Concentration Measures with an Adaptive Algorithm for Processing Sparse Signals

Ljubiša Stanković, Miloš Daković, Stefan Vujović
Faculty of Electrical Engineering, University of Montenegro, Podgorica, Montenegro
E-mail: ljubisa@ac.me, milos@ac.me, stefanv@ac.me

***Abstract*—In the L-estimation and compressive sensing some arbitrarily positioned samples of the signal are either so heavily corrupted by disturbances that it is better to omit them in the analysis or they are unavailable. If the considered signal with missing samples is sparse then we are still able to reconstruct these samples by using the well know reconstruction algorithms. In this paper we will illustrate different measures for the signal concentration and propose a simple adaptive algorithm, applied on these measures, without reformulating the reconstruction problem within the standard linear programming form. Direct application of the gradient on nondifferentiable forms of measures lead to an efficient variable step size algorithm. The results are illustrated on the examples.**

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

I. Introduction

In many signal processing applications a signal that spans over the whole time domain is located within much smaller regions in a transformation domain. Then it is said that this signal is sparse in this transformation domain. The most common case of such a signal is a sum of discrete time sinusoidal signals, when the number of sinusoids is much smaller than the number of signal samples in the time domain. For this kind of signals we do not need all samples in the time domain to reconstruct much smaller number of samples in the frequency domain. Of course, the Fourier domain is just one of possible domains to transform a signal. The samples could be missed due to their physical or measurements unavailability (like in compressive sensing). In applications it could happen that some arbitrarily positioned samples of the signal are so heavily corrupted by disturbances that it is better to omit them in the analysis (for example, using the L-estimation). In both cases the signals could be considered within the framework of missing samples. If the considered signal is sparse then we will still be able to reconstruct the missing/omitted samples. Thus, under some conditions, the processing could be performed with the remaining samples almost as in the case if missing samples were available. Of course, a priori information about the nature of the analyzed signal, its sparsity in a known domain, is used.

Several approaches to the analysis of this kind of signals (based on gradient, homotopy, norm-one (l_1 magic), iterative thresholding, matching pursuit,...) are introduced [1]-[17]. Sparsity of a signal in a transformation domain is related to the number of nonzero samples in that domain. A natural mathematical tool to measure the number of nonzero (significant) samples in a signal transform is the so called zero-norm. However, the zero-norm is very sensitive to any kind of disturbance, more robust norms are used.

In this paper we will first analyze common concentration measures with respect to their possible applications to the processing of sparse signals with missing/omitted samples. Then, a gradient based algorithm for the determination of a large number of unavailable samples is presented. The proposed method belongs to the class of gradient based CS algorithms [5]. However, here the minimization problems is not reformulated within the standard linear programming framework based on the norm-two (l_2) forms. The common adaptive signal processing and the CS algorithms avoid direct use of the measure based on the norm-one, since it is not differentiable and the value of gradient can not be used as measure of the proximity of the algorithm solution. When the iterations are close to the optimal point, gradient value remains the same and oscillate around the true value. Taking sufficiently small step over the whole range would not be a solution, due to extremely large number of iterations over a very large set of variable. This method would not be more efficient than the direct search based one. Here, we present a simple algorithm, applied directly to the appropriately chosen concentration measure. In examples we used norm-one based measure. Since the derivatives are not continuous functions around the minimum, in this algorithm, a variable and self-adaptive step is introduced. The algorithm applied on the measure based on norm-one, with adaptive step, reconstructs a large number of missing samples in a computationally efficient way with arbitrary (computer defined) precision of the results.

The paper is organized as follows. After the introduction, a review and analysis of concentration measures in the processing of sparse signals is done. A gradient based algorithm, with its modifications is presented and illustrated.

II. Measures

Concentration measures of signal transforms were intensively studied in the area of time-frequency analysis. They are used to find an optimal, best concentrated signal representation. Two of these concentration measures will be shortly reviewed before we proceed with the analysis of measure concentration form to reconstruction of missing signal samples.

The most common and the oldest measure introduced to measure concentration of time-frequency representations was in the form of the ratio of the fourth to the second-order norms of the short-time Fourier transforms (defined and used by Jones, Parks, Baraniuk, Flandrin, Williams, *et al.*). In terms of the DFT it reads

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

In general, it has been shown that any other ratio of norms $l_p = \sum_k |X(k)|^p / N$ and l_q, $p > q > 1$, can also be used for measuring the concentration. This kind of concentration measures were inspired by the kurtosis as a measure of distribution peakedness. Similar forms are obtained by using the Rényi measures. When there are two or more components of approximately equal energies, whose concentrations are very different, this norm-based measures will favor the distribution with a "peaky" component, due to raising of distribution values to a high power. It means that if one component is extremely highly concentrated, and all the others are very poorly concentrated, then the measure will not look for a trade-off, when all components are well concentrated. In order to deal with this kind of problems, common in signal analysis, this kind of concentration measures are later defined and applied to smaller, local transformation regions.

Another direction to measure time-frequency representation concentration comes from a classical definition of the time-limited signal duration, rather than measuring signal peakedness. It was used in time-frequency analysis in [21]. If a signal $x(n)$, is time-limited to the interval $[n_1, n_2 - 1]$, that is, $x(n) \neq 0$ only for $n \in [n_1, n_2 - 1]$, then the duration of $x(n)$ is $d = n_2 - n_1$. It can be written as

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

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

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

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

with $p > 1$, where N is total number of samples in signal transform $X(k)$. A lower value of (3) indicates better concentrated distribution. For $p = 1/2$ measure $\mathcal{M}_{1/2}[T[x(n)]]$ reduces to norm 2. For $p = 1$, it is the norm-one form

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

In the next section we will illustrate the influence of measure parameter p on the results.

III. Direct Reconstruction and Measures Performance

The simplest reconstruction algorithm will be based on a direct search over all missing samples values. If we consider the complete set of signal samples

$$\{x(1), x(2), ..., x(N-1)\}$$

and M samples $x(m_1)$, $x(m_2)$,...,$x(m_M)$ are missing then the simplest algorithm will be to search over all possible values of missing samples and find solution that minimize used concentration measure

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

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

One possible approach to reduce the number of operations in the direct search is to use a large step in the first, rough, estimation, then to reduce the step around the rough estimate of $x(m_1)$, $x(m_2)$,..., $x(m_M)$. This can be repeated few times, until the desired accuracy is achieved. For example, for $A = 1$ the accuracy of 0.001 achieved with $L = 2001$. With 7 missing samples that would mean unacceptable number of $2001^7 \sim 128 \times 10^{21}$ measure calculations. However, if the first search is done with $L = 5$, the rough optimal is found, and the procedure is repeated with $L = 5$ values within range determined by rough optimal and previously used step. Repeating the same procedure six more times, the accuracy better than 0.001 is reached with $7 \times 5^7 \sim 6 \times 10^5$ measure calculations. In this way, we were able to analyze (on an ordinary PC, within a reasonable calculation time), signals with up to 10 missing samples.

Although, computationally not efficient, the direct method is very important and helpful in the analysis of various concentration measures with different p, since all more efficient methods produce results with nice values of p only (for example, $p = 1$, $p = 1/2$, or $p = 2$). The direct method can be used with any p. Also, the probability that we find a local minimum is lower in the direct method than when using, for example, the gradient based algorithms.

Example: Consider a discrete signal

$$x(n) = 2.5 \sin(20\pi n/N) \tag{4}$$

for $n = 0, 1, \ldots, N-1$, and $N = 256$ is number of signal samples. The cases with one and two missing samples are analyzed. Direct search is performed over range $[-5, 5]$ with step 0.01. We calculated measure (3) for various values of parameter p. Results are shown in Fig. 1 for one, and Fig. 2 for two missing samples. The measure minimum is located on the true sample values for $p \geq 1$. Case with two missing samples and $p = 1$ is presented in Fig. 3.

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

$$\begin{aligned} x(n) =& 3 \sin(20\pi n/N) + 2 \cos(60\pi n/N) \\ &+ 0.5 \sin(110\pi n/N). \end{aligned} \tag{5}$$

Signal is composed of $N = 256$ samples while the cases of 4 and 7 missing samples are analyzed. The results with 10 and 15 iterations are presented in Fig. 4. We can see that the error for $p \geq 1$ (norms l_q with $q \leq 1$) produce accurate results with MAE depending only on the direct search step. For $p < 1$ (norms l_q with $q > 1$) the bias dominate over the number of iterations, so the results are almost independent from number of iterations. We see that almost same results are obtained for 4 and 7 missing samples cases.

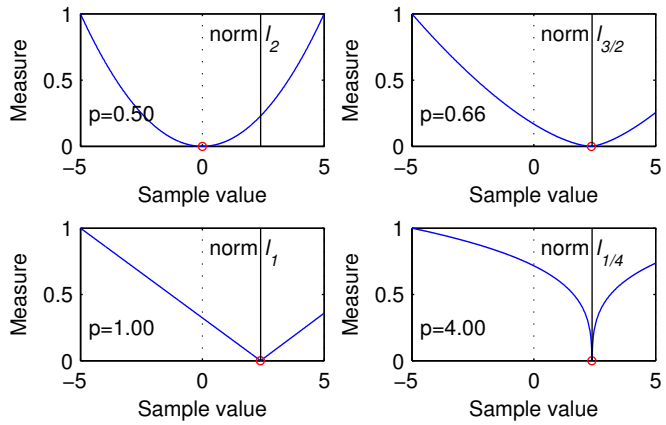

Fig. 1. Measure as function of missing sample value for various p. True value of missing sample is presented with vertical line while measure minimum is denoted by circle. Measure values are normalized.

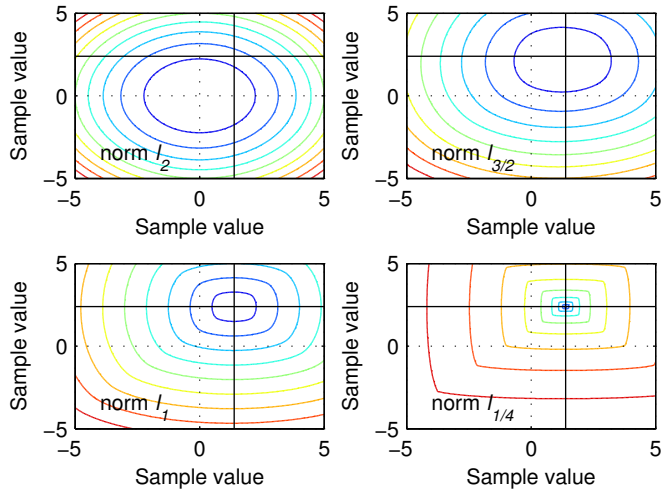

Fig. 2. Measure as function of missing sample values for various p. True values of missing samples are presented with black lines

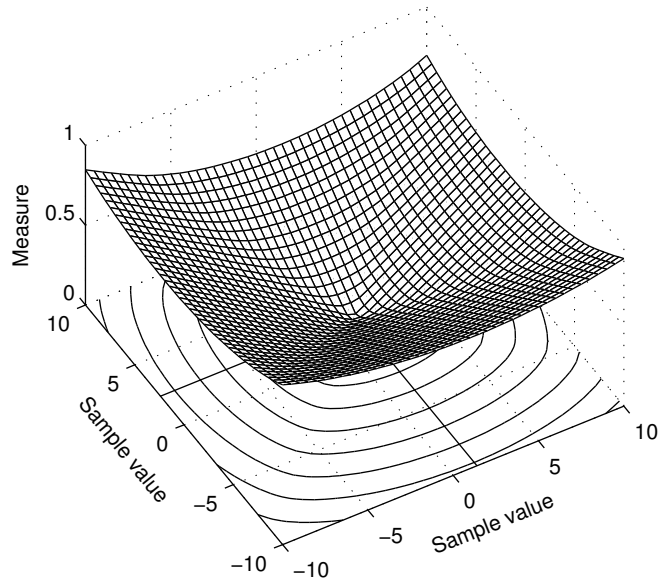

Fig. 3. Measure for $p = 1$ as function of missing sample values

For $p = 0.5$ this measure is equivalent with well-known l_2 norm. For the norm-two (l_2 norm with $p = 1/2$) the MAE is of the signal samples order. As shown in Fig. 1 and Fig. 2 this measure has minimum when the missing signal samples values are set to zero. This result was expected because of the Parserval's theorem stating that the energy of a signal in the time domain is same as the energy in the frequency domain. We know that signal has the lowest energy when its samples are zero-valued. The same holds in the frequency domain. The minimization solution with the l_2 norm is trivial. With this norm, we attempt to minimize

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

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

IV. Adaptive Gradient Based Algorithm

Due to high computational complexity the direct search could be used only if number of missing samples M is small enough. Alternative approach based on concentration measure gradient is presented. Algorithm is form of gradient descent algorithm where missing samples are estimated as the ones producing best concentration in sparse domain. Note that the norms that produce unbiased missing samples values (like for example norm $p = 1$) are not differentiable around the optimal point.

It means that the gradient method directly applied to the measure based on l_1 norm will be able to approach the optimal point, but since the gradient amplitude in the vicinity of the optimal point is almost constant (with changing sign), the algorithm will not improve the accuracy to a level lower than the one defined by the step in the gradient algorithm. This is the reason why appropriate reformulation of the norm-one problem is done within linear programming by using well known and widely used norm-two solutions. Here, we will present an algorithm that is directly applied to the norm-one based measures. As we can see from Fig. 3 measure with $p = 1$ is differentiable and convex everywhere except around the point of minimum (the optimization solution point). Therefore any algorithm applied directly to the measure based on $p = 1$ will oscillate around the solution with amplitude defined by the step and measure form. If we take a very small step for each of a large number of missing samples, it will result in an unacceptable and large number of iterations. Thus, when the steady oscillatory state (steady state in mean) is detected we should reduce the algorithm step, as in the direct search. In order to produce results with a high accuracy, within an acceptable number of iterations, we will present a variable step version of the algorithm, with a possibility of a self-adaptive step. This method will produce the results with an error of the computer precision level. Finally, in addition to the step variation the parameter p (the norm itself) can be changed to improve the initial convergence of the algorithm.

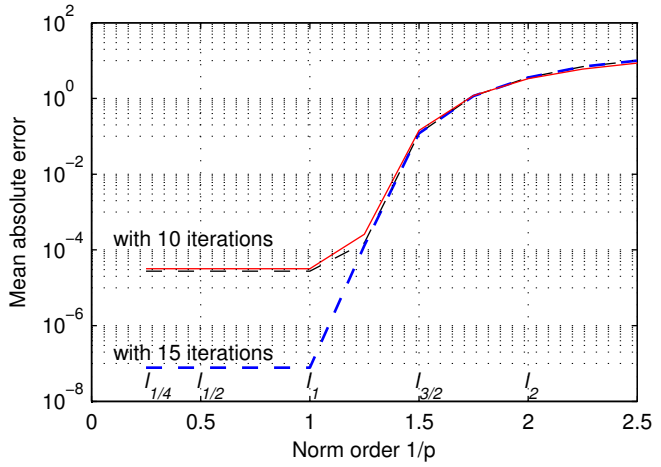

Fig. 4. Mean absolute error in the coefficients estimation as function of concentration measure parameter (various norms l) for 4 missing samples (dashed line) and for 7 missing samples (solid line). MAE is normalized with number of missing samples.

Consider discrete signal $x(n)$ where some samples are unavailable. Assume that signal is sparse in transformation domain T. Algorithm for missing samples reconstruction is implemented as follows:

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

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

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

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

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

where k is iteration number. Constant Δ is used to determine whether sample could be decreased or increased.

Step 2: Estimate differential of measure as

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

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

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

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

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

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

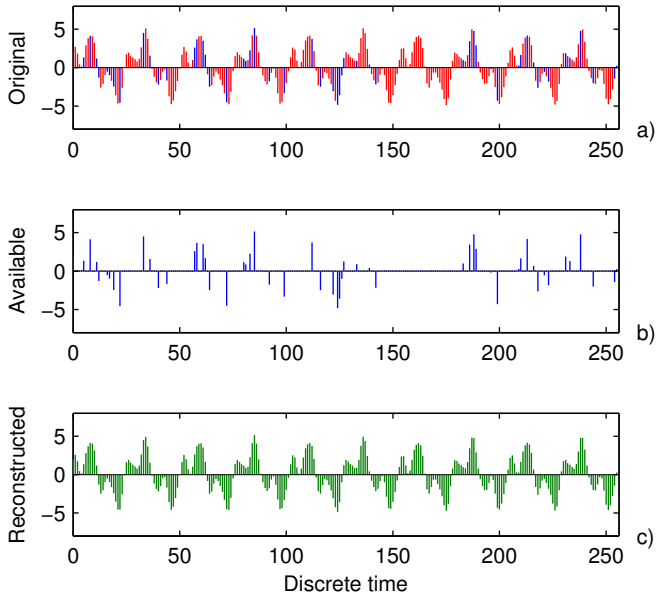

Fig. 5. Reconstruction example for signal with 200 missing samples. (a) original signal; (b) signal with missing samples set to 0 used as input to reconstruction algorithm; (c) reconstructed signal.

Repeating presented iterative procedure, missing values are going to converge to real values. Algorithm performance depend on parameters μ and Δ. Here we use $p \approx 1$.

Varying and Adaptive Step Size: Since we use a difference of 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 precision higher than the step μ multiplied by constant, gradient dependent value. If we try to reduce this bias by using smaller step, then we will face with an unacceptable number of iterations. However, this problem may be solved, by reducing the step size, when we approach the stationary bias zone. By defining self adaptive procedure for reducing the step size, we may achieve error of computer precision order, within a small number of iterations. One possible approach is that for each iteration we calculate D_k as a mean absolute value of $\mu G(n)$ and reduce $\mu^{(k+1)} = \mu^{(k)}/5$ and $\Delta^{(k+1)} = \Delta^{(k)}/5$ whenever small change, $|D_k - D_{k-1}| < 0.025|D_{k-1}|$, is detected.

Example: Consider the signal (5). Signal is composed of $N = 256$ samples while 200 samples are missing or not available. We know their positions, as well as that the signal is sparse in the Fourier domain. Reconstruction is done by the presented algorithm. The reconstruction results are shown in Fig. 5.

Let us now analyze parameters Δ, μ, p and number of iterations in the proposed algorithm for considered example. Mean absolute error calculated as

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

is shown in Fig. 6 for various algorithm setups. It can be concluded that for constant algorithm parameters MAE cannot be improved below some limit by increasing number of iterations. Smaller values of Δ and μ produce lower MAE

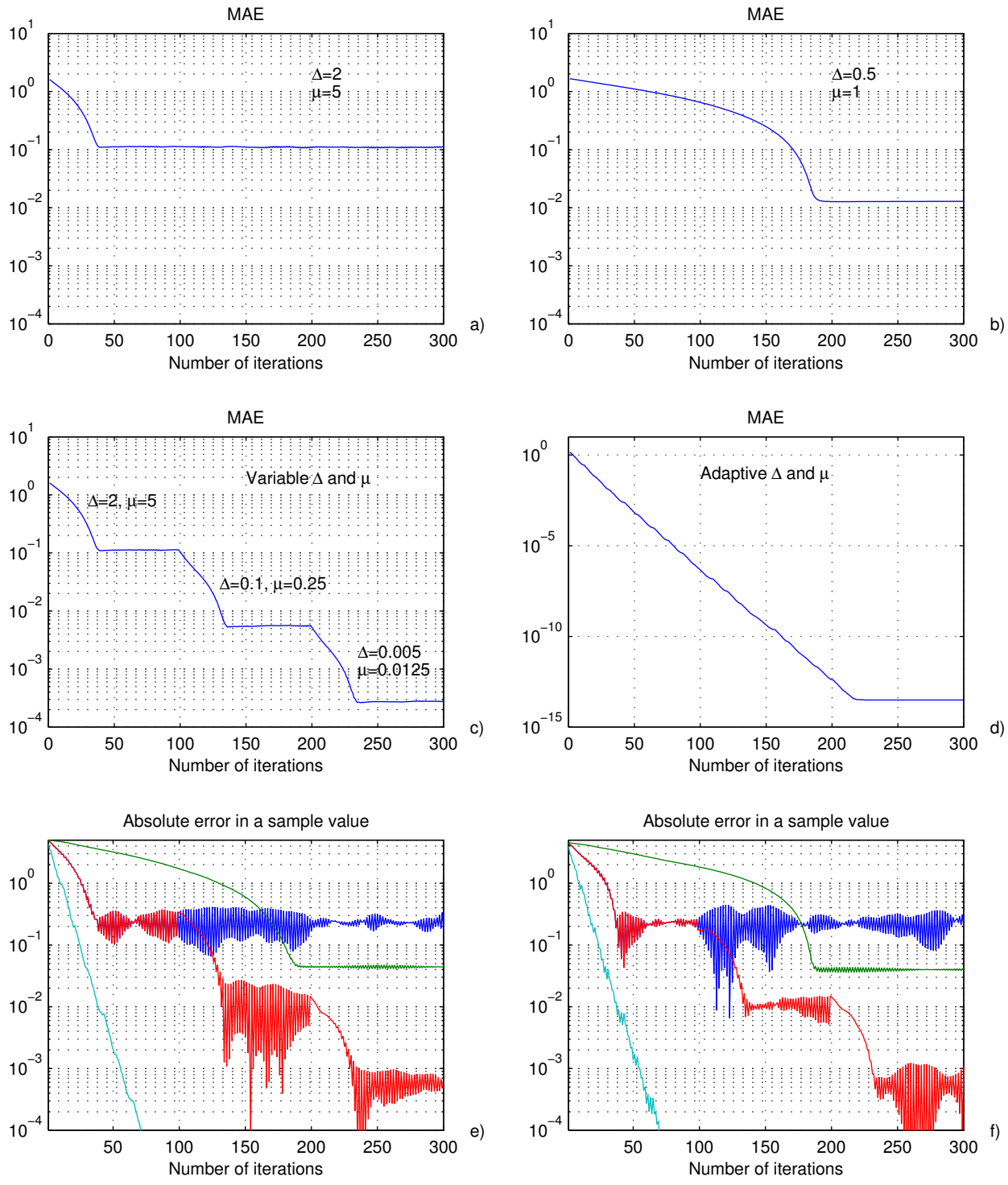

Fig. 6. Mean absolute reconstruction error for constant algorithm parameters Δ and μ (a) and (b); variable parameters and adaptive parameters (c) and (d); absolute errors for two random missing signal samples (e) and (f) for constant algorithm parameters (green and blue line), variable (red line), and adaptive parameters (bright blue).

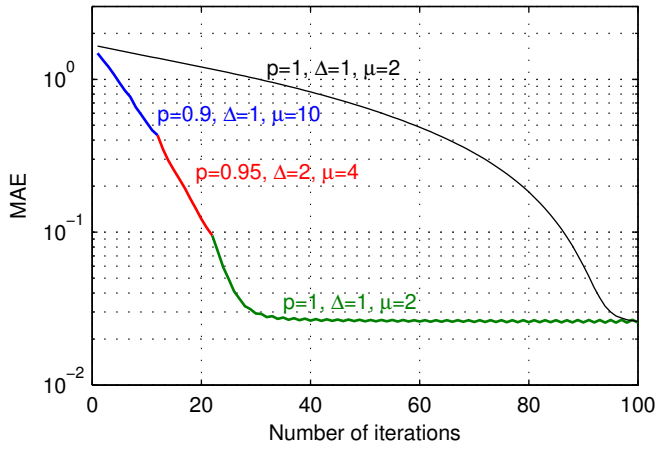

Fig. 7. Reconstruction mean absolute error for constant and varying algorithm parameters.

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

Number of iterations for required accuracy can be further improved by varying measure parameter p. Measures for $p < 1$ are more suitable to gradient based reconstruction. However measures for $p < 1$ do not converge to the true values of missing samples. A possible solution is to use measures with $p < 1$ at the beginning of iterative algorithm and to switch to $p = 1$ afterwards. Figure 7 illustrate the case when $p = 0.9$, $\Delta = 1$ and $\mu = 10$ is used for iterations 1 to 12, $p = 0.95$, $\Delta = 2$ and $\mu = 4$ is used for iterations 13 to 22 and finally $p = 1$, $\Delta = 1$ and $\mu = 2$ is used for iterations form 23 to 100. The case with constant parameters $p = 1$, $\Delta = 1$ and $\mu = 2$ is presented in the same figure. It is notable that the algorithm with varying parameters converges faster.

V. Conclusion

In this paper we have analyzed the signal concentration measures application to missing samples reconstruction problem. It has been shown that the measures close to norm-one can be efficiently used. In addition to the analysis of direct search, an algorithm for the signal reconstruction is presented and analyzed. Various setups of the algorithm parameters are considered including: constant, varying and self-adaptive parameter selection. It has been shown that the algorithm convergence can be significantly improved by using varying measure order.

References

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

[2] 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

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

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

[5] 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

[6] 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.

[7] 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.

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

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

[10] 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.

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

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

[13] 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.

[14] S. Stanković, I. Orović and M.G. Amin, "Compressed Sensing Based Robust Time-Frequency Representation for Signals in Heavy-Tailed Noise," *ISSPA* 2012, Canada

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

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

[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] 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.

[19] D. Donoho, "Denoising by soft thresholding," *IEEE Trans. on Information Theory*, vol. 41, pp. 613–627, May 1995.

[20] D. Donoho and J. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, September 1994.

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

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

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