Convexity of the ℓ_1–norm based Sparsity Measure with Respect to the Missing Samples as Variables

Miloš Brajović, Miloš Daković, Ljubiša Stanković
University of Montenegro
Podgorica, Montenegro
Email: {milosb, milos, ljubisa}@ac.me

***Abstract*—Sparse signal processing and the reconstruction of missing samples of signals exhibiting sparsity in a transform domain have been emerging research topics during the last decade. In this paper, we present the proof of the sparsity measure convexity, when considering the missing samples as minimization variables. The sparsity measure can be directly exploited in the reconstruction procedures, such as in the recently proposed gradient-based reconstruction algorithm. It makes the proof of sparsity measure convexity with respect to the missing samples as minimization variables especially interesting for signal processing. The minimal value of the sparsity measure corresponds to the set of missing sample values representing the sparsest possible solution, assuming that the reconstruction conditions are met. Convexity, along with recently presented proof of the uniqueness of the acquired solution, makes the gradient-based algorithm with missing samples as variables, a complete approach to the signal reconstruction. If the sparsity measure is convex, then we can guarantee that the solution corresponds to the global minimum of the sparsity measure, since the local minima do not exist in that case.**

Keywords*— *Compressive sensing; Concentration measure; Signal reconstruction; Sparse signal processing

I. INTRODUCTION

During the last decade, the sparse signal processing and the compressive sensing (CS) continuously attract a respective research interest [1]-[17]. These areas are related closely since the compressive sensing assumes the sparsity of the considered signal in a transformation domain [1]-[5], [13]. The basic aim, in general, is to recover all signal values from a reduced set of observations. The theoretical foundation of the CS and some representative missing samples reconstruction approaches are presented in [1]-[5], [7]-[12] and [14].

The reconstruction of missing samples is possible under the condition that the analyzed signal is sparse in a transform domain i.e. it has a concise representation with a small number of non-zero coefficients. Missing samples, or simply the reduced set of observations, may arise a consequence of a sampling strategy or their physical unavailability. In the CS area, the samples are omitted with the aim to reduce the data size. On the other side, signal samples may be intentionally omitted due to a high noise corruption, applying robust processing techniques such as the L-estimation [5], [15]. The signal reconstruction process can be interpreted as a procedure of finding the solution of an undetermined system of equations having the sparsest transform representation [1]-[15].

The sparsity can be measured by the ℓ_0-norm of the signal's transform, since it is directly related with the number of non-zero coefficients. However, as discussed in [5], [15] and [16] this norm is very sensitive to noise and small quantization errors, since every small value, even close to zero, equally contributes to the ℓ_0-norm as actual non-zero values. This norm does not allow simple gradient based minimization approaches since it not convex as well. These are the reasons why other norms, for instance the ℓ_1-norm, are exploited as sparsity measures. The idea of using the norms to measure the signal sparsity, i.e. the signal concentration in a transform domain was present for decades in the area of time-frequency signal analysis [18].

It is important to emphasize that the direct application of the ℓ_0-norm in minimization is an NP-hard problem. Relaxing the reconstruction constraint by involving the ℓ_1-norm instead of the ℓ_0-norm has opened the way to the application of linear programming approaches and methods in the reconstruction of missing samples. Representative reconstruction approaches include the convex optimization algorithms, for example, primal-dual interior point methods. Other approaches are the iterative procedures such as Orthogonal Matching Pursuit (OMP), Gradient Pursuit and CoSaMP [1]-[11].

For the topic analyzed in this paper, especially interesting is the gradient reconstruction algorithm presented in [5]. Namely, the reconstruction of missing samples is based on their variation using a steepest descent approach, lead by the sparsity measure. Since the available samples remain fixed during the reconstruction process, thus dictating the constraints for the sparsity measure minimization, the number of minimization variables is equal to the number of missing samples. The uniqueness of the solution is considered in detail in [6]. Detailed discussion and adaptations of this algorithm can be also found in [13], [15], while the initial ideas are presented in [16].

The convexity of the sparsity measure guaranties that the solution of the optimization corresponds to the global minimum of this function. Moreover, the convexity guaranties that the gradient approaches, i.e. converges to the solution corresponding to the global minimum of the optimization function. Herein, we provide a proof of the convexity of the sparsity measure with respect to the missing samples as variables.

In the next section, we provide the theoretical background concerning the sparsity measures, and the reconstruction by varying the missing samples values. As a representative approach which applies these concepts, gradient algorithm is revisited. Section III provides the proof of the sparsity measure convexity. Concluding remarks are given in the end of the paper.

II. Sparsity measures and the reconstruction in the signal domain

The sparse reconstruction can be formulated as a minimization problem, where the sparsity measure is minimized under the constraints defined by available samples. Let us observe the signal $x(n)$ of length N, with M available samples ($N-M$ missing samples) at random positions

$$n_i \in \mathbf{M} = \{n_1, n_2, ..., n_M\} \subset \mathbf{N} = \{0, 1, ..., N-1\}.$$

The set of available samples is given with $\mathbf{y_{cs}} = [x(n_1), x(n_2), ..., x(n_M)]^T$. Let $\mathbf{X} = [X(0), X(1), ..., X(N-1)]^T$ be the vector consisted of transform coefficients $X(k) = \mathrm{T}[x(n)]$. The case study analyzed in this paper is the Discrete Fourier Transform (DFT), and it is assumed that the signal has K non-zero transform coefficients, i.e. the sparsity K. The compressive sensing procedure based on the random selection/acquisition of signal values can be modeled by using a random measurement matrix $\mathbf{\Phi}$ as:

$$\mathbf{y_{cs}} = \mathbf{\Phi x} = \mathbf{\Psi \Phi X} = \mathbf{A_{cs} X}, \tag{1}$$

where $\mathbf{y_{cs}}$ denotes the vector of available samples of the analyzed signal. The matrix $\mathbf{A_{cs}}$ is obtained from the inverse DFT matrix $\mathbf{\Psi}$, by omitting the rows corresponding to the positions of missing samples. The solution of the sparsity minimization:

$$\min \sum_{k=0}^{N-1} |X(k)| \text{ subject to } \mathbf{y_{cs}} = \mathbf{A_{cs} X} \tag{2}$$

corresponds to the values of the missing samples. The function being minimized is the ℓ_1-norm of the DFT coefficients of the analyzed signal.

The simplest reconstruction procedure is a direct search over all unavailable samples, by minimizing the sparsity measure. In the other words, the missing samples can be observed as minimization variables, with fixed values of the available samples. Then, one can search over all possible values of missing samples, aiming to find the combination of their values which minimizes the sparsity measure. However, this is not a computationally feasible problem for a large number of samples. Thus, different CS reconstruction algorithms have been proposed. A simple and efficient procedure for the reconstruction of missing samples, with an arbitrary precision, has been proposed in [5].

A. Review of the gradient reconstruction algorithm

Here, the missing samples are also observed as minimization variables. Their values are varied until the minimum of the sparsity measure is reached. To this aim, the gradient of the sparsity measure is utilized to reach the solution which minimizes the optimization function. The procedure corresponds to the well-known steepest descent (gradient descent) method.

The algorithm starts from the initial signal $y^{(0)}(k)$ containing zeros at missing samples positions, and values of signal $x(n)$ at the positions of available samples:

$$y^{(0)}(n) = \begin{cases} 0 & \text{for } n \in \mathbf{N} \setminus \mathbf{M} \\ x(n) & \text{for } n \in \mathbf{M}. \end{cases}$$

The iterative procedure of the reconstruction algorithm can be summarized as follows:

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

$$y_1^{(k)}(n) = \begin{cases} y^{(k)}(n) + \Delta & \text{for } n \in \mathbf{N} \setminus \mathbf{M} \\ y^{(k)}(n) & \text{for } n \in \mathbf{M}. \end{cases}$$

$$y_2^{(k)}(n) = \begin{cases} y^{(k)}(n) - \Delta & \text{for } n \in \mathbf{N} \setminus \mathbf{M} \\ y^{(k)}(n) & \text{for } n \in \mathbf{M}. \end{cases}$$

where k is the iteration number. Constant Δ is used to determine whether the value of the considered signal sample should be decreased or increased [5], [15].

Step 2: Estimate the difference of the measures as

$$g(n_i) = \mathcal{M}\left[\mathrm{DFT}[y_1^{(k)}(n)]\right] - \mathcal{M}\left[\mathrm{DFT}[y_2^{(k)}(n)]\right]. \tag{3}$$

Step 3: Form a gradient vector $\mathbf{G}^{(k)}$ with the same length as the signal $x(n)$ as follows:

$$G^{(k)}(n) = \begin{cases} g_r(n) & \text{for } n \in \mathbf{N} \setminus \mathbf{M} \\ 0 & \text{for } n \in \mathbf{M}. \end{cases}$$

In other words, only the positions corresponding to the missing samples will be updated, while the remaining signal samples will be unchanged, thus retaining the conditions for the minimization.

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

$$y^{(k+1)}(n) = y^{(k)}(n) - \frac{1}{N} G^{(k)}(n).$$

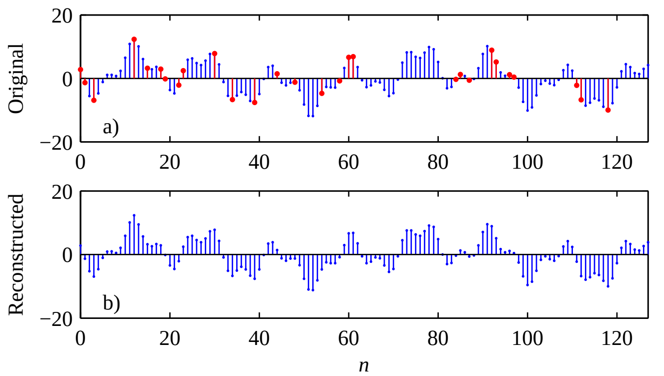

Fig. 1. The reconstruction using the gradient algorithm for signal with 100 missing samples at random positions - a) original signal (blue lines) and available samples (red dots); b) Reconstructed signal.

Repeating the presented iterative procedure, the missing values will converge to the true signal values, producing the minimal concentration measure in the transformation domain [15].

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 an arbitrary precision. Instead of moving toward the optimal point we will obtain oscillations, meaning that the gradient vector completely changes direction in subsequent iterations. This problem can be solved by introducing the concept variable step Δ. Namely, as it is proposed in [13] and [15], when the oscillations are detected, the step size is reduced, such that the arbitrary precision can be reached.

Example: Consider the signal $x(n) = 4\sin(4\pi n/128) + 3\cos(42\pi n/128 + pi/8) + 5.7\sin(240\pi n/128)$ of length $N = 128$ with $M = 28$ available samples, as shown in Fig. 1 a). The signal is sparse in the DFT domain. The result of the reconstruction by using the gradient algorithm is shown in Fig. 1 b).

III. The convexity of the sparsity measure

Consider again the signal $x(n)$ that is sparse in the DFT domain with sparsity K. Assume that the DFT coefficients at frequency positions $k \in \mathbf{K}$ are non–zero. The case when $Q = N - M$ values are missing in the signal $x(n)$, at random positions $q_i \in \mathbf{N}_Q = \mathbf{N} \setminus \mathbf{M},\ i = 0, ..., Q-1$ is considered. The deviations of the signal values at the positions can be modeled with the signal:

$$z(n) = \sum_{i=0}^{Q-1} z_i \delta(n - q_i). \tag{4}$$

Thus, the signal with missing samples can be represented as follows:

$$y(n) = x(n) + z(n), \tag{5}$$

whose DFT is $Y(k) = X(k) + Z(k), k = 0, ..., N-1$. Then the sparsity measure, i.e. the ℓ_1 norm of the DFT of the signal $y(n)$ equals to:

$$\mathcal{M}[Y(k)] = \sum_{k=0}^{N-1} |Y(k)| = \sum_{k=0}^{N-1} |X(k) + Z(k)| = \\ = \sum_{k=0}^{N-1} \left| X(k) + \sum_{i=0}^{Q-1} z_i e^{-j2\pi q_i k/N} \right|. \tag{6}$$

Aim is to prove the convexity of the this function. For the signal with $K = 3$ with two missing samples the sparsity measure is shown in Fig. 2 as the function of missing samples values. We start from the formal definition of the convexity. Consider the function $\mathcal{M} : \mathbf{Z^Q} \to \mathbf{R}$ defined on a convex set. Then, the function $\mathcal{M}(\mathbf{z})$ is convex if for any two points $\mathbf{z}_1 = (z_{1,0}, z_{1,1}, ..., z_{1,Q-1}) \in \mathbf{Z^Q}$ and $\mathbf{z}_2 = (z_{2,0}, z_{2,1}, ..., z_{2,Q-1}) \in \mathbf{Z^Q}$, and for any $0 < \lambda < 1$ holds:

$$\mathcal{M}[\lambda \mathbf{z}_1 + (1-\lambda)\mathbf{z}_2)] \le \lambda \mathcal{M}[\mathbf{z}_1] + (1-\lambda)\mathcal{M}[\mathbf{z}_2]. \tag{7}$$

Left part of (7) can be expressed by using the concentration measure definition in (6). To this aim, let us analyze (6) with more details. It can be easily concluded that (6) has the the largest possible value for given point $\lambda \mathbf{z}_1 + (1-\lambda)\mathbf{z}_2 = (z_1, z_2, ..., z_{Q-1})$ if, for every $k \in \mathbf{K}$, phases of $X(k)$ and $\sum_{i=0}^{Q-1} z_i e^{-j2\pi q_i k/N}$ are the same. Then, the measure (6) becomes:

$$\mathcal{M}(\mathbf{z}) = \sum_{k=0}^{N-1} \left| X(k) + \sum_{i=0}^{Q-1} z_i e^{-j2\pi q_i k/N} \right| \le \\ \sum_{k\in\mathbf{K}} \left| X(k) + \sum_{i=0}^{Q-1} z_i e^{-j2\pi q_i k/N} \right| + (N-K)\sum_{i=0}^{Q-1} |z_i| \le \\ \sum_{k\in\mathbf{K}} \left(|X(k)| + \sum_{i=0}^{Q-1} |z_i| \right) + (N-K)\sum_{i=0}^{Q-1} |z_i| = \\ \sum_{k\in\mathbf{K}} |X(k)| + N\sum_{i=0}^{Q-1} |z_i|. \tag{8}$$

Now consider the convexity condition (7) for the function (6). According to the previous analysis, left part of (7) will have the largest value when the norm (6) is calculated as (8).

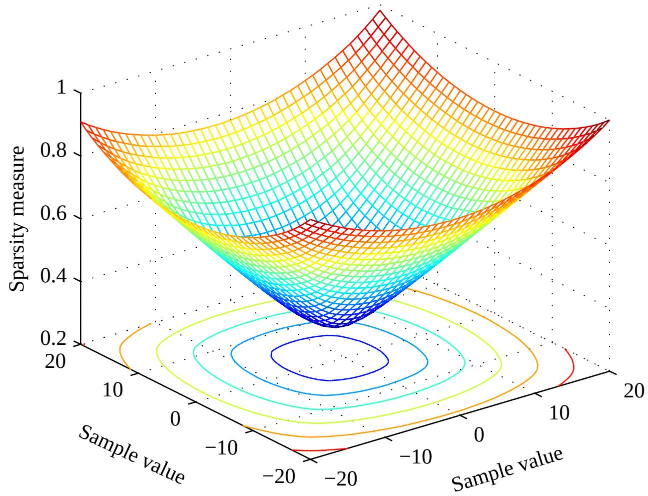

Fig. 2. Sparsity measure as a function of missing samples values

This is the worst case. The convexity condition now becomes:

$$\sum_{k\in\mathbf{K}}|X(k)|+N\sum_{i=0}^{Q-1}|\lambda z_{1,i}+(1-\lambda)z_{2,i}|\leq$$
$$\leq\lambda\left(\sum_{k\in\mathbf{K}}|X(k)|+N\sum_{i=0}^{Q-1}|z_{1,i}|\right)+$$
$$+(1-\lambda)\left(\sum_{k\in\mathbf{K}}|X(k)|+N\sum_{i=0}^{Q-1}|z_{2,i}|\right)$$

After simple rearrangement of the right part, it finally follows:

$$\sum_{k\in\mathbf{K}}|X(k)|+N\sum_{i=0}^{Q-1}|\lambda z_{1,i}+(1-\lambda)z_{2,i}|\leq$$
$$\sum_{k\in\mathbf{K}}|X(k)|+N\lambda\sum_{i=0}^{Q-1}|z_{1,i}|+N(1-\lambda)\sum_{i=0}^{Q-1}|z_{2,i}| \qquad (9)$$

The left side of (9), according to the triangle inequality, has the largest possible value $|\lambda z_{i,1}+(1-\lambda)z_{i,2}|=\lambda\,|z_{i,1}|+(1-\lambda)\,|z_{i,2}|$ for given i, and in the worst case this holds for every i. In that case, left and right side of (9) become equal, which proves the fact that the convexity condition is satisfied.

IV. Conclusion

The ℓ_1 norm of the signal transform coefficients is commonly used as a sparsity measure. Since the reconstruction of missing samples of signals sparse in a transform domain can be done by varying the missing samples values in the signal domain as, for instance, in the reviewed gradient algorithm, the proof of the convexity of the sparsity measure in the context of missing sample values as variables is very important. In this paper, we prove the convexity of the sparsity measure, and revisit the gradient reconstruction algorithm. The case study transform analyzed in this paper is the DFT.

Acknowledgment

This work is supported by the Montenegrin Ministry of Science, project grant: "New ICT Compressive sensing based trends applied to: multimedia, biomedicine and communications".

References

[1] E. Candès, J. Romberg, T. Tao: "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,"*IEEE Trans. on Information Theory*, 2006, 52, (2), pp. 489–509.

[2] E. J. Candès, M. B. Wakin, "An Introduction To Compressive Sampling", *IEEE Signal Processing Magazine*,Vol. 25, Issue 2, pp. 21–30, 2008.

[3] D. Donoho: "Compressed sensing," *IEEE Trans. on Information Theory*, 2006, 52, (4), pp. 1289–1306

[4] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, 24, (4), pp. 118–121, 2007.

[5] 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.

[6] LJ. Stanković, and 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

[7] M. Elad, *Sparse and Redudant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010

[8] H. Rauhut, "Stability Results for Random Sampling of Sparse Trigonometric Polynomials," *IEEE Trans. on Information theory*, 54(12), pp. 5661–5670, 2008

[9] C. Studer, P. Kuppinger, G. Pope, H. Bolcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. on Information Theory*, vol.58, no.5, pp. 3115–3130. , 2012.

[10] M. Davenport, M. Duarte, Y. Eldar, G. Kutyniok, "Introduction to compressed sensing, "*Chapter in Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.

[11] 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

[12] 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

[13] M. Daković, LJ. Stanković, and I. Orović, "Adaptive Gradient Based Algorithm for Complex Sparse Signal Reconstruction, "*22nd Telecommunications Forum TELFOR*, Belgrade, 2014

[14] S. Stanković, I. Orović, and E. Sejdić, *Multimedia signals and systems*, Springer–Verlag, 2012

[15] LJ. Stanković, *Digital Signal Processing with Selected Topics*, CreateSpace Independent Publishing Platform, An Amazon.com Company, November 4, 2015

[16] LJ. Stanković, M. Daković, and S. Vujović, "Concentration measures with an adaptive algorithm for processing sparse signals, "*ISPA 2013*, Trieste, Italy, 4-6 September 2013, pp. 418–423

[17] 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

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