ON COMPRESSIVE SENSING IN AUDIO SIGNALS

Irena Orović, Srdjan Stanković, Branka Jokanović, Andjela Draganić, *irenao@ac.me*
University of Montenegro, Faculty of Electrical Engineering

Abstract – *The reconstruction of musical audio signal by using the Compressive Sensing technique is presented in the paper. Compressive Sensing can provide significant reduction of number of samples required by Shannon-Nyquist theorem. By reducing the number of samples, data compression is achieved along with the data acquisition. In this paper, discrete cosine transform is used in the process of reconstruction. The optimization problem is solved by the primal-dual algorithm. Theory is illustrated by experimental results. Experiments are performed by using musical signal and analysis for number of measurements versus mean absolute error is given as well.*

1. INTRODUCTION

Standard methods for signal sampling are based on the Shannon-Nyquist theorem. According to this theorem, signal must be sampled with frequency at least two times higher than maximum signal frequency. However, many real signals require high Nyquist rate. Sampling such signals results in large number of signal samples. Storing and transmitting these samples through communication channel is a demanding task. Since communication channels have limited bit rate, it is necessary to perform signal compression.

For discrete-time domain signal, two steps can be distinguished in traditional compression algorithms. First is the Shannon-Nyquist sampling of signal in the time domain. Second step involves finding a transform domain where a large number of signal coefficients are close to zero while only few have large values. This distribution of coefficients is called a sparse representation of signal in the transform domain. This fact is used for transform coding based compression. Although the compression algorithms reduce the total amount of data, a large number of samples is processed by the coder in the first step of compression.

Compressed Sensing (CS) method, as an alternative way for signal acquisition, has been used [1]-[7]. It is shown that the number of samples used for signal reconstruction, can be significantly smaller than the number of samples required by the sampling theorem. Instead of collecting large amounts of signal samples and eliminating most of them in the process of compression, CS allows simultaneous acquisition and data compression. The samples in the time domain, which are used in CS technique, are called measurements.

CS is based on two conditions: the signal should have sparse representation in certain transform domain and the measurements have to be incoherent [2].

The majority of the real signals is sparse, which means that most of the information is contained in significantly smaller number of coefficients, compared to the total length of the signal. Incoherence is necessary in order to obtain maximum information from minimum number of independent measurements.

CS approach can be useful in many applications [8]-[17]. The reduction of sampling rate could be used in broadband communications monitoring systems which have large frequency range or in radar systems where Nyquist sampling is not often feasible or is very expensive. Possible application of CS is in image processing (the Compressive Imaging). Another field where CS finds its applicability is medicine. Magnetic Resonance Imaging (MRI) requires high image resolution. Sampling such images with Nyquist frequency would take a significant amount of time. It is recommended to minimize the time exposure of patient to MRI device. This can be achieved by using CS procedure.

The aim of this paper is to apply CS method on musical audio signal. For this purpose, in the process of reconstruction, discrete cosine transform domain and primal dual algorithm are used.

This paper is organized as follows: In Section II basic concepts of CS method are given. The analysis of CS technique for the reconstruction of musical signal is given in Section III. Section IV provides simulation results and the error analysis. Conclusion is given in Section V.

2. THEORETICAL BACKGROUND

Let us consider a discrete–time signal x which can be represented as Nx1 vector. Using a base ψ, signal can be represented as linear weighted sum of basis vectors $\psi_1, \psi_2, ..., \psi_N$:

$$x = \sum_{i=1}^{N} s_i \psi_i . \tag{1}$$

s_i denotes the weighting coefficient vector. Previous relation can be written as:

$$\mathbf{x} = \boldsymbol{\psi}\mathbf{s}, \tag{2}$$

where $\boldsymbol{\psi}$ is NxN transform matrix whose columns are basis vectors. Basis matrix $\boldsymbol{\psi}$ could be Fourier matrix, matrix of wavelet coefficients, discrete cosine transform matrix, etc. If K ($K < N$) signal coefficients from transform domain have large non-zero values, then signal can be considered as K-sparse in the transform domain.

CS approach provides signal reconstruction from a small number of measurements M, by using the fact that the signal is sparse in the transform domain. Procedure of signal acquisition is defined in such a way that important information of the signal is well preserved, despite the dimensionality reduction ($M<N$). If $\mathbf{y}$ denotes measurement vector, then the equation:

$$\mathbf{y}_{M\text{x}1} = \boldsymbol{\phi}_{M\text{x}N}\mathbf{x}_{N\text{x}1}, \tag{3}$$

holds, where $\boldsymbol{\phi}$ denotes measurement matrix. Based on (2) and (3), we get:

$$\mathbf{y} = \boldsymbol{\phi}\mathbf{x} = \boldsymbol{\phi}\boldsymbol{\psi}\mathbf{s} = \boldsymbol{\theta}\mathbf{s}. \tag{4}$$

Reconstruction of signal x requires solving system of M equations with N unknowns. This system is undetermined

($M < N$), and has infinite number of solutions. In order to obtain optimal solution, optimization algorithms are used.

Measurement procedure should satisfy certain conditions, in order to obtain successful reconstruction of signal. Firstly, the measurement matrix $\boldsymbol{\phi}$ must be incoherent with the basis matrix $\boldsymbol{\psi}$. The coherence between two matrices measures the largest correlation between any two elements of matrices. If elements of the two matrices are correlated, coherence is high. The measure of correlation between two matrices is defined as follows:

$$\mu(\phi,\psi) = \sqrt{N} \max_{k\geq 1, j\leq N} \left| \left\langle \phi_k , \psi_j \right\rangle \right|, \tag{5}$$

where N is the signal length, ϕ_k and ψ_j are row vector and column vector of the $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ matrices, respectively. Coherence has values in the range:

$$1 \leq \mu(\phi,\psi) \leq \sqrt{N}. \tag{6}$$

CS approach deals with matrices whose coherence is small. The minimal coherence is achieved for $\mu(\phi,\psi)=1$. Lower coherence between matrices $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ means higher incoherence i.e. smaller number of samples required in the measurement process.

If the incoherence property is satisfied, the number of the required measurements can be obtained as:

$$M \geq cK \log(N / K), \tag{7}$$

where c is a constant. The case of interest is when the number of the required measurements is much smaller than the length of the signal. A related property to incoherence is the *Restricted Isometry Property* (RIP). This property means that any subset of columns of the matrix $\boldsymbol{\theta} = \boldsymbol{\phi\psi}$ should be nearly orthogonal. RIP can be defined as:

$$(1-\delta_K)\|\mathbf{s}\|_{\ell 2}^2 \leq \|\boldsymbol{\theta}\mathbf{s}\|_{\ell 2}^2 \leq (1+\delta_K)\|\mathbf{s}\|_{\ell 2}^2, \tag{8}$$

where δ_K is isometry constant. Signal can be reconstructed with high probability if incoherence property and RIP are satisfied. Since random matrices satisfy both conditions, they are used in the measurement process.

The reconstruction of signal is performed by using the optimization algorithms. There is a number of optimization techniques for finding the sparsest solution of system (3). In addition, reconstruction procedure should be robust. This means that signal should be reconstructed with high probability, even in the presence of noise or when it is not perfectly sparse. It is shown that optimal results provide optimization techniques for l_1-minimization. The optimization problem is defined as:

$$\hat{x} = \min \|\mathbf{x}\|_{l_1} \; \textit{subject to}\; \mathbf{y} = \boldsymbol{\phi}\mathbf{x}, \tag{9}$$

where $\hat{x}$ is solution of the minimization problem and $\|\mathbf{x}\|_{l_1} = \sum_{i=1}^{N} |x_i|$ is l_1-norm of vector $\mathbf{x}$. In the case of noisy signals, (9) is modified as:

$$\min \|\mathbf{x}\|_{\ell 1} \;\; \textit{subject to}\; \|\mathbf{y}-\mathbf{Ax}\|_{\ell 2} \leq \varepsilon, \tag{10}$$

where $\mathbf{y} = \boldsymbol{\phi\psi}\,\mathbf{x} + e$, e is an observation error and $\|e\|_{\ell 2} = \varepsilon$.

3. CS BASED RECONSTRUCTION OF MUSICAL AUDIO SIGNAL

The reconstruction of audio signals of high quality plays an important role in various applications. However, there are constraints such as limited bandwidth and limited storage capacities. In this part, we consider the musical audio signal reconstruction by using CS approach.

Musical signal can be represented as a summation of small number of sinusoids. This implies the sparsity of musical signal in the frequency domain. To transform the musical signal in the frequency domain, discrete cosine transform or Fourier transform can be used. Here, we consider musical signal of N=3000 samples. We assume the non-noisy environment. The discrete cosine transform is used for the frequency representation. Basis matrix $\boldsymbol{\psi}$ is defined as:

$$\boldsymbol{\psi} = \begin{bmatrix} \frac{1}{\sqrt{N}} & \sqrt{\frac{2}{N}}\cos\frac{\pi}{2N} & \cdots & \sqrt{\frac{2}{N}}\cos\frac{2998\pi}{2N} & \sqrt{\frac{2}{N}}\cos\frac{2999\pi}{2N} \\ \frac{1}{\sqrt{N}} & \sqrt{\frac{2}{N}}\cos\frac{3\pi}{2N} & \cdots & \sqrt{\frac{2}{N}}\cos\frac{8994\pi}{2N} & \sqrt{\frac{2}{N}}\cos\frac{8997\pi}{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{\sqrt{N}} & \sqrt{\frac{2}{N}}\cos\frac{5999\pi}{2N} & \cdots & \sqrt{\frac{2}{N}}\cos\frac{17985002\pi}{2N} & \sqrt{\frac{2}{N}}\cos\frac{17991001\pi}{2N} \end{bmatrix}. \tag{11}$$

A simple way for obtaining matrix $\boldsymbol{\theta}$, while satisfying incoherence property and RIP, is to performe random permutations of the matrix $\boldsymbol{\psi}$. Only M =1200 randomly selected rows of the matrix $\boldsymbol{\psi}$ are used in the measurement process. Some of the positions of 1200 rows in matrix $\boldsymbol{\psi}$, which are used to form the matrix $\boldsymbol{\theta}$, are:

$$1945, 1575, 1143, ..., 1878, 1342. \tag{12}$$

Therefore, in our case matrix $\boldsymbol{\theta}$ is:

$$\boldsymbol{\theta} = \begin{bmatrix} \frac{1}{\sqrt{N}} & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1944+1)\pi}{2N} & \cdots & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1944+1)\cdot2998\pi}{2N} & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1944+1)\cdot2999\pi}{2N} \\ \frac{1}{\sqrt{N}} & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1574+1)\pi}{2N} & \cdots & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1574+1)\cdot2998\pi}{2N} & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1574+1)\cdot2999\pi}{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{\sqrt{N}} & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1877+1)\pi}{2N} & \cdots & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1877+1)\cdot2998\pi}{2N} & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1877+1)\cdot2999\pi}{2N} \\ \frac{1}{\sqrt{N}} & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1341+1)\pi}{2N} & \cdots & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1341+1)\cdot2998\pi}{2N} & \sqrt{\frac{2}{N}}\cos\frac{(2\cdot1341+1)\cdot2999\pi}{2N} \end{bmatrix} \tag{13}$$

In order to reconstruct musical signal, we have to solve the following problem:

$$\min \|\mathbf{s}\|_{l_1} \; \textit{subject to}\; \mathbf{y} = \boldsymbol{\theta}\mathbf{s}, \tag{14}$$

i.e.

$$\min \sum_{i=1}^{N} |s_i| \; \textit{subject to}\; \mathbf{y} = \boldsymbol{\theta}\mathbf{s}. \tag{15}$$

$\mathbf{s}$ denotes the vector of discrete cosine transform coefficients, while $\mathbf{y}$ represents measurement vector. For real valued data in (14), l_1-minimization can be recast as linear programming problem which can be solved by using optimization techniques [16]. In our case, $\mathbf{y}, \boldsymbol{\theta}, \mathbf{s}$ are real valued data and (14) can be recast into a linear problem as:

$$\min_u \sum u \;\;\; \textit{subject to}\;\; \mathbf{y} = \boldsymbol{\theta}\mathbf{s},\; \mathbf{s}-\mathbf{u} \leq 0,\; -\mathbf{s}-\mathbf{u} \leq 0. \tag{16}$$

Note that, by introducing variable $\boldsymbol{u}$, absolute value in (15) is avoided and linear programming problem is obtained.

By using the primal – dual interior method, which is an iterative procedure, we solve (16). Firstly, function which we minimize and conditions in (16) are represented with one function which is called Lagrangian:

$$\Lambda(\mathbf{s},\mathbf{u},\mathbf{g},\mathbf{h_1},\mathbf{h_2}) = f(\mathbf{u}) + \mathbf{g}(\boldsymbol{\theta}\mathbf{s}-\mathbf{y}) + \mathbf{h_1}(\mathbf{s}-\mathbf{u}) + \mathbf{h_2}(-\mathbf{s}-\mathbf{u}) .$$

This function introduces additional variables $\boldsymbol{g}$, $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$. Starting from the initial points for $\boldsymbol{s}$, $\boldsymbol{u}$, $\boldsymbol{g}$, $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$, in each iteration we compute new points which are closer to optimal solution. For our reconstruction, starting points are defined as:

$$\begin{aligned}
\mathbf{s_0} &= \boldsymbol{\theta}^T\mathbf{y}, \\
\mathbf{u_0} &= 0.95|\mathbf{s_0}| + 0.1\max\{|\mathbf{s_0}|\}, \\
\mathbf{h_{1_0}} &= -\frac{1}{\mathbf{s_0}-\mathbf{u_0}}, \\
\mathbf{h_{2_0}} &= -\frac{1}{-\mathbf{s_0}-\mathbf{u_0}}, \\
\mathbf{g_0} &= -\boldsymbol{\theta}^T(\mathbf{h_{1_0}}-\mathbf{h_{2_0}}).
\end{aligned}$$

Variables $\boldsymbol{s}$, $\boldsymbol{u}$, $\boldsymbol{g}$, $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are updated by their step direction and step length t. For example, new value for variable $\boldsymbol{s}$ is obtained as:

$$\mathbf{s} = \mathbf{s} + t\triangle\mathbf{s} ,$$

while $\triangle\mathbf{s}$ is step direction.

The iterative procedure can be summarized through the following steps:

Step 1:

Form the Lagrangian $\boldsymbol{\Lambda}$.

Step 2:

Find the first derivatives **D** of the Lagrangian $\boldsymbol{\Lambda}$.

Step 3:

Find the step directions which are used to update the variables. These step directions are computed by solving the following system:

$$\mathbf{D}'(\mathbf{s},\mathbf{u},\mathbf{g},\mathbf{h_1},\mathbf{h_2})\begin{bmatrix}\triangle\mathbf{s}\\ \triangle\mathbf{u}\\ \triangle\mathbf{g}\\ \triangle\mathbf{h_1}\\ \triangle\mathbf{h_2}\end{bmatrix} = -\mathbf{D}(\mathbf{s},\mathbf{u},\mathbf{g},\mathbf{h_1},\mathbf{h_2}) .$$

Step 4:

Find the step length t by using *the backtracking line search* method.

Step 5:

Update the variables.

$$(\mathbf{s},\mathbf{u},\mathbf{g},\mathbf{h_1},\mathbf{h_2}) = (\mathbf{s},\mathbf{u},\mathbf{g},\mathbf{h_1},\mathbf{h_2}) + t(\triangle\mathbf{s},\mathbf{u}\triangle,\triangle\mathbf{g},\triangle\mathbf{h_1},\triangle\mathbf{h_2}) .$$

Step 6:

If maximum number of iterations or sufficient accuracy is obtained, stop the procedure. Otherwise, go to the next iteration.

For the musical signals which have sparse frequency representation, CS approach provides high quality reconstruction within relatively small number of iterations (about 30 iterations).

4. EXPERIMENTAL RESULTS

In this part, we provide the simulation results and error analysis for the CS based reconstruction of musical signal. Signal with total length of 3000 samples, representing the note G4 played on the piano is observed.

The original signal is shown in Fig. 1a, while the reconstructed signal is given in Fig. 1b. CS based reconstruction preserves the details in the signal (Fig. 2a and Fig. 2b).

As it can be seen from the Fig. 3a, discrete cosine transform of the signal consists of small number of non-zero coefficients, and signal can be considered as sparse in the frequency domain. Samples are taken from the time domain because in this domain the signal has dense representation. Signal is reconstructed by using randomly taken 1200 samples, which is 40% of the total signal length. Reconstructed signal does not show any audible distortion (Fig. 3c and Fig. 3d).

In Fig. 4a it is shown that the difference between original and reconstructed signal is significantly smaller compared to the time domain signal amplitudes. Mean absolute error reduces as number of measurements increases, as shown in Fig. 4b. Using 1200 random samples, we get mean absolute error of value 2.7, while the mean absolute value of original signal is 162.141. This error produces no perceptually differences between original and reconstructed signal. By decreasing the number of measurements, error is increased. For M=600 distortion becomes audible (mean absolute error is equal to 7.98).

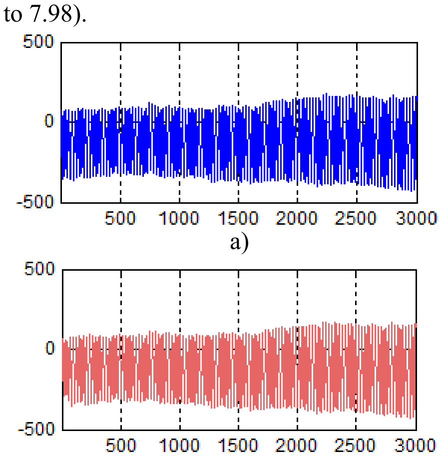

Fig.1. *a) Original signal in time domain; b) Reconstructed signal in time domain.*

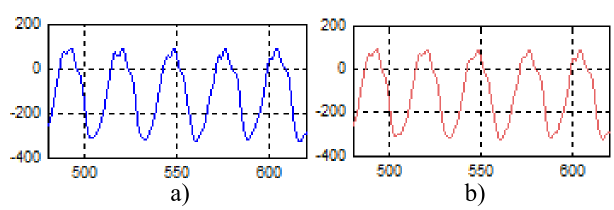

Fig. 2. *Zoomed region of the: a) original and b)reconstructed signal in time domain.*

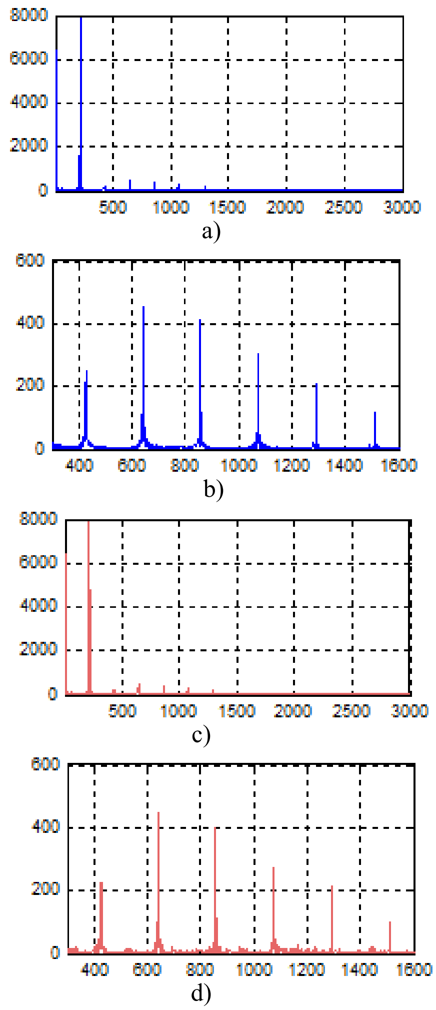

Fig. 3. *a) Discrete cosine transform of the original signal; b)Zoomed region of the discrete cosine transform of the original signal; c) Discrete cosine transform of the reconstructed signal; d) Zoomed region of the discrete cosine transform of the reconstructed signal.*

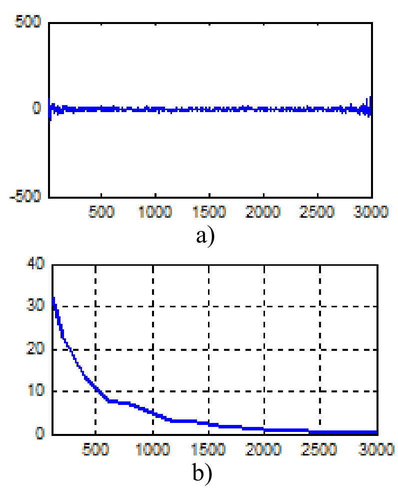

Fig. 4. *a) Reconstruction error; b) Mean absolute error.*

5. CONCLUSION

CS is a new technique that provides signal acquisition with a rate significantly smaller than the Shannon-Nyquist sampling frequency. In this paper, CS is applied to musical audio signal reconstruction. It is shown that with small number of measurements (40% of total signal length) we obtain high quality reconstructed signal, while for 20% the distortion of reconstructed signal becomes audible. For further research, it would be interesting to consider using other transform domains. Besides that, applying other optimization algorithms could be explored.

REFERENCES

[1] R. Baraniuk, "Compressive Sensing", *Lecture Notes in IEEE Signal Processing Magazine Vol. 24, July 2007.*

[2] M. Fornasier, H. Rauhut, "Compressive Sensing" *Handbook of Mathematical Methods in Imaging*, pages 187-228. Springer, 2011.

[3] P. Zhang, Z. Hu, R. C. Qiu, B. M. Sadler, "A Compressed Sensing Based Ultra-Wideband Communication System", *IEEE International Conference on Communications*, 14-18 June 2009.

[4] Y. D. Zhang, M. G. Amin, "Compressive sensing in non-stationary array processing using bilinear transforms", *IEEE Sensor Array and Multichannel Signal Processing Workshop*, Hoboken, NJ, June 2012.

[5] J. Haupt, R. Castro, R. Nowak, G. Fudge, A. Yeh "Compressive sampling for signal classification".

[6] Zhi Tian, G. Leus, V. Lottici, "Detection of sparse signals under finite-alphabet constraints", *IEEE International Conference on Acoustics, Speech and Signal Processing, April 2009, ICASSP 2009, pages 2349 – 2352.*

[7] D. L. Donoho, "Compressed Sensing", *IEEE Transactions on Information Theory,* Vol. 52, No. 4, April 2006.

[8] M. F. Duarte, Y. C. Eldar, "Structured Compressed Sensing: From Theory to Applications", *IEEE Transactions on Signal Processing,* Vol. 59, No. 9, September 2011.

[9] S. Stankovic, I.Orovic, M. Amin, "Compressed sensing based robust time-frequency representation for signals in heavy tailed noise", submitted to *ISSPA 2012*.

[10] A. Divekar, O. Ersoy, "Theory and Applications of Compressive Sensing" (2010). ECE Technical Reports. Paper 402, http://docs.lib.purdue.edu/ecetr/402

[11] M. F. Duarte, Y. C. Eldar, "Structured Compressed Sensing: From Theory to Applications".

[12] E. J. Candès, M. B. Wakin, "An Introduction To Compressive Sampling", *IEEE Signal Processing Magazine*, March 2008.

[13] Y. Tsaig, D. L. Donoho, "Extensions of Compressed Sensing", April 26, 2005.

[14] J. Romberg, "Imaging via Compressive Sampling", *IEEE Signal Processing Magazine*, March 2008.

[15] A. Cohen, W. Dahmen, R. DeVore, "Compressed sensing and best *k*-term approximation", July 24, 2006.

[16] E. Cand`es, J. Romberg, "l1-magic: Recovery of Sparse Signals via Convex Programming", October 2005.

[17] D. L. Donoho, Y. Tsaig, "Fast Solution of l1-norm Minimization Problems When the Solution May be Sparse", *Information Theory, IEEE Transactions on*, Nov. 2008.