On a Gradient-Based Algorithm for Sparse Signal Reconstruction in the Signal/Measurements Domain

Ljubiša Stanković, Miloš Daković

University of Montenegro, Podgorica, ljubisa@ac.me, milos@ac.me

Abstract

Sparse signals can be recovered from a reduced set of samples by using compressive sensing algorithms. In common compressive sensing methods the signal is recovered in the sparsity domain. A method for the reconstruction of sparse signals that reconstructs the missing/unavailable samples/measurements is recently proposed. This method can be efficiently used in signal processing applications where a complete set of signal samples exists. The missing samples are considered as the minimization variables, while the available samples are fixed. Reconstruction of the unavailable signal samples/measurements is preformed using a gradient-based algorithm in the time domain, with an adaptive step. Performance of this algorithm with respect to the step size and convergence are analyzed and a criterion for the step-size adaptation is proposed in this paper. The step adaptation is based on the gradient direction angles. Illustrative examples and statistical study are presented. Computational efficiency of this algorithm is compared with other two commonly used gradient algorithms that reconstruct signal in the sparsity domain. Uniqueness of the recovered signal is checked using a recently introduced theorem. The algorithm application to the reconstruction of highly corrupted images is presented as well.

1. Introduction

A signal is sparse in a transformation domain if the number of nonzero coefficients is much lower than the number of signal samples. For linear signal transforms the signal samples can be considered as linear combinations (measurements) of the signal transform coefficients. Sparse signals can be fully recovered from a reduced set of samples/measurements. Compressive sensing theory is dealing with sparse signal reconstruction [1]-[22]. In common signal processing problems, for signals sparse in a transformation domain, signal samples can be considered as observations since they are linear combinations of the transformation domain coefficients. The goal of compressive sensing is to reduce the set of signal measurements while preserving complete information about the whole signal. In some applications unavailable signal samples result from the system physical constraints including unavailability to perform all signal measurements. Signals may also contain arbitrarily positioned samples that are heavily corrupted. In these cases it is better to remove such samples and consider them as unavailable in the signal recovery [6]. Compressive sensing theory based reconstruction algorithms may be also used when the unavailable signal samples are not a result of an intentional strategy to compress the data. Note that one of the initial compressive sensing theory successes in applications (computed tomography reconstruction) was not related to the intentional compressive sensing strategy but to the physical problem constraints, restricting the set and positions of available data.

Various algorithms for reconstruction of sparse signals from a reduced set of observations are introduced [8]-[17]. Two the most important approaches are based either on the gradient of the sparsity measure [8] or the orthogonal matching pursuit methods [16].

The topic of this paper is the reconstruction of signals with some arbitrary positioned unavailable samples. These samples may result from a compressive sensing strategy or from their unavailability due to various reasons. A method for the unavailable signal samples reconstruction, considering them as variables, has been proposed in [22]. In contrast to the other common methods that recover sparse signals in their sparsity

domain this method reconstructs missing samples/measurements to make the set of samples/measurements complete. The available samples are fixed (invariant) during the reconstruction process. An analysis of the algorithm performance is presented in this paper. A relation between reconstructed signal bias and the algorithm step-size is derived and statistically confirmed. A new criterion for the algorithm step-size adaptation is proposed. The improved version of this algorithm is compared with two other common gradient based algorithms proving its accuracy and computational efficiency. The discrete Fourier transform domain is used as a case study, although the algorithm application is not restricted to this transform [23]. The solution uniqueness is checked after signal reconstruction by using the recently proposed theorem [25].

The paper is organized as follows. After the definitions in the next section, the adaptive gradient algorithm presented. In the comments to the algorithm the step size, and bias are considered and illustrated. A new criterion for the algorithm parameter adaptation is proposed in Section 3 as well. Reconstruction is illustrated on examples and uniqueness of the obtained solution is analyzed. Statistical performance are checked in Section 4. An illustration of the algorithm application to the image reconstruction is given as well.

2. Gradient-based Reconstruction of Missing Samples

Discrete-time signal $x(n)$ with N samples will be considered (with vector notation $\mathbf{x} = [x(0)\ x(1)\ ...\ x(N-1)]^T$). Its transform coefficients will be denoted by $X(k) = T[x(n)]$. If the number of nonzero coefficients $X(k) \neq 0$ for $k \in \{k_1, k_2, ..., k_s\}$ is such that $s \ll N$ then the signal $x(n)$ is sparse in this transformation domain. Next assume that only $M < N$ signal samples are available in the time domain at the discrete-time positions

$$n_i \in \mathbb{N}_A = \{n_1, n_2, ..., n_M\} \subset \mathbb{N} = \{0, 1, 2, ..., N-1\}. \tag{1}$$

Under some conditions, studied within the compressive sensing theory [1]-[3], full signal reconstruction is possible based on the reduced set of available samples if the signal is sparse in a transformation domain. In the signal theory this problem can be formulated as a full recovery of the unavailable samples as well. In general, a reconstruction algorithm is based on the sparsity measure minimization. Nonzero transform coefficients counting is the simplest sparsity measure. Counting can be achieved by the mathematical form $\|\mathbf{X}\|_0$, referred to as the "ℓ_0-norm", [1], [20]. The signal reconstruction problem statement is then

$$\min \|\mathbf{X}\|_0 \ \text{subject to} \ \mathbf{y} = \mathbf{A}\mathbf{X}. \tag{2}$$

Vector $\mathbf{y} = [x(n_1)\ x(n_2)\ ...\ x(n_M)]^T$ contains the available signal samples, while the transform vector $\mathbf{X} = [X(1)\ X(2)\ ...\ X(N)]^T$ elements are the unknown transform coefficients. Matrix $\mathbf{A}$ is the measurement matix. In common signal transforms it corresponds to the inverse transformation matrix where the rows corresponding to the unavailable samples are omitted. If $\mathbf{x} = \mathbf{\Psi}\mathbf{X}$ then the matrix $\mathbf{A}$ is a submatrix of $\mathbf{\Psi}$ containing its rows corresponding to the discrete-time instants (1).

Although there are some approaches to reconstruct the signal using the ℓ_0-norm based formulation [26] in principle, this is an NP-hard combinatorial optimization problem. It is the reason why the closest convex ℓ_1-norm of the signal transform is used instead of the "ℓ_0-norm" in the minimization process,

$$\min \|\mathbf{X}\|_1 \ \text{subject to} \ \mathbf{y} = \mathbf{A}\mathbf{X}, \tag{3}$$

where $\|\mathbf{X}\|_1 = \sum_{k=0}^{N-1} |X(k)|$. Under the conditions defined using the restricted isometry property (RIP), [3], [4], minimization of the ℓ_1-norm can produce the same result as (2), [1, 22, 24]. Common gradient-based algorithms reformulate the problem defined by (3) into a form suitable for application of standard minimization tools. Minimization is done in the sparsity domain. One such a method is briefly reviewed in the Appendix.

2.1. Algorithm

A gradient-based algorithm that minimizes the sparsity measure by varying the missing sample values is presented next. In this gradient-based minimization approach the missing samples are considered as

variables [22]. The missing sample values are varied in an iterative way to produce lower sparsity measure values, and finally to approach to the minimum of the convex ℓ_1-norm based sparsity measure (3) with an acceptable accuracy. Note that if the reconstruction conditions are met for the ℓ_1-norm [3] then the ℓ_1-norm minimum position will correspond to the true values of the missing samples.

The signal

$$x_a^{(0)}(n) = \begin{cases} 0 & \text{for missing samples, } n \in \mathbb{N}_Q \\ x(n) & \text{for available samples, } n \in \mathbb{N}_A \end{cases} \tag{4}$$

will be used as the initial estimate. Note that this signal would follow as a result of the ℓ_2-norm based minimization of the signal transform. The set of missing sample positions is $\mathbb{N}_Q$. Values of the available samples are fixed and will be considered as constants. The missing sample values are varied through iterations. If we denote by $\mathbf{x}_a^{(m)}$ the vector of signal values reconstructed after m iterations, then the goal is to achieve

$$\min \|\mathbf{X}_a\|_1 \quad \text{subject to} \quad x_a^{(m)}(n){=}x(n) \text{ for } n \in \mathbb{N}_A,$$

where the components of $\mathbf{X}_a$ are $X_a(k) = \mathrm{T}[x_a^{(m)}(n)]$. In order to find the position of function $z = \|\mathbf{X}_a\|_1$ minimum, the relation for iterative missing samples (variables) calculation is defined by using the gradient-based descend on the sparsity measure $\|\mathbf{X}_a\|_1$ as

$$\mathbf{y}_c^{(m+1)} = \mathbf{y}_c^{(m)} - \alpha \left. \frac{\partial \|\mathbf{X}_a\|_1}{N \partial \mathbf{y}_c} \right|_{\mathbf{y}_c = \mathbf{y}_c^{(m)}}. \tag{5}$$

Vector $\mathbf{y}_c^{(m)}$ contains the minimization variables (missing signal sample values) in the mth iteration. The gradient vector coordinates $g(n_i) = \partial \|\mathbf{X}_a\|_1 / (\partial \mathbf{y}_c N)$ can be estimated, in the mth iteration, by using the finite differences of the sparsity measure. They are calculated for each missing sample at instants $n_i \in \mathbb{N}_Q$

$$g(n_i) = \frac{\|\mathbf{X}_a^+\|_1 - \|\mathbf{X}_a^-\|_1}{2\Delta N},$$

where

$$X_a^+(k) = \mathrm{T}\{x_a^+(n)\}$$
$$X_a^-(k) = \mathrm{T}\{x_a^-(n)\}$$

and

$$x_a^+(n) = x_a^{(m)}(n) + \Delta\delta(n - n_i)$$
$$x_a^-(n) = x_a^{(m)}(n) - \Delta\delta(n - n_i). \tag{6}$$

Again, for $n_i \in \mathbb{N}_A$ the available signal values are not changed, $g(n_i) = 0$. All gradient vector $\mathbf{g}^{(m)}$ values, including its zero and nonzero coordinates, in the m-th iteration are denoted by $g^{(m)}(n)$.

Through statistical study [22] it has been concluded that an appropriate step parameter value in (5) is related to the finite difference step as $\alpha = 2\Delta$.

3. Comments on the Algorithm

- The algorithm inputs are the signal length N, the set of available samples $\mathbb{N}_A$, and the available signal values $x(n_i)$, $n_i \in \mathbb{N}_A$.
- For the DFT instead of using signals defined by (6) and their transforms for each $n_i \in \mathbb{N}_Q$ it is possible

to calculate

$$\left|X_a^+(k)\right| = \left|X_a^{(m)}(k) + \Delta D_{n_i}(k)\right|$$

$$\left|X_a^-(k)\right| = \left|X_a^{(m)}(k) - \Delta D_{n_i}(k)\right|$$

with $X_a^{(m)}(k) = \mathrm{DFT}[x_a^{(m)}(n)]$ and $D_{n_i}(k) = \mathrm{DFT}[\delta(n - n_i)] = \exp(-j2\pi n_i k/N)$. Note that $D_{n_i}(k)$ values are not dependent on the signal and the iteration number m. They can be calculated only once. Similar simplification can be done for any linear signal transform $\mathrm{T}[\cdot]$.

3.1. Adaptive Step

The influence of step-size Δ to the solution (minimum position) precision is analyzed next. Assume that we have obtained the exact solution for the missing samples. The change of sparsity measure is tested by the change of signal sample $x(n_0)$ for $\pm\Delta$. For a signal of sparsity s in the DFT domain, whose form is $x(n) = \sum_{i=1}^{s} A_i e^{j2\pi n_0 k_i/N}$, the signals with the steps used for the gradient estimate calculation are $x_a^+(n) = x(n) + \Delta\delta(n - n_0)$ and $x_a^-(n) = x(n) - \Delta\delta(n - n_0)$. Their sparsity measures are

$$\left\|\mathbf{X}_a^+\right\|_1 = \sum_{i=1}^{s}\left|A_i + \Delta e^{-j2\pi n_0 k_i/N}\right| + (N - s)\Delta$$

$$\left\|\mathbf{X}_a^-\right\|_1 = \sum_{i=1}^{s}\left|A_i - \Delta e^{-j2\pi n_0 k_i/N}\right| + (N - s)\Delta.$$

In the worst case, amplitudes A_i are in phase with $e^{-j2\pi n_0 k_i/N}$ and $\Delta \leq |A_i|$ when

$$\left\|\mathbf{X}_a^+\right\|_1 = \sum_{i=1}^{s}|A_i| + s\Delta + (N - s)\Delta = \|\mathbf{X}\|_1 + N\Delta$$

$$\left\|\mathbf{X}_a^-\right\|_1 = \sum_{i=1}^{s}|A_i| - s\Delta + (N - s)\Delta = \|\mathbf{X}\|_1 + (N - 2s)\Delta.$$

It means $g(n_0) = \left(\|\mathbf{X}_a^+\|_1 - \|\mathbf{X}_a^-\|_1\right)/(2N\Delta) = s/N \neq 0$. This is an important fact leading to the conclusion that the stationary state of the iterative algorithm is biased. The algorithm moves the solution $x(n_0)$ to a biased value $x(n_0) - b$ in order to produce $g(n_0) = 0$ in the algorithm stationary point. For the zero-value of the gradient we have $x_a^+(n_0) = x(n_0) - b + \Delta$ and $x_a^-(n_0) = x(n) - b - \Delta$. By replacing Δ with $\Delta - b$ in $\|\mathbf{X}_a^+\|_1$ and Δ with $\Delta + b$ in $\|\mathbf{X}_a^-\|_1$ the worst case sparsity measures are $\|\mathbf{X}_a^+\|_1 = \|\mathbf{X}\|_1 + N(\Delta - b)$ and $\|\mathbf{X}_a^-\|_1 = \|\mathbf{X}\|_1 + (N - 2s)(\Delta + b)$. In this case the bias b value can be obtained from $\|\mathbf{X}_a^+\|_1 = \|\mathbf{X}_a^-\|_1$ as

$$N(\Delta - b) = (N - 2s)(\Delta + b)$$

$$b = \frac{s}{N - s}\Delta \cong \frac{s}{N}\Delta \text{ for } s \ll N. \tag{7}$$

The bias limit is proportional to the step-size Δ. Obviously, the bias limit can be reduced by using small values of the step Δ. Calculation with a very small step Δ would be inefficient and time consuming with an extremely large number of iterations. An efficient way to use the gradient-based algorithm is in adapting its step-size. In the initial iteration the step-size of an order of signal amplitude is used

$$\Delta = \max_{n \in \mathbb{N}_A} |x(n)|. \tag{8}$$

When the algorithm reaches a stationary point with this step Δ then the bias will dominate and the mean squared error will be almost constant. The algorithm can improve the solution precision by reducing the step-size.

4

This simple relation between the bias and step-size has been confirmed through statistical analysis on more complex cases, with a large number of missing samples.

3.2. Adaptation Criterion

The gradient-based algorithm convergence improvement by changing steps is analyzed in details in [22]. A criterion that can efficiently detect the event that the algorithm has reached the stationary point with regard to the mean square error (the vicinity of the sparsity measure minimum defined by the bias) described in the previous subsection, is proposed next. The presented criterion is based on the direction change of the gradient vector. When the vicinity of the minimum sparsity measure point is reached within the region defined by the bias value, then the gradient estimate oscillates and changes its direction for almost 180 degrees. The angle between two successive gradient vector estimations $\mathbf{g}^{(m-1)}$ and $\mathbf{g}^{(m)}$, denoted β_m can be calculated as

$$\beta_m = \arccos \frac{\sum_{n=0}^{N-1} g^{(m-1)}(n) g^{(m)}(n)}{\sqrt{\sum_{n=0}^{N-1} \left(g^{(m-1)}(n)\right)^2} \sqrt{\sum_{n=0}^{N-1} \left(g^{(m)}(n)\right)^2}}.$$

If this angle β_m is, for example, above $170°$ then we can conclude that the signal missing samples (variables) reached oscillatory nature around the minimal sparsity measure position. When these values of the angles are detected the algorithm step Δ is reduced, for example, for $\Delta/10 \to \Delta$ or $\Delta/\sqrt{10} \to \Delta$. The iterations are continued starting from the reconstructed signal values reached in the previous step.

When the minimum of the sparsity measure is reached with a sufficiently small Δ, the value of Δ can also be used as an indicator of the solution precision. For example, if the signal amplitudes are of order of 1 and $s/N = 0.1$ taking $\Delta = 1$ in the first iteration will produce the solution with a precision better than 20 [dB]. Then, if the step Δ is reduced to $\Delta = 0.1$ a precision better than 40 [dB] will be obtained, and so on. Value of Δ can be used as the algorithm stopping criterion.

3.3. Stopping Criterion

The precision of the result in iterative algorithms is commonly estimated based on the change of the result values in the last iterations. Therefore an average of changes in a large number of variables is a good estimate of the achieved precision. The value of

$$T_r = 10 \log_{10} \frac{\sum_{n \in \mathbb{N}_Q} |x_a^p(n) - x_a^{(m)}(n)|^2}{\sum_{n \in \mathbb{N}_Q} |x_a^{(m)}(n)|^2}.$$

can be used as a rough estimate of the reconstruction error to signal ratio. In this relation $x_a^p(n)$ is the signal reconstructed before Δ reduction and $x_a^{(m)}(n)$ is the reconstructed signal after iterations done with the reduced step. Value of T_r can be used as the algorithm stopping criterion. If the value of T_r is above the required precision T_{max} (for example, if $T_r > -100 dB$), the calculation procedure should be continued with reduced values of step Δ.

3.4. Large Step Influence

A possible divergence of a gradient-based algorithm is related to large steps $\alpha = 2\Delta$. We can write

$$\left| X_a^+(k) \right| - \left| X_a^-(k) \right| = \left| X_a^{(m)}(k) + \Delta D_{n_i}(k) \right| - \left| X_a^{(m)}(k) - \Delta D_{n_i}(k) \right|$$

$$= \Delta \left| D_{n_i}(k) \right| \left(\left| 1 + \frac{X_a^{(m)}(k)}{\Delta D_{n_i}(k)} \right| - \left| 1 - \frac{X_a^{(m)}(k)}{\Delta D_{n_i}(k)} \right| \right).$$

In general for a complex number $a = X_a^{(m)}(k)/(\Delta D_{n_i}(k))$, with $|a| \ll 1$ and a large Δ, from the problem geometry we can show that the following bounds hold $0 \le ||1 + a| - |1 - a|| \le 2 |a|$. Therefore,

$$0 \le \left| \left| X_a^+(k) \right| - \left| X_a^-(k) \right| \right| \le 2 \left| X_a^{(m)}(k) \right|.$$

Lower limit 0 is obtained if a is imaginary-valued, while the upper limit $2\left|X_a^{(m)}(k)\right|$ follows if a is real-valued. If Δ is large the missing signal values will be adapted for a value independent on Δ. The missing sample values will oscillate within the range of the original signal values until the step Δ is reduced in iterations. The variable values will be arbitrary changed within the signal amplitude order as far as Δ is too large. It will not influence further convergence of the algorithm, when the step Δ assumes appropriately small values.

A pseudo code of this algorithm is presented in Algorithm 1.

Illustrative Example 1:

The effects analyzed within the comments on the algorithm will be illustrated on a simple signal

$$x(n) = 3\cos\left(2\pi\frac{n}{N} - \frac{\pi}{8}\right) - 0.2\sin\left(4\pi\frac{n}{N} + \frac{\pi}{2}\right)$$

with $N = 8$ and sparsity $s = 4$. Assume that the samples at positions $n \in \mathbb{N}_Q = \{1, 6\}$ are missing. Reconstruction of the signal is done in 60 iterations. The initial value of the step $\Delta = 1$ and the initial values of missing samples $x(1) = 0$ and $x(6) = 0$ are used. The missing sample values in the first 20 iterations are given in Fig.1(top) by dots (connected by a line). After 6 iterations the algorithm does not significantly change the missing sample values with given step size $\Delta = 1$ (lower subplot within the figure presents changes of variables in a highly zoomed scale). When the stationary point vicinity is reached with $\Delta = 1$ the gradient estimate is almost zero-valued. Its direction change for almost 180 degrees. The measure values are on the contour with approximately the same value (circles on contour plot). The algorithm resumes its fast approach toward the exact missing sample values after the algorithm step is reduced to $\Delta = 0.1$ in the 20th iteration. When a new stationary state is reached change of Δ to $\Delta = 0.01$ is done and the approach to the correct missing sample values is continued.

The bias upper limit in the stationary state for step $\Delta = 1$, according to (7), is $\frac{s}{N-s}\Delta = 1$. After each reduction of the step size from Δ to $\Delta/10$ the bias caused MSE will be reduced for almost 20 [dB]. The result for the signal reconstruction and the obtained MSE for the calculated missing values $x(1)$ and $x(6)$ are presented in Fig.1.

Illustrative Example 2:

The reconstruction process is repeated using the signal

$$x(n) = 3\sin\left(20\pi\frac{n}{N}\right) + \cos\left(60\pi\frac{n}{N}\right) + 0.7\sin\left(46\pi\frac{n}{N}\right)$$

and $N = 32$. Its sparsity is $s = 6$. Missing sample positions are at $n \in \mathbb{N}_Q = \{0, 1, 2, 5, 7, 8, 11, 15, 18, 27, 28, 30\}$. The reconstruction results for this signal are presented in Fig.2. After the signal is reconstructed the problem is to establish if this reconstruction is unique in the sense that there is no other signal (missing sample values) with the same or lower sparsity. A very simple check of the solution uniqueness, using the missing sample positions and the positions of the reconstructed signal non-zero components in the sparsity domain, is given in [25] by the following:

Theorem *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 $q_m \in \mathbb{N}_Q$. Also assume that the reconstruction is performed and that the DFT of reconstructed signal is of sparsity s. Assume that the positions of the reconstructed nonzero values in the DFT are $k_{0i} \in \mathbb{K}_s = \{k_{01}, k_{02}, ..., k_{0s}\}$. Reconstruction result is unique if the inequality*

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

Algorithm 1 Reconstruction

Require:

- Set of missing/omitted sample positions $\mathbb{N}_Q$

- Available samples $x(n)$, $n \notin \mathbb{N}_Q$

1: Set $x_a^{(0)}(n) \leftarrow x(n)$ $\triangleright$ for $n \notin \mathbb{N}_Q$

2: Set $x_a^{(0)}(n) \leftarrow 0$ $\triangleright$ for $n \in \mathbb{N}_Q$

3: Set $m \leftarrow 0$

4: Set $\Delta \leftarrow \max |x_a^{(0)}(n)|$

5: **repeat**

6:    Set $x_a^p(n) = x_a^{(m)}(n)$ $\triangleright$ for each n

7:    **repeat**

8:      $m \leftarrow m + 1$

9:      **for** $n_i \leftarrow 0$ to $N - 1$ **do**

10:        **if** $n_i \in \mathbb{N}_Q$ **then**

11:          $X_a^+(k) \leftarrow \text{DFT}\{x_a^{(m)}(n) + \Delta\delta(n - n_i)\}$

12:          $X_a^-(k) \leftarrow \text{DFT}\{x_a^{(m)}(n) - \Delta\delta(n - n_i)\}$

13:          $g^{(m)}(n_i) \leftarrow \dfrac{1}{N} \displaystyle\sum_{k=0}^{N-1} |X_a^+(k)| - |X_a^-(k)|$

14:        **else**

15:          $g^{(m)}(n_i) \leftarrow 0$

16:        **end if**

17:        $x_a^{(m+1)}(n_i) \leftarrow x_a^{(m)}(n_i) - g^{(m)}(n_i)$

18:      **end for**

19:      $\beta_m = \arccos \dfrac{\sum_{n=0}^{N-1} g^{(m-1)}(n)g^{(m)}(n)}{\sqrt{\sum_{n=0}^{N-1} \left(g^{(m-1)}(n)\right)^2}\sqrt{\sum_{n=0}^{N-1} \left(g^{(m)}(n)\right)^2}}$

20:    **until** $\beta_m < 170°$

21:    $\Delta \leftarrow \Delta/\sqrt{10}$

22:    $T_r = 10 \log_{10} \dfrac{\sum_{n \in \mathbb{N}_Q} |x_a^p(n) - x_a^{(m)}(n)|^2}{\sum_{n \in \mathbb{N}_Q} |x_a^{(m)}(n)|^2}.$

23: **until** $T_r < T_{max}$

24: **return** $x_a^{(m)}(n)$

Output:
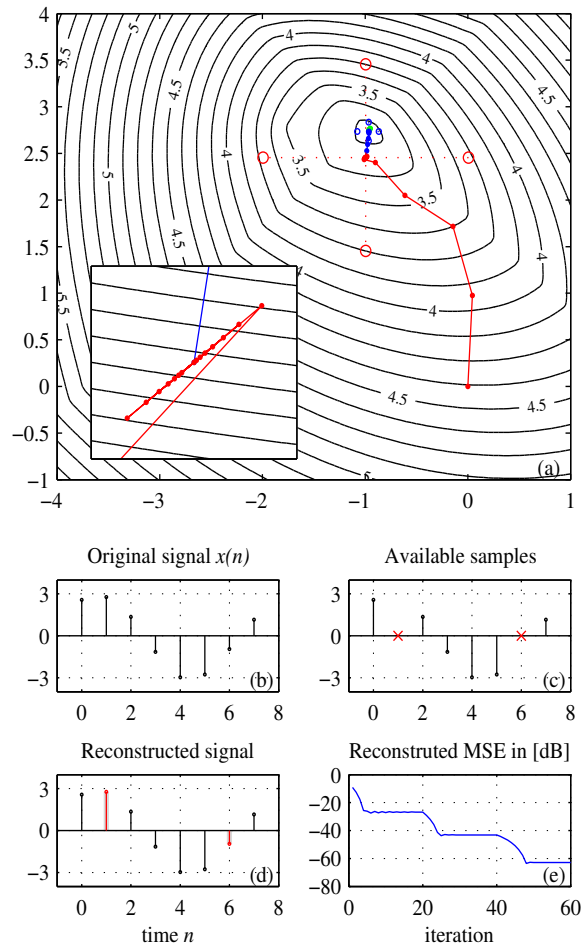
- Reconstructed signal $x_R(n) = x_a^{(m)}(n)$

Figure 1: a) Illustration of a signal reconstruction using variable step gradient algorithm. b) Original signal. c) Signal with two missing samples. d) Reconstructed signal. e) Reconstructed signal MSE as a function of the iteration number.

8

holds. Integers Q_{2^h} and $S_{2^{r-h}}$ are calculated as

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

$$S_{2^{r-h}} = \sum_{l=1}^{Q_{2^h}-1} P_h(l)$$

$$P_h(l) = \underset{b=0,1,\dots,2^{r-h}-1}{\text{sort}} \{\text{card}\{k : k \in \mathbb{K}_s \ and \ \text{mod}(k, 2^{r-h}) = b\}\}$$

where $P_h(1) \leq P_h(2) \leq \dots \leq P_h(2^{r-h})$.
  Signal independent uniqueness corresponds to the worst case signal form, when $S_{2^{r-h}} = 0$.

The answer is obtained almost immediately, since the computational complexity of the Theorem is of order $O(N)$. The proof is given in [25].

For the considered set of missing samples $n \in \mathbb{N}_Q = \{0, 1, 2, 5, 7, 8, 11, 15, 18, 27, 28, 30\}$, from the previous example, we easily get:

$$Q_{2^0} = 8, \quad Q_{2^1} = \max\{6, 6\} = 6, \quad Q_{2^2} = \max\{3, 2, 3, 4\} = 4,$$
$$Q_{2^3} = \max\{2, 1, 2, 2, 1, 1, 1, 2\} = 4, \quad Q_{2^4} = 2.$$

For signal independent uniqueness

$$s < 32 - \max\{1 \times (8-1), 2 \times (2-1), 4 \times (4-1), 8 \times (2-1), 16 \times (2-1)\} - s = 16 - s.$$

It means that $s < 8$, with considered $\mathbb{N}_Q$, guarantees uniqueness of the reconstructed signal for any signal parameters. When a sparse signal is reconstructed, for some sets of $\mathbb{K}_s$ uniqueness may be guarantied for higher values of sparsity s. Additional details and examples may be found in [25].

4. Statistical Analysis and Comparison

Consider a general real-valued form of a signal sparse in the DFT domain

$$x(n) = \sum_{i=1}^{K} A_i \cos(2\pi n k_i / N + \varphi_i). \tag{9}$$

The total number of signal samples is N. The signal sparsity $s = 2K$ is varied from $s = 2$ to $s = N/2$. The signal amplitudes A_i, frequencies k_i, and phases φ_i are random variables. Signal amplitudes A_i are Gaussian random variables with unity variance, the frequency indices k_i are random numbers within $1 \leq k_i \leq N - 1$. The phases are uniform random variables $0 \leq \varphi_i \leq 2\pi$, in each signal realization. Signal is reconstructed in 100 different realizations for each sparsity s and random positions of missing samples. The statistical calculations are performed for $N = 128$. The reconstructed signals are denoted by $x_R(n)$. The reconstruction results are presented in Fig.3(a). The signal-to-reconstruction-error ratio (SRR) is given in [dB]

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

Bright colors indicate the region where the algorithm had fully recovered missing samples (compared to the original samples) in all realizations, while dark colors indicate the region where the algorithm could not recover missing samples in any realization. In the transition region for M slightly greater than $2s$ we have both the cases when the signal recovery is not achieved and the cases of full signal recovery. A stopping criterion for the accuracy of 120 [dB] is used. It corresponds to a precision in the recovered signal of the same order as in input samples, if they are acquired by a 20-bit A/D converter. The case with $N = 64$ is
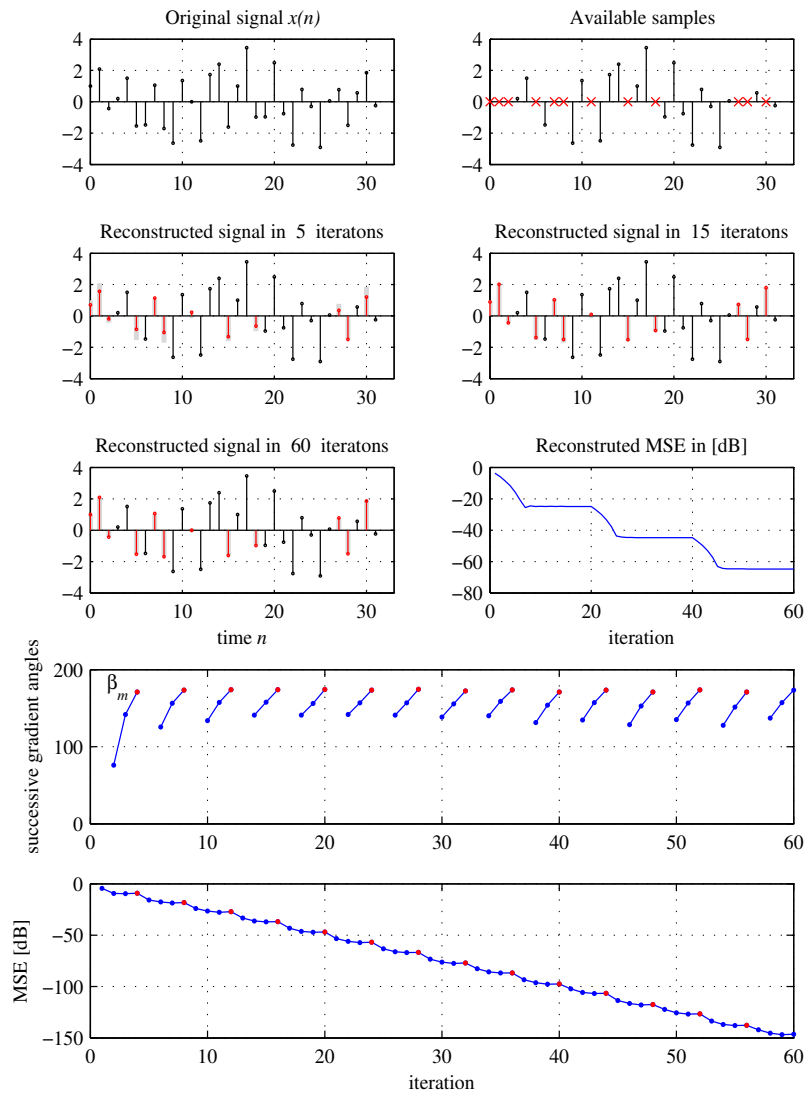
Figure 2: Gradient-based reconstruction of a sparse signal with reconstructed MSE (three top rows). Angle between successive gradient estimations β_m and the MSE as a function of the number of iterations with step changes based on the gradient angles (bottom).
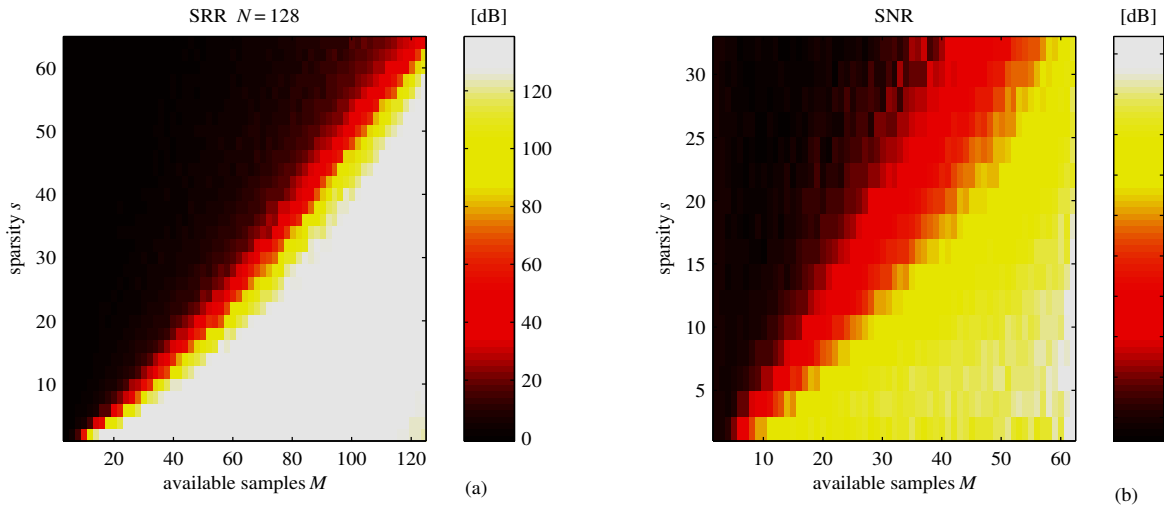
10

Figure 3: Signal-to-reconstruction-error (SRR) obtained by using Algorithm 1, averaged over 100 realizations for various sparsity s and number of available samples M: (a) The total number of samples is $N = 128$. (b) With a Gaussian noise in the input signal, $SNR = 20$ [dB] and $N = 64$.

repeated with an additive input Gaussian noise such that the input signal-to-noise ratio is 20 [dB] in each realization Fig.3(b). The reconstruction error in this case is limited by the input signal-to-noise ratio.

The average reconstruction error in the noise-free cases is related to the number of the full recovery events. For $N = 64$ the number of the full recovery events is checked and presented in Fig.4 (a),(b). The average number of the algorithm iterations to produce the required precision, as a function of the number of missing samples and signal sparsity s, is presented as well, Fig.4(c), along with the corresponding average computation time (in seconds) for the Windows PC with Intel Dual Core processor, Fig.4(d). The average computation time is proportional to the average number of iterations multiplied by the number of missing samples (variables) $Q = N - M$.

The efficiency of the presented algorithm is compared with the standard routines used for the ℓ_1-norm based minimization problems. The performance of the proposed algorithm are compared with the algorithm that recasts the recovery problem (3) into regression framework (LASSO-ISTA) and a linear programming framework with the primal-dual interior point method (L1-magic code in MATLAB). A simple review of the LASSO-ISTA algorithm is presented in the Appendix. All there algorithms are run using 100 sparse signals with random parameters. The algorithm parameters are set in such a way that the average computational time for all cases is about 25 ms. The results are presented in Table 1. Columns notation in the table is: s for sparsity, $Q = N - M$ for the number of missing samples, MAE stands for the mean absolute error. Calculation time using MATLAB is presented in all cases.

We can conclude that in most of the considered cases the proposed algorithm outperforms other two algorithms in both the calculation time and accuracy.

Example 3: The algorithm can be applied to other signal transforms as well [23]. An example with missing (salt-and-pepper noise) pixels in an image with the 2D-DCT as its sparsity domain will be presented. Consider a standard MATLAB test image canoe.tif (Fig.5) and assume that 50% of the samples are corrupted (noisy image). These pixels with extreme values are removed and considered as unavailable. Since the coefficients in the 2D-DCT converges fast we will assume that the sparsity condition is satisfied (without imposing it in an explicit way by making some 2D-DCT coefficients equal to zero). The image is reconstructed by minimizing the 2D-DCT sparsity measure and varying missing pixel values according to the presented algorithm. The original, corrupted and reconstructed images after 4th and 50th iteration are shown in Fig.5.
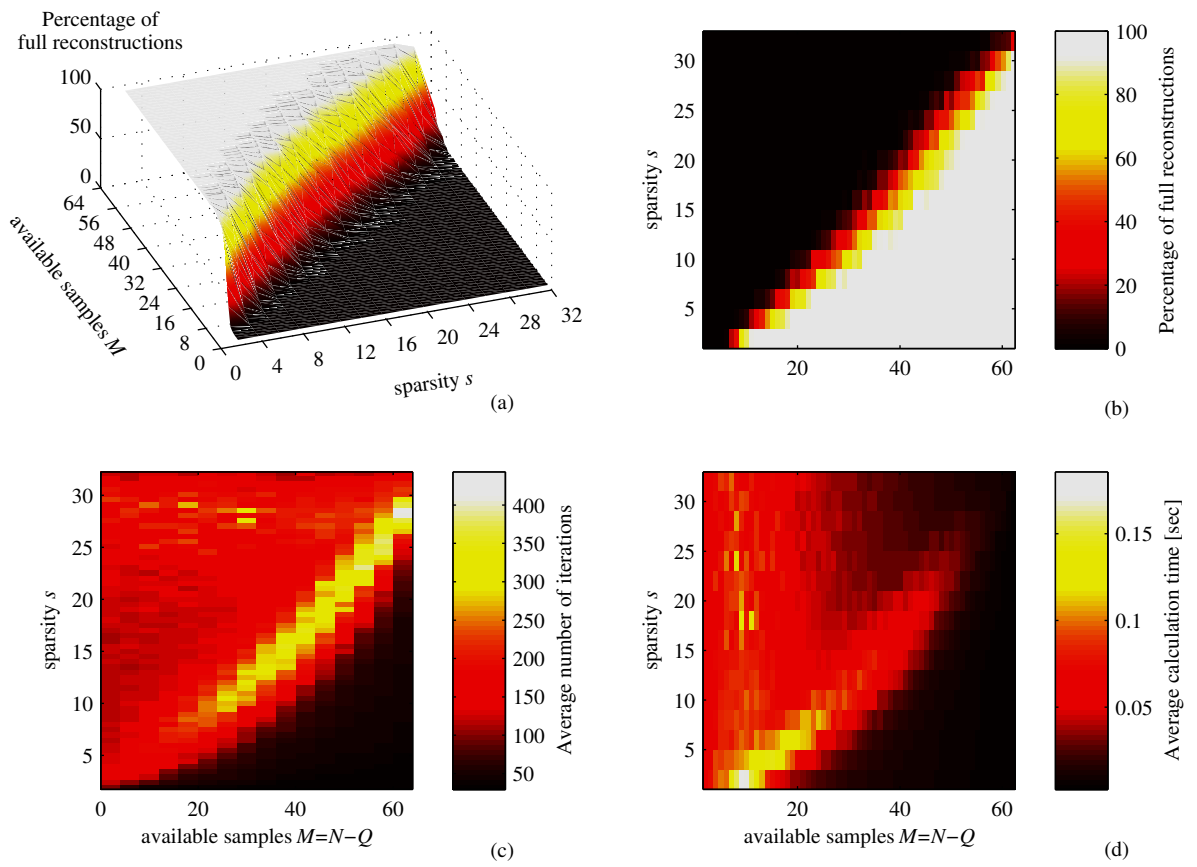
Figure 4: (a)-(b) The percentage of the full recovery events as a function of the number of available samples M and the sparsity s in the case of $N = 64$. (c) The average number of iterations as a function of the number of missing samples and sparsity. (d) The average computation time.

12

Table 1: MAE and computational time for the L1-magic (LP-DP), LASSO-ISTA, and the proposed algorithm

		L1-magic		LASSO		Proposed	
s	$N-M$	MAE	Time	MAE	Time	MAE	Time
8	8	7.94×10^{-11}	29.7	1.24×10^{-4}	25.9	6.05×10^{-12}	7.6
16	8	6.70×10^{-10}	28.4	1.72×10^{-4}	25.9	7.56×10^{-12}	7.9
32	8	1.28×10^{-9}	29.8	2.25×10^{-4}	31.5	4.96×10^{-12}	13.1
8	16	1.64×10^{-10}	31.2	1.30×10^{-4}	30.1	6.92×10^{-12}	15.2
16	16	1.22×10^{-5}	25.4	1.91×10^{-4}	30.6	6.11×10^{-12}	18.9
32	16	5.44×10^{-5}	22.7	7.16×10^{-4}	26.4	1.07×10^{-11}	21.5
8	24	2.83×10^{-10}	23.2	1.37×10^{-4}	25.7	5.25×10^{-12}	20.6
16	24	4.62×10^{-6}	21.6	2.51×10^{-4}	25.4	1.01×10^{-11}	22.8
32	24	2.58×10^{-4}	24.2	1.17×10^{-3}	25.1	1.88×10^{-11}	35.8
8	32	6.66×10^{-10}	20.4	1.51×10^{-4}	25.4	7.83×10^{-12}	28.1
16	32	1.94×10^{-4}	22.4	4.09×10^{-4}	25.0	1.89×10^{-11}	32.8
32	32	1.08×10^{-4}	25.0	8.08×10^{-4}	25.1	1.95×10^{-11}	74.1

Original image

Noisy image

Reconstructed image. Iteration: 4
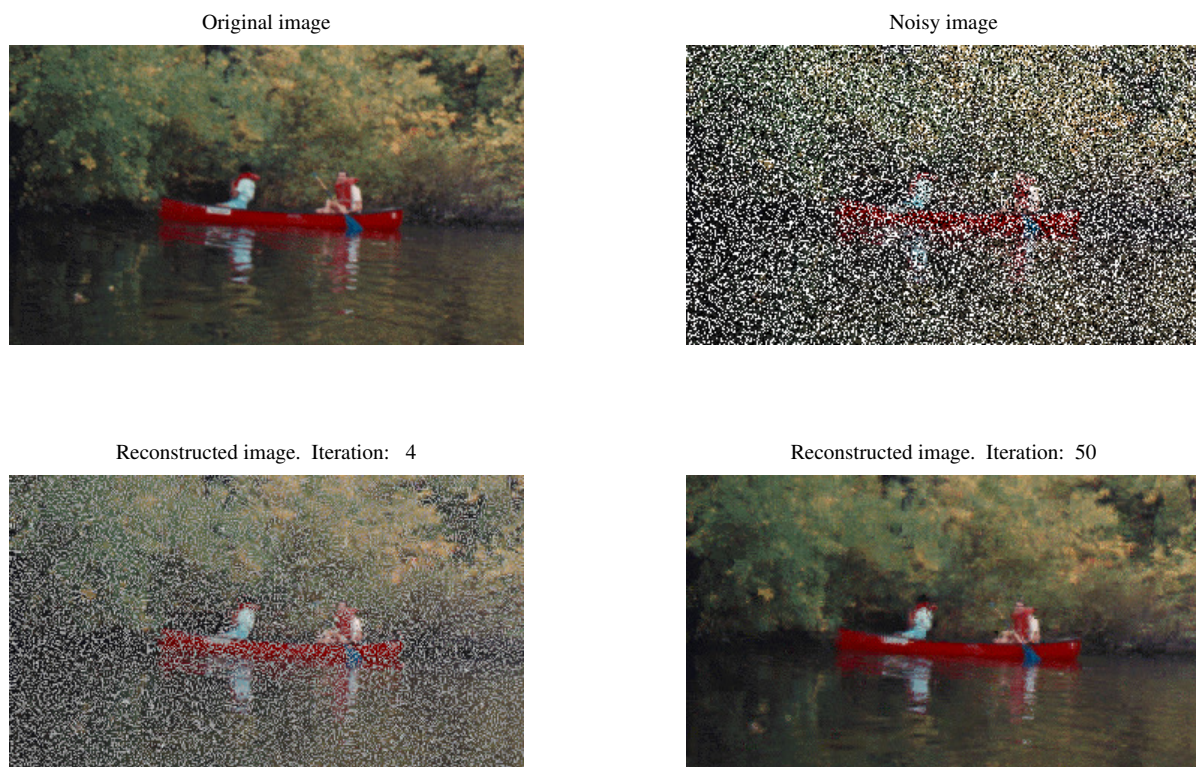
Reconstructed image. Iteration: 50

Figure 5: Original MATLAB image canoe.tif (top-left). Corrupted image with salt-and-pepper noise in 50% of the pixels (top-right). Reconstructed images after 4 and 50 iterations (bottom).

5. Conclusion

Performance of the recently proposed gradient-based signal reconstruction algorithm in the domain of the signal samples/measurements are studied. Application of this algorithm is very efficient in signal processing problems with reduced set of data, since it is easy to define a set of missing signal samples/measurements. The values of signal samples are considered as variables, while the available samples are fixed and not changed in the algorithm. An efficient criterion, based on the gradient directions, is proposed for the algorithm step-size adaptation. Statistical analysis confirms calculation efficiency. The algorithm can easily be adapted to the cases on nonuniform signal samples, when only a reduced set of linear combinations of the signal samples are available.

6. Appendix: Gradient-Based Reconstruction in the Sparsity Domain

One of the forms for formulation of minimization problem (3), in order to find its solution, is based on the Lagrangian

$$F(\mathbf{X}) = \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_1 .$$

The minimization function is denoted by $F(\mathbf{X})$. The signal reconstruction in its sparsity domain is formulated as [9]

$$\mathbf{X} = \arg \min_{\mathbf{X}} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_1 \right\} . \tag{11}$$

This problem can be solved using least absolute selection and shrinking operator (LASSO) minimization. Parameter λ defines a balance between the squared absolute error and sparsity.

The l_1-norm based reconstruction algorithms do not have a closed form solution. In order to define an iterative procedure a non-negative term

$$G(\mathbf{X}) = (\mathbf{X} - \mathbf{X}_s)^T (\alpha \mathbf{I} - \mathbf{A}^T \mathbf{A})(\mathbf{X} - \mathbf{X}_s)$$

is added. It has a zero value at the position of the problem solution, denoted by $\mathbf{X}_s$. This term is added to the function $F(\mathbf{X})$. New minimization function is

$$H(\mathbf{X}) = F(\mathbf{X}) + (\mathbf{X} - \mathbf{X}_s)^T (\alpha \mathbf{I} - \mathbf{A}^T \mathbf{A})(\mathbf{X} - \mathbf{X}_s).$$

The solution follows from

$$\nabla H(\mathbf{X}) = \frac{\partial H(\mathbf{X})}{\partial \mathbf{X}^T} = 0$$

as

$$\mathbf{X} + \frac{\lambda}{2\alpha} \text{sign}\{\mathbf{X}\} = \frac{1}{\alpha} \mathbf{A}^T (\mathbf{y} - \mathbf{A}\mathbf{X}_s) + \mathbf{X}_s.$$

The iterative procedure can be defined by

$$\mathbf{X}_{s+1} + \frac{\lambda}{2\alpha} \text{sign}\{\mathbf{X}_{s+1}\} = \frac{1}{\alpha} \mathbf{A}^T (\mathbf{y} - \mathbf{A}\mathbf{X}_s) + \mathbf{X}_s. \tag{12}$$

The equation of the form

$$x + \lambda \text{sign}(x) = y$$

can be solved by using the soft thresholding rule

$$x = \text{soft}(y, \lambda) = \text{sign}(y)\max\{0, |y| - \lambda\}.$$

The same rule can be used for each coordinate of the sparsity vector in (12). The iterative form for the signal reconstruction in the sparsity domain is

$$\mathbf{X}_{s+1} = \text{soft}(\frac{1}{\alpha} \mathbf{A}^T (\mathbf{y} - \mathbf{A}\mathbf{X}_s) + \mathbf{X}_s, \frac{\lambda}{2\alpha}). \tag{13}$$

This is the iterative soft-thresholding algorithm (ISTA) for problem (11) minimization [26]. The initial estimate is $\mathbf{X}_0 = \mathbf{A}^T \mathbf{y}$, corresponding to the transform of initial signal (4).

7. References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, 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 frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] E. J. Candès, "The restricted isometry property and its implications for compressed sensing", *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589-592, 2008.

[4] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 20, no. 3, pp. 301–321,2009

[5] L. Stanković, I. Orović, S. Stanković, and M. Amin, "Robust Time Frequency Analysis based on the L-estimation and Compressive Sensing," *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 499–502, 2013.

[6] R. E. Carrillo, K. E. Barner, and T. C. Aysal, "Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise," *IEEE Journal of Selected Topics in Signal Processing*, 2010, vol. 4, no. 2, pp. 392–408.

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

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

[9] R.J. Tibshirani, "Regression shrinkage and selection via the lasso: A retrospective," *Journal of the Royal Statistical Society*, Series B, 73:273-282, 2011.

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

[11] 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.

[12] 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.

[13] 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.

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

[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] 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.

[17] 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.

[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, vol. 94, Jan. 2014, pp. 401–408.

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

[21] 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.

[22] L. Stanković, M. Daković and S. Vujović, "Adaptive Variable Step Algorithm for Missing Samples Recovery in Sparse Signals," *IET Signal Processing*, vol. 8, no. 3, pp. 246 -256, 2014, (arXiv:1309.5749v1).

[23] I. Stanković, Recovery of Images with Missing Pixels using a Gradient Compressive Sensing Algorithm, http://arxiv.org/ftp/arxiv/papers/1407/1407.3695.pdf, June 2014.

[24] H. Tsutsu, Y. Morikawa, "An lp Norm Minimization Using Auxiliary Function for Compressed Sensing", in *Proc. IMECS*, March 2012, Hong Kong.

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

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