Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Quantization in Compressive Sensing: A Signal Processing Approach

ISIDORA STANKOVIĆ[1,2], (Member, IEEE), MILOŠ BRAJOVIĆ[1], (Member, IEEE), MILOŠ DAKOVIĆ[1], (Member, IEEE), CORNEL IOANA[2], (Member, IEEE), LJUBIŠA STANKOVIĆ[1], (Fellow, IEEE)

[1]Faculty of Electrical Engineering, University of Montenegro, Džordža Vašingtona bb, 81000 Podgorica, Montenegro
[2]GIPSA Lab, INP Grenoble, University of Grenoble Alpes, 11 Rue des Mathématiques, 38000 Grenoble, France

Corresponding author: Isidora Stanković (e-mail: isidoras@ucg.ac.me).

ABSTRACT The influence of finite-length registers and the corresponding quantization effects on the reconstruction of sparse and approximately sparse signals from a reduced set of measurements is analyzed in this paper. For the nonquantized measurements, the compressive sensing (CS) framework provides highly accurate reconstruction algorithms that produce negligible errors when the reconstruction conditions are met. However, hardware implementations of signal processing algorithms inevitably involve finite-length registers and quantization of the measurements. A detailed analysis of the effects related to the measurement quantization, with an arbitrary number of bits, is provided in this paper. A unified novel mathematical model to characterize the influence of the quantization noise and the signal nonsparsity on the CS reconstruction is introduced. Using this model, an exact formula for the expected error energy in the CS-based reconstructed signal is derived, while in the literature its bounds have been reported only. The theory is validated through various numerical examples with quantized measurements, involving scenarios with approximately sparse signals, noise folding effect, and floating-point arithmetics.

INDEX TERMS compressive sensing, measurements, quantization, signal reconstruction, sparse signal processing

I. INTRODUCTION

COMPRESSIVE sensing (CS) theory provides a rigorous mathematical framework for the reconstruction of sparse signals, using a reduced set of measurements [1]–[10]. Advantages of CS are directly related to the signal transmission and storage efficiency, which is crucial in big data setups. Moreover, the problem of the physical unavailability of measurements, or the problem of significant signal corruption, are also potentially solvable within the CS framework. Since the establishment of CS, phenomena related to the reduced sets of measurements and sparse signal reconstruction have been supported by the fundamental theory and well-defined mathematical framework, while the performances of the reconstruction processes have been continuously improved by newly introduced algorithms, often adapted to perform well in a particular context, or to solve some specific problems [11]–[19]. In real applications, many signals are sparse or approximately sparse in a certain transformation domain. This makes the CS applicable in various fields of signal processing [15].

Ideally, the measurements that are used for the reconstruction should be taken accurately, assuming a very large number of bits in their digital format (providing high precision levels). However, this could be extremely demanding and expensive for hardware implementations [20]. In practice, the measurements are quantized, meaning that they are represented using a limited number of bits. Such measurements bring robustness, memory efficiency and simplicity in the corresponding hardware implementation (particularly in sensor design). This paper investigates the influence of quantization on the CS reconstruction with a simple yet rigorous characterization of the related phenomena, through the derivation of new and exact associated expected squared error expressions. The results are supported by a relevant the-

oretical framework and detailed statistical analysis, through extensive numerical experiments.

The most extreme case of quantization is in limiting the measurements to one bit only. In previous work [20]–[23], one-bit measurements are initially treated as the sign constraints, as opposed to the values to be matched in the mean squared sense during the reconstruction process. Quantization to one-bit measurements is suitable for hardware systems since the quantizers do not suffer from dynamic range issues. However, as the sign of measurements does not provide amplitude information of the signal, it can be recovered up to a constant scalar factor only. Moreover, the number of measurements needed for a successful reconstruction based on such systems exceeds the signal length. In this paper, we focus on the general B-bit quantization of available measurements and its influence on the reconstruction accuracy.

The quantization of measurements undeniably introduces the error in the CS reconstruction result. The effects associated with the quantization have been studied recently [24]–[30]. The results of these studies mainly include the derivation of quantization error bounds and the adaptation of CS algorithms aiming to reduce the distortions caused by the quantization [4], [5]. The upper bound of the reconstruction error, for strictly sparse signals, has been derived in [25]. Other reported results are focused on the worst case analysis [26]. Exact asymptotic distortion rate functions have been derived in [26] for scalar quantization, where the reconstruction strategies have been adapted to accommodate quantization errors. An overview of the quantization phenomena in the compressive sensing context is presented in [27]. Therein, the fundamental analysis provides the performance bounds only, with an additional focus on the Sigma-Delta quantization and the related theory. Recently, the effects of quantization on the estimation of sparsity order, and signal support have been considered with a large number of Monte Carlo simulations in [28]. The most frequently used algorithms in compressive sensing are adjusted to the quantization effect in [29]. For the case of one-bit unlimited sampling quantization approach the bounds of reconstruction error are derived in [30]. The design of quantizer for random measurements that minimize the distortion effects in the reconstruction is considered in [31], [32]. Therein, it is highlighted that minimizing the mean squared error (MSE) of the measurements is not equivalent to minimizing the MSE of the CS reconstruction. The quantization noise was studied in [33], where the lower and upper bound for the ratio of the reconstruction SNR and measurements SNR are derived and related to the noise folding effects in CS on the signal acquisition systems.

Summarizing the above-mentioned literature, there have been only error bounds derived in the previous works. This paper aims to fill the literature gap regarding the exact characterization of the quantization in the CS, by deriving an explicit and exact relation for the mean squared error, instead of the reported error bounds. The error produced by the quantization of measurements is analyzed from a practical signal processing point of view. The paper gives an exact calculation of the error produced by the applied reconstruction procedure. The error appearing when an approximately sparse signal is reconstructed under the sparsity constraint is examined in detail. The analysis is expanded to include the effect of the pre-measurement noise in the sparsity domain coefficients, known as the noise folding [34]. The presented theory is unified by exact relations for the expected squared reconstruction error, derived to take into account all the studied effects. Moreover, we comment on the modifications of the derived relations, required to include the floating-point arithmetics.

In numerical studies, we have performed reconstructions with various numbers of bits, different sparsities, including approximately sparse signals, and noise folding effects. Three different methods of signal reconstruction are used to test the analytic results. In total, for all the considered cases, we performed about $150,000$ realizations with random signal parameters to statistically test the presented theoretic results. The formula for the mean squared error is used for the cases when the reconstruction conditions for the signal are satisfied. We have also tested how the quantization influences the reconstruction conditions by testing the probability of misdetection for various sparsities and the number of available measurements. The misdetection statistical analysis is performed on $10,000$ independent trials.

The paper is organized as follows. In Section II, basic CS concepts and definitions are briefly presented. Section III introduces a common approach to solve the CS reconstruction problem, including a brief overview of relevant properties that characterize possible solutions. Section IV puts the quantization within the compressive sensing framework. In Section V, the concept of nonsparse (approximately sparse) signals reconstructed under the sparsity constraint is analyzed, leading to the reconstruction error equation which unifies the studied effects. The theory is expanded, to take into account the noise folding effect, in Section VI, while Section VII discusses the quantization in floating-point arithmetics. Numerical results verify the presented theory in Section VIII. The probability of misdetection is investigated in Section IX. The paper ends with concluding remarks.

II. BASIC COMPRESSIVE SENSING DEFINITIONS

Definition: A discrete signal $x(n)$, $n = 0, 1, \ldots, N-1$ is sparse in one of its representation domains $X(k)$ if the number K of nonzero coefficients is much smaller than the total number of samples N, that is,

$$X(k) = 0 \text{ for } k \notin \mathbb{K} = \{k_1, k_2, \ldots, k_K\},$$

where $K \ll N$.

Definition: A measurement of a signal is a linear combination of its sparsity domain coefficients $X(k)$,

$$y(m) = \sum_{k=0}^{N-1} a_m(k) X(k), \ m = 1, 2, \ldots, M, \qquad (1)$$

or in matrix form

$$\mathbf{y} = \mathbf{A}\mathbf{X}, \tag{2}$$

where $\mathbf{y}$ is an $M \times 1$ (M-dimensional) column vector of the measurements $y(m)$, $\mathbf{A}$ is an $(M \times N)$-dimensional measurement matrix with the coefficients $a_m(k)$ as its elements, and $\mathbf{X}$ is an $N \times 1$ (N-dimensional) column sparse vector of coefficients $X(k)$. It is common to normalize the measurement matrix such that the energy of its columns is 1. In that case, the diagonal elements of the matrix $\mathbf{A}^H\mathbf{A}$ are equal to 1, where $\mathbf{A}^H$ denotes a Hermitian transpose of $\mathbf{A}$.

By definition, a measurement of a K-sparse signal can be written as

$$y(m) = \sum_{i=1}^{K} X(k_i)a_m(k_i). \tag{3}$$

The compressive sensing theory states that, under certain realistic conditions, it is possible to reconstruct a sparse N-dimensional vector $\mathbf{X}$ from a reduced M-dimensional set of measurements ($M < N$), belonging to the vector $\mathbf{y}$,

$$\mathbf{y} = [y(1), y(2), \ldots, y(M)]^T. \tag{4}$$

The reconstruction conditions are defined in several forms. The most widely used are the forms based on the restricted isometry property (RIP) and the coherence index [1]–[4]. Although providing tighter bounds, the RIP based condition is of high calculation complexity. This is the reason why the coherence based relation will be considered in this paper, along with some comments on its probabilistic relaxation.

The reconstruction of a K-sparse signal representation, $\mathbf{X}$ is unique if $K < (1 + 1/\mu)/2$, where the coherence index, μ, is equal to the maximum absolute off-diagonal element of $\mathbf{A}^H\mathbf{A}$, assuming its unity diagonal elements.

A simple proof will be provided later.

Formally, compressive sensing aims to solve the optimization problem

$$\min \ \|\mathbf{X}\|_0 \ \text{subject to} \ \mathbf{y} = \mathbf{A}\mathbf{X}, \tag{5}$$

or its corresponding relaxed convex form. In this way, the unknown sparse representation, $\mathbf{X}$, of a signal whose dimension is N, is obtained from M measurements, $\mathbf{y}$, by minimizing its sparsity measure $\|\mathbf{X}\|_0$. Amongst many others, an approach based on matching the components corresponding to the nonzero coefficients, can be used to solve (5). It is further assumed that the CS reconstruction is based on a such methodology. The solution is discussed in the next section, since it will be used to model the quantization noise and other studied effects.

III. PROBLEM SOLUTION

To perform the reconstruction, we use an iterative version of the orthogonal matching pursuit algorithm from [11]. Assume first that K nonzero values $X(k)$ are detected at the positions $k \in \mathbb{K} = \{k_1, k_2, \ldots, k_K\}$. The system of measurement equations becomes

$$\mathbf{y} = \mathbf{A}_{MK}\mathbf{X}_K. \tag{6}$$

The system is solved for the nonzero coefficients $X(k)$, $k \in \mathbb{K}$ written in the vector form as $\mathbf{X}_K$, with $K < M$. The matrix $\mathbf{A}_{MK}$ is an $M \times K$ sub-matrix of the $M \times N$ measurement matrix $\mathbf{A}$, where only the columns corresponding to the nonzero elements in $X(k)$ are kept. The solution of the system in (6) is

$$\mathbf{X}_K = (\mathbf{A}_{MK}^H\mathbf{A}_{MK})^{-1}\mathbf{A}_{MK}^H\mathbf{y} = \text{pinv}(\mathbf{A}_{MK})\mathbf{y}, \tag{7}$$

where $\text{pinv}(\mathbf{A}_{MK})$ is the pseudo-inverse of the matrix $\mathbf{A}_{MK}$ and $\mathbf{A}_{MK}^H\mathbf{A}_{MK}$ is known as a $K \times K$ Gram matrix of $\mathbf{A}_{MK}$.

Therefore, the CS problem solution can be split into two steps:

1) detection of the positions of nonzero coefficients in $\mathbf{X}$ and
2) calculation of the unknown coefficient values $X(k)$ at the detected nonzero positions.

A. INITIAL ESTIMATE

Detection of the positions of nonzero coefficients $X(k)$ will be based on the initial estimate concept. An intuitive idea for the initial estimate comes from the fact that the measurements are obtained as linear combinations of the sparsity domain coefficients, with rows of the measurement matrix $\mathbf{A}$ acting as weights. It means that the back-projection of the measurements $\mathbf{y}$ to the measurement matrix $\mathbf{A}$, defined by

$$\mathbf{X}_0 = \mathbf{A}^H\mathbf{y} = \mathbf{A}^H\mathbf{A}\mathbf{X}, \tag{8}$$

can be used to estimate the positions of nonzero coefficients. The back-projection of the available data is present in an implicit or explicit way in all reconstruction algorithms. In most of these algorithms (for example, orthogonal matching pursuit - OMP, least absolute shrinkage and selection operator - LASSO, or Bayesian reconstruction) the back-projection is used as an initial estimate. However, this relation contains more information about the reconstructed signal than serving just as its initial estimate. It has been shown that the crucial criteria for the reconstruction, like, for example, the coherence index (as it will be shown later in Remark 1) and the restricted isometry property can be derived from this back-projection relation. In this paper this relation will be used as a starting point to derive the error in the reconstructed signal, assuming that the reconstruction conditions are met (the accuracy of the main result will be demonstrated on three quite different reconstruction methods).

For the coefficient at the kth position, its initial estimate $X_0(k)$ takes the following form

$$X_0(k) = \sum_{m=1}^{M} y(m)a_m^*(k), \tag{9}$$

or after $y(m)$ is replaced by its value from (3) we get

$$X_0(k) = \sum_{i=1}^{K} X(k_i)\mu(k_i, k), \tag{10}$$

where

$$\mu(k_i,k)=\sum_{m=1}^{M}a_m(k_i)a_m^*(k) \tag{11}$$

are the coefficients of mutual influence (interference) among elements $X(k)$. The coefficients $\mu(k_i,k)$ are equal to the corresponding elements of the matrix $\mathbf{A}^H\mathbf{A}$, with

$$\mu=\max_{k\neq l}|\mu(l,k)|, \tag{12}$$

and $\mu(k,k)=1$. Note that μ is referred to as the coherence index.

For various values of k_i, the off-diagonal elements $\mu(k_i,k)$ of matrix $\mathbf{A}^H\mathbf{A}$ act as random variables, with different distribution for different measurement matrices. For the partial discrete Fourier transform (DFT) matrix, distribution of $\mu(k_i,k)$ tends to a Gaussian distribution for $1\ll M\ll N$, while for an equiangular tight frame (ETF) measurement matrix, $\mu(k_i,k)$ takes only the values such that $|\mu(k_i,k)|=\mu$. Distribution of $\mu(k_i,k)$ for other measurement matrices can also be derived.

The reduced set of measurements (samples) manifests as a noise in the initial estimate, which therefore acts as a random variable, with the mean-value and the variance given by

$$E\{X_0(k)\}=\sum_{i=1}^{K}X(k_i)\delta(k-k_i) \tag{13}$$

$$\mathrm{var}\{X_0(k)\}=\sum_{i=1}^{K}|X(k_i)|^2\mathrm{var}\{\mu(k_i,k)\}\left(1-\delta(k-k_i)\right), \tag{14}$$

where $\delta(k)=1$ for $k=0$ and $\delta(k)=0$ elsewhere.

In the analysis of the reconstruction error, we are interested in the variance of random variable $\mu(k_i,k)$, that is

$$\mathrm{var}\{\mu(k_i,k)\}=\sigma_\mu^2.$$

For the partial DFT matrix, the variance is derived in [6]. For a real-valued ETF measurement matrix, the values $\pm\mu$ are equally probable, producing the variance $\sigma_\mu^2=\mu^2$, where, according to the Welch bound, $\mu^2=(N-M)/(M(N-1))$ holds [14], [35]. For the Gaussian measurement matrix, the variance is $\sigma_\mu^2=1/M$. The same value is obtained for other considered random matrices. The variance σ_μ^2 of $\mu(k_i,k)$ is presented in Table 1 for various measurement matrices [6], [14]–[16].

TABLE 1. Variances in the initial estimate for various types of the measurement matrix for $k\neq k_i$

Measurement matrix	$a_m(k)$	σ_μ^2
Partial DFT	$\frac{1}{\sqrt{M}}e^{j2\pi n_m\frac{k}{N}}$	$\frac{N-M}{M(N-1)}$
Random partial DFT	$\frac{1}{\sqrt{M}}e^{j2\pi t_m\frac{k}{N}}$	$1/M$
Equiangular tight frame	$\mu=\sqrt{\frac{N-M}{M(N-1)}}$	$\frac{N-M}{M(N-1)}$
Gaussian random	$\sim\mathcal{N}(0,\frac{1}{M})$	$1/M$
Uniform random	$\sim\mathcal{U}(0,\frac{1}{M})$	$1/M$
Bernoulli (binary)	$\sim\pm\frac{1}{\sqrt{M}}$	$1/M$

B. DETECTION OF NONZERO ELEMENT POSITIONS

The initial estimate can be used as a starting point for a thorough analysis of the reconstruction performance and its outcomes. Potentially, such analysis can lead to the improvements of the reconstruction process. The detection can be done in one step or in an iterative way.

One-step detection: In an ideal case, matrix $\mathbf{A}^H\mathbf{A}$ should be such that the initial estimate $\mathbf{X}_0$ contains K coefficients higher than the other coefficients. Then, by taking the positions of the highest coefficients in (8) as the set $\mathbb{K}$, the signal is simply reconstructed using (7).

Iterative detection: The condition that all K nonzero coefficients in the initial estimate $\mathbf{X}_0$ are larger than the coefficient values $X_0(k)$ at the original zero-valued positions $k\notin\mathbb{K}$, can be relaxed using an iterative procedure. To find the position of the largest coefficient in $\mathbf{X}$, based on $\mathbf{X}_0$, it is sufficient that the corresponding coefficient $X_0(k)$ has a value larger than the values of the coefficients $X_0(k)$ at the original zero-valued coefficient positions $k\notin\mathbb{K}$.

Remark 1: *Solution uniqueness.* The worst case for the detection of a nonzero coefficient, with a normalized amplitude 1, occurs when the remaining $K-1$ coefficients are equally strong (that is, with unit amplitudes). This is the case of the strongest possible influence of other nonzero coefficients to the initial estimate of the considered largest coefficient. The influence of the kth coefficient on the coefficient at the ith position is equal to $\mu(k_i,k)$, given by (11). Its maximum possible absolute value is the coherence index μ. In the worst case, the amplitude of the considered coefficient in the initial estimate is $1-(K-1)\mu$. At the position where the original coefficient $X(k)$ is zero-valued, in the worst case, the maximum possible contributions μ of all K coefficients sum up *in phase* to produce the maximum possible disturbance $K\mu$. The detection of the strongest coefficient is successful if

$$1-(K-1)\mu>K\mu,$$

producing the well-known coherence condition for the unique reconstruction $K<(1+1/\mu)/2$.

After the largest coefficient position is found and its value is estimated, this coefficient can be subtracted and the procedure can be continued with the remaining $(K-1)$-sparse signal. If the reconstruction condition is met for the K-sparse signal, then it is met for all lower sparsities as well.

The procedure is iteratively repeated for each coefficient. The stopping criterion is that $\mathbf{A}_{MK}\mathbf{X}_K=\mathbf{y}$ holds for the estimated positions $\{k_1,k_2,\ldots,k_K\}$ and coefficients $X(k)$.

The method is summarized in Algorithm 1.

Remark 2: *Solution exactness.* The coherence index condition guarantees that the positions of the nonzero elements in $\mathbf{X}$ will be uniquely determined. Next, we will show that the values of the nonzero coefficients will be exactly recovered.

Algorithm 1 Reconstruction Algorithm

Input: Vector $\mathbf{y}$, matrix $\mathbf{A}$, assumed sparsity K

1: $\mathbb{K} \leftarrow \emptyset$, $\mathbf{e} \leftarrow \mathbf{y}$
2: **for** $i = 1$ **do** K
3: $k \leftarrow$ position of the highest value in $\mathbf{A}^H\mathbf{e}$
4: $\mathbb{K} \leftarrow \mathbb{K} \cup k$
5: $\mathbf{A}_K \leftarrow$ columns of matrix $\mathbf{A}$ selected by set $\mathbb{K}$
6: $\mathbf{X}_K \leftarrow \text{pinv}(\mathbf{A}_K)\mathbf{y}$
7: $\mathbf{y}_K \leftarrow \mathbf{A}_K\mathbf{X}_K$
8: $\mathbf{e} \leftarrow \mathbf{y} - \mathbf{y}_K$
9: **end for**

Output: Reconstructed $\mathbf{X}_R = \mathbf{X}_K$ and positions $\mathbb{K}$.

The system of linear equations in (10), for $k \in \mathbb{K}$, can be written in a matrix form as

$$\mathbf{X}_{0K} = \mathbf{B}\mathbf{X}_K = \mathbf{X}_K + \mathbf{C}_K,$$

where $\mathbf{B}$ is a $K \times K$ matrix with elements $b_{pi} = \mu(k_p, k_i)$, $\mathbf{X}_{0K}$ is the vector with K elements obtained from the initial estimate as $X_{0K}(i) = X_0(k_i)$, and $\mathbf{X}_K$ is the vector with K corresponding coefficients from the original signal. The influence of the other $K-1$ coefficients to the considered coefficient is denoted by $\mathbf{C}_K$.

The reconstructed coefficients $\mathbf{X}_R$, at the nonzero coefficient positions, are obtained by minimizing $\|\mathbf{y} - \mathbf{A}_K\mathbf{X}_R\|_2^2$. They are

$$\mathbf{X}_R = (\mathbf{A}_K^H\mathbf{A}_K)^{-1}\mathbf{A}_K^H\mathbf{y}, \tag{15}$$

where $\mathbf{A}_K$ is the matrix obtained from the measurement matrix $\mathbf{A}$ by keeping the columns for $k \in \mathbb{K}$. Since $\mathbf{A}_K^H\mathbf{y} = \mathbf{X}_{0K}$, according to (8), we can rewrite (15) as

$$\mathbf{X}_R = (\mathbf{A}_K^H\mathbf{A}_K)^{-1}\mathbf{X}_{0K}. \tag{16}$$

Since $\mathbf{X}_{0K} = \mathbf{B}\mathbf{X}_K$, the reconstruction is exact if

$$(\mathbf{A}_K^H\mathbf{A}_K)^{-1} = \mathbf{B}^{-1}$$

holds. Indeed, the elements of matrix $\mathbf{A}_K^H\mathbf{A}_K$ are equal to $\beta_{ij} = \sum_{n=1}^{M} a_n^*(k_i)a_n(k_j) = \mu(k_j, k_i)$, meaning that $\mathbf{A}_K^H\mathbf{A}_K = \mathbf{B}$. Therefore, $\mathbf{X}_R = \mathbf{X}_K$ holds.

The reconstruction algorithm produces the correct coefficient values $X(k)$ at the selected positions $k \in \mathbb{K}$. It means that the influence of other $K-1$ coefficients to each coefficient in the initial coefficient estimate $X_0(k)$, denoted by $C(k)$, is canceled out.

In summary, the reconstruction algorithm for a coefficient at a position $k \in \mathbb{K}$, works as an identity system to the original signal coefficient in $X_0(k)$, eliminating the influence of other coefficient at the same time, Fig. 1.

C. NOISY MEASUREMENTS

Assume next that the measurements are noisy

$$\mathbf{y} + \boldsymbol{\varepsilon} = \mathbf{A}\mathbf{X}, \tag{17}$$

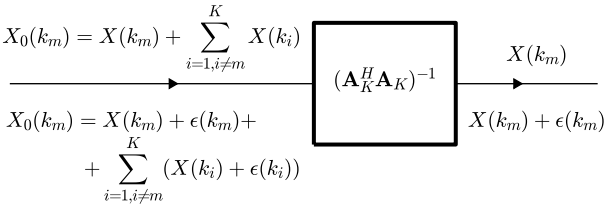

FIGURE 1. Illustration of a system for the reconstruction of a sparse signal $\mathbf{X}_K = [X(k_1), X(k_2), \ldots, X(k_K)]^T$ from the initial estimate $\mathbf{X}_{0K} = [X_0(k_1), X_0(k_2), \ldots, X_0(k_K)]^T$.

with a zero-mean signal independent noise $\boldsymbol{\varepsilon}$. The noise variance of the assumed additive input noise $\boldsymbol{\varepsilon}$ is σ_ε^2 and the covariance is given by

$$\text{E}\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^H\} = \sigma_\varepsilon^2\mathbf{I}.$$

Noisy measurements will result in a noisy initial estimate $\mathbf{X}_0 = \mathbf{A}^H(\mathbf{y} + \boldsymbol{\varepsilon})$. Variance of $X_0(k)$ due to the input noise in measurements, is $\sigma_{X_0(k)}^2 = \sigma_\varepsilon^2$, since it has been assumed that the columns of $\mathbf{A}$ have unite energy,

$$\text{E}\{\mathbf{X}_0\mathbf{X}_0^H\} - |\text{E}\{\mathbf{X}_0\}|^2 = \text{E}\{\mathbf{A}^H\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^H\mathbf{A}\} = \sigma_\varepsilon^2\mathbf{I}.$$

The noise variance in the reconstructed coefficient is (Remark 2 and Fig. 1)

$$\text{var}\{X_R(k)\} = \sigma_\varepsilon^2.$$

Since the noise is independent in each reconstructed coefficient, the total mean squared error (MSE) in K reconstructed coefficients is

$$\begin{aligned}&\text{E}\{\sum_{m=1}^{K} |X_R(k_m) - X(k_m)|^2\}\\ &= \text{E}\{\|\mathbf{X}_R - \mathbf{X}_K\|_2^2\} = K\sigma_\varepsilon^2.\end{aligned} \tag{18}$$

If the partial DFT matrix is formed as a submatrix of the standard inverse DFT matrix (with normalization $1/N$), then we would get $\text{E}\{\|\mathbf{X}_R - \mathbf{X}_K\|_2^2\} = KN^2{\sigma_\varepsilon}^2/M$, as shown, for example, in [16]. The reconstruction error bounds are given in [15].

IV. QUANTIZATION EFFECTS

Traditional CS theory does not consider the limitations in the number of bits used for the representation of measurements. This can affect the reconstruction performance of the standard CS approaches.

The measurement quantization is particularly important in the hardware implementation context. One-bit measurements are the most extreme case, promising simple, comparator-based hardware devices [20]. The one bit used represents the sign of the sample, i.e. $\mathbf{y} = \text{sign}\{\mathbf{A}\mathbf{X}\}$. However, a larger number of samples is required for an accurate reconstruction, which is difficult to achieve using only the sign of measurements [20], [21].

A more general form of the hardware implementation uses a B-bit digital format of the measurements. We will assume

that the measurements are stored into $(B+1)$-bit registers (one one sign bit and B bits for the signal absolute value),

$$\mathbf{y}_B = \text{digital}_B\{\mathbf{AX}\}, \tag{19}$$

whereas the reconstruction of the coefficients $X(k)$ is done in a more realistic sense. The requirement for storage is also significantly reduced for the measurements with a low number of bits, since the total number of bits is reduced. Note that, for a complex-valued signal $x(n)$, the measurements $\mathbf{y}_B$ are also complex, formed as

$$\mathbf{y}_B = \text{digital}_B\{\Re\{\mathbf{AX}\}\} + j\text{digital}_B\{\Im\{\mathbf{AX}\}\} \tag{20}$$

where both the real and the imaginary part of measurement are quantized to $B+1$-bits.

A. QUANTIZATION ERRORS

Quantization influences the results of the compressive sensing reconstruction in several ways:

- Measurement quantization error, described by an additive quantization noise. This influence can be modeled as a uniform noise with values between the quantization level bounds.
- Quantization of the measured sparse signal coefficients.
- Quantization of the results of arithmetic operations. It depends on the way how the calculations are performed.
- Quantization of the coefficients in the algorithm. However, being deterministic for a given measurement matrix, this type of error is commonly neglected from the analysis.

In order to perform an appropriate and exact analysis, some standard assumptions are further made:

- The measurement quantization error is a white noise process with a uniform distribution.
- The quantization errors are uncorrelated.
- The quantization errors are not correlated with the measurement values.

The most important error source is the quantization of the measurements $y(m)$ and the quantization of the measured sparse signal coefficients $X(k)$, referred to as the quantization noise folding. They will be analyzed next.

B. INPUT SIGNAL RANGES

Assume that registers with B bits, with an additional sign bit, are used and that all measurements are normalized to the range

$$-1 \leq y(m) < 1.$$

The total number of bits in a register is $b = B+1$.

In that case, it is important to notice that the sparse signal coefficients $X(k)$ must be within the range $-\min\{\sqrt{M}/K, 1\} < X(k) < \min\{\sqrt{M}/K, 1\}$ so that $\mathbf{y} = \mathbf{A}_{MK}\mathbf{X}_K$ does not produce a value with an amplitude greater or equal to 1. For the partial DFT matrices, this condition holds in a strict sense, while for the Gaussian matrices it holds in a mean sense (all values whose amplitude is greater than 1 are quantized to the closest level with amplitude below 1). Note that the butterfly schemes for the measurements calculation (as in the quantized FFT algorithms) could extend this bounds for $X(k)$ so that the maximum range $-1 < X(k) < 1$ can be used.

C. MEASUREMENTS QUANTIZATION

In order to be stored into registers, the digital measurement values $\mathbf{y}_B$ are coded into a binary format. When the measurement amplitude is quantized to B bits, the difference between the true and the quantized signal value is called the quantization error. This error is bounded by

$$|e(m)| < \Delta/2, \tag{21}$$

where Δ is related to B through

$$\Delta = 2^{-B}. \tag{22}$$

The quantization error of a signal can be defined as an additive uniform white noise affecting the measurements

$$\mathbf{y} = \mathbf{y}_B + \mathbf{e}, \tag{23}$$

where $\mathbf{e}$ is the quantization error vector with elements $e(m)$. The mean and variance of the quantization noise are calculated as [14]

$$\mu_{\mathbf{e}} = \text{E}\{\mathbf{e}\} = 0, \tag{24}$$

$$\sigma^2_{\mathbf{e}} = \Delta^2/12. \tag{25}$$

In many real-world applications, in-phase and quadrature component are used to represent real and imaginary part of the complex valued signal. In mathematical notation and derivation, they are considered as complex-valued, while in the hardware implementation the real-valued (in-phase) part and the imaginary (quadrature) part are stored in separate registers, with appropriate combinations for arithmetic operations. Note that, for a complex-valued signal, both real and imaginary part contribute to the noise. Therefore, in this case, the variance of the quantization noise can be written as

$$\sigma^2_{\mathbf{e}} = 2\Delta^2/12 = \Delta^2/6. \tag{26}$$

Considering $\mathbf{y}$ as noisy measurements, the initial estimate will result in a noisy $X_0(k)$. Since $X_0(k)$ is calculated from (9), with the quantization noise in measurements, its variance will be

$$\sigma^2_{X_0(k)} = \sigma^2_{\mathbf{e}}. \tag{27}$$

Therefore, the noise variance in the output (reconstructed) coefficients, for the system shown in Fig. 1, is equal to the input noise variance

$$\text{var}\{X_R(k)\} = \sigma^2_{\mathbf{e}}. \tag{28}$$

Since only K, out of N, coefficients are used in the reconstruction, the energy of the reconstruction error is

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\sigma^2_{\mathbf{e}}, \tag{29}$$

where for notation simplicity we have used $\|\mathbf{X}_R - \mathbf{X}_K\|_2^2$ to denote the expected value of the squared norm-two of the vector $\mathbf{X}_R - \mathbf{X}_K$. The full and complete notation of the left side of (29) would be $\text{E}\{\|\mathbf{X}_R - \mathbf{X}_K\|_2^2\}$.

D. SPARSITY TO NUMBER OF BITS RELATION

Based on the previous relations, influence of the quantization with B bits can be related to the sparsity K. The error energy in the reconstructed coefficients will remain the same if

$$K\sigma_{\mathbf{e}}^2 = K\frac{2^{-2B}}{6} = \text{const.} \tag{30}$$

It means that the reduction from B to $B-1$ bits will require the sparsity reduction from K to $K/4$. The logarithmic form of the reconstruction error is

$$e^2 = 10\log\left(\|\mathbf{X}_R - \mathbf{X}_K\|_2^2\right) = 3.01\log_2 K - 6.02B - 7.78.$$

V. NONSPARSITY INFLUENCE

Due to many circumstances, majority of signals in real-world scenarios are only approximately sparse or nonsparse. This means that a signal, in addition to the K largest coefficients, has $N-K$ coefficients which are small but nonzero. Assume such an approximately sparse (or nonsparse) signal $\mathbf{X}$. The signal is reconstructed under the K-sparsity constraint using Algorithm 1, with the reconstruction conditions being satisfied in the CS sense, thus allowing that the algorithm can detect the largest K coefficients.

The reconstructed signal $\mathbf{X}_R$ then has K reconstructed coefficients with amplitudes $X_R(k_1),\ X_R(k_2),\ldots,X_R(k_K)$. The remaining $N-K$ coefficients, which are not reconstructed, are treated as a noise in these K largest coefficients. The variance from a nonzero coefficient, according to (14), is $|X(k_l)|^2\sigma_\mu^2$. The total energy of noise in the K reconstructed coefficients $\mathbf{X}_R$ will be

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\sigma_\mu^2 \sum_{l=K+1}^{N} |X(k_l)|^2, \tag{31}$$

where $\mathbf{X}_K$ is the sparse version of the original (nonsparse) signal, that is, a signal with the K largest coefficients from $\mathbf{X}$, and others set to zero. Denoting the energy of the remaining signal, when the K largest coefficients are removed from the original signal, by

$$\|\mathbf{X} - \mathbf{X}_K\|_2^2 = \sum_{l=K+1}^{N} |X(k_l)|^2 \tag{32}$$

we get

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\sigma_\mu^2 \|\mathbf{X} - \mathbf{X}_K\|_2^2. \tag{33}$$

For the partial DFT measurement matrix, the result will be

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\frac{N-M}{M(N-1)} \|\mathbf{X} - \mathbf{X}_K\|_2^2. \tag{34}$$

In the case when the signal is strictly K-sparse, that is, $\mathbf{X} = \mathbf{X}_K$, and when the reconstruction is performed with the non-quantized measurements, the reconstruction is ideal and the error is $\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = 0$ (or negligible). Since the measurements are quantized to B-bits, the error of the form (29) will be introduced.

In the case of a nonsparse signal, a general expression is obtained by combining (29) and (33) to get

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\sigma_\mu^2 \|\mathbf{X} - \mathbf{X}_K\|_2^2 + K\sigma_{\mathbf{e}}^2. \tag{35}$$

This result will be validated by examples in the next section, by calculating the signal-to-noise ratio (SNR) of each result

$$SNR_{th} = 10\log\left(\frac{\|\mathbf{X}_K\|_2^2}{K\sigma_\mu^2\|\mathbf{X} - \mathbf{X}_K\|_2^2 + K\sigma_{\mathbf{e}}^2.}\right) \tag{36}$$

and comparing it with the statistical SNR given by

$$SNR_{st} = 10\log\left(\frac{\|\mathbf{X}_K\|_2^2}{\|\mathbf{X}_R - \mathbf{X}_K\|_2^2}\right). \tag{37}$$

VI. NOISE FOLDING QUANTIZATION

In this section, the results will be extended to the noise-folding effects. Here, it his assumed that the sparse values $\mathbf{X}$ are already sensed and stored within the finite-length registers, before the measurement matrix is applied, [33], [34]. This quantization is modeled by the quantization noise $\mathbf{z}$ in the signal coefficients $\mathbf{X}$, prior to taking the measurements. In this case, the measurements are of the form

$$\mathbf{y}_B + \mathbf{e} = \mathbf{A}(\mathbf{X} + \mathbf{z}), \tag{38}$$

which can be rewritten as

$$\mathbf{y}_B + \mathbf{v} = \mathbf{A}\mathbf{X}, \tag{39}$$

where $\mathbf{v} = \mathbf{e} - \mathbf{A}\mathbf{z}$, and the total quantization noise, affecting the measurements, is denoted by $\mathbf{e}$, with covariance $\sigma_{\mathbf{e}}^2\mathbf{I}$. The quantization noise vector $\mathbf{z}$ is random with covariance $\sigma_{\mathbf{z}}^2\mathbf{I}$ being independent of $\mathbf{e}$. Therefore, the resulting noise $\mathbf{v}$ is characterized by a covariance matrix

$$\mathbf{C} = \sigma_{\mathbf{e}}^2\mathbf{I} + \sigma_{\mathbf{z}}^2\mathbf{A}\mathbf{A}^H. \tag{40}$$

If the considered measurement matrix $\mathbf{A}$ is formed as the partial Fourier matrix, the relation $\mathbf{A}\mathbf{A}^H = \frac{N}{M}\mathbf{I}$ holds. The variance of $\mathbf{v}$ is then

$$\sigma_{\mathbf{v}}^2 = \sigma_{\mathbf{e}}^2 + \frac{N}{M}\sigma_{\mathbf{z}}^2, \tag{41}$$

with the covariance matrix $\mathbf{C} = \sigma_{\mathbf{v}}^2\mathbf{I}$.

Each reconstructed coefficient can be written as

$$X_R(k_m) = X(k_m) + v(k_m),$$

according to Fig. 1, where the variance of noise $v(n)$ is given by (41). For the sparse case, the quantization error is present in K nonzero elements of $\mathbf{X}$, only, yielding

$$\begin{aligned}\mathrm{E}\{\sum_{m=1}^{K} |X_R(k_m) - X(k_m)|^2\} &= \|\mathbf{X}_R - \mathbf{X}_K\|_2^2 \\ = \mathrm{E}\{\sum_{m=1}^{K} |v(k_m)|^2\} &= K\sigma_{\mathbf{v}}^2 = K\sigma_{\mathbf{e}}^2 + \frac{KN}{M}\sigma_{\mathbf{z}}^2.\end{aligned} \tag{42}$$

For the nonsparse signal and the partial DFT matrix, the term $K(N-M)/(M(N-1))\|\mathbf{X} - \mathbf{X}_K\|_2^2$ is added to the right side of (42)

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\sigma_{\mathbf{e}}^2 + \tfrac{K}{M}\sigma_{\mathbf{z}}^2 + K\tfrac{N-M}{M(N-1)}\|\mathbf{X} - \mathbf{X}_K\|_2^2, \tag{43}$$

where it is assumed that the quantization of K the largest elements in $\mathbf{X}$ is dominant in that part of the error. All previous relations, for various measurement matrices, can be applied to this case.

VII. FLOATING POINT REGISTERS

In floating point registers, the quantization error is modeled by a multiplicative error

$$y_B(m) = y(m) + y(m)e(m), \tag{44}$$

where $\mathbf{e}$ is the quantization error vector with elements $e(n)$. As in classical digital signal processing, for the analysis of floating point arithmetics, it will be assumed that the sparse coefficients $X(k_i)$, $i = 1, 2, \ldots, K$, are independent random variables, with the variance σ_X^2 and the mean μ_X. The coefficients $X(k_i)$ are statistically independent from the measurement matrix $\mathbf{A}$ elements $a_m(k)$. The mean value of the quantization error is

$$\mathrm{E}\{y(m)e(m)\} = 0.$$

The variance is

$$\mathrm{var}\{y(m)e(m)\} = \mathrm{var}\{y(m)\}\mathrm{var}\{e(m)\}.$$

For $y(m) = \sum_{i=1}^{K} X(k_i)a_m(k_i)$ we get

$$\mathrm{var}\{y(m)\} = \sigma_X^2 \sum_{i=1}^{K} \mathrm{E}\{|a_m(k_i)|^2\},$$

for all measurement matrices with the normalized energy columns, when their elements $a_m(k)$ are equally distributed. This means that the quantization noise $y(m)e(m)$ has the variance $\sigma_X^2\sigma_{\mathbf{e}}^2 K/M$ and we can write (see Remark 2 and Fig. 1)

$$\sigma_{X_0(k)}^2 = \frac{K}{M}\sigma_X^2\sigma_{\mathbf{e}}^2 \tag{45}$$

with

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = \frac{K^2}{M}\sigma_X^2\sigma_{\mathbf{e}}^2. \tag{46}$$

All formulas, in various considered scenarios, can now be rewritten, including the cases of nonsparse signals and noise folding.

For example, if the measurements are normalized such that $\mathrm{E}\{y^2(m)\} = \sigma_X^2 K/M = 1$ then $\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\sigma_{\mathbf{e}}^2$, that is the floating-point arithmetics produces the same results as the fixed-point arithmetics. However, if the range of the measurement values is lower, for example, $\mathrm{E}\{y^2(m)\} = \sigma_X^2 K/M = 1/10$, then the floating-point arithmetics will produce ten times lower error, $\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\sigma_{\mathbf{e}}^2/10$.

VIII. NUMERICAL RESULTS

In this section, we have performed reconstruction with various number of: bits, sparsities (included approximately sparse signals), and available samples. Various real-valued and complex-valued measurement matrices are considered: (1) Partial DFT, (2) Random partial DFT, (3) Equiangular thigh frame (ETF), (4) Gaussian random, (5) Uniform random, and (6) Bernoulli matrix. Three different methods of reconstruction: (a) orthogonal matching pursuit (OMP), (b) Iterative hard thresholding (IHT), and (c) Bayesian-based reconstruction are used to test the theoretic results. An algorithmic form of all considered reconstruction methods is provided for readers to easily reproduce the results. In total, for all the considered cases, we have performed about $150,000$ realizations to statistically test the presented theoretic results, with random signal parameters.

Example 1: One realization of a sparse and nonsparse signal will be considered as an illustration of the reconstruction. a) Consider an $N = 256$-dimensional signal of sparsity $K = 10$, whose $M = N/2$ available measurements are stored with $B = 6$ bits. The measurements matrix is a partial DFT matrix with randomly selected M out of N rows from the full DFT matrix, with columns being energy normalized (for hardware realization of the DFT see, for example, [41], [42]). The sparsity domain coefficients are assumed in the form

$$X(k_p) = \begin{cases} \frac{\sqrt{M}}{K}(1-\nu(p)), & \text{for } p = 1, \ldots, K \\ 0, & \text{for } p = K+1, \ldots, N, \end{cases} \tag{47}$$

where $\nu(p)$ is a random variable with the uniform distribution from 0 to 0.4.

Since this signal is sparse, the reconstruction error is defined by (29). The SNR is defined by (36) with $\|\mathbf{X} - \mathbf{X}_K\|_2^2 = 0$. The original signal and the reconstructed signal are shown in Fig. 2(top). The statistical SNR is $SNR_{st} = 42.35$ dB and the theoretical value is $SNR_{th} = 42.56$ dB.

b) The signal from a), with $K = 10$ significant coefficients, is considered here. However, we will also assume that the remaining $N - K$ coefficients are small but not zero-valued,

$$X(k_p) = \begin{cases} \frac{\sqrt{M}}{K}(1-\nu(p)), & \text{for } p = 1, \ldots, K \\ \frac{\sqrt{M}}{K}\exp(-p/K), & \text{for } p = K+1, \ldots, N, \end{cases} \tag{48}$$

with $\nu(p)$ being a random variable with uniform distribution from 0 to 0.4 as in a) and $N = 256$. The number of bits in the registers where the measurements are stored is $b = B + 1 = 7$. The original signal and the signal reconstructed under sparsity constraint, using $M = N/2$ measurements, are shown in Fig. 2(bottom). The statistical SNR is $SNR_{st} = 33.33$ dB and the theoretical is $SNR_{th} = 33.68$ dB.

Example 2: Statistical analysis of the sparse signal reconstruction, whose form is given in (47), is performed in this example. Random uniform changes of the coefficient amplitudes $\nu(p)$ are assumed from 0 to 0.2. The numbers of quantized measurements $M = 192$, $M = 170$, and $M = 128$ are considered. Typical cases for the measurements quantization to $B \in \{4, 6, 8, 10, 12, 14, 16, 18, 20, 24\}$ bits are analyzed.

Signals with sparsity levels $K \in \{5, 10, 15, 20, 25, 30\}$ are considered. The average statistical signal-to-noise ratio,

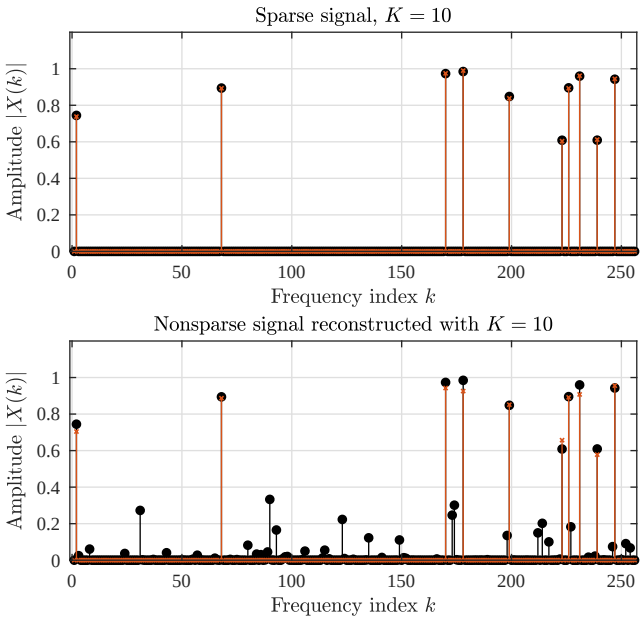

FIGURE 2. Reconstruction results for $N = 256$-dimensional signal whose $M = 128$ measurements are stored into registers with $b = B + 1 = 7$ bits. Reconstruction of a sparse signal with $K = 10$ nonzero coefficients (top). Reconstruction of a nonsparse signal assuming its sparsity $K = 10$ (bottom). The original signal is colored in black, while the reconstructed signal is denoted by red lines and crosses.

SNR_{st}, and the theoretical signal-to-noise ratio, SNR_{th}, values over 100 realizations are presented in Fig. 3(a)-(c). Black dots represent the statistical results, SNR_{st}, and the dash-dot lines show the theoretical results, SNR_{th}. The agreement is high.

For nonsparse signals we used the model in (48). Random changes of the coefficient amplitudes $\nu(p)$ are assumed from 0 to 0.2, while the amplitudes of the coefficients $X(k)$ for $k_p \notin \mathbb{K}$ are of the form $X(k_p) = \exp(-p/(8K))$ in order to reduce its influence to the quantization level. With such amplitudes of the nonsparse coefficients, the quantization error dominates in the reconstruction up to $B = 14$, while the nonsparse energy is dominant for $B \geq 16$, as it can be seen in Fig. 3(d)-(f). Statistics is again in full agreement with the theoretical results.

Finally, the noise folding effect is included, taking into account that the input coefficients $X(k)$ are quantized, in addition to the quantization of measurements $y(m)$. Since the folding part of the quantization error is multiplied by $K/M \ll 1$ in (42), the results do not differ from those presented in Fig. 3(a)-(c). In order to test the influence of noise folding we assumed that the quantized input coefficients $X(k)$ contain an additional noise. An input additive complex-valued i.i.d. Gaussian noise, with the variance $\sigma_z = 0.0001$, is added to these coefficients. This noise is of such a level that it does not influence the quantization error for $B < 14$. However, for $B \geq 14$, it becomes larger than the quantization error and its influence becomes dominant. The results with the quantization and the noise folding, with the additional noise, are shown in Fig. 3(g)-(i).

Example 3: The statistical analysis is extended to other forms of the measurement matrices, namely, to the ETF, the Gaussian, and the uniform random matrix. All three forms of the signal and the quantization error are considered here with $M = 128$ measurements. Sparse and nonsparse signals, described in Example 2, are used in the analysis. The reconstruction error with various number of bits $B = \{4, 6, 8, 10, 12, 14, 16, 18, 20, 24\}$ used in the quantization, and various assumed sparsity levels K is shown in Fig. 4(a)-(c). The results for nonsparse signals, reconstructed with the assumed sparsity, are presented in Fig. 4(d)-(f). The noise folding is analyzed for a reduced number of bits in the quantization of $X(k)$ and presented in Fig. 4(g)-(i).

Example 4: The analysis of quantization effects is done with the assumption that the quantization errors are uncorrelated. This condition is met for all previously considered matrices. However, in the case of the Bernoulli measurement matrix and a small signal sparsity this condition does not hold, meaning that we cannot expect quite accurate estimation of the statistical error using the previous formulas. To explain this effect, we will start with the simplest case of the signal whose sparsity is $K = 1$. The measurements are $y(m) = a_m(k_1)X(k_1)$. For all previously considered matrices $y(m)$ and $y(n)$ are different for $m \neq n$ and the quantization errors are independent. However, for the Bernoulli measurement matrix, we have $y(m) = \pm X(k_1)/\sqrt{M}$. These measurements will produce only two possible quantization errors for all $m = 1, 2, \ldots, M$. It means that $M/2$ errors in the initial estimate will sum up in phase, producing the mean squared error with variance $\text{var}\{X_R(k)\} = \frac{M}{2}\sigma_{\mathbf{e}}^2$. This is significantly higher than $\text{var}\{X_R(k)\} = \sigma_{\mathbf{e}}^2$ in other cases. For $K = 2$, we get the measurements $y(m) = (\pm X(k_1) \pm X(k_2))/\sqrt{M}$, producing four possible values for $y(m)$ and only four possible values of the quantization error. For large K, the number of possible levels increases and the result for the variance converges to the one for uncorrelated quantization errors, $\text{var}\{X_R(k)\} = \sigma_{\mathbf{e}}^2$, obtained under the assumption that all M measurements $y(m)$ are different. The results for the Bernoulli measurement matrix, with the described correction for a small K, are shown in Fig. 5.

Example 5: The results in previous four examples are obtained based on the OMP reconstruction method (Algorithm 1), which is also used for the derivation of the theoretical results. Here we will show that we may expect similar results for other reconstruction methods, as far as the reconstruction conditions are met. The simulations for the reconstruction with the partial DFT measurement matrix with sparse and nonsparse signals, including noise folding, are repeated with the iterative hard thresholding (IHT) reconstruction method, given in Algorithm 2 [13], [29]. The theoretical and statistical errors are shown in Fig. 6(a)-(c), showing high agreement between the statistical and theoretical results.

Example 6: In this example, the reconstruction of sparse and nonsparse signals, including noise folding, is performed

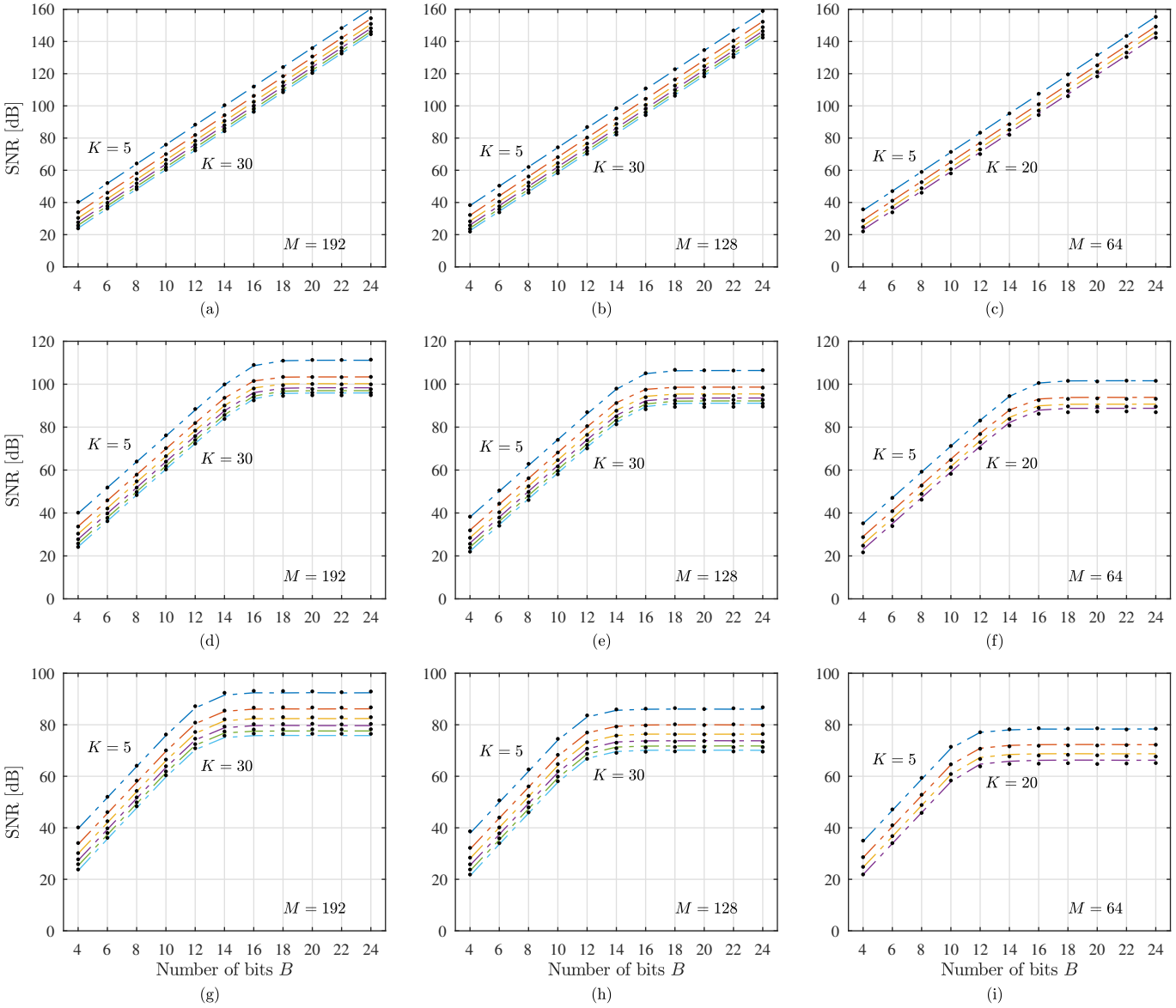

FIGURE 3. Reconstruction error for the measurements quantized to B bits for various sparsity K and the numbers of measurements M. Partial DFT measurement matrix is used. Statistical results are marked by black dots, while the theoretical results are shown by dot-dashed lines. The sparsity value is varied from $K = 5$ to the maximum K indicated in the panels, with a step of 5. (a)-(c) Reconstruction error (theory and statistics) for the sparse signals when only the measurements $\mathbf{y}$ are quantized to B bits (to fit $b = B + 1$ fixed-point registers). (d)-(f) Reconstruction error (theory and statistics) for nonsparse signals when only the measurements $\mathbf{y}$ are quantized to $B + 1$ bit fixed-point registers. (g)-(i) Reconstruction error (theory and statistics) for nonsparse signals when both the measurements $\mathbf{y}$ and the noisy input coefficients $\mathbf{X}$ are quantized to B bits (quantization noise folding with additive input noise).

using the Gaussian measurement matrix and the Bayesian-based method, [36], [37], summarized in Algorithm 3. The theoretical and statistical errors for the signals from Example 1 are shown in Fig. 7(a)-(c). Again, a high agreement between the statistical and theoretical results is obtained.

IX. PROBABILITY OF MISDETECTION

The results for the MSE are derived under the condition that the quantization does not influence the reconstruction condition and that the signal can be uniquely recovered from the available set of measurements. However, the quantization will also degrade conditions for the signal recovery, as indicated in the Appendix. In this section, the probability of misdetection, influenced by the quantization, will be studied. Although the probabilistic approach can also be used to derive a relation between K, N, and M, for a successful reconstruction of $\mathbf{X}$, with a given probability, as thoroughly shown in [6], [38]–[40], we will here use statistical analysis and compare the reconstruction results for the Gaussian (real-valued) and partial DFT measurement matrix, with $N = 256$ and $M = 128$. The reconstruction is performed using two already described robust CS algorithms, a matching pursuit algorithm and the iterative hard thresholding algorithm.

In the experiment, the measurement matrices were used

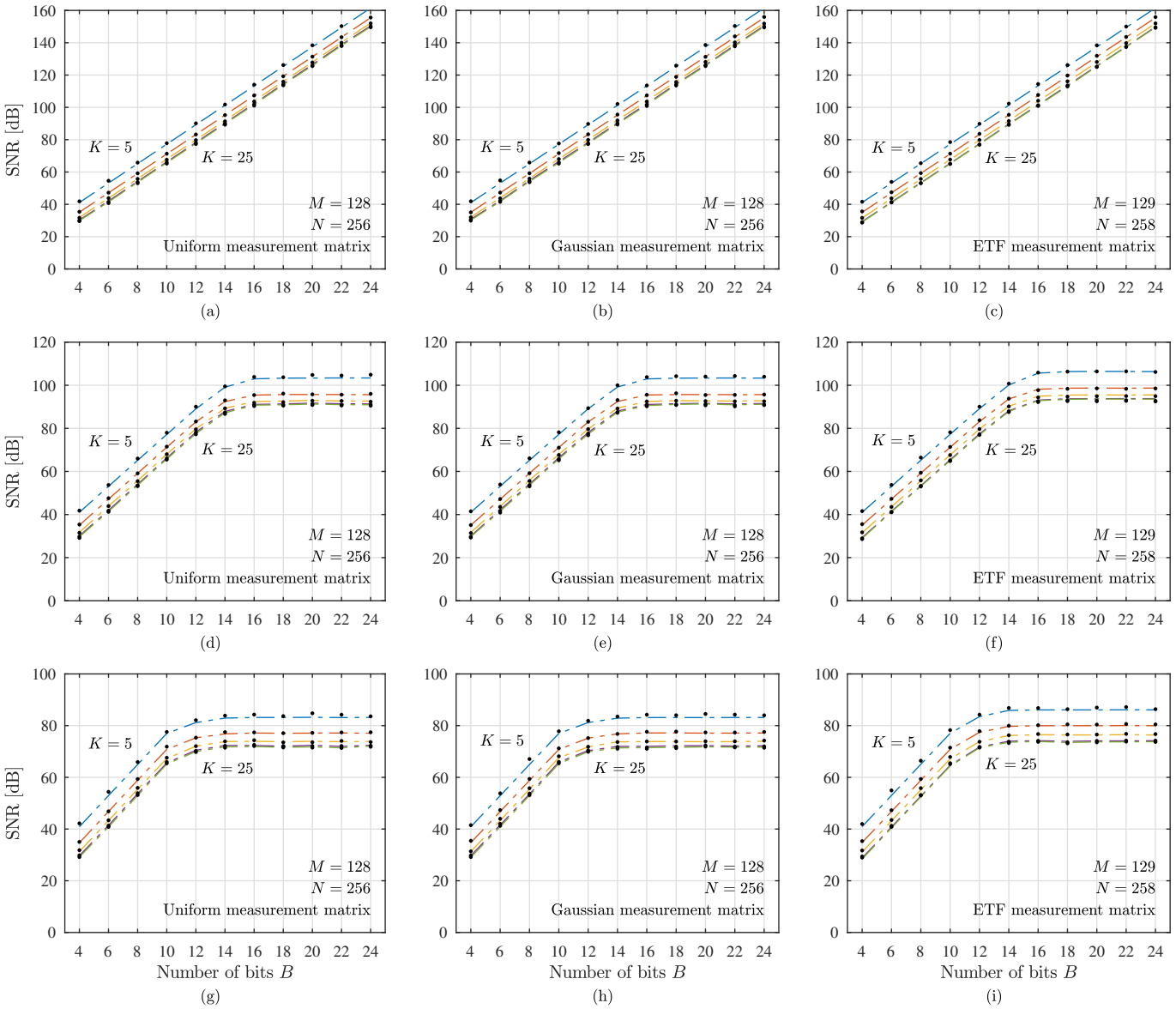

FIGURE 4. Reconstruction error for various measurement matrices. (a)-(c) Sparse signal with measurements quantized to fit the registers with $b = B + 1$ bits for various sparsities using the uniform, Gaussian, and ETF measurement matrices, respectively. (d)-(f) Nonsparse signal with the measurements quantized to fit the registers with $B + 1$ bits for various sparsity K using the uniform, Gaussian, and ETF measurement matrices, respectively. (g)-(i) Nonsparse signals when both the measurements $\mathbf{y}$ and input coefficients $\mathbf{X}$ are quantized to $B + 1$ bit fixed point registers (quantization noise folding)using the uniform, Gaussian, and ETF measurement matrices, respectively.

to reconstruct signals for a range of sparsity degrees, $K = 1, 2, 3, \ldots, 70$. For each sparsity level, K, the problem was solved 10,000 times with random positions of the nonzero coefficients in each signal realization. For each K, the solution was checked against the known positions and the values of the nonzero elements in $X(k)$, and the number of misdetections was recorded. The number of misdetections for each K was then divided by the total number of realizations.

The results for the misdetection probability are shown in Fig. 8. Note that, in the DFT case for $b = 1$, the reconstruction of components is non-unique for small sparsity, meaning that a non-unique solution is obtained with the probability order of 0.001. That is presented by the dots in Fig. 8 (b) and (c).

Observe that at a position of the nonzero coefficients, $k \in \{k_1, k_2, \ldots, k_K\}$, the initial estimate has the characteristics of a random variable with the Gaussian distribution of the form $\mathcal{N}(1, (K-1)\sigma_\mu^2 + \sigma_{\mathbf{e}}^2)$, while at the other positions, $k \notin \{k_1, k_2, \ldots, k_K\}$, the corresponding distribution is also Gaussian, but with the mean-value equal to zero, $\mathcal{N}(0, K\sigma_\mu^2 + \sigma_{\mathbf{e}}^2)$. Based on these probability density functions, an exact probabilistic analysis can be performed, as it has been done in [6], [38], for some specific measurement matrices.

Based on the presented theory, it can be easily determined when the number of bits, B, loses the dominant influence on

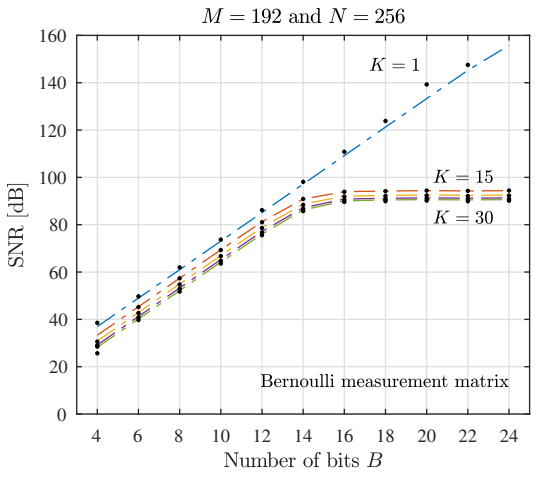

FIGURE 5. Reconstruction error for the Bernoulli matrix for nonsparse signals with the measurements quantized to fit the registers with $B+1$ bits for various sparsities $K \in \{1, 15, 20, 25, 30\}$.

Algorithm 2 Iterative Hard Thresholding (IHT) Reconstruction Algorithm

Input: Vector $\mathbf{y}$, Matrix $\mathbf{A}$, Assumed sparsity K, Number of iterations I_t, and parameter τ.

$\mathbf{X}_0 \leftarrow \mathbf{0}$
for $i = 1$ **do** I_t
  $\mathbf{Y} \leftarrow \mathbf{X}_0 + \tau \mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{X}_0)$
  $\mathbb{K} \leftarrow \text{sort}(|\mathbf{Y}|)$, indices of K largest $|\mathbf{Y}|$
  $\mathbf{X}_0 \leftarrow \mathbf{0}$,
  $X_0(k) \leftarrow Y(k)$ for $k \in \mathbb{K}$ ▷ Hard Thresholding
end for

Output: Reconstructed $\mathbf{X}_R = \mathbf{X}_0$, the set of positions $\mathbb{K}$.

the probability of the positions misdetection. For example, in the case of the Gaussian measurement matrix and $M = 128$ available samples, for the second term in the variance

$$\sigma^2_{X_0(k)} = K\frac{1}{M} + \frac{2^{-2B}}{12},$$

for $B \gg 1$, we have $2^{-2B}/12 \to 0$, meaning that the quantization influence to the reconstruction results is negligible, and

$$\sigma^2_{X_0(k)} \approx K\frac{1}{M}$$

holds. Similar result is obtained for $B = 3$, the second term in $\sigma^2_{X_0(k)}$ is

$$\frac{2^{-2B}}{12} = \frac{1}{768} \ll K\frac{1}{M} = K\frac{1}{128},$$

which means that in this case $\sigma^2_{X_0(k)} \approx K\frac{1}{M}$ also holds. This is confirmed in Fig. 8(a), where for $B = 3$ ($b = B + 1 = 4$) and for $B = 15$ ($b = 16$) curves of the misdetection probability are very close. Observe that for a low number of bits, for example, in the case of $B = 1$ ($b = 2$), the

Algorithm 3 Bayesian-based reconstruction

Input: Vector $\mathbf{y}$, Matrix $\mathbf{A}$

$\alpha_i \leftarrow 1$ ▷ For $i = 1, 2, \ldots, N$
$\sigma^2 \leftarrow 1$ ▷ Initial estimate
$T_h = 10^2$ ▷ Threshold
$\mathbf{p} = [1, 2, \ldots, N]^T$
repeat
  $\mathbf{D} \leftarrow$ diagonal matrix with d_i values
  $\mathbf{\Sigma} \leftarrow (\mathbf{A}^T\mathbf{A}/\sigma^2 + \mathbf{D})^{-1}$
  $\mathbf{V} \leftarrow \mathbf{\Sigma}\mathbf{A}^T\mathbf{y}/\sigma^2$
  $\gamma_i \leftarrow 1 - d_i\Sigma_{ii}$ ▷ For each i
  $d_i \leftarrow \gamma_i/V_i$ ▷ For each i
  $\sigma^2 \leftarrow \dfrac{\|\mathbf{y} - \mathbf{A}\mathbf{V}\|^2}{M - \sum_i \gamma_i}$
  $\mathbb{R} \leftarrow \{i : |d_i| > T_h\}$
  Remove columns from matrix $\mathbf{A}$ selected by $\mathbb{R}$
  Remove elements from array d_i selected by $\mathbb{R}$
  Remove elements from vector $\mathbf{p}$ selected by $\mathbb{R}$
until stopping criterion is satisfied
Reconstructed vector $\mathbf{X}$ nonzero coefficients are in vector $\mathbf{V}$ with corresponding positions in vector $\mathbf{p}$, $X_{p_i} = V_i$

Output:

- Reconstructed signal vector $\mathbf{X}_R = \mathbf{V}$, the set of positions $\mathbb{K} = \mathbf{p}$.

quantization effect has a strong influence on the misdetection probability, as the quantization related term in the variance is equal to $\frac{2^{-2B}}{12} = \frac{1}{12}$ in the first case, and to $\frac{2^{-2B}}{12} = \frac{1}{24}$ in the second case, which is now significant, when compared to the part of the variance induced by the reduced set of measurements, $K\frac{1}{M} = K\frac{1}{128}$. This is numerically confirmed in Fig. 8(a), where the statistical results indicate that the quantization effect overpowers the influence of missing samples and increases the component misdetection probability. Similar analysis can be performed for the partial DFT measurement matrix.

X. CONCLUSIONS

The effects of quantization noise to the reconstruction of signals under sparsity constraint are analyzed in this paper. If the measurements are not quantized, the reconstruction would be ideal and the error negligible. However, the quantization is inevitable, since the hardware realization of systems cannot store the exact values of samples. We have derived the exact error of the reconstruction due to quantization noise. The cases when a signal is not strictly sparse are analyzed, as well as the noise folding effect. The reconstruction performances are validated on the numerical examples, and compared to the statistical error calculation, producing a high agreement between the numerical and theoretical results.

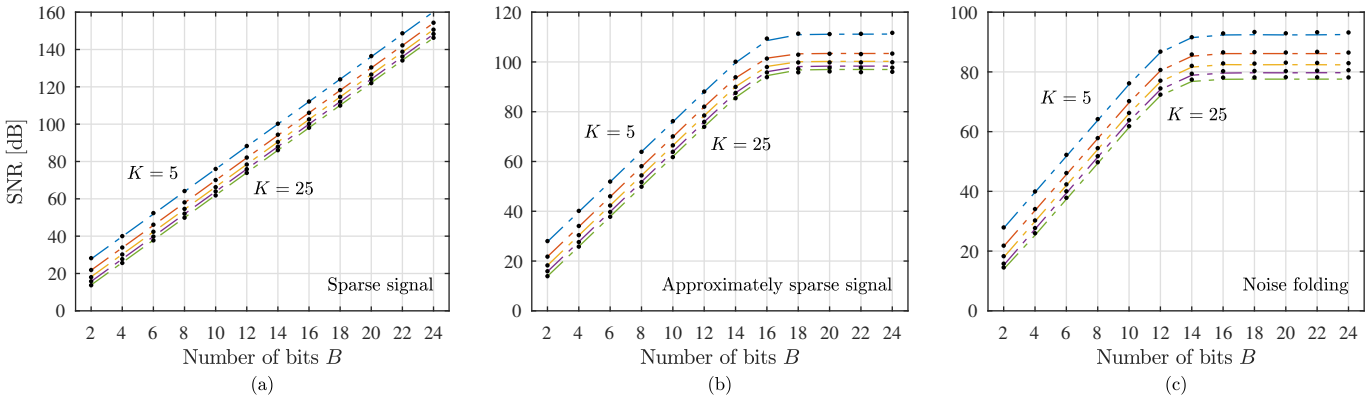

FIGURE 6. Reconstruction error for the partial DFT measurement matrix when the Iterative Hard Thresholding (IHT) reconstruction algorithm is used. (a) Sparse signal with the measurements quantized to fit the registers with $B+1$ bits for various sparsit K. (b) Nonsparse signal with the measurements quantized to fit the registers with $b = B+1$ bits for various sparsity K. (c) Nonsparse signals when both the measurements $\mathbf{y}$ and input coefficients $\mathbf{X}$ are quantized to $B+1$ bit fixed point registers (quantization noise folding).

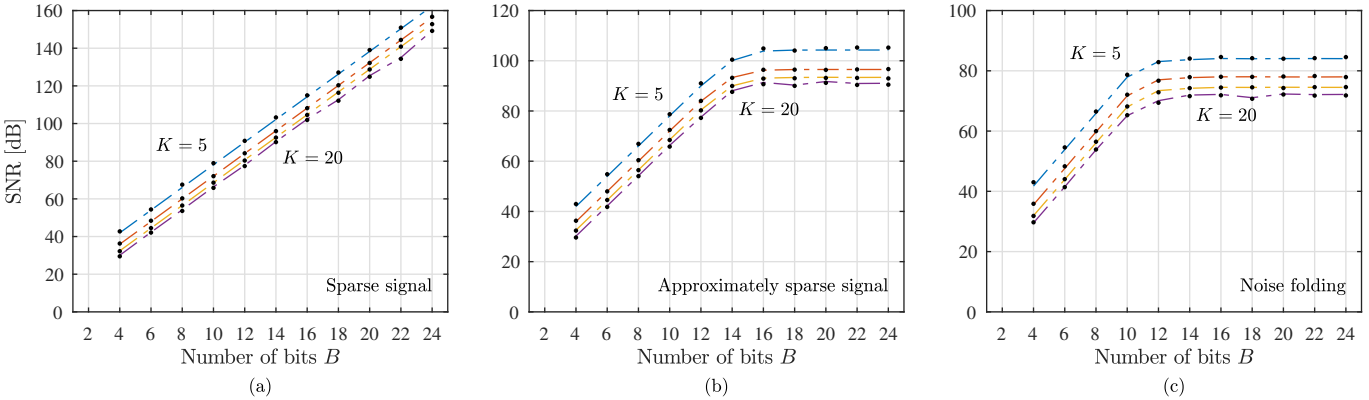

FIGURE 7. Reconstruction error for the partial Gaussian measurement matrix ($M = 128$, $N = 256$) when the Bayesian-based reconstruction algorithm is used. (a) Sparse signal with the measurements quantized to fit the registers with $B+1$ bits for various sparsity K. (b) Nonsparse signal with the measurements quantized to fit the registers with $B+1$ bits for various sparsities. (c) Nonsparse signals when both the measurements $\mathbf{y}$ and the input coefficients $\mathbf{X}$ are quantized to $B+1$ bit fixed point registers (quantization noise folding).

APPENDIX: INFLUENCE OF QUANTIZATION TO THE RECONSTRUCTION CONDITION

Quantization noise can be included in the coherence index based relation for the reconstruction. The worst case amplitude of the considered normalized coefficient in the initial estimate is $1-(K-1)\mu-\nu\Delta/2$ where $\nu = \max_k \sum_{m=1}^{M}|a_m(k)| \leq \sqrt{M}$. This inequality follows from the relation between the norm-two and the norm-one of a vector. For the partial DFT matrix, the random partial DFT matrix and the Bernoulli matrix, the equality $\nu = \sqrt{M}$ holds. Following the same reasoning as in Remark 1, we may conclude that at a position where the original coefficient $X(k)$ is zero-valued, in the worst case, the maximum possible disturbance is $K\mu + \sqrt{M}\Delta/2$. The detection of the strongest coefficient position is always successful if $1-(K-1)\mu-\sqrt{M}\Delta/2 > K\mu+\sqrt{M}\Delta/2$, producing the condition for reconstruction

$$K < \frac{1}{2}\left(1+\frac{1}{\mu}-\frac{\sqrt{M}\Delta}{\mu}\right).$$

Influence of the quantization error to the uniqueness condition will be negligible if $\sqrt{M}\Delta \ll 1$ holds.

Coherence index based values guarantee exact reconstruction, however, they are pessimistic. More practical relations can be obtained by considering the probabilistic analysis [6], [38]. The resulting disturbance in the initial estimate, at the position $k \in \mathbb{K}$, due to the other coefficients and the quantization noise behaves as the Gaussian random variable $\mathcal{N}(1,(K-1)\sigma_\mu^2+\sigma_{\mathbf{e}}^2)$, for $K \gg 1$. The initial estimate at $k \notin \mathbb{K}$ behaves as $\mathcal{N}(0, K\sigma_\mu^2+\sigma_{\mathbf{e}}^2)$. Probabilistic analysis may provide approximative relations among N, M, and K, for a given probability. We have performed the statistical analysis with various measurement matrices. The results of this analysis lead to the conclusion that for high probabilities of the reconstruction we may neglect the quantization effect influence to the reconstruction condition for $B \geq 4$.

Note that for large sparsities K, we have found that the reconstruction probability can be improved by increasing the upper limit for iterations in Algorithm 1 for a few percents, with respect to the expected sparsity K. After the iterations are completed, the expected sparsity K is used in the final

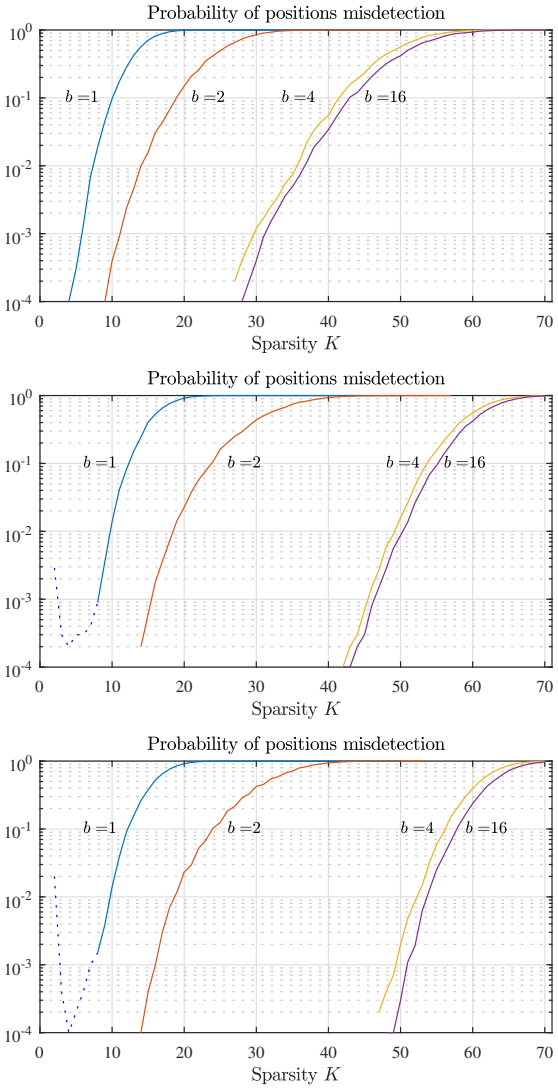

FIGURE 8. Probability of misdetection for: (a) The Gaussian measurement matrix with the OMP. (b) The partial DFT measurement matrix with the OMP. (c) The partial DFT measurement matrix with the IHT.

reconstruction. This solves the problem that the iterative reconstruction in Algorithm 1 cannot produce the exact result if it misses one of the nonzero coefficient positions during the iterative process for large K.

REFERENCES

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

[2] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[3] E. Candes, J. Romberg and T. Tao. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, 2006.

[4] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, 59: 1207-1223., 2006, doi:10.1002/cpa.20124

[5] E. Candes, and J. Romberg. "Encoding the $_p$ from limited measurements," *Data Compression Conference (DCC'06). IEEE*, 2006.

[6] L. Stanković, S. Stanković, and M. Amin, "Missing Samples Analysis in Signals for Applications to L-estimation and Compressive Sensing," *Signal Processing*, vol. 94, pp. 401–408, January 2014.

[7] M. Davenport, M. Duarte, Y. Eldar, and G. Kutyniok, "Introduction to compressed sensing," Chapter in Compressed Sensing: Theory and Applications, Cambridge University Press, 2012.

[8] L. Stanković, and M. Daković, "On the Uniqueness of the Sparse Signals Reconstruction Based on the Missing Samples Variation Analysis," *Mathematical Problems in Engineering*, vol. 2015, Article ID 629759, 14 pages, 2015. doi:10.1155/2015/629759

[9] S. Stanković, L. Stanković, and I. Orović, "A Relationship between the Robust Statistics Theory and Sparse Compressive Sensed Signals Reconstruction,"*IET Signal Processing*, vol. 8, no. 3, May 2014.

[10] G. Dziwoki, G. and M. Kucharczyk, "On a sparse approximation of compressible signals," *Circuits Syst Signal Processing*, 2019. DOI: https://doi.org/10.1007/s00034-019-01287-8

[11] D. Needell, and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis* vol.26, no.3, pp.301–321, 2009.

[12] T. Blumensath, "Sampling and reconstructing signals from a union of linear subspaces," *IEEE Trans. Information Theory* vol.57, no.7, pp.4660–4671, 2011.

[13] T. Bumensath, and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 23, no. 3, pp. 265–274, 2009.

[14] L. Stanković, *Digital Signal Processing with Selected Topics,* CreateSpace Independent Publishing Platform, An Amazon.com Company, Nov.2015.

[15] L. Stanković, E. Sejdić, S. Stanković, M. Daković, and I. Orović "A Tutorial on Sparse Signal Reconstruction and its Applications in Signal Processing" *Circuits, Systems & Signal Processing*, published online 01. Aug. 2018, DOI 10.1007/s00034-018-0909-2.

[16] L. Stanković, M. Daković, I. Stanković, and S. Vujović, "On the Errors in Randomly Sampled Nonsparse Signals Reconstructed with a Sparsity Assumption,"*IEEE Geoscience and Remote Sensing Lett.*, Vol: 14, Issue: 12, pp. 2453–2456 , Dec. 2017, DOI: 10.1109/LGRS.2017.2768664

[17] Z.Zhang, Y. Xu, J. Yang, X. Li, D. Zhang, "A Survey of Sparse Representation: Algorithms and Applications," *IEEE Access*, vol. 3, pp. 490–530, May 2015.

[18] S. Y. Low, "Compressive speech enhancement in the modulation domain" *Speech Communication*, pp. 87–99. (doi:10.1016/j.specom.2018.08.003), September 2018.

[19] E. Sejdic, M. A. Rothfuss, M. L. Gimbel, M. H. Mickle, "Comparative analysis of compressive sensing approaches for recovery of missing samples in implantable wireless Doppler device," *IET Signal Processing*, vol. 8, no. 3, pp. 230-238, 2014.

[20] P. Boufounos, and R. Baraniuk, "1-bit Compressive Sensing," *42nd Conference on Information Sciences and Systems*, Princeton, NJ, USA, 2008.

[21] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings for sparse vectors," *IEEE Trans. Information Theory*, vol. 59. no. 4, pp. 2082–2102, 2011.

[22] P. North, "One-Bit Compressive Sensing with Partial Support Information," *CMC Senior Theses*, paper 1194, [Available online: http://scholarship.claremont.edu/cmc_theses/1194], 2015.

[23] P. T. Boufounos, "Greedy sparse signal reconstruction from sign measurements," *2009 Conference Record of the* 43*rd Asilomar Conference on Signals, Systems and Computers*, CA, USA, 2009.

[24] Laska, Jason N., et al. "Democracy in action: Quantization, saturation, and compressive sensing,"*Applied and Computational Harmonic Analysis* vol. 31, no. 3, pp. 429-443, November 2011.

[25] P. Boufounos and R. Baraniuk, "Quantization of sparse representations," *Data Compression Conf. (DCC) 2007*, Snowbird, UT, USA, March 2007.

[26] W. Dai, and O. Milenkovic, "Information Theoretical and Algorithmic Approaches to Quantized Compressive Sensing," *IEEE Transactions on Communications*, vol. 59, no. 7, July 2011.

[27] P.T. Boufounos, L. Jacques, F. Krahmer, and R. Saab, "Quantization and Compressive Sensing," in: Boche H., Calderbank R., Kutyniok G., Vybial J. (eds) *Compressed Sensing and its Applications. Applied and Numerical Harmonic Analysis*. Birkhäuser, Cham, pp. 193–237, 2013.

[28] Y. Wang, S. Feng, and P. Zhang, "Information Estimations and Acquisition Costs for Quantized Compressive Sensing," *International Conference on Digital Signal Processing (DSP)* 2015, Singapore, Singapore, July 2015.

[29] H.M. Shi, M. Case, X. Gu, S. Tu, and D. Needell, "Methods for quantized compressed sensing," *Information Theory and Applications Workshop (ITA) 2016*, California, USA, 2016.

[30] O. Graf, A. Bhandari, and F. Krahmer, "One-Bit Unlimited Sampling," *ICASSP 2019*, Brighton, United Kingdom, May 2019.

[31] J. Z. Sun and V. K. Goyal, "Optimal quantization of random measurements in compressed sensing," *Proc. IEEE International Symposium on Information Theory (ISIT)*, pp. 6-–10, 2009.
[32] G. K. Vivek, A. K. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 48–56, March 2009.
[33] Davenport, Mark A., et al. "The pros and cons of compressive sensing for wideband signal acquisition: Noise folding versus dynamic range," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4628–4642, 2012.
[34] E. Arias-Castro and Y. Eldar, "Noise folding in compressed sensing," *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 478–481, 2011.
[35] L.R. Welch, "Lower Bounds on the Maximum Cross Correlation of Signals,"*IEEE Transactions on Information Theory*, vol. 20, no. 3, pp. 397–399, May 1974.
[36] M. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, JMLR.org, 1, pp. 211–244, 2001.
[37] S. Ji, Y. Xue, L. Carin, "Bayesian Compressive Sensing", *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.
[38] L. Stanković, and M. Brajović, "Analysis of the Reconstruction of Sparse Signals in the DCT Domain Applied to Audio Signals,"*IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no.7, pp.1216–1231, July 2018.
[39] I. Stanković, M. Brajović, M. Daković, C. Ioana, L. Stanković, "On the Quantization and the Probability of Misdetection in Compressive Sensing," *27th Telecommunications Forum TELFOR 2019*, Belgrade, Serbia, November 2019.
[40] L. Stanković, D. Mandić, M. Daković, and I. Kisil, "Demystifying the Coherence Index in Compressive Sensing," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 152–162, Jan. 2020
[41] R. Ferdian, Y. Hou and M. Okada, "A Low-Complexity Hardware Implementation of Compressed Sensing-Based Channel Estimation for ISDB-T System," in *IEEE Transactions on Broadcasting*, vol. 63, no. 1, pp. 92-102, March 2017. doi: 10.1109/TBC.2016.2617286.
[42] J. Yang, C. Zhang, S. Jin, C. Wen and X. You, "Efficient Hardware Architecture for Compressed Sensing with DFT Sensing Matrix," 2016 *IEEE International Workshop on Signal Processing Systems (SiPS)*, Dallas, TX, 2016, pp. 207-212. doi: 10.1109/SiPS.2016.44.

ISIDORA STANKOVIĆ was born in Podgorica, Montenegro in 1993. She received her BSc degree with First Class Honours in Electrical Engineering at the University of Westminster, London, United Kingdom, in 2014, and her MSc degree in Communications and Signal Processing at the Imperial College London, United Kingdom, in 2015. Currently, she is a PhD student on a double-degree program at the Faculty of Electrical Engineering, University of Montenegro and GIPSA Lab at the INP Grenoble, University of Grenoble Alpes. Her interest includes image processing, time-frequency analysis and compressive sensing algorithms.

MILOŠ BRAJOVIĆ was born in Podgorica, Montenegro, in 1988. He received the B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from the University of Montenegro, Podgorica, Montenegro, in 2011, 2013 and 2019, respectively. He is currently working as a postdoctoral teaching assistant at the University of Montenegro. He is a member of the Time-Frequency Signal Analysis Group (www.tfsa.ac.me), University of Montenegro, where he is involved in several research projects. His research interests include signal processing, time-frequency signal analysis, and compressive sensing.

MILOŠ DAKOVIĆ was born in 1970, Nikšić, Montenegro. He received the B.S. degree in 1996, the M.Sc. degree in 2001 and the Ph.D. degree 2005, all at the University of Montenegro in EE. He is a full professor at the University of Montenegro. His research interests are signal processing, time-frequency signal analysis, compressive sensing, graph signal processing, and radar signal processing. He is a member of the Time-Frequency Signal Analysis Group (www.tfsa.ac.me) at the University of Montenegro where he was involved in several research projects supported by Volkswagen foundation, Montenegrian Ministry of Science and Canadian Government (DRDC).

CORNEL IOANA received the Dipl.-Eng. degree in electrical engineering from the Romanian Military Technical Academy of Bucharest, Romania, in 1999 and the M.S. degree in telecommunication science and the Ph.D. degree in the electrical engineering field, both from University of Brest-France, in 2001 and 2003, respectively. Between 1999 and 2001, he activated as a Military Researcher in a research institute of the Romanian Ministry of Defense (METRA), Bucharest, Romania. Between 2003 and 2006, he worked as Researcher and Development Engineer in ENSIETA, Brest, France. Since 2006, he has been an Associate Professor-Researcher with the Grenoble Institute of Technology/GIPSA-lab. His current research activity deals with the signal processing methods adapted to the natural phenomena. His scientific interests are nonstationary signal processing, natural process characterization, underwater systems, electronic warfare, and real-time systems.

LJUBIŠA STANKOVIĆ (*M'91–SM'96–F'12*) was born in Montenegro in 1960. He received the B.S. degree in EE from the University of Montenegro (UoM), the M.S. degree from the University of Belgrade and the Ph.D. in Theory of Electromagnetic Waves from the UoM. As a Fulbright grantee, he spent 1984-1985 academic year at the Worcester Polytechnic Institute, USA. Since 1982, he has been on the faculty at the UoM, where he has been a full professor since 1995. In 1997-1999, he was on leave at the Ruhr University Bochum, Germany, supported by the AvH Foundation. At the beginning of 2001, he was at the Technische Universiteit Eindhoven. He was vice-president of Montenegro 1989-90. During the period of 2003-2008, he was Rector of the UoM. He was Ambassador of Montenegro to the UK, Ireland, and Iceland from 2010 to 2015. His current interests are in Signal Processing. He published about 400 technical papers, more than 160 of them in the leading journals. He was an Associate Editor of the *IEEE Transactions on Image Processing*, the *IEEE Signal Processing Letters*, *IEEE Transactions on Signal Processing*, and numerous special issues of journals. Prof. Stanković is a member of Editorial Board of *Signal Processing* and a Senior area Editor of the *IEEE Transactions on Image Processing*. He is a member of the National Academy of Science and Arts of Montenegro (CANU) since 1996 and its vice-president since 2016. He is a member of Academia Europea (2012). Stankovic (with coauthors) won the Best paper award from the European Association for Signal Processing (EURASIP) in 2017 for a paper published in the *Signal Processing* journal.