Comparison of a Gradient-Based and LASSO (ISTA) Algorithm for Sparse Signal Reconstruction

Stefan Vujović, Isidora Stanković, Miloš Daković, Ljubiša Stanković
University of Montenegro, Podgorica, Montenegro
Email: [stefanv, isidoras, milos, ljubisa]@ac.me

***Abstract*—Sparse signal reconstruction performed by two different algorithms is considered. First algorithm is the ISTA algorithm for LASSO minimization, while the second one is the gradient-based descent algorithm. Algorithms perform signal reconstruction in a completely different way. The ISTA algorithm reconstructs signals in the sparsity transformation domain. The gradient descent algorithm performs reconstruction in time/measurements domain, considering the missing samples as variables. Both of them use the l_1-norm in minimization. Computational time and mean absolute error are used in comparison analysis presented in this paper.**

***Keywords* - Compressive sensing, Gradient algorithm, Sparse reconstruction, Sparse signals, Lasso minimization, Ista algorithm**

I. Introduction

In signal processing, reconstruction of sparse signals has been an attractive research area in the recent years. A sparse signal is a signal which has only a small number of non-zero components in a transform domain, compared to the total number of components. We can expect that a signal which is sparse in a transformation domain can be reconstructed with less randomly positioned samples compared to the number of samples required by the traditional sampling theorem. This procedure is defined within the field of compressive sensing (CS) [1]–[8]. For a successful reconstruction of a sparse signal, the signal has to meet some conditions. The most important one is that the signal must be sparse. The second important condition is the incoherence between the measurements taken into account and the sparsity basis of the signal [9], [10].

Since the introduction of compressive sensing many methods have been developed for taking measurements and reconstructing sparse signals. These methods have been applied in many everyday applications of signal processing, such as multimedia, radars, biomedicine, communications, etc [11]–[14]. Some of the methods can also be applied for denoising of sparse signals. When some signal samples are heavily corrupted by noise, it is better to exclude them from the calculations and set them to zero. In this case we take less measurements, and use the fact that the signal is sparse, in order to reconstruct the original signal [5].

This research is supported by the project "New ICT Compressive Sensing Based Trends Applied to: Multimedia, Biomedicine and Communications (CS-ICT)" (Montenegro Ministry of Science, Grant No. 01-1002)

In this paper we will consider two algorithms with different approaches to the reconstruction of the sparse signals within compressive sensing framework. The first one is the iterative soft-thresholding algorithm (ISTA) [15]. It is based on the LASSO minimization, i.e. minimization of the signal sparsity in the transformation domain by using the l_1-norm problem formulation. The second one, the gradient descent algorithm, is from the convex relaxation group of the algorithms [16], [17] for which the reconstruction procedure is done in the spatial/measurements domain. In this algorithm [16], the missing samples are considered as variables. Comparison of these two algorithms with respect to computational time and mean absolute error (MAE) will be presented here.

The paper is organized as follows. In Section 2, a review of the minimization problem formulation will be defined. In Section 3 and Section 4, two considered algorithms will be described and in Section 5 the results and comparison of performances of the algorithms are presented. The conclusions are given in Section 6.

II. Theoretical Background

Consider a discrete-time signal $x(n)$, $1 \leq n \leq N$, and its transformation domain coefficients $X(k)$, $1 \leq k \leq N$

$$x(n) = \sum_{k=1}^{N} X(k)\psi_k(n),$$

$$X(k) = \sum_{n=1}^{N} x(n)\varphi_k(n),$$

or in the vector/matrix notation

$$\mathbf{x} = \mathbf{\Psi X} \quad \text{and} \quad \mathbf{X} = \mathbf{\Phi x} \; .$$

Matrix $\mathbf{\Psi}$ is the transformation matrix, $\mathbf{\Phi}$ is its inversion, while $\mathbf{x}$, and $\mathbf{X}$ are signal vector column and transformation coefficients vector column, respectively. A signal is considered to be K-sparse in a transformation domain if the number of non-zero coefficients K is much lower than the total number of coefficients N, i.e. $K \ll N$. Without loss of generality, signals which are sparse in Discrete Cosine Transform (DCT) domain will be considered in this paper. For the DCT transformation matrix and its inverse, the elements are:

$$\psi_k(n) = \phi_k(n) = w(k)\cos\left(\frac{\pi(2n-1)(k-1)}{2N}\right),$$

where $w(k) = 1/\sqrt{N}$ for $k = 1$, and $w(k) = \sqrt{2/N}$ for $2 \le k \le N$.

Random subset of M signal $x(n)$ samples at positions $n_i \in \mathbf{M} = \{n_1, n_2, \ldots, n_M\} \subset \mathbf{N} = \{1, 2, \ldots, N\}$ will be considered as a set of available signal samples and denoted by vector $\mathbf{y}$, while $\mathbf{N}$ is the complete set of all signal $x(n)$ samples. Note that vector $\mathbf{y}$ elements

$$\mathbf{y} = [x(n_1), x(n_2), \ldots, x(n_M)]^T$$

are the measurements of a linear combination of $X(k)$

$$x(n_i) = \sum_{k=1}^{N} X(k)\psi_k(n_i).$$

This system of M equations can be written as

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

where $M \times N$ matrix $\mathbf{A}$ is obtained from $N \times N$ matrix $\mathbf{\Psi}$ in a way to preserve rows at the positions corresponding to the positions of available samples $n_i \in \mathbf{M} = \{n_1, n_2, \ldots, n_M\}$, while the other $N - M$ rows are eliminated from the complete matrix.

The goal of CS reconstruction is to recover the most sparse signal $\mathbf{X}$ from a reduced set of available measurements $\mathbf{y}$. This task can be formulated as:

$$\min \|\mathbf{X}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{AX}. \tag{1}$$

However, since task (1) is an NP-hard combinatorial approach, the common l_1-norm instead of l_0-norm is used. Task (1) is then reformulated as:

$$\min \|\mathbf{X}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{AX}. \tag{2}$$

It is important to note that the solutions of (1) and (2) are the same if signal $x(n)$ and its transform $X(k)$ satisfy restricted isometry property [3], [9]. According to Parseval's theorem, the l_2-norm can not be used since it would produce zero values for all missing samples.

III. LASSO minimization - ISTA algorithm for Reconstruction in the Domain of Sparsity

The standard minimization problem (2) can also be formulated in a Lagrangian form

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

where $F(\mathbf{X})$ is the function that needs to be minimized. Reconstruction task is formulated as

$$\mathbf{X} = \arg\min_{\mathbf{X}} \left\{\|\mathbf{y} - \mathbf{AX}\|_2^2 + \lambda\|\mathbf{X}\|_1\right\} \tag{3}$$

and is titled LASSO (least absolute selection and shrinking operator) minimization. Parameter λ balances between first term (error) and second term (sparsity) in (3).

A. ISTA algorithm

Since the l_1-norm based reconstruction does not have close form solution like the l_2-norm, the iterative procedure is used in order to solve the problem. A non-negative term

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

having zero value at the solution $\mathbf{X}_s$ of the problem is added to the function $F(\mathbf{X})$, where α is a value greater than the maximal eigenvector of $\mathbf{A}^T\mathbf{A}$, what means that the added term is always nonnegative. New function now reads

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

The solution of the equation

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

is

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

The iterative form is then defined as

$$\mathbf{X}_{s+1} + \frac{\lambda}{2\alpha}\text{sign}\{\mathbf{X}_{s+1}\} = \frac{1}{2\alpha}\mathbf{A}^T(\mathbf{y} - \mathbf{AX}_s) + \mathbf{X}_s. \tag{4}$$

The equation

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

is solved by soft thresholding rule as

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

By applying the same rule to (4) we get the final iterative form for reconstruction

$$\mathbf{X}_{s+1} = \text{soft}\left(\frac{1}{\alpha}\mathbf{A}^T(\mathbf{y} - \mathbf{AX}_s) + \mathbf{X}_s, \frac{\lambda}{2\alpha}\right). \tag{5}$$

This is the iterative soft-thresholding algorithm for LASSO (3) minimization, implemented from [9], [18].

IV. Gradient-Based Algorithm for Reconstruction in the Measurements Domain

The main characteristic of this algorithm is that it is based on the reconstruction of the missing samples in time/measurements domain while the available samples remain unchanged. Consider a signal $x(n)$ with $n \in \mathbf{N} = \{1, 2, \ldots, N\}$ which is K-sparse in a transformation domain. Set of M available sample positions will be denoted by $\mathbf{M} = \{n_1, n_2, \ldots, n_M\}$, while the other $N - M$ sample positions which are missing will be denoted by $\mathbf{N}_Q = \mathbf{N} \setminus \mathbf{M}$. The algorithm implementation follows from [16].

Step 0: In the $0th$ (initial) iteration of the algorithm signal $x_r^{(0)}(n)$ is formed. Values of this signal at the available sample positions $\mathbf{M}$ are equal to original signal $x(n)$, while the values at the positions of the missing samples are set to zero

$$x_r^{(0)}(n) = \begin{cases} x(n), & n \in \mathbf{M} \\ 0, & n \in \mathbf{N}_Q \end{cases} \tag{6}$$

Step 1: For each missing sample position $n_i \in \mathbf{N}_Q$, two signal are formed by adding a constant value $\pm\Delta$. New signals are defined as

$$\begin{aligned} x_+(n) &= x_r^{(k)}(n) + \Delta\delta(n - n_i) \\ x_-(n) &= x_r^{(k)}(n) - \Delta\delta(n - n_i) \end{aligned} \tag{7}$$

where k is iteration number. In the initial calculation, constant Δ will be set to $\Delta = \max|x_r^{(0)}(n)|$.

Step 2: The differential of signal transform (DCT) measures [19] is estimated as

$$g(n_i) = \frac{\|DCT\,[x_+(n)]\|_1 - \|DCT\,[x_-(n)]\|_1}{N}. \tag{8}$$

Step 3: Form a gradient vector $\mathbf{G}$ of the same length as the signal $x(n)$. At the positions of available samples this vector has value $G(n) = 0$. Its zero value provides that available samples do not change since they have correct values set in Step 0. At the positions of missing samples this vector has values $g(n_i)$ calculated by (8).

Step 4: Update signal $x_r(n)$ in an iterative way as

$$x_r{}^{(k+1)}(n) = x_r{}^{(k)}(n) - G(n). \tag{9}$$

This procedure is repeated until some desired error is achieved. The parameter Δ is reduced through iterations in order to improve precision, and the algorithm is stopped when the error between two successful iterations is negligible [16]. The theoretical proof of algorithm convergence is presented in [20], while the experimental proof is given in next section.

V. Experimental results

In order to make comparison between two presented algorithms, let us consider a signal $x(n)$ sparse in the DCT domain defined as

$$x(n) = \sum_{i=1}^{K} A_i \cos\left(\frac{\pi(2n-1)(k_i-1)}{2N}\right) \tag{10}$$

where A_i and k_i are random amplitudes and random frequency positions of the signal $x(n)$ components. The algorithms are tested for different sparsity levels, $K = \{1, 3, 5, ..., 63\}$ and for different number of available samples, $M = \{2, 6, 10, ..., 126\}$. The algorithms have been tested in the regions $M \geq 2K$ [1], [6]. For the gradient algorithm, the parameter $\Delta = \max|x_r^{(0)}(n)|$ is used. For the ISTA algorithm, different values of λ were examined, $\lambda = \{0.001, 0.0025, 0.005, 0.01\}$, and the optimum λ was found to be $\lambda = 0.005$. The algorithms are compared in terms of two important parameters: computational time and mean absolute error. Computational time is expressed in seconds, while the MAE was calculated as

$$MAE = mean(|x(n) - x_r(n)|),$$

where $x_r(n)$ is reconstructed signal, and its value expressed in dB as

$$MAE[dB] = 20\log(MAE)$$

is presented. Results are averaged in 10 realizations for each combination of parameters K and M. The same conclusions regarding to reconstruction error can be obtained if MSE (mean square error) were used instead of MAE.

Example 1: Computational time as a function of sparsity K and the number of available samples M is presented in Fig. 1. Top graphic is for the gradient algorithm, while bottom one is for the ISTA algorithm. It is obvious that the gradient algorithm performs faster reconstruction for each combination K and M. The time needed to perform reconstruction with the gradient algorithm is of order 0.02 in comparison to the order of 0.2 for the ISTA algorithm. The mean computational time for the all combinations of K and M is 0.024 while for gradient algorithm and 0.2353 for the ISTA algorithm. The white region from graphics ($M < 2K$) where reconstruction theoretically can not be performed [1], [6] was not considered.

Example 2: The mean absolute error expressed in dB for each combination of K and M for both algorithms is presented in Fig. 2. Top graphic presents results for the gradient algorithm, while the bottom one is for the ISTA algorithm. White region (where reconstruction can not be performed) was not considered here. It can be seen that the gradient algorithm performs reconstruction with a smaller MAE. Mean value of all MAEs for all combinations of K and M (where

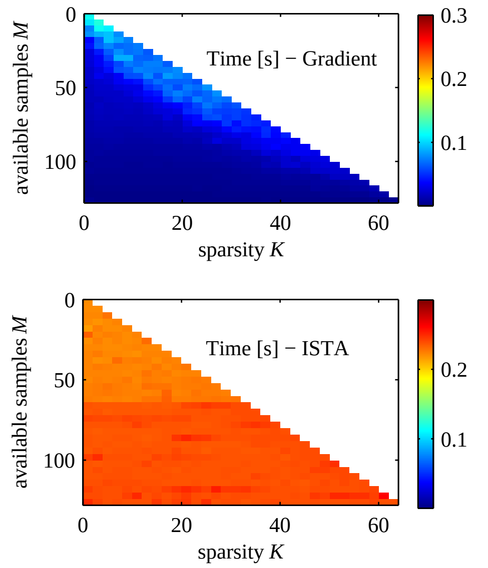

Fig. 1. Time needed for the algorithms to successfully reconstruct: the gradient algorithm (top) and the ISTA algorithm (bottom).

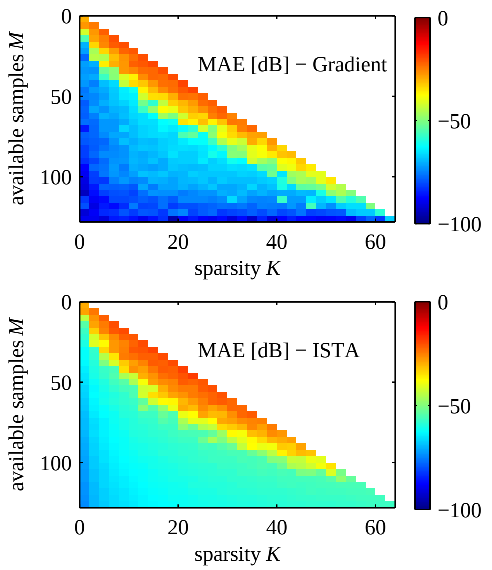

Fig. 2. Mean absolute error of: the gradient algorithm (top) and the ISTA algorithm (bottom) in [dB].

$M \geq 2K$) is -61.88 dB in comparison to -50.88 dB for the ISTA algorithm.

VI. Conclusions

A comparison of two algorithms for sparse signal reconstruction is done. The algorithms have been compared in terms of computational time needed for the reconstruction and the mean absolute error of each algorithm. Two presented algorithms have been chosen because of their different approach to the minimization problem. It is shown that the gradient descent algorithm gives faster and more accurate results for all considered combinations of K and M.

References

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

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

[3] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[4] S. Stanković, I. Orović, LJ. Stanković, "An Automated Signal Reconstruction Method based on Analysis of Compressive Sensed Signals in Noisy Environment," *Signal Processing*, vol. 104, Nov 2014, pp. 43 - 50, 2014.

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

[6] D. Needell, and J. A. Tropp. "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, 26.3 (2009): 301-321.

[7] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems,"*Proceedings of the IEEE* 98.6 (2010): 948-958.

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

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

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

[11] L. Stanković, I. Orović, S. Stanković and M.G. Amin, "Robust Time-Frequency Analysis based on the L-estimation and Compressive Sensing," *IEEE Signal Processing Letters*, May 2013, pp.499–502.

[12] L. Stanković, I. Orović, S. Stanković, and M. Amin, "Compressive Sensing Based Separation of Non-Stationary and Stationary Signals Overlapping in Time-Frequency," *IEEE Transactions on Signal Processing*, Vol. 61, no. 18, pp. 4562 – 4572, Sept. 2013.

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

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

[15] Beck, Amir, and Marc Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems,"*SIAM journal on imaging sciences*, 2.1 (2009): 183-202.

[16] L. Stanković, M. Daković, S. Vujović, "Adaptive variable step algorithm for missing samples recovery in sparse signals," *IET Signal Processing*, vol.8, no.3, 2014, pp.246–256, doi: 10.1049/iet-spr.2013.0385

[17] 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

[18] R.J. Tibshirani, "Regression shrinkage and selection via the lasso: A retrospective,"*Journal of the Royal Statistical Society*, Series B, 73:273282, 2011.

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

[20] M. Brajović, M. Daković, L. Stanković, "Convexity of the l1-norm based Sparsity Measure with Respect to the Missing Samples as Variables, "*Mediterranean Conference on Embedded Computing (MECO) 2016*, submitted.