of CS [41]. The entropy is a measure of the average uncertainty in the random variable. This indicates that our results provide the bounds of the worst case. Different from other literature, we consider the precise reconstruction of the signal instead of just the recovery of the support of the signal. Then we form the lower and upper bound of $P_e$ for general case without specific requirement on the measurement matrix. Furthermore, we provide analysis for the case that the measurement matrix $\Phi$ is Bernoulli matrix, and the related lower and upper bound of the probability for noisy CS are given. Finally, real underwater acoustic sensor network (UWASN) data are compressed to verify the theoretical results derived in this paper.

### C. CONTENT

The rest of this paper is organized as follows. In Section II, we give the related assumptions. The performance bounds of compressive sensing with random noise is shown in Section III. In Section III-A, the lower bounds of the $P_e$ are provided for both general case and the special case of Bernoulli distribution based measurement matrix. In Section III-B, the upper bounds of $P_e$ for CS are provided for both the general case and special case. The detailed proofs are given in Section IV. The experimental results with real underwater acoustic sensor network data are provided in Section V. Conclusions are presented in Section VI.

## II. PROBLEM SETTING

Before diving into the details of the theoretical study, we would like to provide the following assumptions for the problem setting:

1) The source information $\underline{x} \in \mathbb{R}^N$ is $K-$sparse in the sparsity basis $\Psi$, i.e., there are only $K$ non-zero entries in $\underline{\theta} = \Psi^T \underline{x}$, and each element of $\underline{x} \sim \mathcal{N}(0, \sigma_x^2)$. We consider the noise vector $\underline{w} \in \mathbb{R}^M$ and $\underline{w} \sim \mathcal{N}(0, \sigma_w^2)$.
2) This paper provides theoretical analysis of the compressive sensing for the general case without any assumption of the measurement matrix as well as the special case with Bernoulli measurement matrix. To be more specific about the special case, the measurement matrix $\Phi$ is the Bernoulli matrix. The i.i.d entries of $\phi_{m,n}$ have value of either 1 or $-1$ with

$\Pr(\phi_{m,n} = \pm 1) = 0.5$. We also assume that the measurement matrix $\Phi$ is known in advance for both the transmitter and the receiver.

3) For the special case with Bernoulli measurement matrix, the observations can be represented as the following,

$$y_m = \sum_{n=1}^{N} \phi_{m,n} x_n + w_m = \sum_{i=1}^{N_1} (+1) x_i + \sum_{j=1}^{N_2} (-1) x_j + w_m \quad (3)$$

$$N_1 + N_2 = N \quad (4)$$

where $N$ is the length of $x$. Because $\Pr(\phi_{m,n} = \pm 1) = 0.5$, from the strong law of large number (LLN) theorem, as $N \to \infty$, $N_1 = N_2 = \frac{N}{2}$, so that:

$$y_m = \frac{N}{2} \sum_{i=1}^{N_1} x_i - \frac{N}{2} \sum_{j=1}^{N_2} x_j + w_m \quad (5)$$

As both $x_n$ and $w_m$ follow Gaussian distribution with zero mean and different variance $\sigma_x^2$ and $\sigma_w^2$, $y_m$ is also a Gaussian random variable with $\mathcal{N}(0, \sigma_x^2 + \sigma_w^2)$.

## III. PERFORMANCE BOUNDS OF CS WITH RANDOM NOISE: FUNDAMENTAL RESULTS

### A. LOWER BOUNDS

Consider the CS with Bernoulli measurement matrix and additive noise and we assume that both of them follow the Gaussian distribution and $Pr(\phi_{m,n} = \pm 1) = 0.5$. We will have the following theorem in terms of the lower of bound of the $P_e$ as follows:

*Theorem 1: Consider the noisy CS scenario, given the input* $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, *and the elements of* $\phi_{m,n}$ *are i. i. d. with* $Pr(\phi_{m,n} = \pm 1) = 0.5$, *then the lower bound of* $P_e$ *can be derived as:*

$$P_e \geq 1 - \frac{\frac{1}{2}\log(\frac{N\sigma_x^2 + \sigma_w^2}{\sigma_w^2})^M + 1}{\frac{1}{2}\log[(2\pi e \sigma_x^2)^N]} \quad (6)$$

where $N$ stands for the length of the information. We notice that if $\sigma_x^2 \geq 1$, and $N \to \infty$, $\frac{1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \approx 0$, therefore,

1) If $(2\pi e \sigma_x^2)^N \geq (\frac{N\sigma_x^2+\sigma_w^2}{\sigma_w^2})^M$, then $P_{elb} \geq 0$, which means that error always exit so that perfect reconstruction of the information vector is impossible.
2) If $(2\pi e \sigma_x^2)^N \leq (\frac{N\sigma_x^2+\sigma_w^2}{\sigma_w^2})^M$, then $P_{elb} \leq 0$, so that perfect reconstruction of the information vector is possible in this case.

In a nutshell, Theorem 1 demonstrates a special case of the lower bound of the $P_e$ when the measurement matrix $\Phi$ follows Bernoulli distribution with $Pr(\phi_{m,n} = \pm 1) = 0.5$. We can extend to a more general lower bound of the $P_e$ without any special assumption of the measurement matrix, which is given in Theorem 2 as follows:

*Theorem 2: In noisy CS, given the input* $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, *and the measurement matrix is* $\Phi$, *then the lower bound of the* $P_e$ *will be as follows,*

$$P_e \geq \frac{N\log[2\pi e\lambda_{xy_{min}}] + M\log(\frac{\lambda_{xy_{min}}}{\lambda_{y_{max}}})}{N\log[2\pi e\sigma_x^2]} \quad (7)$$

From analysis, we conclude that:

(1). If $(2\pi e\lambda_{xy_{min}})^N \geq (\frac{\lambda_{y_{max}}}{\lambda_{xy_{min}}})^M$, then $P_{elb} \geq 0$, this means that perfect reconstruction of the information vector is impossible, as there will always be certain error.

(2). If $(2\pi e\lambda_{xy_{min}})^N \leq (\frac{\lambda_{y_{max}}}{\lambda_{xy_{min}}})^M$, then $P_{elb} \leq 0$, this means that perfect reconstruction of the information vector is possible.

### B. UPPER BOUNDS

Since a linear MMSE estimator $\underline{y} = \Phi\underline{x} + \underline{w}$ is an affine transformation of Gaussian random vector $\underline{y}$, hence $\hat{\underline{x}}$ is multivariate Gaussian as well, i.e., the estimated expression for linear MMSE estimator also follows Gaussian distribution.

From the reference [41], if $\underline{x}$ and $\hat{\underline{x}}$ are independent and they have the following distribution functions $\underline{x} \sim p(\underline{x})$, and $\hat{\underline{x}} \sim p(\hat{\underline{x}})$, then

$$Pr(\underline{x} = \hat{\underline{x}}) \geq 2^{-H(p(\underline{x})) - D(p(\underline{x})\|p(\hat{\underline{x}}))} \quad (8)$$

where $D(p(\underline{x})\|p(\hat{\underline{x}}))$ is the Kullback-Leibler distance (KLD) between the corresponding two probability mass functions $p(\underline{x})$ and $p(\hat{\underline{x}})$.

Therefore, the $P_e$ between $\underline{x}$ and $\hat{\underline{x}}$ is upper bounded by Theorem 3, with Bernoulli measurement matrix. Theorem 3 provide the upper bound of the $P_e$ in a information-theoretic sense of viewpoint.

*Theorem 3: In noisy CS, if* $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, *and the entries of the measurement matrix* $\phi_{m,n}$ *follows Bernoulli distribution with* $Pr(\phi_{m,n} = \pm 1) = 0.5$, *then the upper bound of* $P_e$ *can be expressed as:,*

$$P_e \leq 1 - (\pi e \sigma_x^2)^{-\frac{N}{2}} 2^{-\frac{N}{2}(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2} - \ln(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2}))} \quad (9)$$

A more general upper bound of the $P_e$ without any special assumption of the measurement matrix is provided in Theorem 4, which is from the estimation theory point of view.

*Theorem 4: In noisy CS, consider* $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, *and the measurement matrix is* $\Phi$, *then the upper bound of the* $P_e$ *will be as follows,*

$$P_e \leq \frac{1}{(2\pi)^{N/2} \cdot det((\frac{1}{\sigma_x^2} I_N + \frac{1}{\sigma_w^2}\Phi^T\Phi)^{-1})^{1/2}} \quad (10)$$

Observe that Theorem 4 is a general case of the upper bound of the $P_e$ for CS with random noise. For the special case when the measurement matrix follows Bernoulli distribution with $Pr(\phi_{m,n} = \pm 1) = 0.5$, the corresponding result is shown in Theorem 5.

*Theorem 5: In noisy CS, if* $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, *and the entries of the measurement matrix* $\phi_{m,n}$ *follows Bernoulli distribution with* $Pr(\phi_{m,n} = \pm 1) = 0.5$, *then the upper bound of the $P_e$ will be as follows,*

$$P_e \leq \frac{1}{(\frac{2\pi\sigma_x^2\sigma_w^2}{\sigma_w^2+M\sigma_x^2})^{N/2}} \tag{11}$$

*Proof:* When the measurement matrix $\Phi$ is Bernoulli matrix, the entries of $\phi_{m,n}$ are i.i.d. with $Pr(\phi_{m,n} = \pm 1) = 0.5$, and if the length $N$ of $x$ is sufficiently large, $\Phi^T\Phi = M \cdot I_N$, from which we could achieve the result in Theorem 5. □

## IV. PROOFS

### A. PROOF OF THEOREM 1

*Proof:* Given that each $x_n$ is a Gaussian i.i.d random variable, and $C_x = \sigma_x^2 I_N$, we employ the definition of the differential entropy of $\underline{x}$ from [41] here as follows:

$$h(\underline{x}) = \frac{1}{2}\log(2\pi e\sigma_x^2)^N. \tag{12}$$

At the receiver side, $\underline{y} = \Phi\underline{x} + \underline{w}$, and also $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, and $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, then we obtain the covariance matrix of $\underline{y}$ as,

$$\begin{aligned} C_{yy} &= E[(\underline{y} - \Phi\underline{\mu_x})(\underline{y} - \Phi\underline{\mu_x})^T] \\ &= E[(\Phi(\underline{x} - \underline{\mu_x}) + \underline{w})(\Phi(\underline{x} - \underline{\mu_x}) + \underline{w})^T] \\ &= \Phi E[\underline{x}\underline{x}^T]\Phi^T + C_w \\ &= \Phi C_{xx}\Phi^T + C_w \\ &\overset{(i)}{=} (N\sigma_x^2)I_M + \sigma_w^2 I_M \\ &= (N\sigma_x^2 + \sigma_w^2)I_M \end{aligned} \tag{13}$$

where equality $(i)$ follows from the fact that $Pr(\phi_{m,n} = \pm 1) = 0.5$, if $N$ is sufficient large, when $m \neq m'$, $E[(\sum_{n=1}^N \phi_{m,n}\phi_{m',n}] = 0$, therefore, $\Phi\Phi^T = N \cdot I_M$.

Since $\underline{y} = \Phi\underline{x} + \underline{w}$, we could notice that the mean of $\underline{y}$ is $\underline{0}$, with covariance matrix $C_{yy}$, i.e., $\underline{y} \sim \mathcal{N}(0, (N\sigma_x^2 + \sigma_w^2)I_M)$.

The joint likelihood function is,

$$p_{y|x}(\underline{y}; \underline{x}) = \frac{1}{2\pi(N\sigma_x^2 + \sigma_w^2)^{\frac{M}{2}}}\exp(-\frac{1}{2}\underline{y}^T \cdot C_{yy}^{-1} \cdot \underline{y}) \tag{14}$$

$$p_{y|x}(\underline{y}; \underline{x}) = \frac{1}{2\pi(N\sigma_x^2 + \sigma_w^2)^{\frac{M}{2}}}\exp\left(-\frac{\| \underline{y} - \Phi\underline{x} \|^2}{2(N\sigma_x^2 + \sigma_w^2)}\cdot\right) \tag{15}$$

**Lemma 1.** For any unbiased estimator $\hat{\underline{x}}$ of compressed signal $\underline{x}$ in noisy compressive sensing with random noise $\underline{x}$, i.e., $\underline{y} = \Phi\underline{x} + \underline{w}$, the CRLB of the estimator $\hat{\underline{x}}$ can be determined as:

$$E\{\| \hat{\underline{x}} - \underline{x} \|^2\} \geq (N\sigma_x^2 + \sigma_w^2)Tr\{(\Phi^T\Phi)^{-1}\} \tag{16}$$

Given the Fisher's Information here from [43]:

$$[I(x)]_{i,j} = -E[\frac{\partial^2 \ln p(\underline{y}; \underline{x})}{\partial x_i \partial x_j}] = \frac{1}{N\sigma_x^2 + \sigma_w^2}\Phi^T\Phi \tag{17}$$

Hence, the CRLB [43] of the estimator $\hat{\underline{x}}$ is

$$E\{\| \hat{\underline{x}} - \underline{x} \|^2\} \geq Tr\{I^{-1}\} = (N\sigma_x^2 + \sigma_w^2)Tr\{(\Phi^T\Phi)^{-1}\} \tag{18}$$

and therefore we complete the proof. □

The entropy of $\underline{y}$ is given by:

$$h(\underline{y}) = \frac{1}{2}\log(2\pi e(N\sigma_x^2 + \sigma_w^2))^M \tag{19}$$

Given the fact that $\underline{x}$ and $\underline{y}$ are joint Gaussian r.v.s, and their covariance matrix $C_{xy}$

$$C_{xy} = \begin{pmatrix} C_{xx} & Cov(\underline{x}, \underline{y}) \\ Cov(\underline{y}, \underline{x}) & C_{yy} \end{pmatrix}$$

$Cov(\underline{x}, \underline{y})$ is the covariance matrix between $\underline{x}$ and $\underline{y}$, where

$$Cov(x_n, y_m) = \phi_{m,n}\sigma_x^2 \tag{20}$$

We will then utilize the following fact: given four matrices $A, B, C, D$ which they serve as a block matrix of the complete matrix, where either $A$ or $D$ is invertible, then the determinant of the block matrix can be given as:

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A)\det(D - CA^{-1}B)$$

or,

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(D)\det(A - BD^{-1}C)$$

As both $C_{xx}$ and $C_{yy}$ are invertible matrices, the determinant of $C_{xy}$ can be obtained as:

$$\begin{aligned} \det(C_{xy}) &= det(C_{xx}) \cdot det(C_{yy} - Cov(\underline{y}, \underline{x})C_{xx}^{-1}Cov(\underline{x}, \underline{y})) \\ &= (\sigma_x^2)^N \cdot det(C_{yy} - Cov(\underline{y}, \underline{x})(\sigma_x^2 I_N)^{-1}Cov(\underline{x}, \underline{y})) \\ &= \sigma_x^2)^N \cdot det(C_{yy} - \frac{1}{\sigma_x^2}Cov(\underline{y}, \underline{x})Cov(\underline{x}, \underline{y})) \\ &\overset{(ii)}{=} (\sigma_x^2)^N \cdot det(C_{yy} - \frac{1}{\sigma_x^2}(N\sigma_x^4)I_M) \\ &= (\sigma_x^2)^N \cdot det((N\sigma_x^2 + \sigma_w^2)I_M - (N\sigma_x^2)I_M) \\ &= (\sigma_x^2)^N(\sigma_w^2)^M \end{aligned} \tag{21}$$

where equality $(ii)$ is obtained due to $Cov(\underline{y}, \underline{x})_{M\times N} = Cov(\underline{x}, \underline{y})_{N\times M}^T$, as $Pr(\phi_{m,n} = \pm 1) = 0.5$, if $N$ of $x$ is sufficient large,

$$Cov(y_m, x_n) \cdot Cov(x_n, y'_m)^T = \begin{cases} \sigma_x^2, & \text{if } m = m' \\ 0, & \text{if } m \neq m' \\ , & \end{cases} \tag{22}$$

and,

$$Cov(\underline{y}, \underline{x})_{M\times N} \cdot Cov(\underline{x}, \underline{y})_{N\times M} = (N\sigma_x^2)I_M \tag{23}$$

Therefore, the joint entropy of $\underline{x}$ and $\underline{y}$ is,

$$\begin{aligned} h(\underline{x}, \underline{y}) &= \frac{1}{2}\log[(2\pi e)^{M+N}det(C_{xy})] \\ &= \frac{1}{2}\log[(2\pi e)^{M+N}(\sigma_x^2)^N((\sigma_w^2)^M] \end{aligned} \tag{24}$$

As the process of CS can be viewed as a Markov chain with $\underline{x} \longrightarrow \underline{y} \longrightarrow \hat{\underline{x}}$ [36], where $\hat{\underline{x}}$ is the estimated information. By the data-processing inequality, $I(\underline{x}; \hat{\underline{x}}) \leq I(\underline{x}; \underline{y})$, therefore, $h(\underline{x}|\hat{\underline{x}}) \geq h(\underline{x}|\underline{y})$ and so we have,

$$h(\underline{x}|\underline{y}) \leq 1 + P_e \cdot h(\underline{x}) \tag{25}$$

From Fano's inequality, we could get the lower bound of the $P_e$, as shown

$$\begin{aligned} P_e &\geq \frac{h(\underline{x} \mid \underline{y}) - 1}{h(\underline{x})} \\ &= \frac{h(\underline{x}, \underline{y}) - h(\underline{y}) - 1}{h(\underline{x})} \\ &= \frac{\frac{1}{2}\log[(2\pi e)^{M+N}(\sigma_x^2)^N((\sigma_w^2)^M]}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \\ &\quad - \frac{\frac{1}{2}\log[(2\pi e)^M(N\sigma_x^2+\sigma_w^2)^M]+1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \\ &= 1 - \frac{\frac{1}{2}\log(\frac{N\sigma_x^2+\sigma_w^2}{\sigma_w^2})^M + 1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \end{aligned} \tag{26}$$

We can revalidate the above bound from another point of view. Since the mutual information can be derived from the entropy as:

$$I(\underline{x}; \underline{y}) = h(\underline{y}) - h(\underline{y}|\underline{x}) \tag{27}$$

Given that $\Phi$ is known in advance so that more conditions reduce mutual information [41]

$$\begin{aligned} I(\underline{x}; \underline{y}|\Phi) &\leq I(\underline{x}; \underline{y}) \\ &= h(\underline{y}) - h(\underline{y}|\underline{x}) \\ &\overset{(iii)}{=} h(\underline{y}) - h(\underline{w}) \\ &= \frac{1}{2}\log(2\pi e(N\sigma_x^2+\sigma_w^2))^M - \frac{1}{2}\log(2\pi e\sigma_w^2)^M \\ &= \frac{1}{2}\log(\frac{N\sigma_x^2+\sigma_w^2}{\sigma_w^2})^M \end{aligned} \tag{28}$$

We finally leverage the Fano's inequality to obtain the $P_e$,

$$\begin{aligned} P_e(\underline{x} \neq \hat{\underline{x}}|\Phi) &\geq \frac{h(\underline{x}|\underline{y}) - 1}{h(\underline{x})} \\ &\overset{(f)}{=} 1 - \frac{I(\underline{x}; \underline{y}|\Phi) + 1}{h(\underline{x})} \\ &\geq 1 - \frac{\frac{1}{2}\log(\frac{N\sigma_x^2+\sigma_w^2}{\sigma_w^2})^M + 1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \end{aligned} \tag{29}$$

### B. PROOF OF THEOREM 2

If the source $\underline{x}$ and the noise $\underline{w}$ follow Gaussian distribution, the observation vector $\underline{y} = \Phi\underline{x} + \underline{w}$ also follows Gaussian distribution. Therefore $h(\underline{y}) = \frac{1}{2}\log[(2\pi e)^M \cdot \det(C_{yy})]$. Furthermore, we have:

$$h(\underline{x}, \underline{y}) = \frac{1}{2}\log[(2\pi e)^{N+M} \cdot det(C_{xy})] \tag{30}$$

For the calculation of determinants of the covariance matrices, we utilizes the results from [36] that $\det(C_{yy}) = \prod_{i=1}^{M}\lambda_y^i$ and $\det(C_{xy}) = \prod_{j=1}^{N+M}\lambda_{xy}^j$ and we can obtain the lower bound of the $P_e$ for general case (without any particular assumption of the measurement matrix) as:

$$\begin{aligned} P_e &\geq \frac{h(\underline{x}, \underline{y}) - h(\underline{y}) - 1}{h(\underline{x})} \\ &= \frac{\frac{1}{2}\log[(2\pi e)^{M+N}\prod_{j=1}^{N+M}\lambda_{xy}^j]}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \\ &\quad - \frac{\frac{1}{2}\log[(2\pi e)^M\prod_{i=1}^{M}\lambda_y^i]+1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \\ &\overset{(g)}{\geq} \frac{\frac{1}{2}\log[(2\pi e)^{M+N}(\lambda_{xy_{min}})^{M+N}]}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \\ &\quad - \frac{\frac{1}{2}\log[(2\pi e)^M(\lambda_{y_{max}})^M]+1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \\ &= (1+\frac{M}{N})\frac{\log(\frac{\lambda_{xy_{min}}}{\lambda_{y_{max}}})}{\log(2\pi e\sigma_x^2)} + \frac{\log(2\pi e\lambda_{y_{max}})}{\log(2\pi e\sigma_x^2)} \\ &\quad - \frac{1}{\frac{N}{2}\log(2\pi e\sigma_x^2)} \\ &\overset{(h)}{\approx} \frac{N\log[2\pi e\lambda_{xy_{min}}] + M\log(\frac{\lambda_{xy_{min}}}{\lambda_{y_{max}}})}{N\log[2\pi e\sigma_x^2]} \end{aligned} \tag{31}$$

where inequality $(g)$ follows from the fact that [36] $\prod_{j=1}^{N+M}\lambda_{xy}^j \geq (\lambda_{xy_{min}})^{M+N}$, and $\prod_{i=1}^{M}\lambda_y^i \leq (\lambda_{y_{max}})^M$. Approximation $(h)$ follows from the fact that when the length $N$ of $\underline{x}$ is sufficient large, and $\sigma_x^2 \geq 1$, then $\frac{1}{\frac{N}{2}\log(2\pi e\sigma_x^2)} = 0$.

### C. PROOF OF THEOREM 3

*Proof:* From a linear MMSE estimator, because $\underline{y} = \Phi\underline{x}+\underline{w}$, $\underline{y} \in \mathbb{R}^M$, $\Phi \in \mathbb{R}^{M\times N}$, and $\underline{w} \in \mathbb{R}^M$, $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, and $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$ are independent.

Since estimator in (49) is an affine transformation of Gaussian random vector $\underline{y}$, hence $\hat{\underline{x}}$ is multivariate Gaussian. Thus the estimated expression for linear MMSE estimator also follows Gaussian distribution, with mean and auto-covariance given by,

$$\hat{\underline{x}} = E[\underline{x}|\underline{y}] = (\frac{1}{\sigma_x^2}I_N + \frac{1}{\sigma_w^2}\Phi^T\Phi)^{-1}\Phi^T\underline{y}/\sigma_w^2 \tag{32}$$

$$E[\hat{\underline{x}}] = E[\underline{x}] = \underline{0} \tag{33}$$

$$\begin{aligned} C_{\hat{\underline{x}}} &= C_{xy}C_{yy}^{-1}C_{yx} \\ &= C_{xx}\Phi^T(\Phi C_{xx}\Phi^T + C_w)^{-1}\Phi C_{xx} \\ &= \sigma_x^4\Phi^T(\sigma_x^2\Phi\Phi^T + \sigma_w^2 I_M)^{-1}\Phi \end{aligned} \tag{34}$$

Therefore, $\hat{\underline{x}} \sim \mathcal{N}(\underline{0}, C_{\hat{\underline{x}}})$, and its probability density function (pdf) is:

$$p(\hat{\underline{x}}) = \frac{1}{(2\pi)^{N/2}det^{1/2}(C_{\hat{\underline{x}}})}exp(-\frac{1}{2}\hat{\underline{x}}^T \cdot C_{\hat{\underline{x}}}^{-1} \cdot \hat{\underline{x}}) \tag{35}$$

As we know, the source signal $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, its pdf is,

$$p(\underline{x}) = \frac{1}{(2\pi\sigma_x^2)^{N/2}}exp(-\frac{1}{2}\underline{x}^T \cdot (\sigma_x^2 I_N)^{-1} \cdot \underline{x}) \tag{36}$$

From information theory [41], we know that if $\underline{x}$ and $\hat{\underline{x}}$ are independent with $\underline{x} \sim p(\underline{x})$, and $\hat{\underline{x}} \sim p(\hat{\underline{x}})$, then

$$Pr(\underline{x} = \hat{\underline{x}}) \geq 2^{-H(p(\underline{x}))-D(p(\underline{x})\|p(\hat{\underline{x}}))} \tag{37}$$

where $D(p(\underline{x})\|p(\hat{\underline{x}}))$ is the KLD between two probability mass functions $p(\underline{x})$ and $p(\hat{\underline{x}})$ which is defined as

$$D(p(\underline{x})\|p(\hat{\underline{x}})) = \sum p(\underline{x}) \log \frac{p(\underline{x})}{p(\hat{\underline{x}})} \tag{38}$$

Therefore, the probability of error between $\underline{x}$ and $\hat{\underline{x}}$ is upper bounded by

$$\begin{aligned} P_e &= 1 - Pr(\underline{x} = \hat{\underline{x}}) \\ &\leq 1 - 2^{-H(p(\underline{x}))-D(p(\underline{x})\|p(\hat{\underline{x}}))} \\ &= 1 - 2^{-\frac{1}{2}\log(2\pi e\sigma_x^2)^N} \cdot 2^{-D(p(\underline{x})\|p(\hat{\underline{x}}))} \\ &= 1 - (2\pi e\sigma_x^2)^{-N/2} 2^{-D(p(\underline{x})\|p(\hat{\underline{x}}))} \end{aligned} \tag{39}$$

Consider the case where measurement matrix is Bernoulli matrix and the entries of $\phi_{m,n}$ are i.i.d. with $Pr(\phi_{m,n} = \pm 1) = 0.5$, and suppose dimension of the source signal $N$ is sufficiently large, $\Phi^T\Phi = M \cdot I_N$, and $\Phi\Phi^T = N \cdot I_M$, therefore,

$$\begin{aligned} C_{\hat{\underline{x}}} &= \sigma_x^4 \Phi^T (\sigma_x^2 \Phi\Phi^T + \sigma_w^2 I_M)^{-1} \Phi \\ &= \frac{M\sigma_x^4}{N\sigma_x^2 + \sigma_w^2} I_N \end{aligned} \tag{40}$$

and

$$det(C_{\hat{\underline{x}}}) = (\frac{M\sigma_x^4}{N\sigma_x^2 + \sigma_w^2})^N \tag{41}$$

As $M < N$, $\sigma_x^2 > 0$ and $\sigma_w^2 > 0$, we could conclude that:

$$\frac{M\sigma_x^4}{N\sigma_x^2 + \sigma_w^2} < \frac{M\sigma_x^4}{N\sigma_x^2} = \frac{M}{N}\sigma_x^2 < \sigma_x^2 \tag{42}$$

From the fact that the KLD between two $N$ dimensional multivariate normal distributions with the mean vectors $\boldsymbol{\mu_0}, \boldsymbol{\mu_1}$ and their corresponding nonsingular covariance matrices $\boldsymbol{\Sigma_0}, \boldsymbol{\Sigma_1}$ is:

$$\begin{aligned} D_{KL}(\mathcal{N}_0\|\mathcal{N}_1) = \frac{1}{2}(tr(\boldsymbol{\Sigma_1}^{-1}\boldsymbol{\Sigma_0}) + (\boldsymbol{\mu_1} - \boldsymbol{\mu_0})^T \boldsymbol{\Sigma_1}^{-1} \cdot \\ (\boldsymbol{\mu_1} - \boldsymbol{\mu_0}) - N - ln(\frac{det\boldsymbol{\Sigma_0}}{det\boldsymbol{\Sigma_1}})) \end{aligned} \tag{43}$$

Therefore, the KLD between the source signal $\underline{x}$ and the estimated signal $\hat{\underline{x}}$ could be obtained as:

$$\begin{aligned} &D(p(\underline{x})\|p(\hat{\underline{x}})) \\ &= \frac{1}{2}(N\left(\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2} - 1\right) - \ln(\frac{\sigma_x^{2N}}{(\frac{M\sigma_x^4}{N\sigma_x^2+\sigma_w^2})^N})) \\ &= \frac{1}{2}(N\left(\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2} - 1\right) - N\ln(\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2})) \\ &= \frac{N}{2}(\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2} - 1 - \ln(\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2})) \end{aligned} \tag{44}$$

Since

$$\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2} = \frac{N}{M} + \frac{\sigma_w^2}{M\sigma_x^2} > \frac{N}{M} > 1 \tag{45}$$

When $M$ and $N$ are sufficiently large, $\frac{\sigma_w^2}{M\sigma_x^2} \to 0$.

We notice that when $a > 1$, then $a - 1 > \ln a$, therefore, from the analysis above, $D(p(\underline{x})\|p(\hat{\underline{x}})) > 0$ is concluded, which satisfy the requirement of the KLD.

Hence, the upper bound of the $P_e$ will be:

$$\begin{aligned} P_e &\leq 1 - (2\pi e\sigma_x^2)^{-N/2} 2^{-D(p(\underline{x})\|p(\hat{\underline{x}}))} \\ &= 1 - (2\pi e\sigma_x^2)^{-N/2} 2^{-\frac{N}{2}(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2} - 1 - \ln(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2}))} \\ &= 1 - (\pi e\sigma_x^2)^{-N/2} 2^{-\frac{N}{2}(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2} - \ln(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2}))} \end{aligned} \tag{46}$$

□

### D. PROOF OF THEOREM 4

*Proof:* $\underline{y} = \Phi\underline{x} + \underline{w}$. As each of $x_n$ is an i. i. d. Gaussian random variable with zero mean and variance $\sigma_x^2$, and each noise $w_m$ also follows Gaussian with zero mean and variance $\sigma_w^2$, we know that $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, and $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$.

Using Bayesian estimator [43],

$$\begin{aligned} B_{MSE}(\hat{\underline{x}}) &= E[(\underline{x} - \hat{\underline{x}})^2] \\ &= \iint (\underline{x} - \hat{\underline{x}})^2 p(\underline{y}; \underline{x}) \, d\underline{y} \, d\underline{x} \end{aligned} \tag{47}$$

The aim of the estimator is to find $\hat{\underline{x}}$ that minimizes $B_{MSE}(\hat{\underline{x}})$,

$$\hat{\underline{x}} = arg\min_x B_{MSE}(\hat{\underline{x}}) = E[\underline{x} \mid \underline{y}] \tag{48}$$

For the assumed CS scheme, the estimated $\hat{\underline{x}}$ can be expressed as:

$$\begin{aligned} \hat{\underline{x}} &= E[\underline{x}|\underline{y}] \\ &= E[\underline{x}] + C_{xy}C_{yy}^{-1}(\underline{y} - E[\underline{x}]) \\ &= (\frac{1}{\sigma_x^2} I_N + \frac{1}{\sigma_w^2}\Phi^T\Phi)^{-1}\Phi^T \underline{y}/\sigma_w^2 \end{aligned} \tag{49}$$

By leveraging the following conclusion of the covariance matrix [43]

$$\begin{aligned} C_{\underline{x}|\underline{y}} &= E[(\underline{x} - E[\underline{x}|\underline{y}])(\underline{x} - E[\underline{x}|\underline{y}])^T |\underline{y}] \\ &= (\frac{1}{\sigma_x^2} \cdot I_N + \frac{1}{\sigma_w^2}\Phi^T\Phi)^{-1} \end{aligned} \tag{50}$$

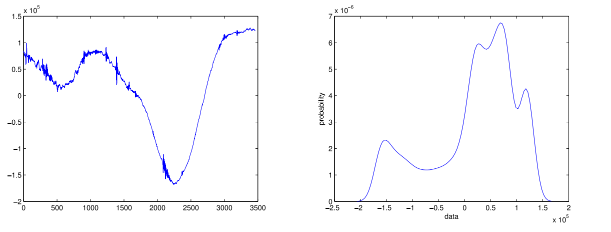
The error between $\underline{x}$ and $\hat{\underline{x}}$ is

$$\underline{\varepsilon} = \underline{x} - \hat{\underline{x}} = \underline{x} - E[\underline{x}|\underline{y}] \tag{51}$$

Since estimator in (49) is an affine transformation of Gaussian random vector $\underline{y}$, hence $\hat{\underline{x}}$ is multivariate Gaussian. Estimation error $\underline{\varepsilon} = \underline{x} - \hat{\underline{x}}$ is an affine transformation of jointly Gaussian vectors $\underline{x}$ and $\underline{y}$, hence $\underline{\varepsilon}$ is also Gaussian.

The mean of $\underline{\varepsilon}$ is calculated as

$$E_{x,y}[\underline{\varepsilon}] = \int [\int (\underline{x} - E[\underline{x}|\underline{y}]) p(\underline{y}|\underline{x}) \, d\underline{x}] p(\underline{y}) \, d\underline{y} = 0 \tag{52}$$



(a) (b)

**FIGURE 1.** underwater acoustic sensor data. (a) UWASN data. (b) UWASN data PDF.

The covariance of $\underline{\varepsilon}$ is

$$E[\underline{\varepsilon}\underline{\varepsilon}^T] = E_{\underline{x},\underline{y}}[(\underline{x} - E[\underline{x}|\underline{y}])(\underline{x} - E[\underline{x}|\underline{y}])^T] = C_{\underline{x}|\underline{y}} \quad (53)$$

The error vector's probability distribution function (pdf) [43] will be derived as,

$$p(\underline{\varepsilon}) = \frac{1}{(2\pi)^{N/2} \cdot \det(C_{\underline{x}|\underline{y}})^{1/2}} exp(-\frac{1}{2}\underline{\varepsilon}^T \cdot C_{\underline{x}|\underline{y}} \cdot \underline{\varepsilon}) \quad (54)$$

To be more specific, for Gaussian distribution, we have the following:

$$\begin{aligned} p(\underline{\varepsilon}) \leq p(\underline{0}) &= \frac{1}{(2\pi)^{N/2} \cdot det(C_{\underline{x}|\underline{y}})^{1/2}} \\ &= \frac{1}{(2\pi)^{N/2} \cdot det((\frac{1}{\sigma_x^2} I_N + \frac{1}{\sigma_w^2}\Phi^T\Phi)^{-1})^{1/2}} \end{aligned} \quad (55)$$

□

For more details of the deviation of this case, the readers can be referred to [30] for more details.

## V. EXPERIMENTAL RESULTS OF UNDERWATER ACOUSTIC SENSOR NETWORK

In our experiment, we assume the source signal follows Gaussian distribution, and provide the corresponding results. However, in real world, not all the source information follows Gaussian distribution [11]. For example, for underwater acoustic sensor networks, the data gathered looks like in Fig. 1a, with the pdf shown as in Fig. 1b.

From information theory, we know that the Gaussian distribution maximizes the entropy over all distributions with the same variance [41]. As the entropy can be regarded as the metric for uncertainty, which indicates that our results provide the bound of the worst case. For real application of compressive sensing, even though the source does not follow Gaussian distribution, our results could be used as an reference to judge the performance of compressive sensing.

Figure 2 shows the compression performance of real UnderWater Acoustic Sensor Networks (UWASN) data with different SNR using compressive sensing. The real UWASN data we used in this paper was obtained by Prof. Jin Chen from Qilihai, Tianjin, China. We thank Prof. Jin Chen for sharing these UWASN data into this research. We have

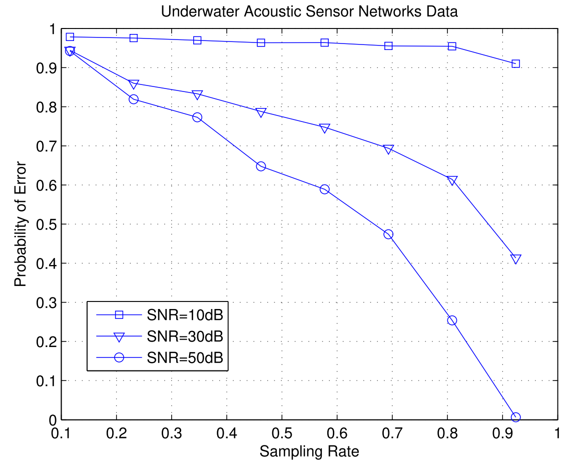



**FIGURE 2.** UWASNs Data - probability of error.

found that the increasing of sampling rate $M/N$ leads to the decrease of the $P_e$. This property indicates that in order to get better reconstruction performance with lower probability of error, the sampling rate should be chosen as high as possible. However, higher sampling rate means higher computation complexity. To balance this, in the real application of compressive sensing, Fig. 2 could be referred for the chosen of certain sampling rate at an acceptable probability of error. It is obvious that the probability of error is always above the lower bound of the theoretical $P_e$ with the same SNR.

These results based on UWASNs data verify the effectiveness of our theoretical results. Despite that we employ the basic Gaussian distribution here, these results still could be used as theoretical reference of real application of compressive sensing, as we explained in earlier part of this paper that the Gaussian distribution. For the worst case, we can maximize the entropy over all pdfs with same variance. The results provided in this paper could be used as theoretical references when apply compressive sensing to real data compression and reconstruction.

## VI. CONCLUSIONS

The theoretical bounds and performance analysis of noisy CS under different scenarios were proposed. It is assumed that the source information and the additive noise follow the Gaussian distribution and we provide the lower and upper bound of the $P_e$ for noisy CS setting. We have demonstrated that for certain conditions (e.g. $(2\pi e\sigma_x^2)^N \geq (\frac{N\sigma_x^2+\sigma_w^2}{\sigma_w^2})^M$), the lower bound of the error is always large than zero so that the we cannot reconstruct the signal completely. To be more specific, we further consider the case where the measurement matrix is a Bernoulli matrix, the related lower and upper bound of the $P_e$ were expressed as a clear closed-form solution. We provide an application study of the UWASN data and demonstrate the performance under different SNR with the increasing fo the sampling rate, which the previous theoretical results are proved accordingly.

## REFERENCES


[1] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–124, Jul. 2007.

[2] E. J. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, Madrid, Spain, Aug. 2006, pp. 1433–1452.

[3] R. G. Baraniuk, E. J. Candès, R. Nowak, and M. Vetterli, "Compressive sampling [from the guest editors]," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 12–13, Mar. 2008.

[4] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[5] G. Mateos and G. B. Giannakis, "Robust PCA as bilinear decomposition with outlier-sparsity regularization," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5176–5190, Oct. 2012.

[6] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, "Online adaptive estimation of sparse signals: Where RLS meets the $\ell_1$-norm," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3436–3447, Jul. 2010.

[7] I. D. Schizas and G. B. Giannakis, "Covariance eigenvector sparsity for compression and denoising," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2408–2421, May 2012.

[8] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[9] V. Kekatos and G. B. Giannakis, "From sparse signals to sparse residuals for robust sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3355–3368, Jul. 2011.

[10] L. Jacques, D. K. Hammond, and J. M. Fadili, "Dequantizing compressed sensing: When oversampling and non-Gaussian constraints combine," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 559–571, Jan. 2011.

[11] J. Chen, Q. Liang, J. Paden, and P. Gogineni, "Compressive sensing analysis of synthetic aperture radar raw data," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 6362–6366.

[12] E. J. Candès 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.

[13] V. K. Goyal, A. K. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 48–56, Mar. 2008.

[14] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[15] E. J. Candés and Y. Plan, "Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2342–2359, Apr. 2011.

[16] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.

[17] E. J. Candès and B. Recht, "Simple bounds for recovering low-complexity models," *Math. Program.*, vol. 141, pp. 577–589, Oct. 2012.

[18] E. Arias-Castro, E. J. Candès, and M. A. Davenport, "On the fundamental limits of adaptive sensing," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 472–480, Jan. 2013.

[19] W. Wang, M. J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2967–2979, Jun. 2010.

[20] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1383–1399, Mar. 2010.

[21] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2210–2219, Apr. 2008.

[22] E. J. Candès and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7235–7254, Nov. 2011.

[23] M. Akçakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.

[24] W. Dai and O. Milenkovic, "Information theoretical and algorithmic approaches to quantized compressive sensing," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 1857–1866, Jul. 2011.

[25] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[26] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.

[27] B. Mulgrew and M. E. Davies, "Approximate lower bounds for rate-distortion in compressive sensing systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2008, pp. 3849–3852.

[28] J. Chen and Q. Liang, "Rate distortion performance analysis of compressive sensing," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, Houston, TX, USA, Dec. 2011, pp. 1–5.

[29] J. Chen, Q. Liang, B. Zhang, and X. Wu, "Information theoretic performance bounds for noisy compressive sensing," in *Proc. IEEE ICC Workshop Radar Sonar Netw. (RSN)*, Budapest, Hungary, Jun. 2013, pp. 972–976.

[30] J. Chen and Q. Liang, "Theoretical performance limits for compressive sensing with random noise," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3400–3405.

[31] Z. Ben-Haim and Y. C. Eldar, "Performance bounds for sparse estimation with random noise," in *Proc. IEEE/SP 15th Workshop Stat. Signal Process.*, Aug. 2009, pp. 225–228.

[32] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5111–5130, Oct. 2010.

[33] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér–Rao bound for noisy compressive sampling," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1233–1236, Mar. 2009.

[34] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramér–Rao bound in noisy compressed sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 518–526, Jan. 2012.

[35] M. Capdevila and O. W. M. Flórez, "A communication perspective on automatic text categorization," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 7, pp. 1027–1041, Jul. 2009.

[36] D. Kirachaiwanich and Q. Liang, "Compressive sensing: To compress or not to compress," in *Proc. Conf. Rec. 45th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2011, pp. 809–813.

[37] H. Gish and N. J. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inf. Theory*, vol. 14, no. 5, pp. 676–683, Sep. 1968.

[38] V. Sheinin and A. Jagmohan, "Low rate uniform scalar quantization of memoryless Gaussian sources," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 793–796.

[39] P. Zador, "Development and evaluation of procedures for quantizing multivariate distributions," Ph.D. dissertation, Stanford Univ., Stanford, CA, USA, 1964.

[40] D. Hui and D. L. Neuhoff, "Asymptotic analysis of optimal fixed-rate uniform scalar quantization," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 957–977, Mar. 2001.

[41] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[42] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY, USA: McGraw-Hill, 2000.

[43] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.


**JUNJIE CHEN** received the B.S. degree from Xidian University, China, in 2007, the M.S. degree from the Beijing University of Posts and Telecommunications (BUPT), in 2010, and the Ph.D. degree from the Department of Electrical Engineering, The University of Texas at Arlington, in 2014. Her research interests include wireless sensor networks, wireless communications and networking, compressive sensing, sparse sampling, signal processing, and information theory.

**FANGQI ZHU** received the B.Sc. and M.Sc. degrees from the University of Electronic Science and Technology of China, Chengdu, in 2013 and 2016, respectively. He is currently the Ph.D. degree from The University of Texas at Arlington, all in electrical engineering. His current research interests include deep learning, statistical inference and optimization, signal processing, computational intelligence, and wireless communication. He is a Reviewer of several conferences and journals.

**QILIAN LIANG** received the Ph.D. degree from the Department of Electrical Engineering, University of Southern California, in 2000. Prior to joining the faculty of The University of Texas at Arlington (UTA), in 2002, he was a Member of Technical Staff with Hughes Network Systems, Inc., San Diego, CA, USA. He is currently a Distinguished University Professor with the Department of Electrical Engineering, UTA. He has published more than 300 journal and conference papers. His research interests include wireless sensor networks, wireless communications, signal processing, and computational intelligence. He received the 2002 IEEE TRANSACTIONS ON FUZZY SYSTEMS Outstanding Paper Award, the 2003 U.S. Office of Naval Research Young Investigator Award, 2007, 2009, the 2010 U.S. Air Force Summer Faculty Fellowship Program Award, the 2012 UTA College of Engineering Excellence in Research Award, the 2013 UTA Outstanding Research Achievement Award, and was inducted into UTA Academy of Distinguished Scholars, in 2015.