Comparison of the L1-magic and the Gradient Algorithm for Sparse Signals Reconstruction

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

Abstract — **Reconstruction of the sparse signals, performed by two algorithms which belongs to the class of convex optimization algorithms, is considered in this paper. Widely used algorithm implemented by the *l_1-magic* code packet is used as the first algorithm. Its realization is based on the primal dual interior point algorithm for convex optimization. The second considered algorithm also belongs to the convex optimization group of algorithms. It is a recently proposed adaptive step gradient-based algorithm. The reconstruction of missing data is based on the direct adaptation of the signal values by minimizing the concentration/sparsity measure of the signal in the transformation domain, where the signal is sparse. Comparison of these two algorithms is done here.**

Keywords — **Compressive sensing, Gradient algorithm, *l_1-magic*, Sparse reconstruction, Sparse signals**

I. Introduction

SIGNAL is considered to be sparse in a transformation domain if the number of significant (non-zero) coefficients in that domain is much smaller than the number of all coefficients. Reconstruction of this kind of signals attracts significant research interest in the last decade since the sparse signals are present in many applications [1]-[21]. The problem of reconstruction is minutely studied within the theory of compressive sensing (CS) [1]-[11].

In order to perform sparse signal reconstruction many algorithms are introduced [2]-[14]. There are many groups of algorithms that are widely used in reconstruction like pursuit methods, convex relaxation, nonconvex relaxation, brute force based methods [11], etc. Here, we will focus on one commonly used class of the algorithms, the convex relaxation class of algorithms. The convex relaxations algorithms are based on the l_1-norm optimization. Two methods of convex relaxation algorithms, the interior point and the adaptive gradient based one, are considered in this paper. The software which is based on the interior point methods (commonly used for reconstruction) today, is the *l_1-magic* software packet [12]. The other algorithm that will be considered here is the one that belongs to the class of gradient descent algorithms [14]. The gradient descent algorithms are used for convex optimizations problems. The goal of this paper is to make comparison between these two algorithms since both of them are from the same class of CS reconstruction methods (convex relaxation).

Detailed analysis of the computational performance of each of these two algorithms will be done. Two very important parameters that will be analyzed are the reconstruction time and the reconstruction accuracy. The accuracy will be measured by the mean absolute error (MAE). Since the *l_1-magic* software tool is widely used in research, we found interesting to compare it with a recently proposed algorithm dealing with the same problem of reconstructing sparse signals. Both of these two algorithms will be analyzed under the same conditions and the conclusions about the speed and the MAE of each of them will be presented.

The paper is organized as follows. After introduction, reconstruction definitions are introduced in Section II. Section III provides overview of both algorithms. After that, comparison of these two algorithms is done in Section IV. At the end of the paper there is a concluding Section V.

II. Definitions

Let $\mathbf{x}$ be discrete-time domain signal of length N. Suppose that this signal has sparse transformation in some domain. Here we will consider the case of the discrete Fourier domain (DFT) as a study case. The sparse transform of the signal $\mathbf{x}$ will be denoted as $\mathbf{X}$. The relation between these two domains is

$$\mathbf{X}=\mathbf{W}\mathbf{x}\ (\mathbf{x}=\mathbf{W}^{-1}\mathbf{X}),$$

where $\mathbf{W}$ is NxN DFT matrix with elements $W(n,k)=exp(-2j\pi nk/N)$.

Assume that the signal $\mathbf{x}$ has only M available samples, while the other N-M samples are missing or are unavailable due to a physical constraint or measurement unavailability. The vector of M available signal samples will be denoted as $\mathbf{y}$. Its values are

$$y(i)=x(n_i),\text{ where } i=1,2,...,M\ .$$

The reconstruction task can be defined as:

$$\min\|\tilde{\mathbf{X}}\|_o \text{ subject to } \mathbf{y}=\mathbf{A}\tilde{\mathbf{X}}\ , \tag{1}$$

where $\boldsymbol{A}$ is MxN matrix obtained from matrix $\boldsymbol{W}^{-1}$ by eliminating rows corresponding to missing samples

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) and by the European Union project Fore-Mont as a part of Seventh Framework Programme (Grant Agreement No. 315970 FP7-REGPOT-CT-2013).

Stefan Vujović, (contact author), Elektrotehnički fakultet, Džordža Vašingtona bb, 81000 Podgorica, Montenegro (phone: 382-67-302313, e-mail: stefanv@ac.me)

Miloš Daković, Elektrotehnički fakultet, Džordža Vašingtona bb, 81000 Podgorica, Montegro (e-mail: milos@ac.me)

Ljubiša Stanković, Elektrotehnički fakultet, Džordža Vašingtona bb, 81000 Podgorica, Montenegro (e-mail: ljubisa@ac.me)

(preserve rows with indexes corresponding to instants n_i, where $i=1,2,\ldots,M$). Since (1) is an NP-hard combinatorial approach commonly l_1-norm instead of l_0-norm is used. In practical signal processing applications l_0-norm is sensitive to small values, even for a values of computer precision order too [21]. It is one more reason to use l_1-norm. The considering reconstruction task (1) is then

$$\min \|\tilde{\mathbf{X}}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\tilde{\mathbf{X}} \ . \quad (2)$$

It is important to note that solutions of (1) and (2) are the same if the signal $\mathbf{x}$ and its transform $\mathbf{X}$ satisfy restricted isometry property (RIP) [2].

After the sparse presentation of signal $\tilde{\mathbf{X}}$ is approximated according to (2), a reconstructed signal is obtained as

$$\mathbf{x_r} = \mathbf{W}^{-1}\tilde{\mathbf{X}} \ .$$

When the reconstructed signal is obtained, the reconstruction quality can be measured using, for example, the mean absolute error of the reconstructed signal:

$$MAE = \frac{\sum_{n=1}^{N} |x(n) - x_r(n)|}{N} \ . \quad (3)$$

III. Algorithm Definitions

As it has been mentioned in Section I, there are many classes of algorithms that deal with sparse signals reconstruction. Algorithms that belong to the convex optimization class are the widely used ones. Here, we will deal with two types of algorithms that solve convex optimization problems. They will be presented next.

A. *l_1-magic*

Software packet *l_1-magic* is a tool intensively used to solve the CS reconstruction problems. This algorithm belongs to the class of interior point algorithms. It is a primal dual interior point method for linear programming [13, chapter 11]. The standard form of a linear program is:

$$\min_{\mathbf{X}} \langle \mathbf{c_0}, \mathbf{X} \rangle \text{ subject to } \mathbf{y} = \mathbf{AX}$$
$$f_i(\mathbf{X}) \le 0 \ ,$$

where $\mathbf{X}\in R^N$, $\mathbf{y}=R^K$, A is KxN matrix, and each of f_i , $i=1,\ldots,m$ is a linear functional:

$$f_i(\mathbf{X}) = \langle c_i, \mathbf{X} \rangle + d_i$$

for some $c_i\in R^N$, $d_i\in R$. At the optimal point $\tilde{\mathbf{X}}$, there will exist dual vectors $\mathbf{v}^*\in R^K$, and $\boldsymbol{\lambda}^*\in R^m$ that satisfy Karush-Kuhn-tucker conditions:

$$\begin{gathered} c_0 + \mathbf{A}^T\mathbf{v}^* + \sum_i \boldsymbol{\lambda}_\mathbf{i}^* c_i = 0, \\ \boldsymbol{\lambda}_\mathbf{i}^* f_i(\tilde{\mathbf{X}}) = 0, \ \ i = 1,\ldots,m, \\ \mathbf{A}\tilde{\mathbf{X}} = \mathbf{y}, \\ f_i(\tilde{\mathbf{X}}) \le 0, \ \ i = 1,\ldots,m \end{gathered} \ .$$

Algorithm finds search vector $\tilde{\mathbf{X}}$ (optimal dual vector) by solving the system of nonlinear equations. At the interior point ($\tilde{\mathbf{X}}, \mathbf{v}^*$ and $\boldsymbol{\lambda}^*$), the system is linearized and solved, for example, by a widely used software package know as the *l_1-magic*.

B. *Gradient descent algorithm*

One more algorithm that can reconstruct sparse signals is the adaptive gradient descent one. It also belongs to convex optimization class of algorithms. It is very simple and efficient algorithm reconstructing discrete-time domain samples of the signal by optimizing the concentration of sparse presentation of signal.

Implementation of algorithm follows [14]:

Step 0: Form the signal $x_{r(0)}(n)$, where (0) means that this is the first iteration of the algorithm, defined as $x_{r(0)}(n)=x(n)$ for available samples, and 0 for missing samples.

Step 1: Two signals $x_{r1}(n)$ and $x_{r2}(n)$ for each missing sample at n_i , in each next iteration are formed as:

$$x_{r1}^{(k)}(n) = x_r(n) + \Delta\delta(n-n_i) \quad x_{r2}^{(k)}(n) = x_r(n) - \Delta\delta(n-n_i)$$

where k is the iteration number.

Step 2: The differential of the signal transform (here the DFT) measure is then estimated as:

$$g(n_i) = \frac{\frac{1}{N}\left(\sum\left|DFT\left[x_{r1}^{(k)}(n)\right]\right| - \sum\left|DFT\left[x_{r2}^{(k)}(n)\right]\right|\right)}{2\Delta} \quad (4)$$

Step 3: Form a gradient vector $\mathbf{G}$ with the same length as the signal $x(n)$. At the positions of the available samples, this vector has value $G(n) = 0$. At the positions of the missing samples its values are $g(n_i)$, calculated by (4).

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

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

where μ is a step size that affects the performances of the algorithm (the error and the speed of convergence).

By repeating the presented iterative procedure, the missing values will converge to the true signal values which produce a minimal concentration measure in the transformation domain.

IV. Experimental Results

Consider a signal

$$x(n) = \sum_{i=1}^{K} A_i \cos(2\pi k_i n / N + \phi_i) \quad (5)$$

with $N = 128$. Sparsity (number of signal components) of this signal $s=2K$ is changed from 2 to $N/2$. Randomly chosen amplitudes, frequencies and phases were within $1 \le A_i \le 2$, $1 \le k_i \le 63$ and $1 \le \phi_i \le 2\pi$. Results are averaged in 100 realizations for each combination of the sparsity s and the number of missing samples M.

Example 1: Results shown in Fig. 1. present the MAE logarithm as a function of the sparsity s and the number of available samples M. White color indicates the region where the algorithms were not able to reconstruct the signal in a such way that *MAE<0.01*. It is obvious that the reconstruction was performed only if the condition *M≥2s* was satisfied [1], [10]. As we can see from graphics, the adaptive gradient algorithm has precisely reconstructed missing samples in a wide region. The achieved precision is higher than with the l_1*-magic* algorithm. As it has been shown in [14], this algorithm has the option to reconstruct the signal samples with the computer precision. Also it is very important to note that this algorithm has successfully reconstructed signal in cases when l_1*-magic* could not reconstruct signal with *MAE<0.01* (blue colored area in the top graphic is larger than the one in the bottom graphic). In almost all cases where the adaptive gradient algorithm has performed reconstruction, the logarithm of *MAE* was about -6 corresponding to low error $MAE=10^{-6}$.

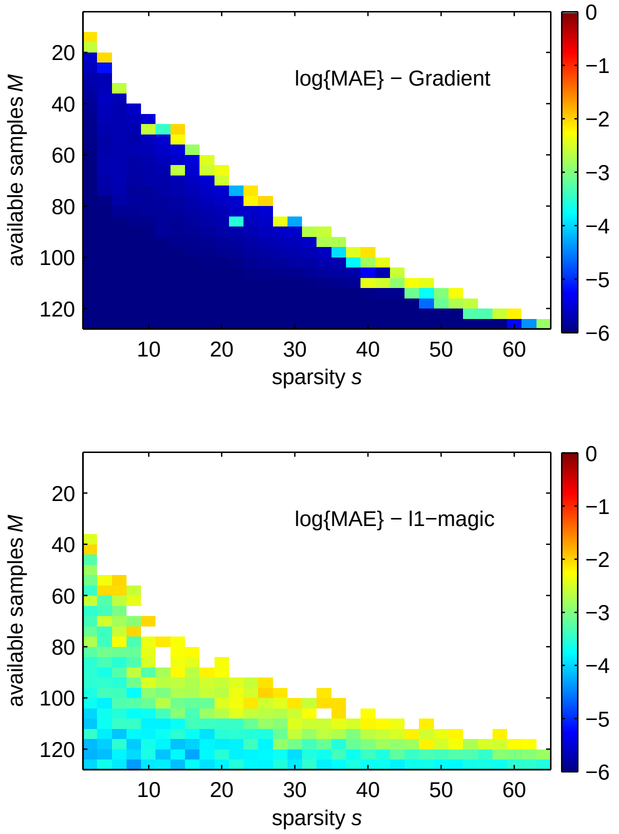

Fig. 1. Logarithm of MAE as a function of the sparsity *s* and the number of available samples *M* for the adaptive gradient algorithm (top) and the l_1*-magic* algorithm (bottom). White color corresponds to the region where there was not reconstruction with algorithms

Example 2: One more very important parameter in each algorithm is the speed of convergence. We have analyzed this parameter for both algorithms. We can see that the reconstruction was performed in less than 0.1s in all cases when successful reconstruction was performed. In many cases, time was significantly lower (order of 0.02) in both algorithms. Results shown in Fig.2 are presented for the cases when a successful reconstruction was achieved with any of the algorithm.

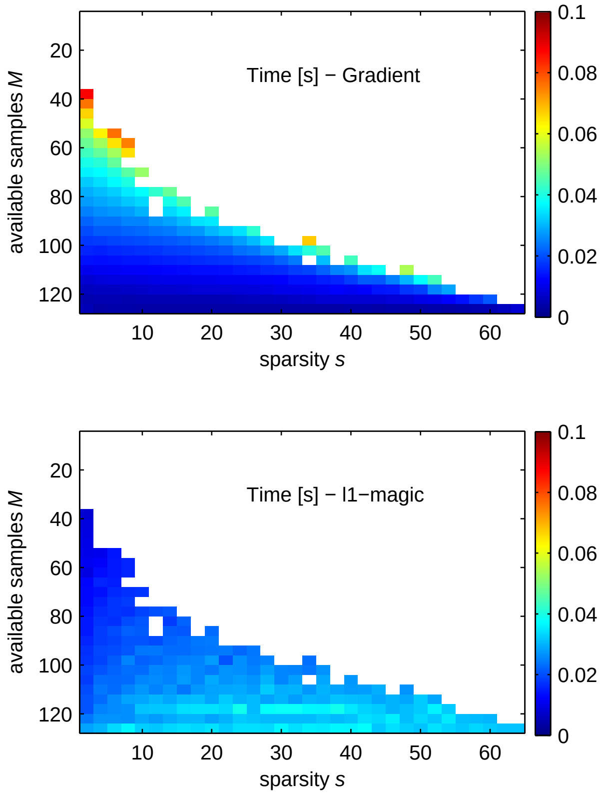

Fig. 2. Time as a function of sparsity *s* and number of available samples *M* for adaptive gradient algorithm (top) and l_1*-magic* algorithm (bottom). White color corresponds to the region where there was not reconstruction with algorithms

Top graphic of Fig.3. presents comparison in speed of convergence of two mentioned algorithms. White color indicates region where there was no reconstruction in both cases. Positive values correspond to situation when l_1*-magic* algorithm has faster convergence, while negative values correspond to the situation when the adaptive gradient algorithm has faster convergence. It is important to note that in both of these two situations gradient algorithm had better accuracy (Fig.1.). Bottom graphic of Fig.3. is just an illustration which shows regions where each of algorithms had faster convergence. As we can see, for large number of available samples, the adaptive gradient algorithm produces faster reconstruction, while the situations corresponding to a small sparsity are on the side of l_1*-magic* algorithm, since it had faster convergence. Regions denoted by blue and red circles correspond to the situation when the adaptive gradient algorithm produced the signal reconstructed, while the l_1*-magic* hadn't or l_1*-magic* had and gradient hadn't perform reconstruction, respectively.

V. Conclusion

Accuracy of the reconstructed signal and the speed of convergence for the adaptive gradient and the l_1*-magic* algorithm are considered. Both of algorithms are from the class of convex optimization algorithms. Cases when the MAE of reconstructed signal was below 0.01 are considered as successful reconstruction cases. The adaptive gradient algorithm is very simple and has better accuracy in all cases when successful reconstruction was performed. The adaptive gradient algorithms had faster convergence in some regions, as well.

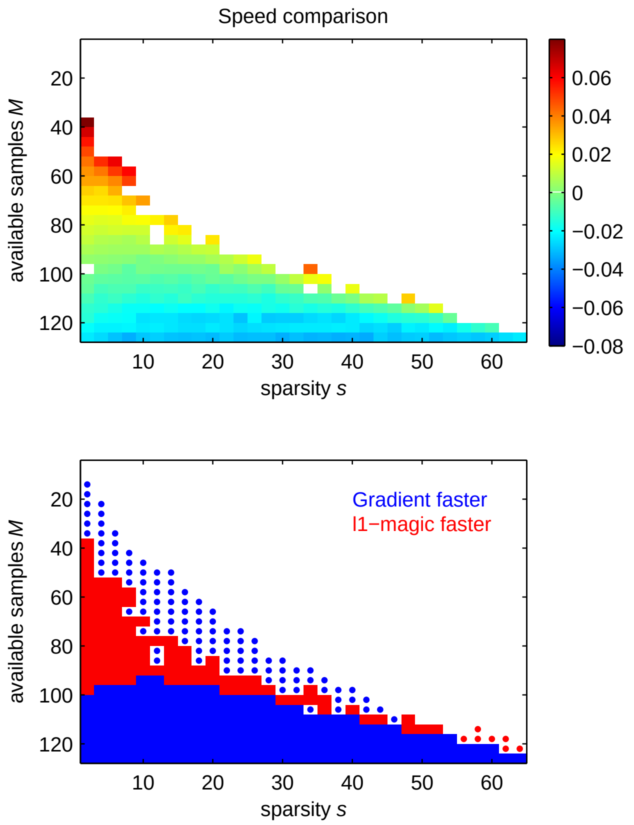

Fig. 3. Speed comparison of the adaptive gradient and the l_1-*magic* algorithms. Positive values on the top graphic correspond to the situation when the l_1-*magic* algorithm was faster. White color corresponds to the region where there was no reconstruction with none of algorithms in both cases.

References

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

[2] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, 51.12 (2005): 4203 – 4215.

[3] E. J. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[4] 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.

[5] C. Studer, P. Kuppinger, G. Pope, and H. B¨olcskei, "Recovery of sparsely corrupted signals," *CoRR*, abs/1102.1621, 2011.

[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*, vol. 4, no. 2, pp. 392–408, 2010.

[7] 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.

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

[9] 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.

[10] 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

[11] J.A. Tropp and S. J. Wright. "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE 98*, no. 6, (2010): 948 – 958.

[12] E. Candès and J. Romberg, "l1-magic: Recovery of sparse signals via convex programming," URL: http://users.ece.gatech.edu/~justin/l1magic/downloads/l1magic.pdf

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004

[14] LJ. 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).

[15] LJ. Stanković, M. Daković, and S. Vujović, "Reconstruction of Sparse Signals in Impulsive Disturbance Environments", *IEEE Transaction on Signal Processing*, submitted.

[16] LJ. Stankovic, S. Stankovic, 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

[17] S. Stankovic, I. Orovic and E. Sejdic, *Multimedia signals and Systems*, Springer, 2012.

[18] LJ. 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*, vol. 20, no. 5, pp. 499–502, 2013.

[19] LJ. Stanković, S. Stanković, I. Orović, and M. G. 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, 2013.

[20] LJ. Stanković, M. Daković, and T. Thayaparan, *Time-Frequency Signal Analysis with Applications*, Artech House, Boston, 2013.

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