Noname manuscript No.
(will be inserted by the editor)

Reconstruction of Sparse Signals in Impulsive Disturbance Environments

Ljubiša Stanković · Miloš Daković · Stefan Vujović

the date of receipt and acceptance should be inserted later

Abstract Sparse signals corrupted by impulsive disturbances are considered. The assumption about disturbances is that they degrade the original signal sparsity. No assumption about their statistical behavior or range of values is made. In the first part of the paper it is assumed that some uncorrupted signal samples exist. A criterion for selection of corrupted signal samples is proposed. It is based on the analysis of the first step of a gradient based iterative algorithm used in the signal reconstruction. An iterative extension of the original criterion is introduced to enhance its selection property. Based on this criterion the corrupted signal samples are efficiently removed. Then the compressive sensing theory based reconstruction methods are used for signal recovery, along with an appropriately defined criterion to detect a full recovery event among different realizations. In the second part of the paper a case when all signal samples are corrupted by an impulsive disturbance is considered as well. Based on the defined criterion the most heavily corrupted samples are removed. The presented criterion and the reconstruction algorithm are applied on the signal with a Gaussian noise.

Keywords— Sparse signals, Robust signal processing, Impulsive noise, Compressive sensing, Sample selection, DFT

1 Introduction

A signal is considered to be sparse in a transformation domain if the number of nonzero coefficients is much smaller than the number of signal samples. Signals

Ljubiša Stanković · Miloš Daković · Stefan Vujović
University of Montenegro, 81000 Podgorica, Montenegro, tel: +382 67 234 430
E-mail: ljubisa@ac.me, milos@ac.me, stefanv@ac.me

having a sparse representation can be reconstructed from a reduced subset of randomly positioned samples. Processing of these signals with a large number of missing/unavailable samples attracted significant interest in the recent years within the theory of compressive sensing (CS) [2–19, 22, 30, 31, 33, 34, 36]. The number of samples required to reconstruct the signal is related to the number of nonzero coefficients in the sparse domain [4, 12, 17]. If some samples of a sparse signal are heavily corrupted by a disturbance it has been shown that it is better to omit them from the analysis or processing [31, 32].

Topic of this paper is the recovery of signals which are sparse in a transformation domain, with samples corrupted by a disturbance in the time domain, at unknown instants. Positions and the number of the nonzero coefficients of the signal transform are considered to be unknown. In contrast to the methods that combine the robust estimation and the CS based signal recovery, here we assume that we cannot distinguish the corrupted samples from the uncorrupted ones based on their values. This assumption, along with signal sparsity assumption, makes this approach different from the common robust filtering techniques [1,8,11,20,21,35,37]. Furthermore, the proposed approach can provide reconstruction with a user-defined precision, rather than to obtain its filtered approximation. The only fact that we assume about the disturbance is that it degrades the signal sparsity in the transformation domain.

A gradient-based algorithm for reconstruction of sparse signal with missing samples has been proposed in [28]. In this paper the idea of gradient calculation is reviewed in Section 2 and used as a criterion for the detection of signal samples corrupted by impulsive noise. These samples are then removed and considered as unavailable.

Three approaches to obtain noise free subsets of samples are proposed in Section 3. One is based on random selection of subsets of signal samples and detection of the event when a disturbance-free subset is selected. In the second approach a simple criterion is used in order to select disturbance-free subset of samples, while in the third approach an iterative variant of samples selection is proposed. Uniqueness of the reconstruction from the remaining subset of signal samples, after the corrupted samples are removed using one of the proposed approaches, can be considered from the theoretical point of view using the restricted isometry property or coherence index analysis. In this paper we will use a simple and computationally feasible result for the reconstruction uniqueness of signals sparse in the DFT domain, whose details are given in [27], [29].

Analysis is extended to the case when all signal samples are corrupted by a disturbance within Section 4. In this case a subset of disturbance-free samples does not exist, but we can select samples with small disturbances and reconstruct heavily dis-

turbed samples. It has been shown that the recovery process results in a significant disturbance rejection and the signal to noise improvement.

2 Definitions and Reconstruction Algorithm

Consider a signal $x(n)$ with N samples in the discrete-time domain. Assume that the sparsity domain of the signal is the discrete Fourier transform (DFT) domain. The signal and the DFT coefficients are related via

$$\mathrm{DFT}[x(n)] = X(k) = \sum_{n=0}^{N-1} x(n)\varphi_k^*(n), \tag{1}$$

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)\varphi_k(n) \tag{2}$$

or $\mathbf{X} = \mathbf{W}\mathbf{x}$ and $\mathbf{x} = \mathbf{W}^{-1}\mathbf{X}$, where for the DFT holds $\varphi_k(n) = \exp(j2\pi nk/N)$. In a matrix notation the N-dimensional vector $\mathbf{X}$ has elements $X(k)$, vector $\mathbf{x}$ has elements $x(n)$, and the $N \times N$ matrix $\mathbf{W}$ is with elements $\varphi_k^*(n)$. The time presentation of a signal which is sparse in this transformation domain is

$$x(n) = \sum_{i=1}^{s} A_i \varphi_{k(i)}(n)., \tag{3}$$

where A_i are nonzero coefficients and $k(i)$ are their positions in the transformation domain for $i = 1,2,\ldots,s$. The sparsity of this signal is $s \ll N$.

Assume that there are M available/uncorrupted samples at the instants $n_i \in \mathbb{N}_{xc}$, $i = 1,2,...,M$ ($n_1 < n_2 < \cdots < n_M$). The vector of available signal samples, whose length is M, is denoted by $\mathbf{y}$. Its elements are

$$y(i) = x(n_i), \;\; i = 1,2,...,M.$$

The task is to eliminate the corrupted samples $x(n_i), \;\; i = M+1, M+2, ..., N$ and to reconstruct the signal so that the number of nonzero transform coefficients $X(k)$ is minimal, subject to the available/uncorrupted sample values. The task can be defined as to find the unavailable/corrupted signal samples from

$$\min \|\mathbf{X}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{X}, \tag{4}$$

where $\mathbf{y} = \mathbf{A}\mathbf{X}$ is a matrix notation for the system of M linear equations

$$x(n_i) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)\varphi_k(n_i) \tag{5}$$

for $i = 1, 2, ..., M$. The matrix $\mathbf{A}$ is obtained from matrix $\mathbf{W}^{-1}$ by eliminating the rows corresponding to the unavailable/corrupted samples.

Simple counting of the nonzero values of $X(k)$ is achieved using so called l_0-norm $\|\mathbf{X}\|_0$. However, the l_0-norm based solution is an NP-hard combinatorial optimization problem. Its calculation complexity is of order $\binom{N}{s}$. In theory, the NP-hard problems can be solved by an exhaustive search. However, as the problem parameters N and s increase the computational time increases and the problem becomes unsolvable in realistic time. Moreover, in practical signal processing applications, the l_0-norm can not be used even for small values of parameters N and s, when the combinatorial approach could be used within a reasonable computational time. In all real-life signals at least A/D quantization noise (very small but not zero) exists. Even if all values of an originally sparse signal in the DFT domain ($s \ll N$) are known, the l_0-norm of $X(k)$ for a signal whose samples are stored into a computer with a finite precision will be equal to N.

These are the reasons why commonly the l_1-norm of the signal transform is used. It is convex and more robust to the noise. Minimization problem is then

$$\min \sum_{k=0}^{N-1} |X(k)| \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{X}. \tag{6}$$

It has been shown that if the signal and its transform satisfy the restricted isometry property, with appropriate isometry constants, then the solutions of (4) and (6) are the same [5], [29].

For a reconstruction of unavailable/corrupted samples in the time domain we will use a very simple and efficient algorithm, based on the gradient of a sparsity measure [25, 28, 29]. This algorithm is inspired by the adaptive signal processing methods with an adaptive step size. It is based on the gradient descent, where the missing samples are considered as variables and reconstructed as the values producing minimal sparsity measure of the signal transform in the sparsity domain. Based on the properties of this algorithm, a criterion for selection of corrupted samples will be introduced. The reconstruction algorithm belongs to the wide class of gradient based CS algorithms [15, 23]. Its pseudocode is given in Algorithm 1.

Comments on Algorithm 1: The Algorithm 1 is iterative with m as an iteration index. The initial iteration for the reconstructed signal is

$$x_r^{(0)}(n) = \begin{cases} x(n) & \text{for available samples, } n \in \mathbb{N}_{xc} \\ 0 & \text{for unavailable/corrupted samples, } n \in \mathbb{N}_x \end{cases},$$

where set $\mathbb{N}_x$ contains the positions of unavailable/corrupted samples. It is the complement of $\mathbb{N}_{xc}$.

Algorithm 1 Reconstruction procedure GRADREC

Require:

- Set of unavailable/corrupted sample positions $\mathbb{N}_x$
- Available samples $x(n)$

function GRADREC($x(n)$,$\mathbb{N}_x$)
  Set $x_r^{(0)}(n) \leftarrow x(n)$ ▷ for $n \notin \mathbb{N}_x$
  Set $x_r^{(0)}(n) \leftarrow 0$ ▷ for $n \in \mathbb{N}_x$
  Set $m \leftarrow 0$
  Set $\Delta \leftarrow \max |x_r^{(0)}(n)|$
  repeat
    repeat
      $x_r^{(m+1)}(n) \leftarrow x_r^{(m)}(n)$ ▷ for each n
      for $n_i \in \mathbb{N}_x$ **do**
        $X_1(k) \leftarrow \mathrm{DFT}\{x_r^{(m)}(n) + \Delta\,\delta(n-n_i)\}$
        $X_2(k) \leftarrow \mathrm{DFT}\{x_r^{(m)}(n) - \Delta\,\delta(n-n_i)\}$
        $g(n_i) \leftarrow \frac{1}{N}\sum_{k=0}^{N-1} |X_1(k)| - |X_2(k)|$
        $x_r^{(m+1)}(n_i) \leftarrow x_r^{(m)}(n_i) - g(n_i)$
      end for
      $m \leftarrow m+1$
    until stopping criterion is satisfied
    $\Delta \leftarrow \Delta/3$
  until required precision is achieved
  return $x_r^{(m)}(n)$
end function

Output:

- Reconstructed signal $x_R(n) = x_r^{(m)}(n)$

In the next iterations, for each missing sample at $n_i \in \mathbb{N}_x$ we consider two signals $x_1(n)$ and $x_2(n)$ which are defined as

$$x_1(n) = \begin{cases} x_r^{(m)}(n) + \Delta & \text{for } n = n_i \\ x_r^{(m)}(n) & \text{for } n \neq n_i \end{cases}$$

$$x_2(n) = \begin{cases} x_r^{(m)}(n) - \Delta & \text{for } n = n_i \\ x_r^{(m)}(n) & \text{for } n \neq n_i \end{cases}.$$

The parameter Δ is used to determine whether the considered signal sample $x_r^{(m)}(n_i)$ should be decreased or increased. Then we estimate a correction for the considered sample based on the finite difference of the sparsity measure in the transformation domain as

$$g(n_i) = \frac{1}{N}\sum_{k=0}^{N-1} \big[\, |X_1(k)| - |X_2(k)| \,\big], \tag{7}$$

where $X_1(k)$ and $X_2(k)$ are the transformations of $x_1(n)$ and $x_2(n)$ for a given n_i.

We should stop the inner repeat loop in Algorithm 1 (lines 7–16) when we approach the stationary error zone, i.e., when the reconstruction result does not change any more for the considered Δ, [29].

Outer loop (lines 6–18) repeats the reconstruction with a smaller Δ, for example $\Delta \leftarrow \Delta/3$. The stopping criterion for this loop can be based on the minimal value of Δ or on the sparsity measure of the reconstructed signal. The rate of algorithm convergence is considered in detail in [28, 29].

Convergence analysis in a gradient-based algorithm is related to its behavior for large values of step Δ. Small step value influence only the rate of the algorithm convergence, [28, 29]. In order to examine the algorithm behavior for a large Δ note that

$$\begin{aligned} &|X_1(k)| - |X_2(k)| \\ &= \left|X_r^{(m)}(k) + \Delta D_{n_i}(k)\right| - \left|X_r^{(m)}(k) - \Delta D_{n_i}(k)\right| \\ &= \Delta\,|D_{n_i}(k)| \left(\left|1 + \frac{X_r^{(m)}(k)}{\Delta D_{n_i}(k)}\right| - \left|1 - \frac{X_r^{(m)}(k)}{\Delta D_{n_i}(k)}\right| \right). \end{aligned}$$

For any complex-valued $a = X_r^{(m)}(k)/(\Delta D_{n_i}(k))$, when $|a| \ll 1$ for a large value of Δ, we can easily derive the bounds $0 \leq ||1+a| - |1-a|| \leq 2\,|a|$. It means

$$0 \leq \left| \,|X_1(k)| - |X_2(k)|\,\right| \leq 2\left|X_r^{(m)}(k)\right|.$$

This is an important conclusion stating that the difference $|X_1(k)| - |X_2(k)|$ does not depend on the algorithm step Δ, for a large Δ. This difference is used in the adaptation of the missing sample values in the algorithm. It means that the adaptation of variables will be done for a value that does not depend on Δ if Δ is large. In the realization it results in the missing sample values oscillations around the original signal values (being of order $\left|X_r^{(m)}(k)\right|/N$) until the step Δ is reduced in the next iterations. After the value of Δ is reduced within few iterations the missing samples will start convergence toward the positions of the sparsity measure minimum [26], [29].

Efficiency of the presented algorithm will be tested on a sparse signal with various sparsity coefficients, sets of available samples, and random signal parameters.

Example 1: Consider a signal

$$x(n) = \sum_{i=1}^{K} A_i \cos(2\pi k_i n/N + \phi_i) \tag{8}$$

with $N = 128$. The sparsity of this signal is $s = 2K$. It has been varied from $s = 2$ to $s = N/2$ with step 2 (i.e., values $s = 2, 4, \ldots, N/2$ are used). The amplitudes, frequencies,

and phases were taken randomly within $1 \leq A_i \leq 2$, $1 \leq k_i \leq 63$ and $0 \leq \phi_i \leq 2\pi$, for each realization. The reconstruction is performed by using 100 realizations for each number of missing samples $Q = N - M$ from 4 up to 124 (with step 4) and the reconstructed signals $x_R(n)$ are obtained. The results are presented in Fig.1 in a form of the signal-to-reconstruction-error ratio (SRR) in [dB]

$$SRR = 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\sum_{n=0}^{N-1}|x(n)-x_R(n)|^2}. \tag{9}$$

Bright colors indicate the region where the algorithm had fully recovered missing samples, while the dark colors indicate the region where the algorithm could not recover missing samples. The minimal possible number of available samples for the reconstruction can roughly be estimated as $M \geq 2s$ [12, 19]. In the transition region for M slightly greater than $2s$ we have cases when the signal recovery is not achieved as well as cases of full signal recovery. A practical way to make a decision on the recovery of a sparse signal in this region is in considering the values of the sparsity measure close to the l_0-norm. For the case when the recovery is not achieved the sparsity measure value will be high (in theory equal to N). For the event when the recovery of a sparse signal is achieved its value is close to the sparsity $s \ll N$. A theorem to check if the recovered sparse signal is unique will be presented and discussed later.

In this example, a stopping criterion is set to the precision of about 120 [dB]. This level of precision corresponds to an input quantization noise in the signal obtained from a high precision 20 bit A/D converter. By increasing the number of iterations, with a sufficient number of available samples (Fig.1), the value of SRR could be of the computer precision order, [28].

3 Disturbance in some Signal Samples

Consider a signal $x(n)$, $0 \leq n \leq N-1$ that is sparse in a transformation domain with a sparsity $s \ll N$. Assume that I samples of the signal $x(n)$, at unknown positions $n \in \mathbb{N}_I$, are corrupted with a signal $\varepsilon(n)$. The disturbing signal $\varepsilon(n)$ is modeled as: $\varepsilon(n) = 0$ for $n \notin \mathbb{N}_I$ and $\varepsilon(n)$ assumes arbitrary values for $n \in \mathbb{N}_I$. The original signal can be fully recovered if a sufficient number of uncorrupted samples exists.

The sufficient number of uncorrupted samples is directly related to the full recovery conditions studied in the CS theory [4, 12, 19]. A rough estimation of this number can be made based on the statistical results presented in Fig. 1, while a precise uniqueness check is provided in Section III.D.

The three methods for signal recovery from such corrupted samples are described below.

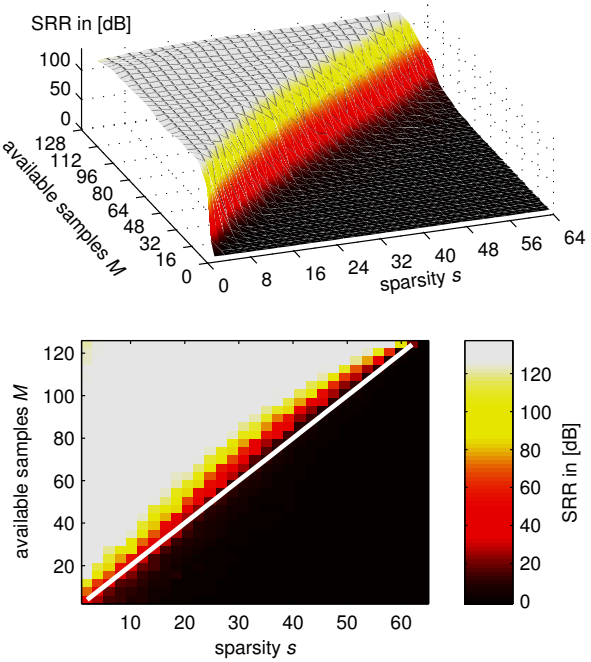

Fig. 1 Signal-to-reconstruction-error (SRR) obtained by using Algorithm 1, averaged over 100 realizations for various sparsity s and number of available samples M.

3.1 Direct Search Procedure

In order to solve the stated problem we will begin with a simple idea of eliminating random subsets of samples and performing the reconstruction using the remaining samples. For each realization the sparsity measure of the recovered signal is used for detection of the full recovery event. By using a sparsity measure close to the l_0-norm all realizations containing disturbed samples will produce a value of sparsity measure close to the total number of samples N. In the case when only the uncorrupted samples are used in the reconstruction, the measure value is of order s. It is much lower than the total number of samples N. Thus by setting a threshold T_μ within $s < T_\mu < N$ we can detect a full recovery event.

We have assumed that the signal length is N and the number of arbitrary corrupted samples is up to I. For each realization the observation set of the signal positions is denoted as $n \in \mathbb{N}_{xc}$, with $M < N$ randomly positioned signal samples. Since the considered signal is sparse in the considered transformation domain it can be recovered based on $2s \leq M \leq N$ uncorrupted samples if the recovery conditions are meet. In order to estimate the computational complexity of this method, we will find the prob-

ability that within M randomly selected observation samples there are no corrupted samples.

Probability that the first randomly chosen sample is not affected by disturbance is $(N-I)/N$ since there are N samples in total and $N-I$ of them are disturbance-free. Then we continue the process of random samples selection. The probability that both the first and second chosen samples are not affected by disturbance is $\frac{N-I}{N}\frac{N-I-1}{N-1}$. In this way we can calculate the probability that all of M randomly chosen samples at the positions $n \in \mathbb{N}_{xc}$ are not affected by a disturbance. This probability is

$$P(M,N) = \prod_{i=0}^{M-1} \frac{N-I-i}{N-i}. \tag{10}$$

Since $\frac{N-I-i}{N-i} < 1$ we can see that the probability $P(M,N)$ decreases as the number of terms in the product increases. Thus, in this kind of reconstruction it is important to keep the number of samples M in the observation set as low as possible, while satisfying the reconstruction conditions.

In general, for an expected number of pulses I, the expected number of random realizations to achieve at least one disturbance-free reconstruction using a subset of M samples is $1/P(M,N)$.

Example 2: Consider a signal $x(n)$ from the previous example corrupted by an impulsive disturbance $\varepsilon(n)$. Impulse disturbance is expected in about 12% of the signal values, corresponding to $I = 15$ corrupted signal samples, Fig.2(a). A signal $x(n)$ of sparsity $s = 6$ is used in this example. Thus, by using $M = 32$ samples in the reconstruction we are well within the full recovery region, with a high probability (Fig.1). The probability that a randomly chosen subset of M out of N samples will be disturbance free is $P(32,128) = 0.0099$. It means that in 1000 realizations, we can expect about 10 full recovery realizations. The calculation is performed with an impulsive disturbance of the form $\varepsilon(n) = \varepsilon_1(n)/\varepsilon_2(n) + \varepsilon_3(n)/\varepsilon_4(n) + 10\varepsilon_5(n)$ where $\varepsilon_i(n)$, $i = 1,2,3,4,5$, for $n \in \mathbb{N}_I$, are the unite variance Gaussian noises. It is important to note that the results do not depend on the disturbance amplitude values or their distribution. The l_1-norm is used as sparsity measure in the reconstruction process.

As we can see from Fig.2(b) there are some realizations of sparse signal recovery corresponding to the algorithm precision value of SRR. Now we have to define an appropriate procedure for their detection. Since the assumption is that the signal is sparse in the considered transformation domain, the measure of reconstructed signal sparsity is an obvious criterion for sparse signal recovery detection. Here, the task is in a posteriori measuring of the signal sparsity. In this case the measures based on the norms close to the l_0-norm, will give much clearer threshold for the detection

of a sparse recovery event the l_1-norm. What is a weak point of the l_0-norm in the optimization process, it may be an advantage in the detection. The l_0-norm can clearly distinguish the cases of the full sparse signal recovery from nonrecovery cases. It will produce N as the measure of sparsity value for every signal reconstructed based on a subset containing some corrupted samples, while its value will be $s \ll N$ for the full recovery event, when the samples without disturbance are used.

In practice, a measure calculated using an l_p-moment, with a value of p close to 0 should be used. In the calculation with a finite precision, a sparse recovery will produce very small (but nonzero) transformation coefficients values $X(k)$ at the positions where they should be zero. Value of p should be such that $|X(k)|^p$ at these positions is much lower than the value of $|X(k)|^p$ at the original nonzero signal positions. Thus, instead of the l_0-norm, in the detection we will use sparsity measure

$$\mu\{X(k)\} = \sum_{k=0}^{N-1} \left| \frac{1}{N} X(k) \right|^p . \tag{11}$$

With a small p the behavior of this sparsity measure is close to the l_0-norm, but robustness to the small errors in the reconstructed signal is achieved, Fig.2(d). If the input data or the reconstruction precision is of order 10^{-c} with the signal values of order 10^0 then the value p should be such that $10^{-pc} \ll 10^0$. For example, if the reconstruction algorithm works with a precision of 10^{-6}, then the sparsity measure for the case of full recovery will be of order s if p is such that $pc > 1$. Thus, the value $p = 1/4$ will keep the original zero-valued transform coefficients $X(k)$ at a level of an order bellow the nonzero coefficient values, in the full recovery case. Quite similar conclusion can be made even if we assume $c = 16$. The sparsity measure values for the cases when a nonsparse signal is recovered will be much higher than s. Therefore, a low value of sparsity measure corresponds to a high SRR, Fig.2(b). In Fig.2(c), the realization with the smallest sparsity measure is used to reconstruct the signal $x(n)$. The reconstructed signal is $x_R(n)$. In reality we need only one full recovery realization. Calculation efficiency can be improved if we set a threshold for the sparsity measure $\mu\{X(k)\}$, and stop further random selections of sample subsets when the sparsity measure threshold T_μ is reached. For example, in this case a threshold level of $\mu\{X(k)\} < T_\mu = 20$ would stop search after just a few realizations.

Note that in simulations the reconstructed signal is compared with the original signal for all instants n. In theory, there is a probability that the available samples do not guarantee a unique solution for a given signal sparsity. In such cases, other solutions may exist. The uniqueness check of the solution will be discussed later.

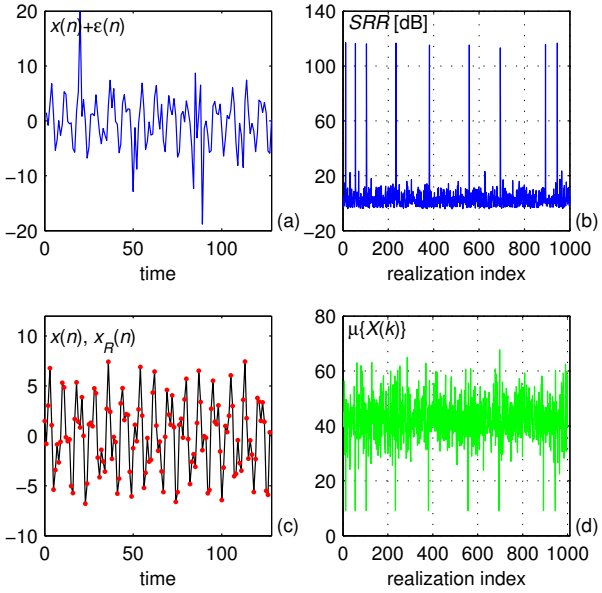

Fig. 2 Reconstruction of a signal with $I = 15$ out of $N = 128$ samples being affected by an impulsive disturbance. In each realization 96 randomly chosen samples are removed. Total number of realizations is 1000. a) The available corrupted signal; b) The SRR for each of 1000 realizations; c) The original (line) and the reconstructed (dots) signal for the best realization; d) The sparsity measure for each of 1000 realizations.

3.2 Procedure with a Criterion for Selecting Preferred Samples

The direct search procedure can be used on signal with a small number of corrupted samples. An average number of the random realizations required to have an uncorrupted subset of signal samples increases with the number of corrupted samples. In the case when the disturbance is much stronger than the signal, then the trimmed L-statistics can be used to eliminate the corrupted signal samples, without any search procedure. Then, a direct application of the recovery algorithm on the remaining samples is possible, [32] (or in image processing [24]). This kind of approach can be used, in general, when a knowledge about the disturbance behavior can help us to detect the positions of the corrupted samples [1, 8, 11, 20, 21, 37].

The L-statistics method can also be used in combination with the presented direct search method when some of the corrupted samples may be eliminated based on their values, while a small number of corrupted samples at unknown positions remains. The extension is straightforward.

With this kind of approaches we have to know the signal range or behavior in order to eliminate heavily corrupted signal samples. In general, some of the highly corrupted and most of the moderately and low corrupted signal samples of unknown behavior cannot be easily distinguished from the uncorrupted signal values.

Our next analysis will be focused on a criterion that will mark some signal samples as probably more corrupted than the others, without assuming any particular distribution or the number of corrupted samples. Then these signal samples will be marked as unavailable and reconstructed in the process of signal recovery.

Some insight into the basic idea of favorable samples selection in the recovery process can be obtained from a simplified analysis. If just one sample is considered as corrupted at $n = n_1$ then we can form the signals $x_1(n) = x(n) + (\varepsilon(n) + \Delta)\delta(n - n_1)$ and $x_2(n) = x(n) + (\varepsilon(n) - \Delta)\delta(n - n_1)$. The pulses $(\varepsilon(n) \pm \Delta)\delta(n - n_1)$ will spread over all coefficients in the transformation domain $X_1(k) = X(k) + (\varepsilon(n_1) + \Delta)e^{j2\pi n_1 k/N}$ and $X_2(k) = X(k) + (\varepsilon(n_1) - \Delta)e^{j2\pi n_1 k/N}$. The changes at the positions of the nonzero coefficients $X(k)$ are of the same order as the changes at the zero coefficients $X(k)$. Since the signal is sparse, the number of nonzero coefficients is much smaller than the total number of coefficients, $s \ll N$. Therefore, in the approximative analysis, we may neglect the total value of changes in the nonzero coefficients $X(k)$. Then, we may approximately write $\mu_1 = \sum_{k=0}^{N-1} |X_1(k)| \cong \mu_0 + |\varepsilon(n_1) + \Delta| N$ and $\mu_2 = \sum_{k=0}^{N-1} |X_2(k)| \cong \mu_0 + |\varepsilon(n_1) - \Delta| N$, where μ_0 is the sparsity measure of the uncorrupted signal $x(n)$ and $N - s \cong N$. Therefore, the difference of measures, normalized with N, is (7)

$$g(n_1) \cong |\varepsilon(n_1) + \Delta| - |\varepsilon(n_1) - \Delta|.$$

For $\Delta > |\varepsilon(n_1)|$ we get

$$g(n_1) = 2\varepsilon(n_1). \tag{12}$$

It means that the finite difference value, for a large Δ, can be used as an indicator of the signal value deviation from its correct value, i.e., an indicator of the disturbance $\varepsilon(n_1)$ intensity.

Of course, this is not an exact criterion about the significance of a disturbance in a signal with many corrupted samples. For a large number of corrupted samples, the signal deviation significance follows this criterion in the most cases, as it will be presented and statistically checked in the rest of the paper.

The idea for this criterion is illustrated on a two-dimensional case, Fig.3, for two corrupted samples $x(n_1) + \varepsilon(n_1)$ and $x(n_2) + \varepsilon(n_2)$. The first signal sample $x(n_1)$ possible values are presented on the horizontal axis, while the second sample $x(n_2)$ possible values are on the vertical axis. The true signal values are presented by horizontal and vertical lines (passing through the measure minimum). The cross indicates

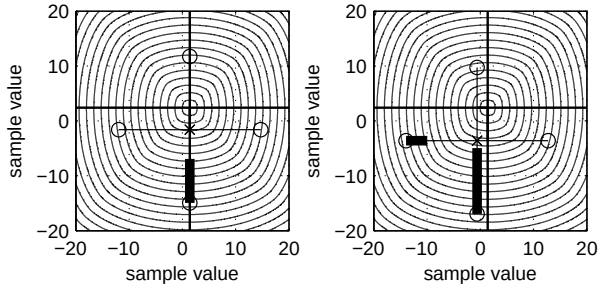

Fig. 3 Sparsity measure as a function of the signal values in the case of two corrupted samples (horizontal and vertical lines represent true values). Crosses indicate the position on the measure function for the corrupted samples $x(n_1)$ and $x(n_2)$. Difference of measures (thick lines) calculated at $x(n_1)+\varepsilon(n_1)+\Delta$ and $x(n_1)+\varepsilon(n_1)-\Delta$ or $x(n_2)+\varepsilon(n_2)+\Delta$ and $x(n_2)+\varepsilon(n_2)-\Delta$ (positions are indicated on the contour plot by circles) are proportional to the disturbance (displacement of the crosses from the horizontal and vertical lines representing the true signal values). The cases with one corrupted sample (left) and with both corrupted samples (right) are shown.

position of the measure function for corrupted samples $x(n_1)$ and $x(n_2)$. Horizontal and vertical displacement of the cross from intersection of lines representing the true signal values are equal to the disturbances $\varepsilon(n_1)$ and $\varepsilon(n_2)$ respectively. On the left side of Fig.3 a case with $\varepsilon(n_1)=0$ is shown, while on the right side of Fig.3 a case with $|\varepsilon(n_1)|<|\varepsilon(n_2)|$ is shown. Measures are calculated at $x(n_1)+\varepsilon(n_1)+\Delta$ and $x(n_1)+\varepsilon(n_1)-\Delta$ and then at $x(n_2)+\varepsilon(n_2)+\Delta$ and $x(n_2)+\varepsilon(n_2)-\Delta$. Their values are indicated by circles on the contour plot in all cases. The measure contour plot is calculated for all possible values of these two samples for a given signal. Differences $g(n_i)$ of the corresponding measures are presented by thick lines. Obviously the differences of the corresponding measures are proportional to the deviations $\varepsilon(n_1)$ and $\varepsilon(n_2)$ of the signal samples from their true values, (12).

A procedure of choosing samples which are probably less corrupted than the others, according to the presented criterion, is summarized into Algorithm 2.

Comments on Algorithm 2: A corrupted signal $x(n)$ is assumed. For each instant $m=0,\ldots,N-1$, the signals $x_1(n)=x(n)+\Delta\delta(n-m)$ and $x_2(n)=x(n)-\Delta\delta(n-m)$ are formed. The difference of measure values is calculated as

$$g(m)=\frac{1}{N}\left(\sum_{k=0}^{N-1}|X_1(k)|-\sum_{k=0}^{N-1}|X_2(k)|\right), \tag{13}$$

where $X_1(k)=\mathrm{DFT}[x_1(n)]$ and $X_2(k)=\mathrm{DFT}[x_2(n)]$. The signal values with the largest $|g(m)|$ are eliminated and considered as unavailable in the reconstruction. Parameter Δ, used in algorithm, should be higher than the disturbance magnitude. We can, for example, put a value of an order of $\Delta=\max|x(n)|$.

Algorithm 2 Disturbance-free Samples Selection Procedure

Require:

- Signal $x(n)$
- Parameter Δ
- Number of required samples M

1: **for** $m \leftarrow 0$ **to** $N-1$ **do**
2: $X_1(k) \leftarrow \text{DFT}\{x(n)+\Delta\delta(n-m)\}$
3: $X_2(k) \leftarrow \text{DFT}\{x(n)-\Delta\delta(n-m)\}$
4: $g(m) \leftarrow \dfrac{1}{N}\displaystyle\sum_{k=0}^{N-1} |X_1(k)| - |X_2(k)|$
5: **end for**
6: Sort $|g(m)|$ in nonincreasing order

$$|g(m_1)| \leq |g(m_2)| \leq \cdots \leq |g(m_N)|$$

Output:

- Positions of selected samples $m_1, m_2, \ldots, m_M$.

Example 3: Consider $N = 128$ samples of signal (8) with sparsity $s = 6$, and $A_1 = 4$, $A_2 = 3$, $A_3 = 0.7$, where 64 of them (50%) are affected by an impulsive disturbance of the form $40\varepsilon(n)$, where $\varepsilon(n)$ is a Gaussian unit variance noise. The term impulsive is used here for the disturbance existing in some signal samples only. The amplitude of this disturbance is of signal amplitude order and many corrupted signal samples are within the signal amplitude range. For example, a threshold that would eliminate disturbance whose amplitudes are certainly outside the signal range $2A = 15.4$ (if the amplitude were known), would leave 30% of the corrupted samples. This number of corrupted samples would require a large number of reconstruction realizations in direct search.

The selection procedure described by Algorithm 2 is applied to this corrupted signal. Just in order to illustrate the trivial case if the algorithm is applied on the disturbance only case (without signal), we present results in Fig.4(left). In this case, when there is no signal, the criterion works as an exact sorting procedure. It means that Algorithm 2 applied on the described disturbance will perform exact sorting of its values. However, in all cases of interests the signal is also present. The result of procedure described by Algorithm 2 applied on corrupted signal is presented in Fig.4(right). Disturbance amplitudes in samples selected and sorted by the algorithm are presented. Assume that we mark the first 64 samples as corrupted ones. A few corrupted samples in the remaining 64 samples (sorting indexes 65-128), Fig.4(right) are still left. This is the consequence of the fact that the criterion is not able to locate all corrupted samples. In this example, we will continue with a direct search proce-

dure. We will use 32 randomly positioned samples in the reconstruction, from the set of 64 remaining samples. The results are presented in Fig.5. A disturbance-free realization is detected based on the sparsity measure of the reconstructed signal. If we set a threshold for measure at $\mu\{X(k)\} < T_\mu = 20$ (approximately $s < 20$ complex signal components), we would achieve sparse signal recovery within a small number of realizations. In Fig.5(d) the sparse recovery event was detected in the 58th realization. Further direct search is then stopped. The recovered result was with the reconstruction Algorithm 1 precision of about $SRR = 120$ [dB], Fig.5(b).

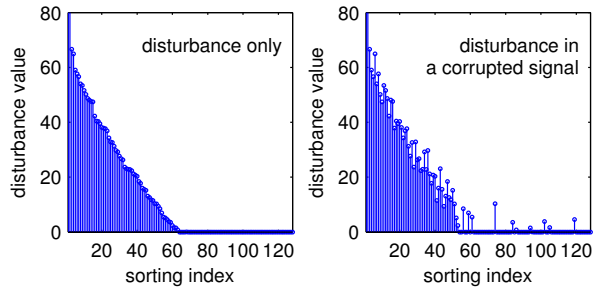

Fig. 4 Disturbance values in the signal sorted according to the introduced significance criterion. Criterion applied to the disturbance only (left). Criterion applied to the signal corrupted by the same disturbance (right).

3.3 Iterative Procedure for Remaining Noise

Further improvement can be achieved if the direct search is replaced with a "smart" iterative procedure for choosing samples to be removed and reconstructed in each iteration. It is presented in Algorithm 3.

Comments on Algorithm 3:

The criterion from Algorithm 2 could be applied to remove an initial set $\mathbb{N}_x$ of the most heavily corrupted samples. We can also start the selection with the criterion defined by Algorithm 3 only, when the initial set is $\mathbb{N}_x = \emptyset$.

Then we perform the reconstruction under the assumption that samples at the positions from the set $\mathbb{N}_{xc}$ are available. The reconstructed signal is denoted by $x_R(n)$. For each signal sample from the set of selected samples $\mathbb{N}_{xc}$ the reconstruction is performed, assuming that only the considered sample is unavailable. The sparsity measure difference is calculated before and after the reconstruction. It is denoted by $g(m)$. This difference is used for the selection of samples which are probably the

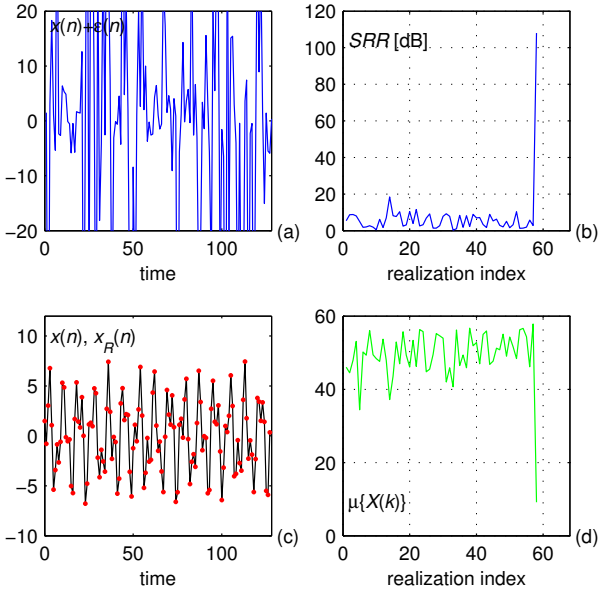

Fig. 5 Reconstruction of a corrupted signal with 64 out of $N = 128$ samples being removed using the criterion. The signal to reconstruction error ratio (SRR) corresponds to the concentration measure values for each realization. a) Input noisy signal; b) The SRR as a function of iteration index; c) The original (line) and the reconstructed (dots) signal for the last realization; d) The concentration measure for each realizations.

most corrupt. A highly corrupted sample will give a high $|g(m)|$, while a sample with a low or no corruption will produce small $|g(m)|$. In the next step r samples with the highest measure difference $g(m)$ are selected, marked as the corrupted samples, and added to the set of unavailable sample positions $\mathbb{N}_x$.

The presented procedure is repeated with the new sets $\mathbb{N}_{xc}$ and $\mathbb{N}_x$. At the end of N_{it} iterations, in total rN_{it} most corrupted samples will be removed.

Example 4: Consider $N = 128$ samples of a signal defined by (8), for two different sparsity values $s = 6$ and $s = 10$. The number of samples affected by a disturbance is $I = 64$. The disturbance of form $\varepsilon(n) = 40(\varepsilon_1(n) - 0.5) + 40(\varepsilon_2(n) - 0.5)$ is used, where $\varepsilon_1(n)$ and $\varepsilon_2(n)$ are white uniform noises. This kind of disturbance has a large number of values within the signal amplitude range at the positions where the disturbance exits. In the reconstruction, the iterative removal procedure is used with $r = 4$ sample positions being added to the set of unavailable/corrupted sample positions $\mathbb{N}_x$ in each iteration. The iterative procedure detects the remaining most corrupted samples in each next iteration. In an ideal case if there is no miss-detection, all corrupted samples from this example will be removed in 16 iterations. Of course, it can happen

Algorithm 3 Iterative Denoising

Require:

- Noisy signal $x_\varepsilon(n) = x(n) + \varepsilon(n)$, $n = 0, 1, \ldots, N-1$. It is known that nonnoisy signal $x(n)$ is sparse in the DFT domain with unknown sparsity.
- Number of samples r to be removed in each iteration.
- Reconstruction procedure GRADREC defined by Algorithm 1.

Set $\mathbb{N}_{xc} = \{0, 1, \ldots, N-1\}$
Set $\mathbb{N}_x = \emptyset$
repeat
  $x_R(n) \leftarrow$ GRADREC($x_\varepsilon(n)$,$\mathbb{N}_x$)
  $X_R(k) \leftarrow$DFT$\{x_R(n)\}$
  for $m \in \mathbb{N}_{xc}$ **do**
    $x_{R1}(n) \leftarrow$ GRADREC($x_R(n)$,$\{m\}$)
    $X_{R1}(k) \leftarrow$ DFT$\{x_{R1}(n)\}$
    $g(m) \leftarrow \frac{1}{N}(\sum_{k=0}^{N-1} |X_R(k)| - \sum_{k=0}^{N-1} |X_{R1}(k)|)$
  end for
  Select r samples with highest $|g(m)|$
  Remove selected sample positions from the set $\mathbb{N}_{xc}$
  Add selected sample positions to the set $\mathbb{N}_x$
until stopping criterion is satisfied
return $x_R(n)$

Output:

- Reconstructed signal $x_R(n)$.

that an uncorrupted signal value is marked as a corrupted one in an iteration. Results of reconstruction are presented in Fig. 6. We can see that the number of omitted samples for the considered signals is $Q = 68$, meaning that just a few miss-detections existed. Disturbance values, in the order as they are detected and omitted by the algorithm, are presented in Fig. 7. It can be seen that the algorithm followed quite well the significance order of the disturbance in their omission. The same process is repeated 100 times with arbitrary signal amplitudes and frequencies. In the case of signal with sparsity $s = 6$, in all realizations the corrupted samples where among 68 samples selected by the criterion. A sparse signal recovery was achieved in all cases.

3.4 Uniqueness of the Obtained Solution

After a sparse signal is reconstructed form a reduced set of samples its uniqueness should be confirmed. The restricted isometry property is used as a tool to define the uniqueness of a sparse signal reconstruction. With appropriate isometry constant for l_1-norm based minimization it guaranties that the solution is the same as if l_0-

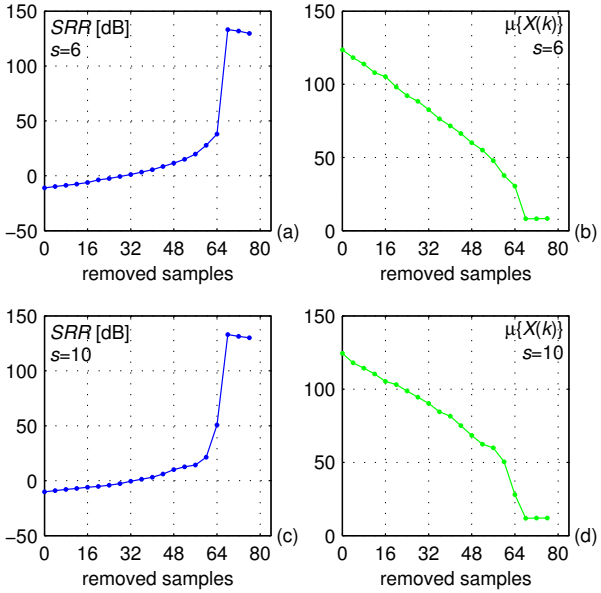

Fig. 6 Reconstruction of a sparse signal when corrupted samples are removed by using the criterion in iterative way. In each iteration $r = 4$ samples are removed. a) The SRR during the iterations for a signal of sparsity $s = 6$; b) The sparsity measure during the iterations for a signal of sparsity $s = 6$; d) The SRR during the iterations for a signal of sparsity $s = 10$; b) The sparsity measure during the iterations for a signal of sparsity $s = 10$;.

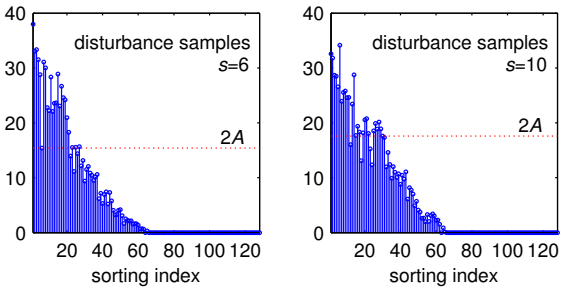

Fig. 7 Disturbance values in the signal, sorted according to the introduced significance criterion, with signal range in amplitude $2A$. Sorting of the disturbances in corrupted signals of sparsity $s = 6$ and $s = 10$ are presented.

norm were used and that the solution is unique. However, the check of the restricted isometry property requires a combinatorial approach, which is an NP hard problem. In addition, for a specific measurement matrix it produces quite conservative bounds for nonuniqueness.

A theorem for the solution uniqueness [27] will be presented and used here. The gradient-based algorithm considers omitted samples as variables

$$x_a(n) = x(n) + z(n).$$

For the available sample positions $z(n) \equiv 0$, $n \in \mathbb{N}_x$. This variable takes arbitrary values at the missing sample positions $n = q_m \in \mathbb{N}_x = \{q_1, q_2, ..., q_Q\}$. The DFT of this signal is

$$X_a(k) = X(k) + Z(k)$$
$$= \sum_{i=1}^{s} \sigma_i \delta(k - k_{0i}) + \sum_{m=1}^{Q} z(q_m) e^{-j2\pi q_m k/N}.$$

In the process of sparsity measure minimization the missing samples $y(n) = x_a(n)$, $n \in \mathbb{N}_x$, are variables. The goal is to obtain $x_a(n) = x(n)$, with $z(n) = 0$ for all n. The uniqueness means that if a sparse signal, with the transform $X(k)$, is reconstructed using the set of available sample positions then there is no other signal transform of the same or lower sparsity that can be obtained by varying values of the missing samples. Solution uniqueness depends on the number of missing samples $Q = N - M$, their positions $\mathbb{N}_x$, and the available signal values [27].

Theorem 1 *Consider a signal $x(n)$ that is sparse in the DFT domain with unknown sparsity. Assume that the signal length is $N = 2^r$ samples and that Q samples are missing at the instants $q \in \mathbb{N}_x$. Also assume that the reconstruction is performed and that the DFT of reconstructed signal is of sparsity s. Reconstruction result is unique if the inequality*

$$s < N - \max_{h=0,1,...,r-1} \left\{ 2^h (Q_{2^h} - 1) - s \right\}$$

holds. Integers Q_{2^h} are calculated as

$$Q_{2^h} = \max_{b=0,1,...,2^h-1} \{ \text{card}\{ q : q \in \mathbb{N}_x \text{ and } \text{mod}(q, 2^h) = b \} \}.$$

The answer is obtained almost immediately, since the computational complexity of the Theorem is of order $O(N)$. The proof is given in [27].

Here we will illustrate the uniqueness test on the signal from the Example 4, whose sparsity is $s = 10$ and the algorithm has removed $Q = 68$ out of $N = 128$ samples. Using the theorem for the specific set of removed samples $\mathbb{N}_x$ we obtained the sparsity limit $s < 16$. It means that the reconstruction is unique.

For the same number of missing samples the theorem is run $100,000$ times with arbitrary possible distribution of $Q = 68$ removed sample positions. The probability that a signal with sparsity s is unique, with randomly removed $Q = 68$ samples is presented in Fig.8. Probability that the worst case signal with sparsity $s = 10$ is unique for $Q = 68$ is 0.9188.

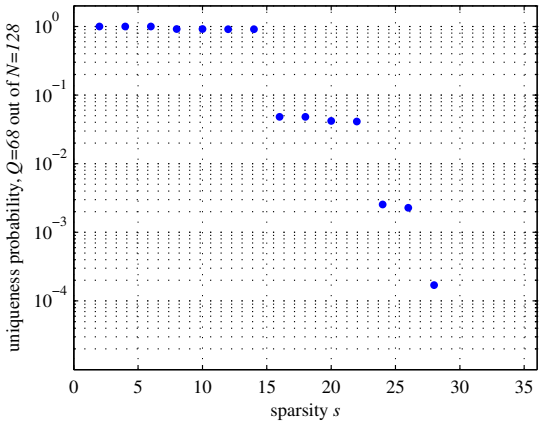

Fig. 8 Sparsity limit probability distribution for the worst possible case of signal with $Q = 68$ out of $N = 128$ samples in 100,000 random realizations.

4 Disturbance in all Signal Samples

The results from the previous section will be generalized to the case when all signal samples are corrupted.

Consider a signal $x(n)$, $0 \leq n \leq N-1$ that is sparse in a transformation domain with unknown sparsity $s \ll N$. Assume that all samples of signal $x(n)$ are corrupted with an arbitrary signal $\varepsilon(n)$. The original signal can be recovered with the signal-to-noise ratio

$$SNR = SNR_i - 10\log\left(C_Q \frac{s}{M}\right) \tag{14}$$

where SNR_i is the input signal-to-noise ratio, M is the number of samples used in the reconstruction, and the disturbance suppression ratio is

$$C_Q = \frac{\frac{1}{M}\sum_{i=1}^{M} |\varepsilon(n_i)|^2}{\frac{1}{N}\sum_{n=0}^{N-1} |\varepsilon(n)|^2}, \tag{15}$$

where n_i, $i = 1,2,\ldots,M$ are positions of the samples used in reconstruction.

By using a criterion for selective choice of samples, the average disturbance energy in the reconstructed samples is much smaller than the average energy in the whole signal. Then the disturbance suppression ratio is small, $C_Q \ll 1$. A significant signal to-noise ratio improvement can be achieved in this way. In an ideal case, when the remaining samples do not contain disturbance a full recovery can be achieved, with $SNR \to \infty$ as $C_Q \to 0$ (the case presented in Section 3). If no selection of the signal samples is used in the recovery process then $C_Q = 1$ with $SNR = SNR_i - 10\log(s/M)$.

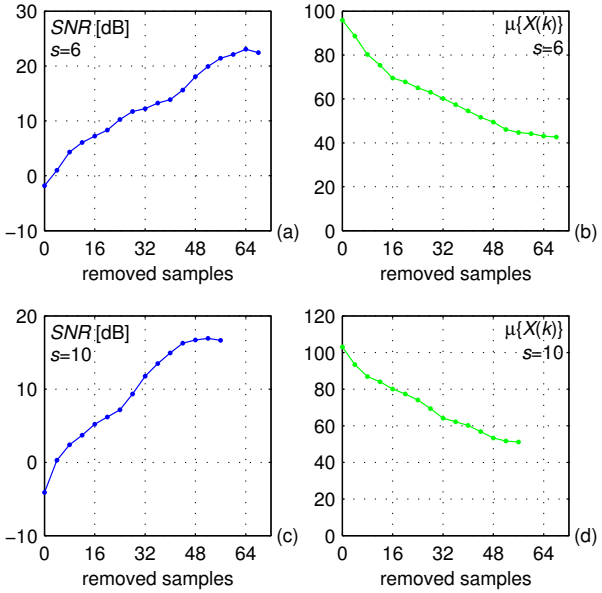

Fig. 9 Reconstruction of signals corrupted in all samples. Using the criterion $r = 4$ the most corrupted samples are removed in each iteration. (a)-(b) The SNR and concentration measure during the iterations for $s = 6$. (c)-(d) The SNR and concentration measure during the iterations for $s = 10$.

We will illustrate the case of disturbance in all signal samples on an example.

Example 5: Let us consider signal (8) with additive impulsive noise $1.5\varepsilon^3(n)$, where $\varepsilon(n)$ is the Gaussian noise with $\sigma_\varepsilon = 1$ and two sparsity values $s = 6$ and $s = 10$. In the reconstruction, the iterative removal procedure is used, with $r = 4$ signal sample positions being added to the set of unavailable/corrupted sample positions N_x in each iteration. The SNR increases in each iteration, as shown in Fig.9. The final reconstruction is with an SNR improvement greater than 20 [dB] in both cases, Figs.9(a),(c). In the case when the noise existence is expected in all remaining samples we should avoid to use the lowest possible number of signal samples sufficient for reconstruction. The influence of residual disturbance increases when the number of samples in the reconstruction approaches to the reconstruction limit.

The same procedure is repeated for signal (8) with random parameters and various s in each realization. Theoretical and statistical results for the signals of sparsity $s = 6, 10, 14, 20$ and 30, obtained in 100 realizations, are presented in Table 1. Statistical and theoretical results for the output SNR without using information about the sparsity s and assuming that the sparsity s is known or properly estimated are presented. The results are given for an average input SNR of about -5.3 [dB]. Number of samples was $N = 128$. The average number of omitted samples, denoted by

Table 1 Statistical and theoretical results for the SNR as a function of the sparsity s

Sparsity s	6	10	14	20	30
SNR_i	-5.37	-5.43	-5.41	-5.34	-5.27
$\langle N-M\rangle$	62	75	84	97	105
C_Q	0.003	0.0095	0.019	0.042	0.081
SNR	24.64	18.72	15.00	10.34	6.89
SNR_s	30.57	23.71	19.77	15.01	10.85
SNR_T	30.04	23.53	19.61	15.27	11.09

Input signal to noise ratio SNR_i; Average number of omitted samples $\langle N-M\rangle$ in 100 realizations; The average disturbance reduction ratio C_Q; Output signal to noise ratio SNR improvement without using information about sparsity s; Output signal to noise ratio SNR_s assuming that the sparsity s is known or properly estimated; Theory expected output signal to noise ratio SNR_T with the sparsity s being known or properly estimated.

$\langle N-M\rangle$, in the presented procedure and the average disturbance reduction ratio C_Q, obtained by using the proposed selection criterion in 100 realizations, are given for each sparsity s.

4.1 Analysis of the Residual Disturbance Influence

The signal-to-noise ratio (SNR) in the input signal, $x(n)+\varepsilon(n)$, is

$$SNR_i = 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\sum_{n=0}^{N-1}|\varepsilon(n)|^2} = 10\log\frac{E_x}{E_\varepsilon}.$$

Assume that by using the described procedures (the criterion based sorting, random search, and/or their combinations) we have found M samples that are the favorable samples for the signal reconstruction (to be used as "available" samples). Then the reconstruction is performed based on

$$\begin{aligned} y(n_i) &= x(n_i)+\varepsilon(n_i) \\ i &= 1,2,...,M. \end{aligned} \tag{16}$$

The same holds for the original compressive sensing formulation when M samples are available.

If the noisy signal contains at least M noise-free samples, where M is a number of samples that guarantees a successful signal reconstruction, then the perfect reconstruction can be achieved, in theory, up to the computer precision. Then the signal to reconstruction error ratio tends to infinity, $SNR = SRR \to \infty$. The matter here is only the number of calculations that have to be performed to find and correctly detect the case when only the disturbance-free samples are used in the reconstruction.

In the case when we cannot provide that there are M disturbance-free samples, then the influence of the residual disturbance to the reconstruction error is analyzed. Assume that in the remaining M samples the disturbance energy is

$$E_{\varepsilon A} = \sum_{i=1}^{M} |\varepsilon(n_i)|^2 . \tag{17}$$

For the beginning consider the simplest case when the signal is reconstructed in a direct way, by using the signal transform calculated with the available M samples at the instants n_i

$$X_R(k) = \sum_{i=1}^{M} y(n_i)\varphi_k^*(n_i). \tag{18}$$

Omitting the unavailable samples in summation is the same as assuming that their values are zero. This kind of calculation corresponds to the result that would be achieved for the signal transform if the l_2-norm, i.e., $\min \sum_{k=0}^{N-1} |X(k)|^2$, is used in the minimization, [28, 30, 32]. Then the available signal values $y(n_i)$, $i = 1, 2, ..., M$, for a frequency corresponding to a signal component are summed in phase, while the unavailable samples are omitted in summation. This summation produces value MA_i at a frequency of the ith component of signal (3), where A_i is the component amplitude in the time domain. The total disturbance energy in both the signal and the signal transform is equal to the disturbance energy in available samples (according to Parseval's theorem).

The true amplitude in the signal transform at the frequency k_i, in the case if all signal samples were used, would be NA_i. To compensate the resulting transform for the known bias in amplitude when only M available samples are used we should multiply the coefficient by N/M. The same is done by any reconstruction algorithm that produces (reconstructs) the correct amplitude of the signal component. It means that in a full recovery, a signal transform coefficient should correspond to the coefficient of the original signal with all signal samples being used. The disturbance in the transform coefficients will be also multiplied by the same factor. Therefore, its energy would be increased to $E_{\varepsilon A}N^2/M^2$. The signal to noise ratio in the resulting signal would be

$$SNR = 10\log \frac{\sum_{n=0}^{N-1} |x(n)|^2}{\frac{N^2}{M^2} \sum_{i=1}^{M} |\varepsilon(n_i)|^2} \tag{19}$$

If the distribution of the disturbance in the available samples is the same as in other signal samples (no selection criterion is used), then $\sum_{i=1}^{M}|\varepsilon(n_i)|^2 = M\sigma_\varepsilon^2$ and

$$SNR = 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\frac{N^2}{M^2}M\sigma_\varepsilon^2}$$
$$= 10\log\frac{E_x}{E_\varepsilon} - 10\log\left(\frac{N}{M}\right) \tag{20}$$
$$= SNR_i - 10\log\left(\frac{N}{M}\right). \tag{21}$$

Therefore, this kind of signal reconstruction, using (18), would worsen SNR, since $N > M$.

An improvement in the direct reconstruction can be expected only if we remove the corrupted samples in a selective manner so that the remaining samples are less disturbed than the omitted samples. If a selection criterion, for example the one proposed in this paper, can remove the disturbance in such a way that

$$\frac{N^2}{M^2}\sum_{i=1}^{M}|\varepsilon(n_i)|^2 < \sum_{n=0}^{N-1}|\varepsilon(n)|^2,$$

then the reconstruction can be improved, even by the direct calculation (18).

The direct calculation is used in a combination with the assumption that the signal is sparse, with a sparsity $s < M$. Positions of the nonzero coefficients in transformation domain are assumed to be known (or estimated based on (18) by using a threshold or an iterative procedure) for the s signal components. In an ideal case all the DFT values other than the values at the known s frequencies are then set to zero. The following system,

$$\frac{1}{N}\sum_{l=1}^{s}X(k_l)\varphi_{k_l}(n_i) = x(n_i), \tag{22}$$

of M linear equations (for all available signal samples $x(n_i)$, $i = 1,2,...,M$), based on (2), is solved for s unknown coefficients $X(k_l)$, $l = 1,2,...,s$, $M \geq s$. A matrix notation of this relation is $\Phi\mathbf{X}_s = \mathbf{y}$. Its solution follows from $\Phi^{\mathrm{H}}\Phi\mathbf{X}_{\mathrm{s}} = \Phi^{\mathrm{H}}\mathbf{y}$ as $\mathbf{X}_s = (\Phi^{\mathrm{H}}\Phi)^{-1}\Phi^{\mathrm{H}}\mathbf{y}$. In this notation $\mathbf{y}$ is the vector of available signal samples, $\mathbf{X}_s$ is the vector with s unknown nonzero transform coefficients ($X(k_l)$, $l = 1,2,...,s$) and Φ is the inverse transformation matrix with omitted rows corresponding to the unavailable signal samples and omitted columns corresponding to zero transform coefficients.

In this way the energy of the reconstruction error is reduced for the factor of s/N, since only s out of N DFT coefficients could be used in the signal reconstruction. The disturbance energy in the signal reconstructed in this way, is

$$E_{\varepsilon R} = \frac{s}{N}\frac{N^2}{M^2}\sum_{i=1}^{M}|\varepsilon(n_i)|^2.$$

The overall signal to noise ratio in the reconstructed signal, with s components at known (or estimated) frequency positions, is

$$SNR = 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\frac{sN}{M^2}\sum_{i=1}^{M}|\varepsilon(n_i)|^2}. \tag{23}$$

Consider two special cases of this kind of reconstruction:

(a) The special case of random distributed noise, without any criterion for corrupted samples selection, when

$$\sum_{i=1}^{M}|\varepsilon(n_i)|^2 = M\sigma_\varepsilon^2. \tag{24}$$

Then SNR can be calculated as

$$\begin{aligned} SNR &= 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\frac{sN}{M^2}M\sigma_\varepsilon^2} \\ &= 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\frac{s}{M}\sigma_\varepsilon^2 N} = SNR_i - 10\log\frac{s}{M}. \end{aligned} \tag{25}$$

In this case the improvement depends on the ratio of available samples and the number of signal components whose frequencies we know. This relation is statistically checked for reconstructions with various N, M and s and the agreement with the presented theory is exact.

(b) If a criterion for corrupted samples selection is used then the mean disturbance energy (variance) in the remaining samples is lower than the mean disturbance energy in all samples, and $0 \le C_Q < 1$ where C_Q is the criterion selection efficiency (15). With the coefficient $0 \le C_Q \le 1$ the overall improvement is

$$\begin{aligned} SNR &= 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\frac{sN}{M^2}\sum_{i=1}^{M}|\varepsilon(n_i)|^2} \\ &= 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\frac{s}{M}C_Q\sum_{n=0}^{N-1}|\varepsilon(n)|^2} \\ &= SNR_i - 10\log\left(C_Q\frac{s}{M}\right). \end{aligned} \tag{26}$$

This relation holds for the case when s DFT values are used. If we do not have any knowledge about the signal sparsity s then

$$\begin{aligned} SNR &= 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\frac{N^2}{M^2}\sum_{i=1}^{M}|\varepsilon(n_i)|^2} \\ &= 10\log\frac{\sum_{n=0}^{N-1}|x(n)|^2}{\frac{N^2}{M^2}\frac{M}{N}C_Q\sum_{n=0}^{N-1}|\varepsilon(n)|^2} \\ &= SNR_i - 10\log\left(\frac{N}{M}C_Q\right). \end{aligned} \tag{27}$$

Improvement in the reconstruction algorithms, without using knowledge about s, is achieved if the criterion for the noisy samples selection is such that $C_Q < M/N$.

4.2 Gaussian Noise Analysis

In the Gaussian noise case the best scenario would be to eliminate signal samples with higher noise values and to keep for the reconstruction the samples with lower noise values. For the case of N signal samples and the reconstruction based on M samples we can find the interval of amplitudes A_L for the lowest M noisy samples based on

$$\frac{1}{\sqrt{2\pi}}\int_{-A_L}^{A_L} e^{-\xi^2/2}d\xi = \text{erf}\left(\frac{A_L}{\sqrt{2}}\right) = \frac{M}{N}.$$

The calculation of A_L value is easily related to the inverse $\text{erf}(x)$ function denoted by $\text{erfinv}(x)$. For a given M/N, the amplitude is $A_L = \sqrt{2}\text{erfinv}(\frac{M}{N})$. For example, for $M = N/2$ a half of the lowest noise samples will be within the interval $[-0.6745\sigma_\varepsilon, 0.6745\sigma_\varepsilon]$ since $A_L = \sqrt{2}\text{erfinv}(0.5) = 0.6745$. The variance of this new noise (formed from the Gaussian noise after the largest $N/2$ values are removed) is much lower than the variance of the whole noise. It is

$$\sigma_L^2 = \frac{1}{\frac{M}{N}\sigma_\varepsilon\sqrt{2\pi}}\int_{-\sqrt{2}\text{erfinv}(\frac{M}{N})\sigma_\varepsilon}^{\sqrt{2}\text{erfinv}(\frac{M}{N})\sigma_\varepsilon} \xi^2 e^{-\xi^2/(2\sigma_\varepsilon^2)}d\xi. \tag{28}$$

The factor M/N in (28) comes from the probability density normalization. For the new noise being formed using the Gaussian noise values below a certain value, the integral over its pdf is 1. For $M/N = 1/2$ the variance of the new noise that consists of 50% the lowest input Gaussian noise values is $\sigma_L^2 = 0.1426\sigma_\varepsilon^2$.

Thus, the input noise to the reconstruction algorithm will have lower variance if a sorting criterion is applied. The mean energy of noise (variance) of M out of N the lowest values of the Gaussian noise with respect to the input signal mean energy is shown in Fig.11 (line with + marks). It is calculated theoretically using (28) and statistically for various M. The agreement is complete. For example, for $M = N/2$ the noise mean energy is reduced by sorting for

$$10\log\frac{\sigma_L^2}{\sigma_\varepsilon^2} = 10\log(0.1426) = -8.46 \text{ [dB]}.$$

If the criterion for selecting samples is applied on the noise only with a large Δ it produces the same result as the noise sorting. Then the ideal noise reduction would be achieved with SNR improvements described in the previous section.

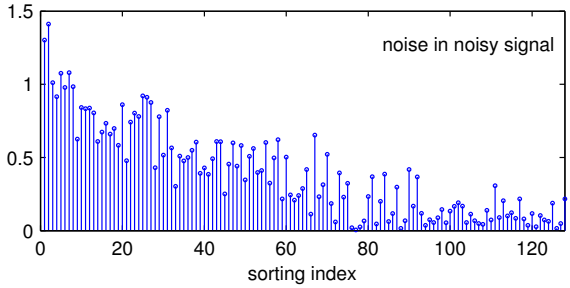

Fig. 10 A realization of small Gaussian noise in all signal samples, sorted according to the criterion.

The presented algorithm reconstructs the signal, with a tendency to reduce the noise in the reconstructed signal in such a way that the sum of absolute DFT values is as small as possible (minimization of the sum of absolute values is completely different than the minimization of the sum of squared absolute values). The improvement in the mean noise energy in the reconstructed signal with respect to the input noise energy, in the noise-only case, for different number of omitted samples $N-M$ is calculated statistically over 100 realizations and presented in Fig.11 (line with x marks). In the case of large number of omitted samples, close to N, the remaining noise energy tends to zero. However, this region can not be used in the reconstruction since there is not sufficient number of samples to reconstruct the signal, Fig. 1. Thus, this region should be avoided. Based on the way the presented algorithm works, it is interesting that a slight noise reduction will result even for a random selection of the noise samples. For example, $M = N/2$ and a random selection of samples the output variance of the reconstructed signal is $\sigma_R^2 = \frac{2}{3}\sigma_\varepsilon^2$ corresponding to an improvement of $10\log(2/3) = -1.76$ [dB]. This kind of a slight improvement was reported in [28].

In the case of noisy signal the criterion deviates from pure sorting. However in average the criterion produces the noise values with greater amplitudes first, while the lower amplitudes are at the end of the criterion preferred values Fig. 10. The improvement with criterion is calculated for various number of omitted samples and presented in Fig.11 (line with o marks). For $M = N/2$ the improvement is

$$10\log\frac{\sigma_Q^2}{\sigma_\varepsilon^2} = 10\log(0.31) = -5.1 \text{ [dB]}.$$

It is just about 3 [dB] worse than the best possible case that could be achieved if the noise values could be sorted in perfect order and removed accordingly from the noisy signal (line with + mark).

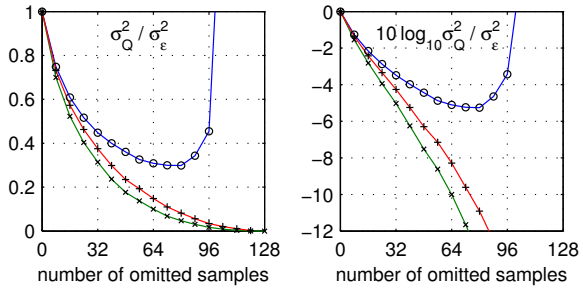

Fig. 11 The improvement in the mean noise energy in the reconstructed signal with respect to the input noise energy, in the noise-only case (line denoted by x) and the noisy signal (line with o marks), for different number of omitted samples $N - M$ and the criterion based selection. The line with + marks represents the mean energy of input noise after the criterion for selection is applied in noise-only case (corresponding to pure noise sorting). Ratio values in [dB] are presented as well (right).

5 Conclusion

Reconstruction of sparse signals in impulsive disturbance environments is considered. In the case of a small number of disturbing samples the direct search over a disturbance-free subset of samples can be performed. The sparsity measure can be used to detect the event of disturbance-free based recovery. For a large number of corrupted samples, with most of the noise being within the signal amplitude order a criterion to detect and eliminate the corrupted samples is defined. This criterion used alone or in a form of a iterative procedure, turn out to be a powerful tool for the analysis, even in the case when all signal samples are corrupted by impulsive noise. Some improvements may be achieved in the case of Gaussian noise as well. All methods are illustrated by examples and the theory was checked by simulations.

References

1. Ahmad, M.S., Kukrer, O., Hocanin, A.: Robust recursive inverse adaptive algorithm in impulsive noise. Circuits, Systems, and Signal Processing **31**(2),703–710,(2012)
2. Baraniuk, R.G.: Compressive sensing. IEEE Signal Processing Magazine **24**(4) (2007)
3. Blumensath, T.: Sampling and reconstructing signals from a union of linear subspaces. IEEE Transactions on Information Theory **57**(7), 4660–4671 (2011)
4. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory **52**(2), 489–509 (2006)
5. Candès, E.J., Tao, T.: Decoding by linear programming. IEEE Transactions on Information Theory **51**(12), 4203–4215 (2005)
6. Candès, E.J., Tao, T.: Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Transactions on Information Theory **52**(12), 5406–5425 (2006)

7. Carrillo, R.E., Barner, K.E., Aysal, T.C.: Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise. IEEE Journal of Selected Topics in Signal Processing **4**(2),392–408 (2010)
8. Dadouchi, F., Gervaise, C., Ioana, C., Huillery, J., Mars, J.I.: Automated segmentation of linear time-frequency representations of marine-mammal sounds. The Journal of the Acoustical Society of America **134**(3), 2546–2555 (2013)
9. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on pure and applied mathematics **57**(11), 1413–1457 (2004)
10. Davis, G., Mallat, S., Avellaneda, M.: Adaptive greedy approximations. Constructive approximation **13**(1), 57–98 (1997)
11. Djurović, I., Lukin, V.V., Simeunović, M., Barkat, B.: Quasi maximum likelihood estimator of polynomial phase signals for compressed sensed data. AEU-International Journal of Electronics and Communications **68**(7), 631–636 (2014)
12. Donoho, D.L.: Compressed sensing. IEEE Transactions on Information Theory **52**(4), 1289–1306 (2006)
13. Donoho, D.L., Elad, M., Temlyakov, V.N.: Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Transactions on Information Theory **52**(1), 6–18 (2006)
14. Fadili, M.J., Starck, J.L., Murtagh, F.: Inpainting and zooming using sparse representations. The Computer Journal **52**(1), 64–79 (2009)
15. Figueiredo, M.A., Nowak, R.D., Wright, S.J.: Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. IEEE Journal of Selected Topics in Signal Processing **1**(4), 586–597 (2007)
16. Flandrin, P., Borgnat, P.: Time-frequency energy distributions meet compressed sensing. IEEE Transactions on Signal Processing **58**(6), 2974–2982 (2010)
17. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing **41**(12), 3397–3415 (1993)
18. Moré, J.J., Toraldo, G.: On the solution of large quadratic programming problems with bound constraints. SIAM Journal on Optimization **1**(1), 93–113 (1991)
19. Needell, D., Tropp, J.A.: CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. Applied and Computational Harmonic Analysis **26**(3), 301–321 (2009)
20. Ramadan, Z.M.: Efficient restoration method for images corrupted with impulse noise. Circuits, Systems, and Signal Processing **31**(4), 1397–1406 (2012)
21. Roenko, A.A., Lukin, V.V., Djurović, I.: Two approaches to adaptation of sample myriad to characteristics of SαS distribution data. Signal Processing **90**(7), 2113–2123 (2010)
22. Sejdić, E., Cam, A., Chaparro, L., Steele, C., Chau, T.: Compressive sampling of swallowing accelerometry signals using time-frequency dictionaries based on modulated discrete prolate spheroidal sequences. EURASIP Journal on Advances in Signal Processing **101**, 1–14 (2012)
23. Serafini, T., Zanghirati, G., Zanni, L.: Gradient projection methods for quadratic programs and applications in training support vector machines. Optimization Methods and Software **20**(2-3), 353–378 (2005)
24. Stanković, I.: Recovery of images with missing pixels using a gradient compressive sensing algorithm. arXiv preprint arXiv:1407.3695 (2014)
25. Stanković, L.: A measure of some time-frequency distributions concentration. Signal Processing **81**(3), 621–631 (2001)
26. Stanković, L., Daković, M.: On the reconstruction of randomly sampled sparse signals using an adaptive gradient algorithm. http://arxiv.org/abs/1412.0624
27. Stanković, L., Daković, M.: On the uniqueness of the sparse signals reconstruction based on the missing samples variation analysis. Mathematical Problems in Engineering, Article ID 629759, 14 pages **2015** (2015). doi.10.1155/2015/629759

28. Stanković, L., Daković, M., Vujović, S.: Adaptive variable step algorithm for missing samples recovery in sparse signals. IET Signal Processing **8**(3), 246–256 (2014)
29. Stanković, L.: Digital Signal Processing with Selected Topics, CreateSpace Independent Publishing Platform, An Amazon.com Company, (2015)
30. Stanković, L., Orović, I., Stanković, S., Amin, M.G.: Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency. IEEE Transactions on Signal Processing **61**(18), 4562–4572 (2013). doi.10.1109/TSP.2013.2271752
31. Stanković, L., Stanković, S., Amin, M.G.: Missing samples analysis in signals for applications to L-estimation and compressive sensing. Signal Processing **94**, 401–408 (2014)
32. Stanković, L., Stanković, S., Orović, I., Amin, M.G.: Robust time-frequency analysis based on the L-estimation and compressive sensing. IEEE Signal Processing Letters **20**(5), 499–502 (2013)
33. Studer, C., Kuppinger, P., Pope, G., Bolcskei, H.: Recovery of sparsely corrupted signals. IEEE Transactions on Information Theory **58**(5), 3115–3130 (2012)
34. Turlach, B.A.: On algorithms for solving least squares problems under an L1 penalty or an L1 constraint. In: 2004 Proceedings of the American Statistical Association, Statistical Computing Section [CD-ROM], pp. 2572–2577 (2005)
35. Vaseghi, S.V.: Advanced digital signal processing and noise reduction. John Wiley & Sons (2008)
36. Wright, S.J.: Implementing proximal point methods for linear programming. Journal of optimization Theory and Applications **65**(3), 531–554 (1990)
37. You, G.h., Qiu, T.s., Song, A.m.: Novel direction findings for cyclostationary signals in impulsive noise environments. Circuits, Systems, and Signal Processing **32**(6), 2939–2956 (2013)