Adaptive Variable Step Algorithm for Missing Samples Recovery in Sparse Signals

Ljubiša Stanković, Miloš Daković, Stefan Vujović

Abstract—**Recovery of arbitrarily positioned samples that are missing in sparse signals recently attracted significant research interest. Sparse signals with heavily corrupted arbitrary positioned samples could be analyzed in the same way as compressive sensed signals by omitting the corrupted samples and considering them as unavailable during the recovery process. The reconstruction of missing samples is done by using one of the well known reconstruction algorithms. In this paper we will propose a very simple and efficient adaptive variable step algorithm, applied directly to the concentration measures, without reformulating the reconstruction problem within the standard linear programming form. Direct application of the gradient approach to the nondifferentiable forms of measures lead us to introduce a variable step size algorithm. A criterion for changing adaptive algorithm parameters is presented. The results are illustrated on the examples with sparse signals, including approximately sparse signals and noisy sparse signals.**

Keywords— **Sparse signals, Compressive sensing, Robust signal processing, Concentration measure, Signal reconstruction, L-estimation**

I. INTRODUCTION

In many signal processing applications a signal that spans over the whole time domain is located within much smaller regions in a transformation domain. If we consider a discrete time-limited signal, it could contain much smaller number of nonzero samples (coefficients) in an arbitrary transformation domain (Fourier domain, Discrete cosine domain, Discrete Wavelet domain,...). It is said that this signal is sparse in this transformation domain. Remaining transformation coefficients of the signal are assumed to be equal to zero or could be approximated by zero without making significant error. If this condition is satisfied we can reconstruct the signal without using the whole data set required by the Shannon-Nyquist sampling theorem. Processing of sparse signal with a large number of missing/unavailable samples attracted significant interest in recent years. This research area interacts with many other research areas like signal processing, statistics, machine learning,

Address: University of Montenegro, 81000 Podgorica, Montenegro
E-mail: ljubisa@ac.me, milos@ac.me, stefanv@ac.me

coding. Compressive sensing/sampling (CS) is a field which refers to a sparse approximation [1], [2]. The crucial parameter in the approximation is the number of available samples/measurements used in the reconstruction. It is directly related to the number of non-zero sparse coefficients [1], [2], [3].

Signal samples may be missing due to their physical or measurements unavailability. Also, if some arbitrary positioned samples of signal are so heavily corrupted by disturbance, it was shown that it is better to omit them in the analysis or processing (by L-estimation, for example [4], [5], [6]). Both of these situations corresponds to the CS approximation problems if the signal to be analyzed is sparse and could be considered within the framework of missing samples. Under the conditions defined within the CS framework, the processing of a signal could be performed with the remaining samples almost as in the case if all missing/unavailable samples were available.

Several approaches to reconstruct these kind of signals are introduced [7]-[27]. One group of them is based on the gradient [27] and the other is based on the matching pursuit approaches [25]. A common approach to this problem is based on redefining it within the linear programming (LP) as the bound constrained quadratic program (BCQP). A measure of signal sparsity is used as a minimization function in the sparse signal reconstruction. This measure is related to the number of nonzero transformation coefficients. This kind of measure was also used, especially in time-frequency analysis, for measuring the concentration of a signal representation. Since, the sparsity of a signal in a transformation domain is related to the number of nonzero samples, a natural mathematical form to measure the number of nonzero (significant) samples in a signal transform is the norm-zero (l_0 norm). This norm is sum of the signal transformation absolute values raised to a zeroth power. Since this power produces value one for any nonzero transformation coefficient, the norm just counts the number of nonzero coefficients. However, this norm is very sensitive to any kind of disturbance that can make zero transformation values to be small but different from zero. Thus more robust norms are used. The norm that may be used for measuring the transformation concentration,

being also less sensitive to disturbances, is the norm-one (l_1 norm). Since the norm-one is not differentiable around the optimal point, the CS algorithms reformulate problem under the norm-two (l_2 norm) conditions before the optimization is done.

In this paper we will present a gradient based algorithm for reconstruction of sparse signals. Presented algorithm uses an arbitrarily concentration measure in a direct way, without redefining the problem to a quadratic form and using liner programming tools. Since the signal reconstruction is required in the time domain, the proposed algorithm performs a search over the time domain coefficients. The presented algorithm can reconstruct a large number of missing samples in computationally efficient way. The proposed method belongs to the class of gradient based CS algorithms [10]. However, common adaptive signal processing and the CS algorithms avoid direct use of the measure based on the norm-one (or similar norms between norm-zero and norm-one) since it is not differentiable. The value of gradient cannot be used as an indicator of the proximity of iteration values to the algorithm solution. When the iterations are close to the optimal point, gradient value remains the same in norm-one and oscillate around the true value. Taking sufficiently small step over the whole range would not be a solution, due to extremely large number of iterations over a very large set of variable. Here, the adaptive gradient based approach is directly applied to an appropriately chosen concentration measure. Since for commonly used norm-one (or any other norm between norm-zero and norm-one) based concentration measure the derivatives are not continuous functions around the minimum, a variable and self-adaptive step in the algorithm is introduced. The presented algorithm, with this adaptive step, reconstructs a large number of missing samples in a simple and computationally efficient way with arbitrary (computer defined) precision of the results.

The paper is organized as follows. After the introduction, a review and analysis of concentration measures in the processing of sparse signals is done. A gradient based algorithm, with its modifications is presented and illustrated. The presented algorithm efficiency is demonstrated on several examples with large number of missing samples, including the samples missing in blocks and the noisy signals. The basic idea for this algorithm was presented in [28].

II. MEASURES AND DIRECT RECONSTRUCTION

Concentration measures of signal transforms were intensively studied and used in the area of time-frequency signal analysis and processing. They are used to find an optimal, best concentrated time-frequency representation

of a signal. The most common and the oldest measure introduced to measure concentration of time-frequency representations is defined by Jones, Parks, Baraniuk, Flandrin, Williams, *et al.* Concentration of a signal transform $X(k)$ is measured by

$$\mathcal{M}^{(4/2)}[X(k)] = \frac{\sum_k |X(k)|^4}{\left(\sum_k |X(k)|^2\right)^2}. \qquad (1)$$

In general, it has been shown that any other ratio of norms $l_p = \sum_k |X(k)|^p$ and $l_q = \sum_k |X(k)|^q$, $p > q > 1$, can also be used for measuring the concentration. This kind of concentration measures were inspired by the kurtosis as a measure of distribution peakedness. Similar forms are obtained by using the Rényi measures.

Another direction to measure time-frequency representation concentration comes from a classical definition of the time-limited signal duration, rather than measuring signal peakedness. It was used in time-frequency analysis in [24]. If a signal $x(n)$ is time-limited, $x(n) \neq 0$ only for $n \in [n_1, n_2 - 1]$, then the duration of $x(n)$ is $d = n_2 - n_1$. It can be written as

$$d = \lim_{p \to \infty} \sum_n |x(n)|^{1/p} = \|x(n)\|_0, \qquad (2)$$

where $\|x(n)\|_0$ denotes the norm-zero l_0 of signal. In reality, there is no sharp edge between $x(n) \neq 0$ and $x(n) = 0$, so the value of d in (2) could, for very large p (close to norm-zero), be sensitive to small values of $|x(n)|$. The robustness may be achieved by using lower-order forms, with $1 \leq p < \infty$ (norms from l_1 to l_0).

Therefore, the concentration of a signal transform $X(k) = T[x(n)]$ can be measured with the function of the form

$$\mathcal{M}_p[T[x(n)]] = \frac{1}{N} \sum_k |X(k)|^{1/p}, \qquad (3)$$

with $1 \leq p < \infty$, where N is total number of samples in signal transform $X(k)$. A lower value of (3) indicates better concentrated distribution. For $p = 1$, it is the norm-one form

$$\mathcal{M}_1[T[x(n)]] = \frac{1}{N} \sum_k |X(k)| = \frac{1}{N} \|X(k)\|_1.$$

Minimization of the norm-one of the short-time Fourier transform (norm 1/2 of the spectrogram) is used in [24] to optimize the window width and to produce the best concentrated signal representation. The norm-one is also the most commonly used in the CS algorithms for measuring signal sparsity/concentration [1], [2], [17], [18]. Here we will illustrate the influence of measure parameter p on the results, including explanation why the norms greater than one ($p < 1$, including l_2 case) cannot be used for measuring concentration. These norms could

be used in the ratio forms (1) only, but not as the stand-alone transformation measures, like in the case of $1 \leq p < \infty$.

The simplest reconstruction algorithm will be based on a direct search over all unavailable/missing samples values, by minimizing the concentration measure. If we consider a complete set of signal samples $\{x(1), x(2), ..., x(N-1)\}$ and M samples $x(m_1)$, $x(m_2)$,...,$x(m_M)$ are missing then the simplest algorithm will be to search over all possible values of missing samples and find solution that minimizes the concentration measure

$$\min_{x(m_1), x(m_2),...,x(m_M)} \{\mathcal{M}_p[T[x(n)]]\}.$$

From the remaining samples we can estimate the range limits for the missing samples, $|x(m_k)| \leq A$. In the direct search approach we can vary each missing sample value from $-A$ to A with a step $2A/(L-1)$ where L is the number of considered values within the selected range. It is obvious that the reconstruction error is limited by the step $2A/(L-1)$ used in the direct search. Number of analyzed values for M coefficients is L^M. Obviously, the direct search can be used only for a small number of missing samples, since for any reasonable accuracy value of L is large.

One possible approach to reduce the number of operations in the direct search is to use a large step (small L) in the first (rough) estimation, then to reduce the step around the rough estimate of unavailable/missing values $x(m_1)$, $x(m_2)$,..., $x(m_M)$. This can be repeated several times, until the desired accuracy is achieved. For example, for $A = 1$ the accuracy of 0.001 is achieved in one iteration if $L = 2001$. With, for example, 7 missing samples that would mean unacceptable number of $2001^7 \sim 10^{23}$ measure calculations. However, if the first search is done with, for example $L = 5$, the rough optimal values are found, and the procedure is repeated with $L = 5$ values within the range determined by the rough optimal in the first step. Repeating the same procedure six more times, the accuracy better than 0.001 is reached with $7 \times 5^7 \sim 10^5$ measure calculations. In this way, we have been able to analyze (on an ordinary PC, within a reasonable calculation time), signals with up to 10 missing samples.

Although, computationally not efficient, the direct search method is very important and helpful in the analysis of various concentration measures with different p, since all more advanced and efficient methods from literature produce results with nice values of p only (for example, $p = 1$, $p = 1/2$, or $p = 2$). The direct method can be used with any p. Also, the probability that we find
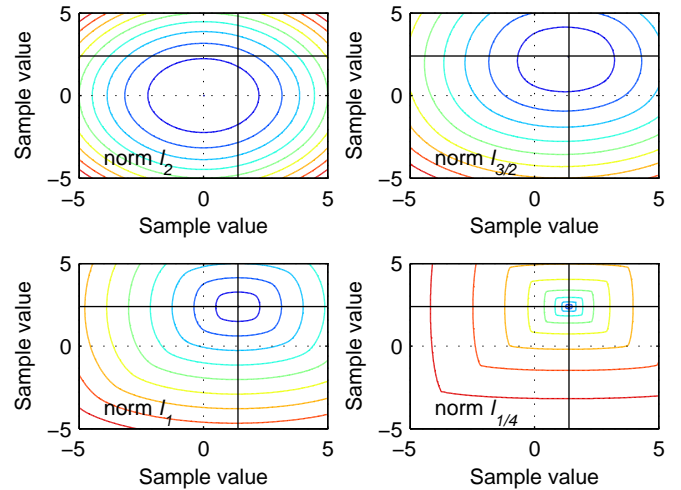

Fig. 1. Measure as a function of two missing sample values for various norms. True values of missing samples are presented with black lines

and stay in a local minimum is lower in the direct method than when using other algorithms. Thus, we will use the direct search to illustrate how the solution depends on the chose norm (concentration measure form).

Example: Consider a discrete signal

$$x(n) = 2.5 \sin(20\pi n/N) \qquad (4)$$

for $n = 0, 1, \ldots, N-1$, and $N = 256$ is number of signal samples. The case of two missing samples is presented first, as the one appropriate for graphical illustration. Direct search is performed over range $[-5, 5]$ with step 0.01. The measure (3) is calculated for various values of parameter p. Results are shown in Fig. 1. The measure minimum is located on the true sample values for $p \geq 1$ (norms l_1 and lower). The measure minimum for $p < 1$ (including norm-two, l_2, for $p = 1/2$) is not located at the true signal values. Important case with two missing samples and $p = 1$ is presented in Fig. 2.

In order to illustrate the measure influence on the mean absolute error (MAE) the direct search is also performed on the signal

$$x(n) = 3 \sin(10\pi n/N) + 2 \cos(30\pi n/N). \qquad (5)$$

Signal is composed of $N = 256$ samples while the cases with 4 and 7 missing samples are analyzed. The results with 10 and 15 iterations (to reduce the step size) are presented in Fig. 3. We can see that:

1) $p \geq 1$ (norms l_q with $q = 1/p \leq 1$, including l_1) produces accurate results with the MAE depending on the direct search step only. The MAE can be further reduced to the computer precision by reducing the step size with more iterations.
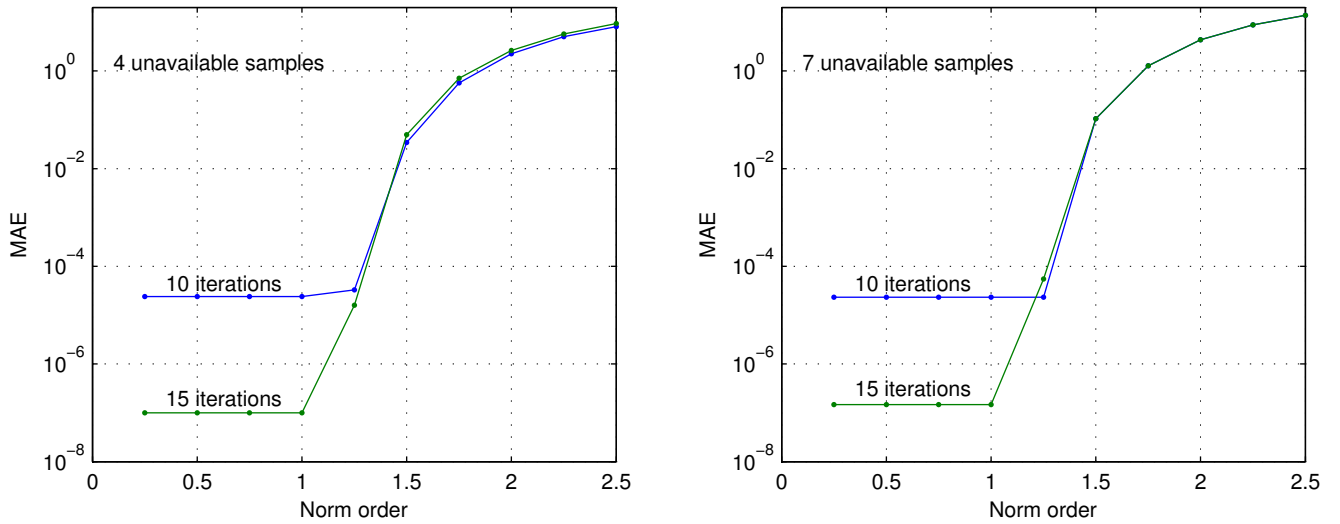
Fig. 3. Mean absolute error (MAE) in the coefficients estimation as a function of the norm l order for 4 missing samples (left) and 7 missing samples (right). The MAE is normalized with the number of missing samples.
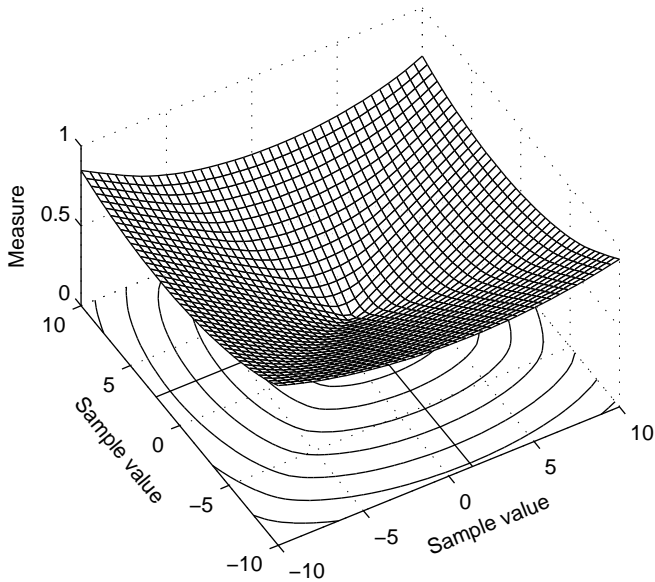

Fig. 2. Measure for $p = 1$ (norm l_1) as a function of the missing sample values

2) For $p < 1$ (norms l_q with $q > 1$, including l_2) the bias dominate over the number of iterations, so the results are almost independent from number of iterations.

Almost the same results are obtained for 4 and 7 missing samples cases.

Minimization Using l_2 Norm: For $p = 0.5$ this measure is equivalent with the well-known l_2 norm used in definitions of standard signal transforms [6]. For the norm-two (l_2 norm with $p = 1/2$) the MAE is of the signal samples order, as shown in Fig. 3. **The measure with l_2 norm has a minimum when the missing signal samples values are set to zero.**

This result was expected and can be proven for any number of missing samples for signals and its transforms satisfying the Parseval's theorem. The Parserval's theorem states that the energy of a signal in the time domain is the same as the energy of the Fourier transform in the frequency domain. We know that signal has the lowest energy when its missing samples are zero-valued. Adding any other value of missing samples than zero will increase the energy. The same holds in the frequency domain since the energy in the frequency domain equals the energy in the time domain, for this norm. The minimization solution with the l_2 norm is therefore trivial. With this norm, we attempt to minimize

$$\|\mathbf{X}\|_{l_2} = \sum_{k=0}^{N-1} |X(k)|^2.$$

According to Parseval's theorem we have $\|\mathbf{X}\|_{l_2} = N \sum_{n=0}^{N-1} |x(n)|^2$. Since any value other than $x(n) = 0$ for the unavailable/missing signal samples, would increase $\|\mathbf{X}\|_2$, then the solution for the non-available sample values, with respect to the l_2 norm, are all zero values. This was the reason why this norm was not used as a concentration measure. This is also the reason why this norm can not be used in the CS based algorithms for the missing samples recovery.

III. ADAPTIVE GRADIENT BASED ALGORITHM

Due to a high computational complexity the direct search could be used only if the number of missing samples M is small enough. It is the reason why many other, more sophisticated, CS algorithms have been proposed. Here, we will present one very simple and efficient

algorithm, based on the direct use of the concentration measure gradient. This algorithm is inspired by the adaptive signal processing methods with a variable step size. This algorithm is the form of gradient descent algorithm where missing samples are estimated as the ones producing best concentration measure of the signal transformation in the sparse domain.

The norms that produce unbiased missing samples values (like for example norms with $p \geq 1$) are not differentiable around the optimal point. It means that the gradient method, if directly applied to the measure based on, for example, the l_1 norm (or any other norm with $p \geq 1$), will have a problem when approaching the optimal point. Since the gradient amplitude in the vicinity of the optimal point is almost constant for $p = 1$ (with a changing sign), the algorithm will not improve the accuracy to a level lower than the accuracy defined by the step in the gradient algorithm. This is the reason why this approach has not been used and why appropriate reformulation of this problem is done in literature. These reformulations are done within the linear programming by using the well known and available norm-two based solutions. Here, we will not try to reformulate the problem based on the l_1 norm (or any other norm with $p \geq 1$) within the linear programming l_2 framework, but to use the gradient based adaptive algorithm, with the step being appropriately adjusted (in a simple way) around the optimal point. The algorithm presented here will be a simple (and efficient) direct application of the gradient based adaptive approach to the measures based on norms that are not differentiable around the optimal point, like the l_1 norm.

As we can see from Fig. 2 measure with $p = 1$ is differentiable and convex everywhere except around the point of minimum (the optimization solution point). Therefore any algorithm applied directly to the measure based on $p = 1$ will oscillate around the solution with an amplitude defined by the step and measure form (this will be illustrated within examples). If we take a very small step for each of a large number of missing samples, it will result in an unacceptable and large number of iterations. Thus, when the steady oscillatory state (steady state in mean absolute error) is detected we should reduce the algorithm step, as we presented in the direct search. In this way, the results with a high accuracy, within an acceptable number of iterations, are achieved with a variable self-adaptive step. This simple method is able to produce the results with an error of the computer precision level. Finally, in addition to the step variation, this kind of algorithms enables that the parameter p (the norm form itself) is changed to improve the initial convergence of the algorithm.

A. Algorithm

Consider a discrete signal $x(n)$ with some samples that are not available. Assume that signal is sparse in a transformation domain $T[x(n)]$. The algorithm for missing samples reconstruction is implemented as follows:

Step 0: Form the initial signal $y^{(0)}(n)$, where (0) means that it is first iteration of algorithm, as:

$$y^{(0)}(n) = \begin{cases} x(n) & \text{for available samples} \\ 0 & \text{for missing samples} \end{cases}$$

Step 1: For each missing sample at n_i we form two signals $y_1(n)$ and $y_2(n)$ in each next iteration as

$$y_1^{(k)}(n) = \begin{cases} y^{(k)}(n) + \Delta & \text{for } n = n_i \\ y^{(k)}(n) & \text{for } n \neq n_i \end{cases}$$

$$y_2^{(k)}(n) = \begin{cases} y^{(k)}(n) - \Delta & \text{for } n = n_i \\ y^{(k)}(n) & \text{for } n \neq n_i \end{cases}$$

where k is the iteration number. Constant Δ is used to determine whether the considered signal sample should be decreased or increased.

Step 2: Estimate the differential of the signal transform measure as

$$g(n_i) = \frac{\mathcal{M}_p\left[T[y_1^{(k)}(n)]\right] - \mathcal{M}_p\left[T[y_2^{(k)}(n)]\right]}{2\Delta}, \quad (6)$$

where $\mathcal{M}_p$ is defined by (3). The differential of measure is proportional to the error $(y^{(k)}(n) - x(n))$.

Step 3: Form a gradient vector $\mathbf{G}$ with the same length as signal $x(n)$. At positions of available samples, this vector has value $G(n) = 0$. At the positions of missing samples it has values are $g(n_i)$ calculated by (6).

Step 4: Correct the values of signal $y(n)$ iteratively by

$$y^{(k+1)}(n) = y^{(k)}(n) - \mu G(n)$$

where μ is a constant that affect performances of algorithm (error and speed of convergence).

Repeating the presented iterative procedure, the missing values are going to converge to the true signal values that produce minimal concentration measure in the transformation domain. The algorithm performance depend on parameters μ and Δ. Here we will use measure that are close to norm-one based measure.

B. Varying and Adaptive Step Size

Since we use a difference of the measures to estimate the gradient, when we approach to the optimal point, the gradient with norm l_1 will be constant and we will not be able to approach the solution with a precision higher than the step μ, multiplied by constant (gradient dependent value). If we try to reduce the oscillations around the true value by using smaller step from the beginning, then

we will face with an unacceptable number of iterations. However, this problem may be solved, by reducing the step size, when we approach the stationary oscillations zone. The best solution is to use an adaptive step size in the algorithm. A large step size should be used when the concentration measure is not close to its minimum (in the starting iterations). The step is reduced as we approach the concentration measure minimum. Next we will present a method for adaptive parameters adjustment that could be applied to the proposed algorithm in order to reduce the error and to increase accuracy.

When the algorithm with constant parameters is close to the optimal point the concentration measure tends to have constant value in a few consecutive iterations. This behavior will be detected by checking the difference between two consecutive measure calculations $\mathcal{M}_p^{(k-1)} - \mathcal{M}_p^{(k)}$, where $\mathcal{M}_p^{(k)} = \mathcal{M}_p\left[T[y^{(k)}(n)]\right]$ is the sparsity measure of the reconstructed signal in the k-th iteration. When it is smaller than, for example one percent of the highest previously calculated measure difference,

$$\mathcal{M}_p^{(k-1)} - \mathcal{M}_p^{(k)} \leq P \max_{m=1,2,...,k-1} \left| \mathcal{M}_p^{(m-1)} - \mathcal{M}_p^{(m)} \right|, \tag{7}$$

where $P = 0.01$, the algorithm parameters Δ and μ should be reduced, for example 10 times.

IV. NUMERICAL EXAMPLES

Example 1: Consider the signal

$$x(n) = 3\sin(20\pi n/N) + 2\cos(60\pi n/N) + 0.5\sin(110\pi n/N). \tag{8}$$

The total number of signal samples is $N = 256$. We assume that 200 samples are missing or are not available. Two cases will be considered. One when missing samples are randomly positioned and other when samples are missing in randomly positioned blocks. We know their positions, as well as that the signal is sparse in the Fourier domain. Here we will perform the signal reconstruction with constant algorithm parameters $\Delta = 2$, $\mu = 3$, and $p = 1$.

The reconstruction results are shown in Fig. 4 and Fig. 5. Since the constant algorithm parameters are used the achieved error is small but still notable, Fig. 4(d), Fig. 5(d). The residual error value is determined by the algorithm parameters. Next we will analyze influence of parameters Δ, μ, p on the number of iterations and mean absolute error (MAE). The MAE calculated as

$$\text{MAE}(k) = \frac{1}{N} \sum_n |x(n) - y^{(k)}(n)|$$

is shown in Fig. 6 for various algorithm setups. It can be concluded that for constant algorithm parameters MAE cannot be improved by increasing number of iterations below some limit. Smaller values of Δ and μ produce lower MAE but with an increased number of iterations, as presented in Fig. 6(a) and (b). The results obtained for varying Δ and μ are presented in Fig. 6(c). Here, the parameters are changed at iterations $k = 100$ and $k = 200$. We can see that with the same number of iterations a smaller MAE is achieved. Therefore, the parameters Δ and μ should be adaptive, resulting in MAE presented in Fig. 6(d). Here we detect that after some number of iterations the gradient algorithm does not further improve sparsity of the reconstructed signal and then we use smaller values of Δ and μ for next iterations, as described in the previous section. In the Fig. 6(e) and (f) the absolute errors in two signal samples, during iteration process, are shown for all previous cases of the algorithm setup. It can be seen that this absolute errors behave in a similar manner as the MAE in the above subplots, with difference that they oscillate around the steady value, due to nondifferentiable measure around the solution. It is expected, as previously explained.

Example 2: From the previous example and theoretical considerations we have concluded that adaptive algorithm parameters should be used. The results obtained by the proposed method for self-adaptive parameter adjustment are presented in Fig. 7. The criterion (7) for adaptive step size is applied on signal (8), with 150 randomly positioned missing signal samples. The starting parameters for the adaptive algorithm were $\Delta = 20$ and $\mu = 20$. The graphics in Fig. 7(a) illustrates the MAE as a function of the iteration number. Each color on these graph matches one set of the parameters Δ and μ. The parameters were divided by 10 (and color in graph is changed) when condition (7) is met.

The dashed color lines represent the MAE when the algorithm with constant parameters is run from the same initial point. The solid red line is for the MAE when constant $\Delta = 20$ and $\mu = 20$ are used. The solid green line is for the MAE when $\Delta = 20$ and $\mu = 20$ are used at the beginning, while the algorithm has changed the parameters to $\Delta = 2$ and $\mu = 2$ when condition (7) is met. The dashed green line represents the MAE if $\Delta = 2$ and $\mu = 2$ are used from the first iteration. The solid blue line is for the MAE when the algorithm parameters at the beginning were $\Delta = 20$ and $\mu = 20$, then changed to $\Delta = 2$ and $\mu = 2$, and finally changed to $\Delta = 0.2$ and $\mu = 0.2$. The dashed blue line represents the MAE if $\Delta = 0.2$ and $\mu = 0.2$ were used from the first iteration. This process continues in the same way two more times for Fig. 7(a) and 12 more times for Fig. 7(b).
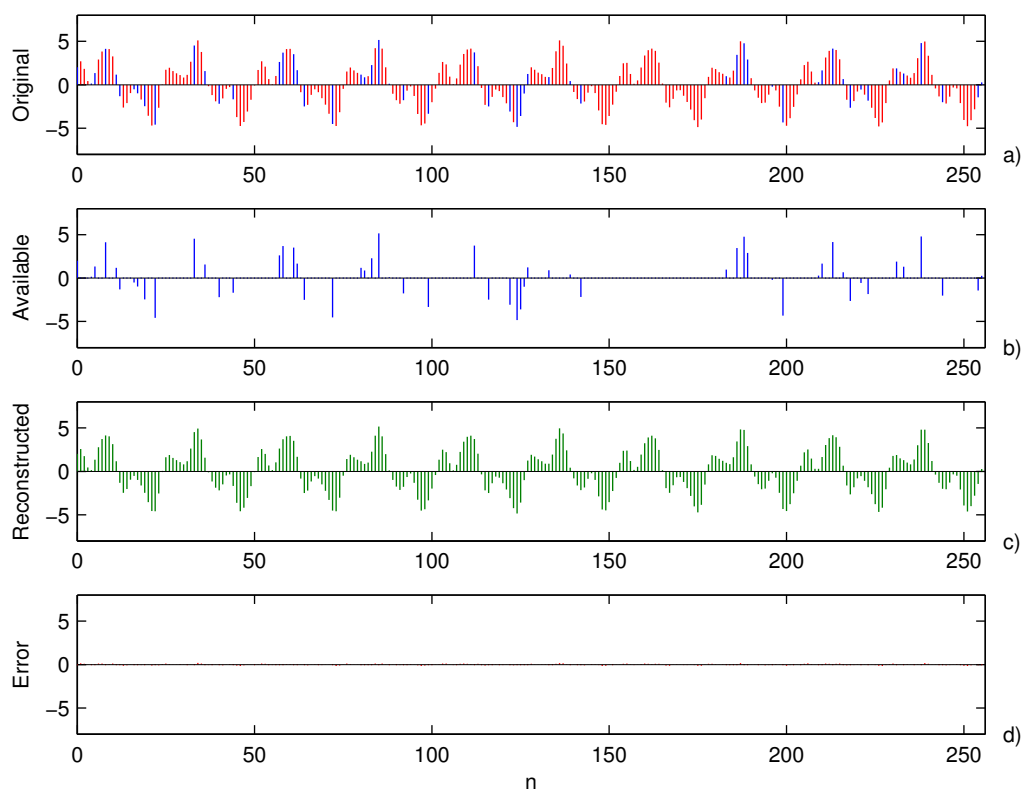
Fig. 4. Reconstruction example for signal with 200 missing samples at random positions. (a) original signal; (b) signal with missing samples set to 0 and used as an input to the reconstruction algorithm; (c) reconstructed signal; (d) reconstruction error.
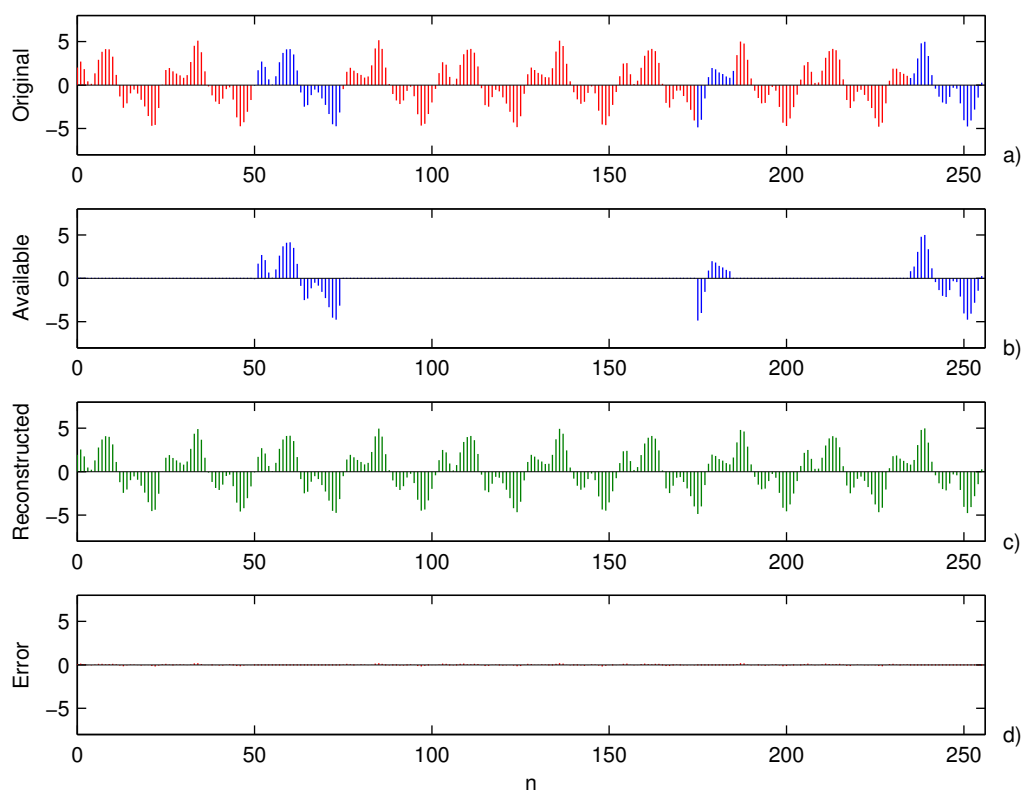

Fig. 5. Reconstruction example for a signal with 200 missing samples grouped into three blocks. (a) original signal; (b) signal with missing samples set to 0 and used as an input to the reconstruction algorithm; (c) reconstructed signal; (d) reconstruction error.
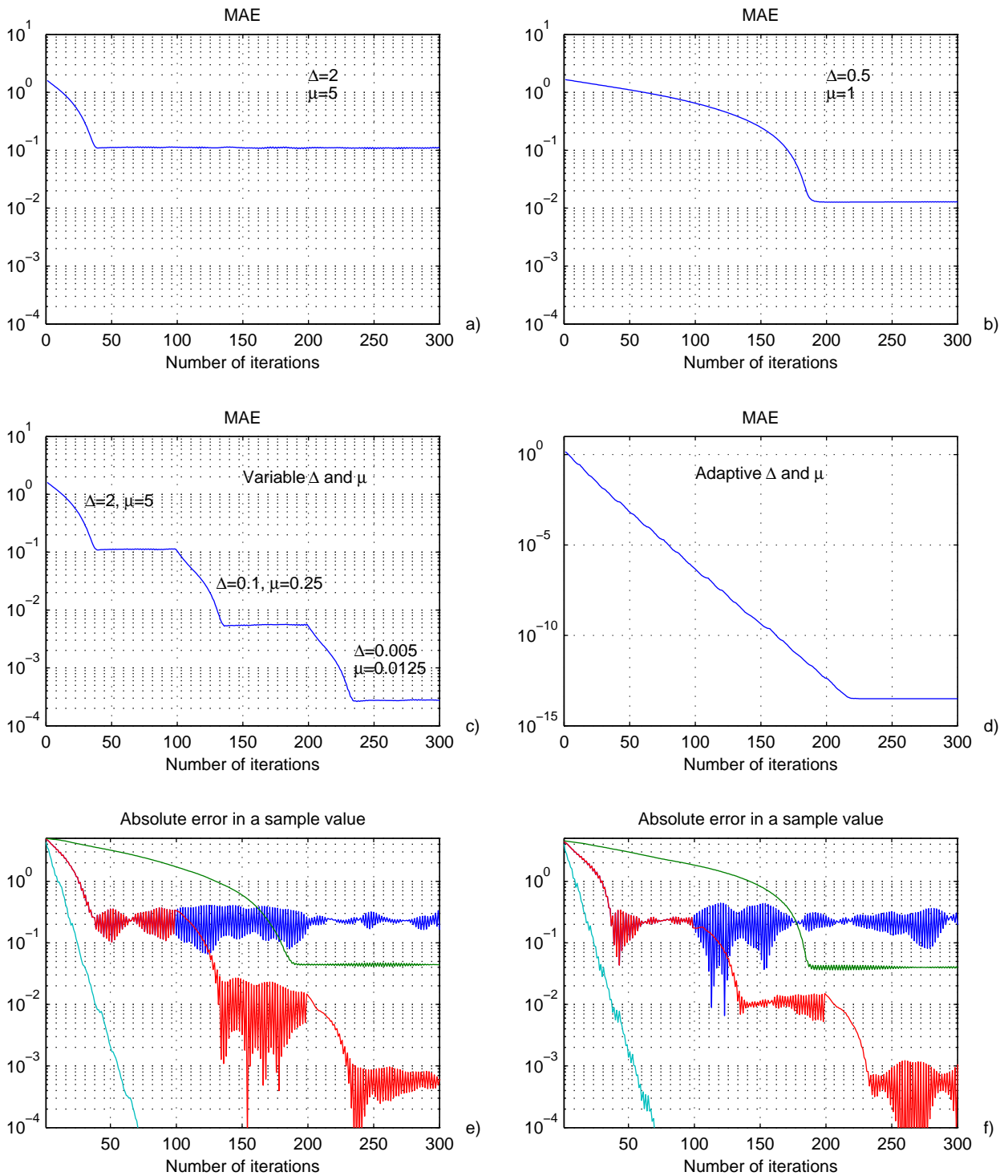
Fig. 6.   Mean absolute error (MAE) for constant algorithm parameters Δ and μ (a) and (b); variable parameters and adaptive parameters (c) and (d); absolute errors for two randomly chosen missing signal samples (e) and (f) for constant algorithm parameters (green and blue line), variable (red line), and adaptive parameters (bright blue).
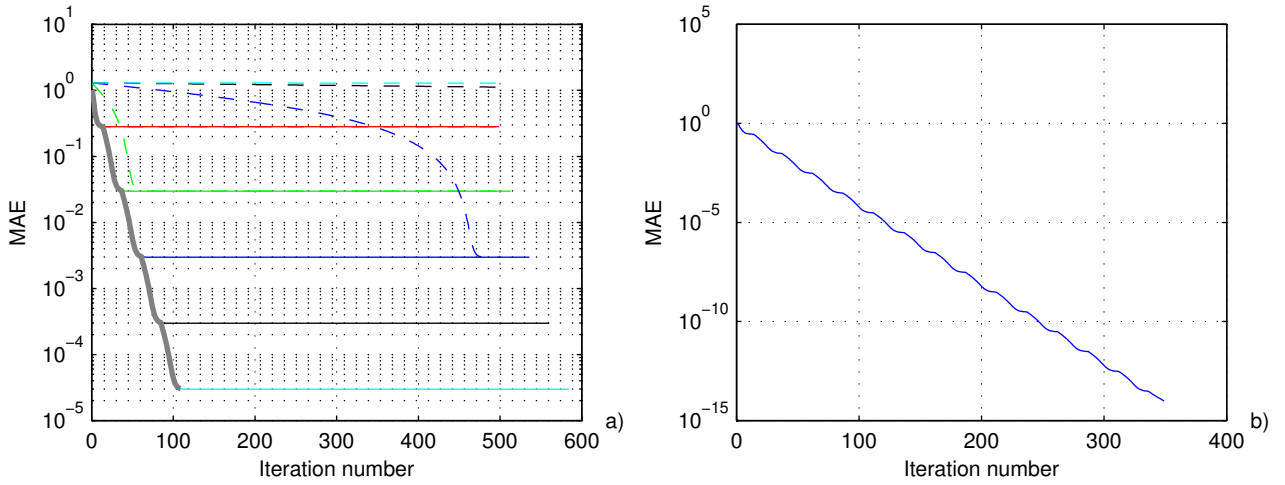
Fig. 7.   MAE for adaptive parameters using the measure-based criterion. Each color on graphics a) presents one set of parameters Δ and μ. Gray line presents the MAE with adaptive parameters. In b) the MAE is presented for the case when the algorithm parameters adaptation is done up to the computer precision.

Note that from the behavior of dashed lines (constant parameters) we can conclude that they achieved their stationary state in 10 times more iterations than by using the previous larger parameters. The green dashed line achieved stationary MAE in about 50 iterations, blue in about 500 iterations. So we can expect that the dashed gray line will achieve its stationary MAE in about 5000 iterations, and so on.

We may conclude that the MAE would achieve 10^{-15} (what is the standard computer double precision error) in about 10^{14} iterations with corresponding constant parameters. Of course this is not acceptable in practical calculations. As we can see, the same order of MAE is achieved by the presented adaptive algorithm in a relatively small number of iterations (about 350).

Reconstruction of Approximately Sparse Signals: Consider a signal:

$$x(n) = 3\sin(11.2\pi n/N) + 5\sin(50.6\pi n/N)$$
$$+ 3\cos(160.8\pi n/N). \quad (9)$$

whose frequencies do not match frequency grid in the DFT. By definition this signal is not sparse in the common DFT domain. We will apply the presented gradient algorithm with $\Delta = 3$ and $\mu = 4$ on this signal, when 70 signal samples are missing. Although the analyzed signal is not sparse in a strict sense, satisfactory reconstruction results are obtained. Fig. 8(a), (b) and (c) present the original signal, the available signal samples, and the reconstructed signal, respectively. In Fig. 8(d), (e) and (f) the DFT coefficients of the original signal, the available samples, and the reconstructed signal are shown, respectively. As we can see, although the original

signal is not sparse (its frequencies do not match to the frequency grid), the reconstruction is good.

Noisy Signals: The proposed algorithm is used for analysis of a noisy signal. It has been assumed that a sparse signal is corrupted by an additive Gaussian noise. From the reconstruction based on a limited number of samples we can come to a conclusion that in the case of sparse noisy signals certain improvement can be achieved if a number of signal samples are purposely omitted and the reconstruction is performed. Reconstruction for 200 omitted samples is presented in Fig. 9.

From Fig. 10 we can see that the SNR improvement higher than 2dB is achieved if we randomly omit 150 out of 256 signal samples and perform the reconstruction. Improvement could be significantly higher if we were able to selectively remove the most damaged samples [5].

Varying Concentration Measure: The number of iterations for required accuracy can be further improved by varying measure parameter p. Measures for $p < 1$ are more suitable to gradient based reconstruction. However measures for $p < 1$ do not converge to the true values of missing samples. A possible solution is to use measures with p slightly lower than 1 at the beginning of iterative algorithm and to switch to $p = 1$ afterwards. Figure 11 illustrate the case when $p = 0.9$, $\Delta = 1$ and $\mu = 10$ is used for iterations 1 to 12, $p = 0.95$, $\Delta = 2$ and $\mu = 4$ is used for iterations 13 to 22 and finally $p = 1$, $\Delta = 1$ and $\mu = 2$ is used for iterations form 23 to 100. The case with constant parameters $p = 1$, $\Delta = 1$ and $\mu = 2$ is presented in the same figure.
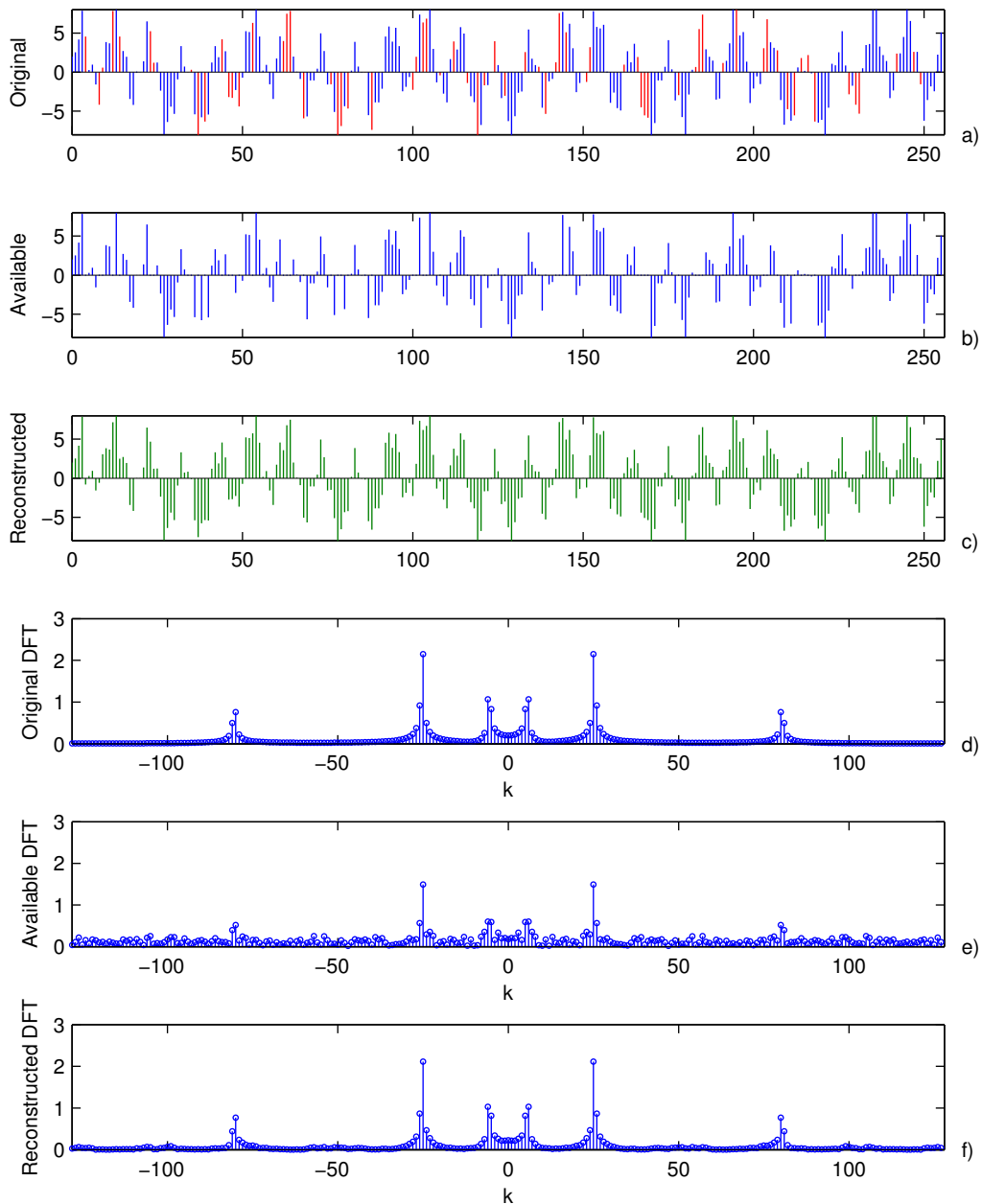
Fig. 8. Reconstruction of an approximately sparse signal whose frequencies do not match the frequency grid. a) original signal, b) available signal, c) reconstructed signal, d) DFT of the original signal, e) DFT of the available signal, f) DFT of the reconstructed signal.

V. CONCLUSION

In this paper we have presented an algorithm for unavailable/missing samples reconstruction in the sparse signals. The algorithm is based on the concentration measures used to quantify the signal sparsity. Since the commonly used measures are not differentiable around the optimal point, a criterion for variable algorithm parameters is introduced. The presented, gradient-based, adaptive step size algorithm is able to achieve the computer precision accuracy in a simple and numerically efficient way. The algorithm is applied to sparse signals, including the signals that are only approximately sparse. An example with a noisy signal is considered, demonstrating that some improvements in SNR may be achieved by omitting a number of samples. This algorithm can be applied on any concentration/sparsity measure form. A simple example on this topic is also presented.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.
[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete
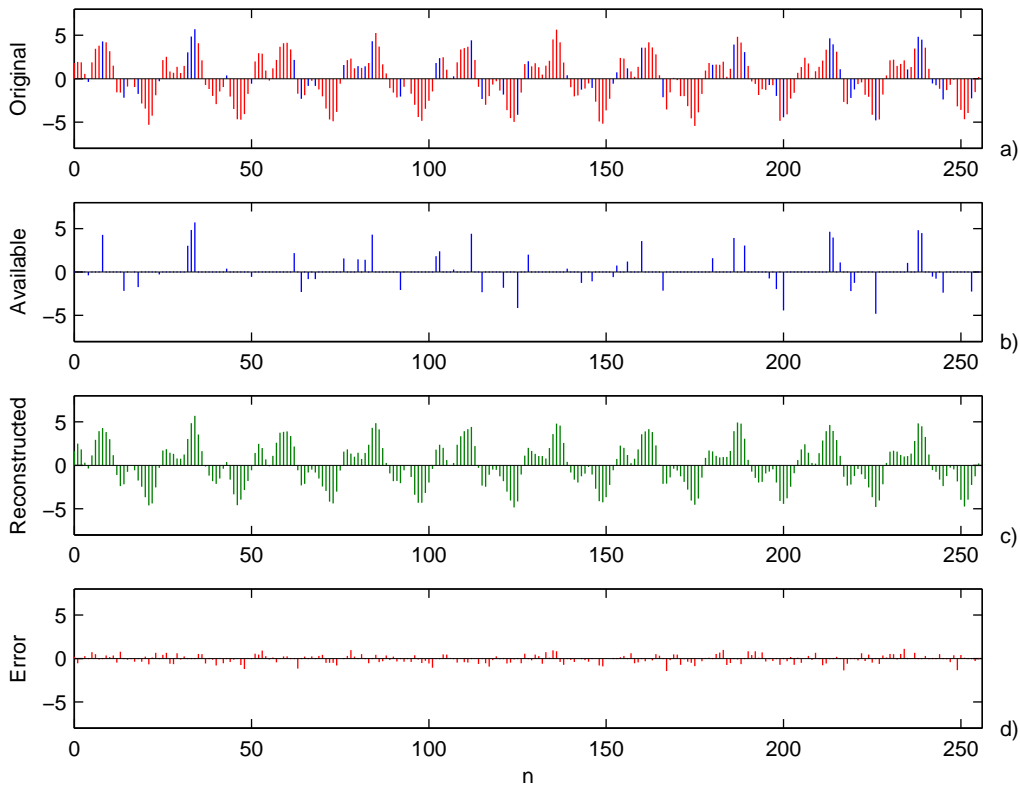
Fig. 9. Reconstruction example for a noisy signal with 200 omitted samples: (a) original signal; (b) signal with omitted samples set to 0 and used as an input to the reconstruction algorithm; (c) reconstructed signal; (d) reconstruction error (residual noise).
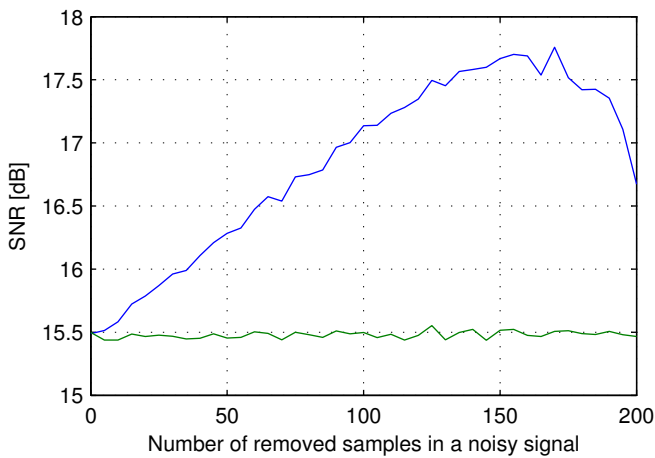

Fig. 10.   SNR of the reconstructed signal as a function of number of omitted samples (blue line). SNR of the original signal (green line).
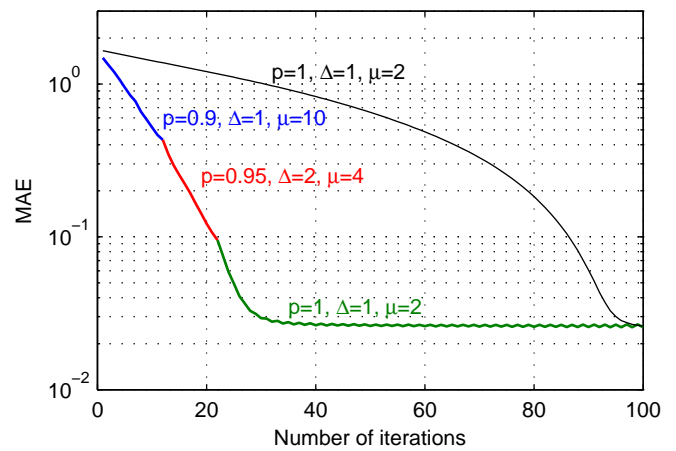

Fig. 11.   Reconstruction mean absolute error for the constant and varying measure and other algorithm parameters.

frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[3] S. Mallat, *A wavelet tour of signal processing*, Academic press, 1999.

[4] I. Djurović, L. Stanković and J. F. Böhme, "Robust L-estimation based forms of signal transforms and time-frequency representations", *IEEE Trans. on SP*, vol. 51, no. 7, pp. 1753-1761, 2003.

[5] L. Stanković, I. Orović, S. Stanković and M. G. Amin, "Robust Time-Frequency Analysis based on the L-estimation and Com-

pressive Sensing," *IEEE Signal Processing Letters*, May 2013, pp.499–502.

[6] L. Stanković, M. Daković and T. Thayaparan, *Time–Frequency Signal Analysis with Application*, Artech House, 2013

[7] M. A. T. Figueiredo, R. D. Nowak and S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," *IEEE Journal on Selected Topics in Signal Processing*, 2007

[8] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.

[9] S. J. Wright. "Implementing proximal point methods for linear programming," *Journal of Optimization Theory and Applications*, vol. 65, pp. 531–554, 1990.

[10] T. Serafini, G. Zanghirati and L. Zanni, "Gradient projection methods for large quadratic programs and applications in training support vector machines," *Optimization Methods and Software*, vol. 20, no. 2–3, pp. 353–378, 2004

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

[12] J. More and G. Toraldo, "On the solution of large quadratic programming problems with bound constraints," *SIAM Journal on Optimization*, vol. 1, pp. 93–113, 1991.

[13] G. Davis, S. Mallat and M. Avellaneda, "Greedy adaptive approximation," *Journal of Constructive Approximation*, vol. 12, pp. 57–98, 1997.

[14] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. on Information Theory*, vol. 52, pp. 6–18, 2006.

[15] B. Turlach, "On algorithms for solving least squares problems under an L1 penalty or an L1 constraint," *Proc. of the American Statistical Association; Statistical Computing Section*, pp. 2572–2577, Alexandria, VA, 2005.

[16] D. Angelosante, G. B. Giannakis and E. Grossi, "Compressed sensign of time-varying signals," *Int. Conf. on DSP*, 2009, pp.1–8.

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

[18] P. Flandrin and P. Borgnat, "Time-Frequency Energy Distributions Meet Compressed Sensing," *IEEE Trans. on Signal Processing*, vol. 58, no. 6, 2010, pp. 2974–2982.

[19] L. Stanković, S. Stanković and M. G. Amin, "Missing Samples Analysis in Signals for Applications to L-Estimation and Compressive Sensing", *Signal Processing*, Elsevier, Volume 94, Jan. 2014, Pages 401–408.

[20] F. Ahmad and M.G. Amin, "Through-the-wall human motion indication using sparsity-driven change detection," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 50, no. 12, 2012.

[21] S. Stanković, I. Orović and E. Sejdić, *Multimedia signals and Systems*, Springer, 2012.

[22] Y. Yoon and M. G. Amin, "Compressed sensing technique for high-resolution radar imaging," *Proc. SPIE*, vol. 6968, 2008, pp. 6968A–69681A–10.

[23] E. Sejdić, A. Cam, L. F. Chaparro, C. M. Steele and T. Chau, "Compressive sampling of swallowing accelerometry signals using TF dictionaries based on modulated discrete prolate spheroidal sequences," *EURASIP Journal on Advances in Signal Processing*, 2012:101 doi:10.1186/1687–6180–2012–101

[24] L. Stanković, "A measure of some time–frequency distributions concentration," *Signal Processing*, vol. 81, pp. 621–631, 2001

[25] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.

[26] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[27] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,"*Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 586–597, 2007.

[28] L. Stanković, M. Daković and S. Vujović "Concentration Measures with an Adaptive Algorithm for Processing Sparse Signals," in *Proceedings of ISPA 2013*, Sept. 4–6, 2013, Trieste, Italy, pp. 418–423