International Journal of Computer Applications (0975 - 8887)
Volume 153 - No.11, November 2016

Speech Signal Reconstruction using Two-Step Iterative Shrinkage Thresholding Algorithm

Rachit Saluja
Department of EEE
PESIT, Bangalore-560064
India

Susmita Deb
Department of EEE
PESIT, Bangalore-560064
India

ABSTRACT

The idea behind Compressive Sensing(CS) is the reconstruction of sparse signals from very few samples, by means of solving a convex optimization problem. In this paper we propose a compressive sensing framework using the Two-Step Iterative Shrinkage/Thresholding Algorithms(TwIST) for reconstructing speech signals. Further, we compare this framework with two other convex optimization algorithms, l_1 Magic and Gradient Projection for Sparse Reconstruction(GPSR). The performance of our framework is demonstrated via simulations and exhibits a faster convergence rate and better peak signal-to-noise ratio(PSNR).

Keywords

Compressive Sensing, Convex Optimization, Two-Step Iterative Shrinkage/Thresholding Algorithms, l_1 Magic, Gradient Projection for Sparse Reconstruction

1. INTRODUCTION

Speech is an information rich signal which has become the primary means of communication among humans. Digitizing real world signals help to achieve more compact representations and provides better utilization of available resources. With an ever increasing demands for system capacities, compression of all real world signals has become a necessity. In many applications, huge amount of data is generated at the sensing stage, as high sampling rate is preferred for better quality of signals. This in turn, demands more space and increases the need for data compression before storage or transmission.

CS is a relatively new field in digital signal processing [1, 2]. CS theory asserts that one can recover certain signals from far fewer samples or measurements than conventional methods for recovering signals, which are presently being used for transmission and storage [3, 4, 5]. The conventional approach of sampling a Nyquist rate (twice the bandwidth) requires a lot of signal processing at transmitter end [6].

Compressed Sensing, which provides a framework for simultaneous sensing and compression has gained much attention in literature due to its diverse applications in a variety of fields. Also, it exploits the sparsity notion which is minimally explored, inherent characteristic present in almost all real world signals. CS has been applied to strictly sparse as well as compressible signals. Most of the real world signals are compressible in some domain or the other [7, 8]. This is true for speech signals too. The major applications where CS theory has been applied are image compression, image de-noising, image fusion, content based image retrieval, compressed medical imaging (MRI etc), radar imaging, face recognition etc and recently to speech and audio processing. In resource limited scenarios, CS can facilitate efficient utilization of the available resources with substantial performance gains. Since sparsity is the main principle behind CS, effective sparse representations of signals play a major role in the success of CS based applications. The motivation behind this paper is the fact that speech signals are sparse in transform domain. i.e, they are compressible and CS theory which is based on sparsity of signals, can be applied to speech signals.

In this work, we propose a framework in the Fourier transform domain which uses the advantage of CS in acquiring lesser samples. This framework uses TwIST [9] to solve the convex optimization problem and is then compared with l_1 Magic [10] and GPSR [11]. The rate of convergence, PSNR and the Mean-squared error (MSE) is observed for each of the optimization techniques.

2. COMPRESSIVE SENSING - AN OVERVIEW

To analyse mathematically, let us start with a real-valued, finite-length, one-dimensional, discrete-time signal x. Signal x can be represented as an $NX1$ column vector, situated in a vector space R^N populated with elements $x[n], n = 1, 2, .., N$. Any signal of a higher dimension can be represented into a one dimensional signal by vectorizing it. For a signal situated in signal space R^N, it can be represented in terms of a basis of $NX1$ vectors $\Psi_{i=1}^N$. We assume that the basis is orthonormal, for the purpose of simplicity. The signal x can then be further represented as $\Psi = [\Psi_1|\Psi_2|...|\Psi_N]$ using the set of NXN basis vectors defined above, with the Ψ_i as columns. It can be represented as:

$$x = \sum_{n=1}^{N} s_i \Psi_i \tag{1}$$

or

$$x = \Psi s \tag{2}$$

here s is the $NX1$ column vector of weighting coefficients $s_i = \langle x, \Psi_i \rangle$. Clearly, x and s are equivalent representations of the signal, with x in the time or space domain and s in the Ψ domain. We

define the signal x as K-sparse, as the signal can be represented as a linear combination of only K vectors. Thus, we begin with a signal x, which is said to be compressible since its representation in the Ψ basis has only a few large coefficients [12].
Now, what we have a compressible signal with us, but we do not want to compute all the coefficients for compression, since we know most of them will be discarded anyway. So, we need a method to integrate its dimensionality reduction within the sensing process itself, so that we aren't faced with those many coefficients in the first place. Compressive sensing seeks to directly acquire a compressed representation of the signal without computing all the N coefficients.
A general linear measurement process is considered which computes $M < N$ inner products between x and a collection of vectors $\phi_{j=1}^{M}$ as in $y_j = \langle x, \phi_j \rangle$. The vectors y_j are then arranged in a column vector with dimensions $MX1$, referred to as Y. Similarly, vectors ϕ_j^T are arranged in an MXN matrix, called Φ, also known as the measurement matrix.
This implies from the equation written above that $y = \Phi x = \Phi\Psi x = \Theta s$. Here, Θ is the term used to represent $\Phi\Psi$. The measurement process is not adaptive, meaning that Φ is fixed and does not depend on the signal x. This ensures that we have a robust system for sensing and reconstructing the signal. No priory information is required.
The main aim of an appropriate reconstruction algorithm would be to retrieve an N-value signal from M measurements in the Y matrix, the random measurement matrix Φ and the transform matrix Φ. If the signal is K-sparse, then there will be infinitely many s' that will satisfy the equation $\Theta s' = Y$. This is because if $\Theta s' = Y$, then there will most certainly be another $\Theta(s + g) = Y$ for any vector g in the null space of Θ. To prevent this issue from obstructing the hunt for the solution, the reconstruction algorithm searched for the signal's sparse coefficient vector in the $(N - M)$ translated vector space. Convex optimisation methods are applied to extract the solution from the problem at hand. Specifically, minimization methods are used to zero-in on to the solution plane, coupled with other constraints added as per requirement.

3. CONVEX OPTIMIZATION TECHNIQUES

(1) l_2 **norm minimisation technique:**
Let the l_p norm of a vector s be defined a $(||s||_p)^p = \sum_{i=1}^{N} |s_i|^p$. The conventional way to solve inverse problems of this type is to find the vector in the translated null space which has the smallest energy by optimising

$$\hat{s} = argmin||s'||_2 \tag{3}$$

such that $\Theta s' = Y$. While it may seem that this optimisation has a convenient closed-form solution $\hat{s} = \Theta^T(\Theta\Theta^T)^{-1}Y$, this leads nowhere unfortunately because almost without exception, the optimisation gives a solution which is non-sparse with non-zero elements.

(2) l_0 **norm minimisation technique:**
l_2 norm theoretically finds only the energy of the signal and minimises it, and not its sparsity. The sparsity of a K-sparse signal can be measured by the l_0 norm of the signal. That is, the l_0 of a K-sparse signal will be K itself, since it will simply count the number of non-zero coefficients present in the signal. The modified optimisation in this case would be

$$\hat{s} = argmin||s'||_0 \tag{4}$$

such that $\Theta s' = Y$. Even though theoretically this optimisation can result in a K-sparse solution with a very high probability using only $M = K + 1$ samples, it still is not the ideal candidate for our purpose. Solving this optimisation problem is numerically unstable and NP-complete. Moreover, we would need to identify each of the $\binom{N}{k}$ combinations of the non-zero coefficients which becomes computationally exhaustive and redundant.

(3) l_1 **norm minimisation technique:**
l_1 minimisation is a perfect candidate to find a K-sparse solution which would be numerically stable as well. It can accurately recover K-sparse compressible signals using only $M > cKlog(N/K)$ samples with a high probability. This type of convex optimisation problem resolves into a linear program which can be solved by various algorithms like basis pursuit, orthogonal matching pursuit, and so on. The optimisation in this case becomes

$$\hat{s} = argmin||s'||_1 \tag{5}$$

such that $\Theta s' = Y$.

4. COMPRESSIVE SENSING FRAMEWORK

Keeping in mind how compressive sensing works, it is time to perform the three different convex optimization techniques using a common framework developed to demonstrate which optimization technique has a better PSNR and a faster convergence rate.
Consider a signal x of length N. Since speech signals are long, the signal is split into L smaller speech signals that are of equal length N'. Each of the smaller signals are then subjected to a Fourier transform consecutively, as speech signals are sparse in the Fourier domain i.e. $X = \Psi x_L$, where $\Psi = F$, F being the Fourier transform matrix.
Assuming that the speech signal is K - Sparse we use K random measurements from N' to build the measurement vector Y, where $Y = \Theta x_L$, Y being a $KX1$ vector. Θ is a matrix of KXN' and is obtained by considering K orthobasis of the Ψ^{-1} matrix.
To obtain the signal with K number of samples we solve the optimization problem using l_1 Magic, GPSR and TwIST.

(1) l_1 **Magic minimisation technique:** For the l_1 Magic reconstruction algorithm, we minimise

$$||x_L||_1 \tag{6}$$

s.t. $\Theta x_L = Y$, where the x'_L we obtain is row vector of dimension $NX1$, a vectorised version of our desired signal [13].

(2) **GPSR minimization technique:** For the GPSR reconstruction algorithm, we minimise

$$\frac{1}{2}||Y - \Theta x_L||_2 + \tau||x_L||_1 \tag{7}$$

where the x'_L we obtain is row vector of dimension $NX1$, and τ is a non-negative factor.

(3) **TwIST minimization technique:** For the TwIST reconstruction algorithm, we minimise

$$\frac{1}{2}||y - \Theta x_L||_2^2 + \tau\Phi(x_L), \tag{8}$$

where the x'_L we obtain is row vector of dimension $NX1$, τ is a non-negative factor and $\Phi(x_L)$ is the total variation norm regularization function [14].

After obtaining the x'_L, the reconstructed signal in the time domain is recovered by taking an inverse Fourier transform,i.e $\tilde{x_L} = \Psi^{-1}x'_L$. Now all the $\tilde{x_L}$ are concatenated to obtain the complete reconstructed speech signal using CS [15].

5. OBSERVATIONS

The performance of each algorithm is tested by comparing the MSE, PSNR and the rate of convergence by taking different number of samples. Figure 2 represents the plotting of the reconstructed speech signal by applying TwIST using 30,000 samples (75% less samples). From the figures we note that it is a near perfect reconstruction.

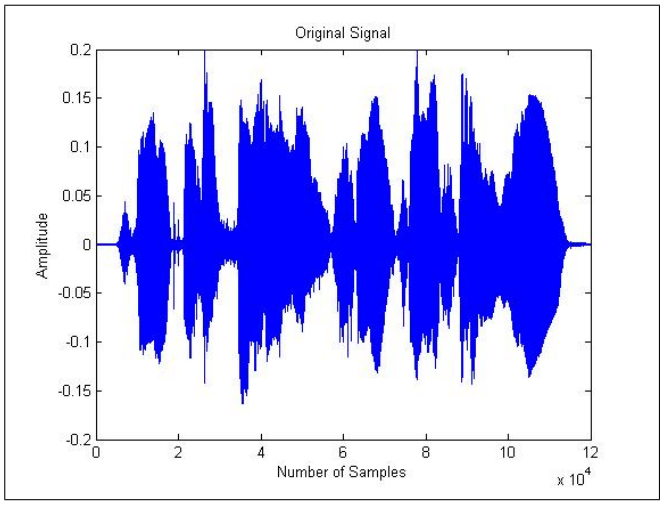

Fig. 1. Graph plotting Amplitude (y-axis) against number of samples for original signal.

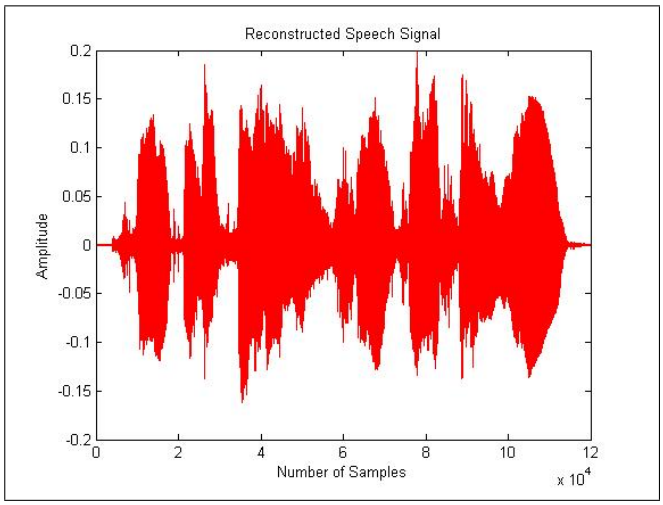

Fig. 2. Graph plotting Amplitude (y-axis) against number of samples utilised for reconstruction using TwIST.

First we observe the MSE obtained for the three techniques. From Figure 3 we infer that the least MSE is for TwIST and that the MSE decreases with the increase in number of samples. We see that GPSR has maximum error and that the MSE for TwIST almost tends to zero at 30,000 samples.

Table 1. Time of Convergence, MSE and PSNR for the reconstructed signal using 30,000 samples

Technique	Time of Convergence(Seconds)	MSE	PSNR
TwIST	397.412	8.28231e-05	40.8185
l_1 Magic	939.59	0.000371619	34.299
GPSR	418.92	0.00157627	28.0237

The PSNR values obtained for the TwIST technique are significantly higher than the PSNR values obtained by l_1 Magic and GPSR, as shown in the Figure 4. We note that for TwIST and l_1 Magic, the PSNR values increases with increase in number of samples, however for GPSR, the PSNR nearly remains constant. The above two results obtained are due to the regularization function that is used in TwIST, which helps in denoising the speech signal. We take the original signal shown in Figure 1 as the reference.

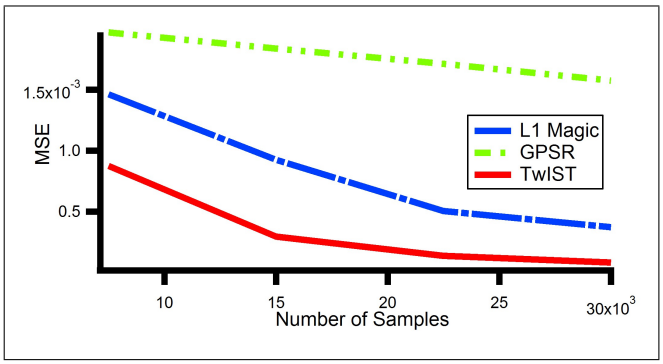

Fig. 3. Graph plotting MSE (y-axis) against number of samples utilised for reconstruction using TwIST, l_1 Magic and GPSR.

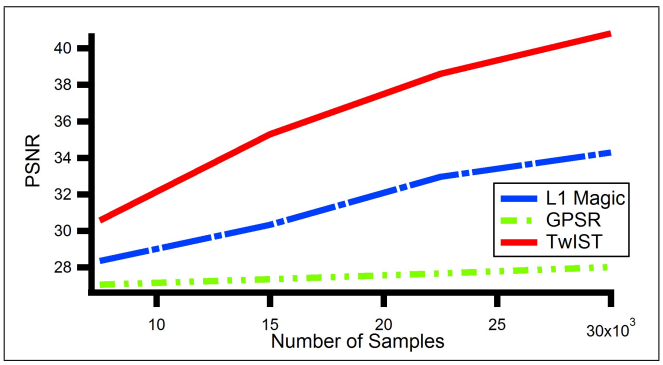

Fig. 4. Graph plotting PSNR (y-axis) against number of samples utilised for reconstruction using TwIST, l_1 Magic and GPSR.

Finally, we compare the time taken by each method to converge, i.e to obtain the reconstructed signal. From Figure 5 we observe that the rate of convergences of TwIST is better than GPSR slightly and is significantly better than l_1 Magic. As the number of samples increases, the time taken to converge also increases. In case of l_1 Magic it drastically increases.

From the observations made above, we infer that TwIST is a better reconstruction algorithm as it has the fastest rate of convergence,

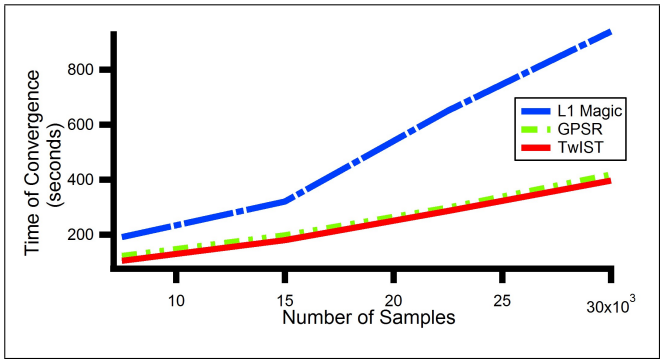

Fig. 5. Graph plotting Time of Convergence (y-axis) against number of samples utilised for reconstruction using TwIST, l_1 Magic and GPSR.

the highest PSNR and the lowest MSE. Though l_1 Magic has low MSE and a high PSNR, the time taken to converge is very high. Whereas, using GPSR would reconstruct a speech signal with poor PSNR and a high MSE. Hence, the work has successfully demonstrated that TwIST would be a better technique to use.

6. CONCLUSIONS

From the previous section we observe that TwIST is a better algorithm to use for the reconstruction of a speech signal. The implications of compression in speech signals are as follows:

(1) Reduction in bit rate thereby achieving reduction in bandwidth and memory storage requirement.

(2) Reduction in transmission power requirement because after compression there are less bits (hence less energy) per second to transmit.

(3) Immunity to noise, as error control coding methods can be introduced in place of some of the saved bits per sample in order to protect speech parameters from channel noise and distortion.

(4) Encryption of source information.

With the successful demonstration of the application of compressive sensing in speech signals, new avenues have been opened for future research in this direction. A more efficient sparse representation using other sparsifying transforms or dictionaries and better reconstruction algorithms, which might improve both the reconstructed speech signal quality and the compression ratios, can also be explored.

7. REFERENCES

[1] Richard G Baraniuk. Compressive sensing. *IEEE signal processing magazine*, 24(4), 2007.

[2] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Springer, 2013.

[3] Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007.

[4] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[5] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.

[6] Thippur V Sreenivas and W Bastiaan Kleijn. Compressive sensing for sparsely excited speech signals. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4125–4128. IEEE, 2009.

[7] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006.

[8] Emmanuel J Candès et al. Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Madrid, Spain, 2006.

[9] José M Bioucas-Dias and Mário AT Figueiredo. A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image processing*, 16(12):2992–3004, 2007.

[10] Emmanuel Candes and Justin Romberg. l1-magic: Recovery of sparse signals via convex programming. *URL: www. acm. caltech. edu/l1magic/downloads/l1magic. pdf*, 4:14, 2005.

[11] Robert D Nowak, Stephen J Wright, et al. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of selected topics in signal processing*, 1(4):586–597, 2007.

[12] Wei Dai and Olgica Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, 55(5):2230–2249, 2009.

[13] Elaine T Hale, Wotao Yin, and Yin Zhang. A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing. *CAAM TR07-07, Rice University*, 43:44, 2007.

[14] Thomas Blumensath and Mike E Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.

[15] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.