On the Errors in Randomly Sampled Nonsparse Signals Reconstructed with a Sparsity Assumption

Ljubiša Stanković, *Fellow, IEEE,* Miloš Daković, *Member, IEEE,*
Isidora Stanković, *Student Member, IEEE,* Stefan Vujović, *Student Member, IEEE*

Abstract—An analysis of errors in the reconstruction of approximately sparse and nonsparse noisy signals in the discrete Fourier transform domain is considered in this paper. Signal reconstruction is performed from a reduced set of data, using compressive sensing methods and the sparsity assumption. Random sampling positions in time are considered. Reconstruction results are compared with those obtained with a subset of uniformly sampled signals. A random subset of uniformly sampled data produces better reconstruction results. Theoretical results are statistically confirmed.

Index Terms—Sparse signals, DFT, Compressive sensing, Random sampling

I. INTRODUCTION

COMPRESSIVE sensing (CS) methods are used to reconstruct sparse signals from reduced sets of measurements [1]–[6]. In general, measurements are linear combinations of the signal coefficients in the sparsity domain. A common transformation and sparsity domain in signal processing is the Discrete Fourier Transform (DFT) domain. Since the signal samples are linear combinations of the DFT coefficients they can be considered as CS measurements. In many real applications, signals that are processed as sparse are only approximately sparse. Even the simplest form of a sinusoidal signal, whose frequency is not on the grid, in theory, is not a sparse signal in the DFT domain. Presence of additive noise is also unavoidable in all practical applications. In order to increase randomness in the measurements and to reduce coherence in the CS approach, random DFT measurement matrices, resulting from random signal sampling, are often preferred and used.

The main result of this paper is a formula relating the expected error energy in the reconstructed coefficients with the energy of the remaining coefficients in a nonsparse signal and the measurement noise. This result is presented in the form of a theorem. Proof of the theorem is composed of three parts. First, the reconstruction of sparse signals is considered. Secondly, the analysis of reconstruction using noisy measurements is added. Finally, the signal sparsity is relaxed to approximately sparse and nonsparse signals.

II. RANDOMLY SAMPLED NONSPARSE SIGNAL WITH ADDITIVE NOISE

Consider a signal $x(t)$ of a duration T and its samples $x(n\Delta t)$ satisfying the sampling theorem. The periodic extension of this signal can be written in a Fourier series (FS) form

$$x(t) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi k \frac{t}{T}}, \qquad (1)$$

where the FS coefficients $X(k)$ are equal to the DFT coefficients if we use the notation $x(n)$ for $x(n\Delta t)$, and $\Delta t = T/N$ as the sampling interval. When the sampling theorem is satisfied then

$$X(k) = \frac{N}{T} \int_0^T x(t) e^{-j2\pi k \frac{t}{T}} dt = \sum_{n=0}^{N-1} x(n) e^{-j2\pi k \frac{n}{N}}. \quad (2)$$

A signal is sparse with sparsity K (K-sparse signal) in the DFT domain if the number of nonzero coefficients $X(k)$, at $k \in \mathbb{K} = \{k_1, k_2, \dots, k_K\}$, is much smaller than the total number of coefficients, that is, $K \ll N$. In general, signals of interest are more likely to be only approximately sparse, or even nonsparse in the transformation domain, than exactly K-sparse [2], [7]. A signal will be referred to as approximately K-sparse in the DFT domain if the amplitudes of $X(k)$ for $k \notin \mathbb{K}$ are much smaller than the amplitudes of $X(k)$ for $k \in \mathbb{K}$. If the most significant amplitudes of $X(k)$ for $k \in \mathbb{K}$ are of the same order as some of the amplitudes of $X(k)$ for $k \notin \mathbb{K}$, the signal is not sparse with sparsity K.

The error in approximately sparse and nonsparse signals, reconstructed from noisy measurements using CS methods with a K-sparsity assumption, will be discussed next.

Consider a signal and its DFT $X(k)$ of length N, in vector notation $\mathbf{X} = [X(0), X(1), \dots, X(N-1)]^T$. Assume that the signal samples $x(t_n)$ are sensed at random positions, $0 \leq t_n < T$. The special case of uniform sampling, when sampling instants t_n are integer multiples of Δt, is considered in [9]. Assume that an input additive white noise $\varepsilon(t_n)$, with variance σ_ε^2, exists in samples $x(t_n)$. Assume that the number of available signal samples is N_A. The signal samples are at the positions defined by $t_n \in \mathbb{T}_A = \{t_1, t_2, \dots, t_{N_A}\}$. The DFT coefficients $X(k)$ are reconstructed from the available set of samples under the assumption that the signal is K-sparse, i.e., that the vector $\mathbf{X}$ has only K nonzero elements.

General bounds for the reconstruction error for nonsparse signals, reconstructed with the sparsity assumption, are given in [2]. An analysis of the FFT algorithms with sparse signals and a reduced number of samples, along with the error bounds

in the reconstruction of sparse signals, has been done in [7], [8]. The exact relation for the expected squared error in the reconstruction of approximately sparse or nonsparse signals from a reduced set of samples, reconstructed under the K-sparsity assumption, is given by the theorem. It will be defined after the notation is introduced next.

Notation:

- $\mathbf{X} = [X(0), X(1), \ldots, X(N-1)]^T$ is the signal transform vector (in the DFT domain of sparsity);
- $\mathbb{T}_A = \{t_1, t_2, \ldots, t_{N_A}\}$ is a set of measurement instants;
- $\mathbf{A}$ is a measurement matrix with elements

$$a_{nk} = \frac{1}{N} \exp(j2\pi k t_n/T), \qquad (3)$$

  for $k = 0, 1, \ldots, N-1$, and $n = 1, 2, \ldots, N_A$;
- $\mathbf{y} = [x(t_1), x(t_2), \ldots, x(t_{N_A})]^T = \mathbf{AX}$ is the measurement vector;
- $\mathbb{K} = \{k_1, k_2, \ldots, k_K\}$ is the set of estimated positions of nonzero coefficients in a K-sparse $\mathbf{X}$;
- $\mathbf{X}_K = [X(k_1), X(k_2), \ldots, X(k_K)]^T$ are the original signal coefficients at the positions $k \in \mathbb{K}$;
- $\mathbf{X}_{K0}$ is equal to $\mathbf{X}_K$ at the positions used for reconstruction, $X_{K0}(k_p) = X_K(p) = X(k_p)$ for $k_p \in \mathbb{K}$ and $X_{K0}(k) = 0$ for $k \notin \mathbb{K}$;
- $\mathbf{A}_K$ is obtained from matrix $\mathbf{A}$ keeping the columns for $k \in \mathbb{K}$, corresponding to the estimated nonzero coefficient positions; and
- $\mathbf{X}_R$ is the vector of K reconstructed nonzero coefficients $X_R(k)$ at $k \in \mathbb{K}$.

Theorem: *The mean squared error energy in the reconstructed coefficients, with respect to K corresponding coefficients in the original signal, is*

$$\|\mathbf{X}_K - \mathbf{X}_R\|_2^2 = \frac{K}{N_A} \|\mathbf{X} - \mathbf{X}_{K0}\|_2^2 + K\frac{N^2}{N_A}\sigma_\varepsilon^2, \quad (4)$$

where $\|\mathbf{X}\|_2^2$ denotes the expected value of squared norm-two, $\|\mathbf{X}\|_2^2 = \mathrm{E}\{\sum_k |X(k)|^2\}$.

Proof: The proof will be performed in three steps. First we will consider the case when the signal is exactly K-sparse in the DFT domain, without input noise (noise-free signal). Next, the measurement noise will be added into consideration. In the final step we will relax the sparsity assumption for the original signal.

1) Based on (1) we can write the measurement relation as

$$\mathbf{y} = \mathbf{AX},$$

where the elements a_{nk} of an $N_A \times N$ measurement matrix $\mathbf{A}$ are defined as in (3).

The initial estimate of coefficients $X(k)$ is calculated using the available samples only,

$$\hat{\mathbf{X}} = N\mathbf{A}^H\mathbf{y}. \qquad (5)$$

Using (5) and (1) we get:

$$\hat{X}(k) = \sum_{n=1}^{N_A} x(t_n)e^{-j2\pi k \frac{t_n}{T}} = \sum_{i=0}^{N-1} X(i)\mu_{ik}, \qquad (6)$$

where

$$\mu_{ik} = \frac{1}{N}\sum_{n=1}^{N_A} e^{j2\pi(i-k)\frac{t_n}{T}} = N\sum_{n=1}^{N_A} a_{ni}a_{nk}^*.$$

For $i = k$ we have $\mu_{kk} = N_A/N$. For $i \neq k$ the coefficient μ_{ik} behaves as a random variable with zero-mean and variance $\mathrm{var}\{\mu_{ik}\} = N_A/N^2$.

The mean and variance of $\hat{X}(k)$ for random sampling are:

$$\mathrm{E}\{\hat{X}(k)\} = \frac{N_A}{N}X(k), \qquad (7)$$

$$\mathrm{var}\{\hat{X}(k)\} = \frac{N_A}{N^2}\sum_{i=0}^{N-1} |X(i)|^2 \left(1 - \delta(k-i)\right). \qquad (8)$$

We will assume that the positions $k_1, k_2, \ldots, k_K$ of nonzero values in a K-sparse $\mathbf{X}$ are estimated from $\hat{\mathbf{X}}$. This estimation can be done in various ways, depending on the reconstruction algorithm, for example in an iterative way using the matching pursuits (MP) algorithms (as we did in the presented example), or in one step using an appropriate threshold [10]–[12]. We also assume that the measurement matrix satisfies the CS theory conditions for the exact recovery for a given sparsity K and the available measurements.[1]

The initial estimate at the detected nonzero positions is

$$\hat{X}(k_p) = \sum_{i \in \mathbb{K}} X(i)\mu_{ik_p} \quad \text{for } p = 1, 2, \ldots, K. \qquad (9)$$

This system of linear equations can be written in matrix form:

$$\hat{\mathbf{X}}_K = \mathbf{BX}_K = \frac{N_A}{N}\mathbf{X}_K + \mathbf{C}_K,$$

where $\mathbf{B}$ is a $K \times K$ matrix with elements $b_{ij} = \mu_{k_i k_j}$, $\hat{\mathbf{X}}_K$ is the vector with K elements obtained from the initial estimate as $\hat{X}_K(i) = \hat{X}(k_i)$, and $\mathbf{X}_K$ is the vector with K corresponding coefficients from the original signal. The influence of other $K - 1$ components to the considered component is denoted by $\mathbf{C}_K$.

---

[1]The exact reconstruction conditions are commonly defined by the restricted isometry property (RIP), using the spark, or the coherence of the measurement matrix [15]–[17]. Since the presented theory holds when the exact recovery conditions are met for the assumed sparsity K, in numerical tests we followed [17] (Sec. 5) where "as a set of practical guidelines for situations where one can expect perfect recovery from partial Fourier information using convex optimization" the authors suggested, "For $K \leq N_A/8$, the recovery rate is practically 100%."

Consider a K-sparse signal. If we want to estimate the position of the strongest DFT component, then the worst case (the highest influence of other components) occurs when the remaining $K-1$ components are equally strong (assume unity amplitudes). The worst case for the detection of this component is when all other components maximally reduce the value of the considered component in the initial estimate. The influence of the kth component on the ith position is equal to μ_{ik}. Its maximal possible normalized value is the coherence index μ. In the worst case the amplitude of the considered normalized component in the initial estimate is $1 - (K - 1)\mu$. At the position where there is no component, in the worst case, the maximal possible contributions μ of all K components sum up in phase to produce the maximal possible disturbance $K\mu$. The detection of the strongest component is always successful if $1 - (K - 1)\mu > K\mu$, producing the well known coherence condition for reconstruction $K < 0.5(1 + 1/\mu)$. If the strongest component is reconstructed and removed then this relation will guarantee that remaining $K - 1$ components will be reconstructed, since the sparsity is reduced.

The reconstructed DFT 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

$$\mathbf{X}_R = (\mathbf{A}_K^H\mathbf{A}_K)^{-1}\mathbf{A}_K^H\mathbf{y}, \qquad (10)$$

where $\mathbf{A}_K$ is a matrix obtained from measurement matrix $\mathbf{A}$ by keeping the columns for $k \in \mathbb{K}$. Since $\mathbf{A}_K^H\mathbf{y} = \frac{1}{N}\hat{\mathbf{X}}_K$, according to (5), we can rewrite (10) as

$$\mathbf{X}_R = \frac{1}{N}(\mathbf{A}_K^H\mathbf{A}_K)^{-1}\hat{\mathbf{X}}_K. \qquad (11)$$

Since $\hat{\mathbf{X}}_K = \mathbf{B}\mathbf{X}_K$ the reconstruction is exact if

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

Indeed, the elements of matrix $\mathbf{A}_K^H\mathbf{A}_K$ are equal to $\beta_{k_ik_j} = \sum_{n=1}^{N_A} a_{k_in}^* a_{nk_j} = \mu_{k_ik_j}/N$ meaning that $\mathbf{A}_K^H\mathbf{A}_K = \mathbf{B}/N$. Therefore, $\mathbf{X}_R = \mathbf{X}_K$ holds.

The reconstruction algorithm produces correct coefficient values $X(k)$ at the selected positions $k \in \mathbb{K}$. It means that the influence of other $K-1$ components to each component in the initial coefficient estimate $\hat{X}(k)$, denoted by $C(k)$, is cancelled out, resulting in $X_R(k) = \frac{N}{N_A}(\hat{X}(k) - C(k)) = \frac{N}{N_A}(\frac{N_A}{N}X(k))$.

In summary, the reconstruction algorithm for a coefficient at a position $k \in \mathbb{K}$, works as an amplifier for N/N_A to the original signal component in $\hat{X}(k)$, eliminating the influence of other components at the same time.

2) Assume next that the observations are noisy

$$\mathbf{y} + \varepsilon = \mathbf{A}\mathbf{X}.$$

Variance of the assumed additive input noise ε is σ_ε^2. Noisy measurements will result in a noisy estimate $\hat{X}(k)$. Variance in $\hat{X}(k)$, caused by the measurements input noise, is $\sigma_{\hat{X}(k)}^2 = N_A\sigma_\varepsilon^2$. Since the initial estimate is multiplied by N/N_A in the reconstruction, the noise variance in the reconstructed component is

$$\mathrm{var}\{X_R(k)\} = N_A\sigma_\varepsilon^2\left(\frac{N}{N_A}\right)^2 = \frac{N^2}{N_A}\sigma_\varepsilon^2.$$

This simple result can easily be checked either statistically [10] or by comparing it with the final covariance in the Bayesian based approach [13].

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

$$\|\mathbf{X}_R - \mathbf{X}_K\|_2^2 = K\frac{N^2}{N_A}\sigma_\varepsilon{}^2. \qquad (12)$$

3) Next, we will use the assumption that the signal is approximately sparse or nonsparse and that it is reconstructed with the K-sparsity assumption. Then the remaining $N-K$ components will not be reconstructed. They will behave as additional noise in the initial estimate and in the reconstructed components. Properties of this noise are defined by (7) and (8). Each nonreconstructed component $X(k_i)$, $k_i \notin \mathbb{K}$, contributes to the noise in the initial estimate with variance $\frac{N_A}{N^2}|X(k_i)|^2$. In the reconstruction process, this variance is scaled by a factor of $(N/N_A)^2$. The final variance value is $\frac{1}{N_A}|X(k_i)|^2$.

The total energy of noise in all reconstructed components of $\mathbf{X}_R$ is K times greater than the variance in one reconstructed component. Since there are $N-K$ nonreconstructed components, the total error energy is

$$\|\mathbf{X}_K - \mathbf{X}_R\|_2^2 = K\frac{1}{N_A}\sum_{i=K+1}^{N}|X(k_i)|^2. \qquad (13)$$

The energy of nonreconstructed coefficients is:

$$\|\mathbf{X} - \mathbf{X}_{K0}\|_2^2 = \sum_{i=K+1}^{N}|X(k_i)|^2. \qquad (14)$$

From (13) and (14) the total error energy follows:

$$\|\mathbf{X}_K - \mathbf{X}_R\|_2^2 = \frac{K}{N_A}\|\mathbf{X} - \mathbf{X}_{K0}\|_2^2. \qquad (15)$$

The error caused by a sparse approximation of an approximately sparse or nonsparse signal and the error caused by additive noise in the measurements are independent. It means that the total expected error energy in the reconstructed signal coefficients is the sum of these two error energies:

$$\|\mathbf{X}_K - \mathbf{X}_R\|_2^2 = \frac{K}{N_A}\|\mathbf{X} - \mathbf{X}_{K0}\|_2^2 + K\frac{N^2}{N_A}\sigma_\varepsilon^2. \qquad (16)$$

This concludes the theorem proof. $\qquad\square$

For the special case of uniformly sampled signal the formula for the error can easily be derived from the two-dimensional case presented in [14]:

$$\|\mathbf{X}_K - \mathbf{X}_R\|_2^2 = K\frac{N - N_A}{NN_A}\|\mathbf{X} - \mathbf{X}_{K0}\|_2^2 + K\frac{N^2}{N_A}\sigma_\varepsilon^2. \qquad (17)$$

This result can be also derived following the steps in the previous proof, with the fact that the expression for variance (8) in the case of uniformly sampled signals has the terms of form $|X(i)|^2 N_A(N - N_A)/N$.

III. Example

Consider a nonsparse signal

$$x(t_n) = \frac{1}{N}\sum_{i=1}^{N}X(k_i)e^{j2\pi k_it_n/T} + \varepsilon(t_n),$$

with $N = 256$ and $N_A = 192$ available randomly positioned samples. Frequency values k_i are random $0 \leq k_i < N$. Noise $\varepsilon(t_n)$ is Gaussian additive zero-mean white noise with a standard deviation of $\sigma_\varepsilon = 0.1/N$. Amplitudes of the signal are defined in the DFT domain as $X(k_i) = 1 + \nu(i)$, for $i = 1, 2, \ldots, S$ (where $\nu(i)$ is a uniform random variable from 0 to 0.2) and $X(k_i) = \exp(-3i/(2S))$ for $i = S+1, S+2, \ldots, N$. The signal is reconstructed using a matching pursuit (MP) algorithm, assuming various sparsity levels denoted by K. Depending on the assumed sparsity K in the reconstruction we can consider this signal as:

– an approximately K-sparse signal for $K \geq S = 10$, or
– a non K-sparse signal for $K < 10$.

Assumed sparsity values K above and bellow the approximate sparsity threshold $K = S$ are considered in the
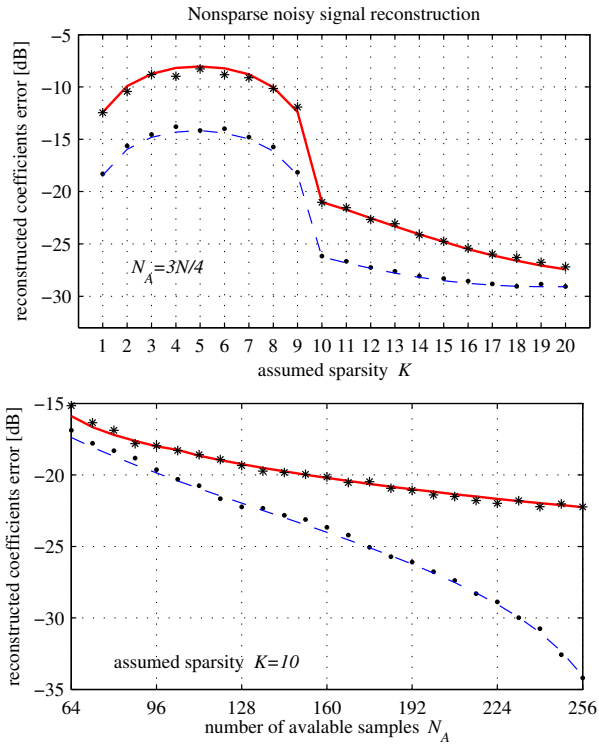
Fig. 1. Total error energy in the reconstructed coefficients of a nonsparse signal with $N = 256$, obtained statistically (100 realizations) and theoretically, with various assumed sparsities K (upper subplot) and various numbers of available samples (lower subplot) when: the signal is randomly sampled (marks '*'); the signal is uniformly sampled (marks "o"). Lines represent theoretical values, marks represent statistical ones.

numerical implementations. The assumed sparsity K in the reconstruction was varied from 1 to 20. The average error energies in 100 realizations, with random frequency values and random available samples positions, are calculated according to (4) as

$$E_{statistics} = 10 \log \left(\|\mathbf{X}_K - \mathbf{X}_R\|_2^2 \right)$$

$$E_{theory} = 10 \log \left(\frac{K}{N_A} \|\mathbf{X} - \mathbf{X}_{K0}\|_2^2 + K \frac{N^2}{N_A} \sigma_\varepsilon^2 \right).$$

In the case of uniform sampling the factor K/N_A in the first term in E_{theory} is replaced by $K(N-N_A)/(NN_A)$ as in (17). The results are given in Fig. 1. Data for random sampling are denoted by "*", while the results with a subset of uniformly sampled data are denoted by a dots. The marks represent the averaged statistical values, while the lines represent the theoretical values.

Results for both cases with assumed sparsity at the approximate sparsity threshold $K = 10$ and varying number of available samples N_A are presented in Fig.1 (lower subplot).

IV. CONCLUSION

Reconstruction of randomly sampled approximately sparse and nonsparse signals is considered. Sparsity constraint is used in the reconstruction. A formula for the expected error energy in the reconstructed coefficients is derived. From the presented results we may conclude that a random subset of uniformly

sampled data produces better results, although the CS theory favors random sampling and random partial DFT matrices.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no.4, pp.1289–1306, 2006.
[2] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.
[3] H. Rauhut, "Stability results for random sampling of sparse trigonometric polynomials," *IEEE Trans. Inf. Theory*, vol.54, no.12, pp.5661–5670, 2008.
[4] E. Sejdić, M. A. Rothfuss, M. L. Gimbel, and M. H. Mickle, "Comparative analysis of compressive sensing approaches for recovery of missing samples in implantable wireless doppler device," *IET Signal Process.*, vol. 8, no. 3, pp. 230–238, 2014.
[5] B. Jokanović, M. G. Amin, Y. D. Zhang, and F. Ahmad, "Multi-window time–frequency signature reconstruction from undersampled continuous-wave radar measurements for fall detection," *IET Radar, Sonar & Navigation*, vol. 9, no. 2, pp. 173–183, 2014.
[6] G. Hua, Y. Xiang, and G. Bi, "When compressive sensing meets data hiding," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 473–477, 2016.
[7] H. Hassanieh, P. Indyk, D. Katabi, and E. Price, "Simple and practical algorithm for sparse Fourier transform," in *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, pp. 1183–1194, 2012.
[8] K. D. Ba, P. Indyk, E. Price, and D. P. Woodruff, "Lower bounds for sparse recovery," in *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, pp. 1190–1197, 2010.
[9] L. Stanković, S. Stanković, and M. G. Amin, "Missing samples analysis in signals for applications to L-estimation and compressive sensing," *Signal Process.*, vol. 94, pp. 401–408, 2014.
[10] S. Stanković, I. Orović, and L. Stanković, "An automated signal reconstruction method based on analysis of compressive sensed signals in noisy environment," *Signal Processing*, vol. 104, pp. 43–50, 2014.
[11] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol.41, no.12, pp.3397–3415, 1993.
[12] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
[13] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, 2008.
[14] L. Stanković, I. Stanković, and M. Daković, "Nonsparsity influence on the ISAR recovery from a reduced set of data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 6, 2016.
[15] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.
[16] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ1 minimization." *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003
[17] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.