Adaptive Compressed Sensing of Speech Signals

Manipal Reddy Kuchakuntla
Masters Student
Department of Electrical Engineering
University of South Florida, Tampa Florida, USA.

Abstract - **Compressed Sensing (CS) is an emerging signal acquisition theory that provides a universal approach for characterizing signals which are sparse or compressible on some basis at sub-Nyquist sampling rate. Based on an over-complete data as the sparse basis specialized for speech signals, CS Sampling and reconstruction of speech signal are realized. Furthermore, we propose to choose the sensing matrix adaptively, according to the energy distribution of original speech signal. Experimental results show significant improvement of speech reconstruction quality by using such adaptive approach against traditional random sensing matrix. The key objective in compressed sensing (also referred to as sparse signal recovery or compressive sampling) is to reconstruct a signal accurately and efficiently from a set of few non-adaptive linear measurements.**

The compressed sensing field has provided many recovery algorithms, most with provable as well as empirical results. There are several important traits that an optimal recovery algorithm must possess. The algorithm needs to be fast, so that it can efficiently recover signals in practice. The algorithm should provide uniform guarantees, meaning that given a specific method of acquiring linear measurements, the algorithm recovers all sparse signals (possibly with high probability). Ideally, the algorithm would require as few linear measurements as possible. However, recovery using only this property would require searching through the exponentially large set of all possible lower dimensional subspaces, and so in practice is not numerically feasible. Thus in the more realistic setting, we may need slightly more measurements. Finally, we wish our ideal recovery algorithm to be stable.

This means that if the signal or its measurements are perturbed slightly, then the recovery should still be approximately accurate. This is essential, since in practice we often encounter not only noisy signals or measurements, but also signals that are not exactly sparse, but close to being sparse. The conventional scheme in signal processing, acquiring the entire signal and then compressing it, was questioned by Donoho. Indeed, this technique uses tremendous resources to acquire often very large signals, just to throw away information during compression.

I. OBJECTIVE

Compressed sensing (CS) is an emerging signal acquisition theory that directly collects signals in a compressed form if they are sparse on some certain basis. It originates from the idea that it is not necessary to invest a lot of power into observing the entries of a sparse signal in all coordinates when most of them are zero anyway. Rather it should be possible to collect only a small number of measurements that still allow for reconstruction. This is potentially useful in applications where one cannot afford to collect or transmit a lot of measurements but has rich resources at the decoder.

Observing that different kind speech frames have different intra-frame correlations, a frame-based adaptive compressed sensing framework for speech signals has been proposed. The objective of this project is to further improve the performance of the existing compressed sensing process that uses non adaptive projection matrix, by using the adaptive projection matrix based on frame analysis. Average-frame signal-to-noise ratio (AFSNR) is calculated to evaluate the performance of the frame-based adaptive CS with the non-adaptive CS.

Compressed Sensing

In a typical communication system, the signal is sampled at least at twice the highest frequency contained in the signal. However, this limits efficient ways to compress the signal, as it places a huge burden on sampling the entire signal while only a small number of the transform coefficients are needed to represent the signal. On the other hand, compressive sampling provides a new way to reconstruct the original signal from a minimal number of observations. CS is a sampling paradigm that allows to go beyond the Shannon limit by exploiting the sparsity structure of the signal. It allows to capture and represent the compressible signals at a rate significantly below the Nyquist rate.

The signal is then reconstructed from these projections by using different optimization techniques. During compressive sampling only the important information about a signal is acquired, rather than acquiring the important information plus the information of a signal which will be eventually discarded at the receiver. The key elements that need to be addressed before using compressive sensing are the following

1. how to find the transform domain in which the signal has a sparse representation
2. how to effectively sample the sparsely signal in the time domain,
3. how to recover the original signal from the samples by using optimization techniques.

Signal Sparsity

Signal Sparsity represents the presence of signal transform coefficients less densely. Sparsity allows to reconstruct the signal with less number of projections (samples).The procedure used to ensure the sparsity of the signal is called transform coding, which is performed by the following four steps

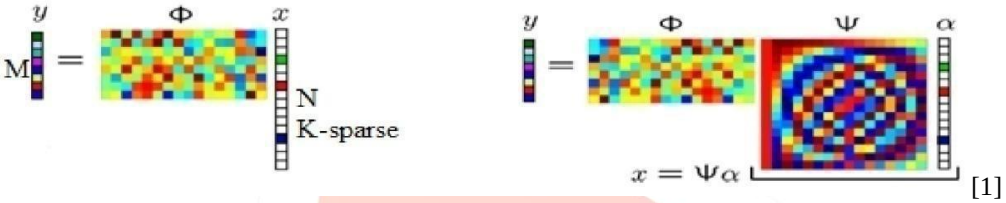
1. Full N-points of a signal x is obtained using the Nyquist rate ,
2. Complete set of transform coefficients (DFT) is obtained,
3. Locate the K largest coefficients and throw away the smallest coefficients
4. Multiply the signal by the measurement matrix to obtain the observation vector of length M.

Measurement Matrix

In compressed sensing special emphasis is given to represent the signal with an incoherent basis. The linear measurement process that computes M < N inner products between x and the collection of vectors

$$\{\emptyset_j\}_{j=1}^{M} \quad \text{via} \quad y_j = <x, \emptyset_j>| \qquad [1]$$

Where phi is an M×N measurement matrix with each row been a measurement vector. It has been seen that some of the measurement matrices can be used in any scenario, in the sense that they are incoherent with any fixed basis Ψ such as Gabor, spikes, sinusoidal and wavelets. The compressive sensing measurement process with K-sparse coefficient vector x is depicted in Figure 1



[1]

Figure 1 Compressive sensing measurement process

The measurement matrix plays a vital role in the process of recovering the original signal. There are two types of measurements matrices that can be used in compressive sensing. The Random measurement matrix and the predefined measurement matrix. The fundamental revelation is that, if a signal x composed of N samples is sparse then the actual signal can be reconstructed using below formula

$$M \geq O\,(\text{Log}\,(N/K))[1]$$

Furthermore, x can be perfectly reconstructed using different optimization techniques. If Φ is a structurally random matrix, its rows are not stochastically independent because they are randomized from the same random seed vector. The random matrix is transposed and then orthogonalized. This will have the effect of creating a matrix that represents an ortho-normal basis. In a predefined measurement matrix, the matrix is created by using function like the Dirac functions and Sine functions.

In this case, the signal is multiplied by several Dirac functions cantered at different locations to obtain the observation vector. Then the speech signal can be reconstructed using the l1 normalization method by using the observation vector and the predefined measurement matrix.

Signal Reconstruction in Compressive Sensing

Recent developments in signal theory have shown that a sparse signal is a useful model in areas such as communications, radar and image processing. Therefore the assumption that every signal can be represented in a sparse form has helped in the compression of the signal of interest. The perfect reconstruction of a signal x depends on the measurement matrix Φ and the measurement vector y.

The compressive sensing theory tells that when the matrix ΦΨ has the Restricted Isometric Property (RIP) which are nearly ortho-normal then it is possible to recover the K largest significant coefficients from a similar size set of M=O(K log(N/K)) [1] measurements of y. As a result, the sparse signal can be reconstructed by different optimization techniques such as l1 norm and convex optimization. The first minimization technique which has been used to reconstruct the signal is the l1 minimization

$$(P1)\ \min \| x \|l1 \text{ Subject to } \Phi x = y \ [3]$$

This is also known as basis pursuit (P1). The goal of this technique is to find the vectors with the smallest l1-norm

$$\|X\|_1 = \sum_{i=1}^{n} | X_i | \qquad [3]$$

Orthogonal Matching Pursuit (OMP) Optimizing Technique [3]

In this project, signal is reconstructed frame by frame using OMP method. OMP uses sub Gaussian measurement matrices to reconstruct sparse signals. If Φ is such a measurement matrix, then Φ*Φ is in a loose sense close to the identity. Therefore the largest coordinate of the observation vector y= Φ*Φx is expected to correspond to a non-zero entry of x. Thus one coordinate for the support of the signal x is estimated. Subtracting of that contribution from the observation vector y and repeating eventually yields the entire support of the signal x. OMP is quite fast, both in theory and in practice, but its guarantees are not as strong as

those of Basis Pursuit. The algorithm's simplicity enables a fast runtime. The algorithm iterates s times and each iteration does a selection though d elements, multiplies by Φ*Φx and solves a least squares problem.

Adaptive Compressed Sensing

In conventional compressed sensing process, the projection matrix which is used to generate the required compressed signal is generated randomly and considered to be fixed during the entire conversion process. That means the projection matrix is non-adaptive. Though this process results in better performance when compared to conventional sampling process, even better results can be obtained by using adaptive projection matrix.

Adaptive Projection Matrix

Most work in CS research focus on random projection matrix which is constructed by considering only the signals sparsity rather than other properties. In other word, the construction of projection matrix is non-adaptive. Observing that different kind speech frames have different intra-frame correlations, a frame-based adaptive compressed sensing framework for speech signals has been proposed , which applies adaptive projection matrix. To do so, the neighboring frames are compared to estimate their intra-frame correlation, every frame is classified into different categories and the number of projections are adjusted accordingly.

The experimental results show that the adaptive projection matrix can significantly improve the speech reconstruction quality. Intra-frame correlation of speech signals is explored to achieve efficient sampling. Because different kind speech signals may have different intra-frame correlations a frame-based adaptive CS framework that uses different sampling strategies in different kind speech frames, has been proposed.

II. FRAME ANALYSIS

Each speech sequence is divided into non-overlapping frames of size 1x n and all frames in a speech sequence are processed independently. The projection matrix is initialized by Gaussian random matrix Φ which has been proven to be incoherent with most sparse basises at high probability.

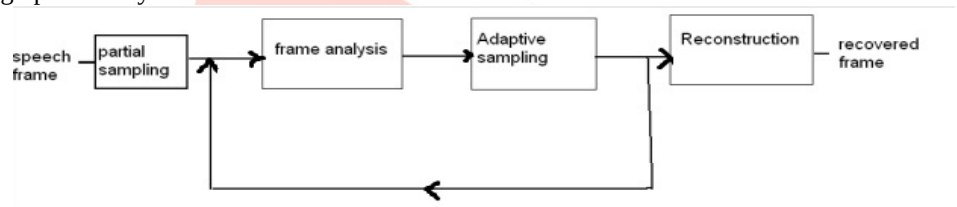


Figure 2 The frame-based adaptive CS framework for speech

As shown in Figure 2, for each frame in a speech sequence, a small number of projections is collected and compared these projections with the projections collected for the previous frame. Based on the comparison results, the correlation between these two frames is estimated and the correlation is classified into different categories. Then the sampling strategy is adjusted according to the correlation type and different number of samples for the current frame are collected.

For the current t th frame original speech signal, we represent it as x(t) its previous frame t-1 is represented using x(t-1). The difference between x(t) and x(t-1) reflects the correlation between the two neighboring frames and can be used to classify the correlation. We use the collected measurements to estimate the correlation instead. The same projection matrix Φ is applied to all frames in the partial sampling stage and we have y(t)-y(t-1)= Φx(t)- Φx(t-1)[1], where y(t) and y(t-1) are the projection vectors of x(t) and x(t-1) respectively. As each sample in y(t)-y(t-1) is a linear combination of x(t)-x(t-1) the difference between the two projection vectors also reflects the intensity changes in the two frame.

Therefore, we can estimate the amount of intensity changes in the two frames using only a small number of projections. Let ΦM0 be a matrix containing the first M0 rows of the Gaussian random matrix Φ. For the current frame t, we first use ΦM0 to collect M0 measurements y (t) ^ M0= ΦM0 .x(t) in the partial sampling stage. Then, we compare it with the first M0 measurements in y (t-1) and calculate the difference y (t) ^d=y (t) ^M0-y (t-1) ^M0. In the frame analysis module, given y (t) ^d, we calculate its l^2 norm normalized by M0 and compare with two thresholds T1 and T2 (T1<T2).

If y (t) ^d/M0<=T1, the current frame is almost the same as its previous frame. We consider the two neighboring frames may be both surd and label the intra-frame correlation as surd vs. surd. If T1< y (t) ^d/M0<=T2, it indicates that these two neighboring frames undergo small changes. In this situation, the two neighboring frames may be both sonant at high probability and the intra-correlation is labeled as sonant vs. sonant. If y(t)^d/M0>T0, the two frames are significantly different from each other, which is most likely due to the change of the frame type, and we label the correlation as surd vs. sonant.

Partial Sampling

For each frame in a speech sequence, we first collect a small number of projections, and compare it with the projections collected for the previous frame. Based on the comparison results, we estimate the correlation between these two frames and classify the correlation into different categories. We then adjust the sampling strategy according to the correlation type and collect different number of samples for the current frame. The next sections discuss details of each step in the above framework.

For the current 't'th frame original speech signal, we represent it as xt . Its previous frame t-1 is represented using xt-1. The difference between xt and xt -1 reflects the correlation between the two neighboring frames and can be used to classify the correlation. Since xt*xt-1 is not available at the sampling stage. We use the collected measurements to estimate the correlation instead. The same projection matrix Φ is applied to all frames in the partial sampling stage and we have yt-yt-1=φxt-φxt-1 , where yt and yt-1 are the projection vectors of xt and xt-1 respectively. As each sample in is a linear combination, the difference

between the two projection vectors also reflects the intensity changes in the two frame. Therefore, we can estimate the amount of intensity changes in the two frames using only a small number of projections.

Adaptive Sampling

Depending on their classified intra-frame correlation types, different number of projections is used for the speech frames. We consider the frame as surd frame if its intra-frame correlation type is surd vs. surd. A surd frame contains the least new information in the speech. Thus, the M0 measurements collected in the partial sampling stage are sufficient and we do not need additional sampling. When its intra-frame correlation is sonant vs. sonant, the frame is considered as sonant and contains some new information, which requires more measurements to be collected.

For such frames, we collect M1 (M1 >M0) measurements. We use the (M0+1)th to the M1th rows of the Gaussian random matrix Φ and combine with M0 to generate the final projection vector yt . The frames that experience large changes must contain the most new information. Therefore, we collect a total of M2 (M2>M1>M0) measurements during the sampling process. The total projection matrix is the first M2 rows of the Gaussian random matrix Φ.

Orthogonal Matching Pursuit Reconstruction

Orthogonal Matching Pursuit (OMP), put forth by Mallrat and his collaborators and analyzed by Gilbert and Troop .OMP uses sub Gaussian measurement matrices to reconstruct sparse signals. If Φ is such a measurement matrix, then Φ *Φ is in a loose sense close to the identity. If Φ is such a measurement matrix, then Φ *Φ is in a loose sense close to the identity. Therefore one would expect the largest coordinate of the observation vector y=Φ8Φ to correspond to a non-zero entry of x.

III. SIMULATION RESULTS

To compare the performance of this proposed adaptive compressed sensing and the conventional non-adaptive CS, some experiments are conducted. As a part of that, an arbitrary speech signal has been chosen which is 10 kHz sampled and 16 bits quantized for each sample. Adaptive CS and CS sampling and reconstruction are performed frame by frame, with a frame length of N=320 samples. Threshold values T1 and T2 are chosen as 0.08 and 0.04 respectively which is tested through a great number of experiments. Average Frame Signal to Noise Ratio (AFSNR) is calculated and used to evaluate the reconstruction quality of speech signal. Average Frame Signal to Noise Ratio (AFSNR) is calculated using the formula shown below

$$AFSNR=1/K\sum_{k=1}^{K} 10Log10(\|x_k\|)^2/(\|x_k-\widehat{x_k}\|^2) \quad [2]$$

Where K is the total frame number of a speech sequence xk and $\widehat{xk}$ represent the frame speech and the frame reconstructed speech. Under different compressed ratio(r=0.2, r=0.4 and r=0.6), which is defined as r=M/N, the different test results are obtained based on the proposed frame-based adaptive CS using OMP reconstruction algorithm.

Compression Ratio r=0.2

Here the compression ratio r is equal to 0.2 which indicates that the number of projections M=64 for the frame of samples N = 320.The below figures 3 shows the time domain waveform of the original speech signal and adaptive CS reconstructed speech with compressed ratio of 0.2.

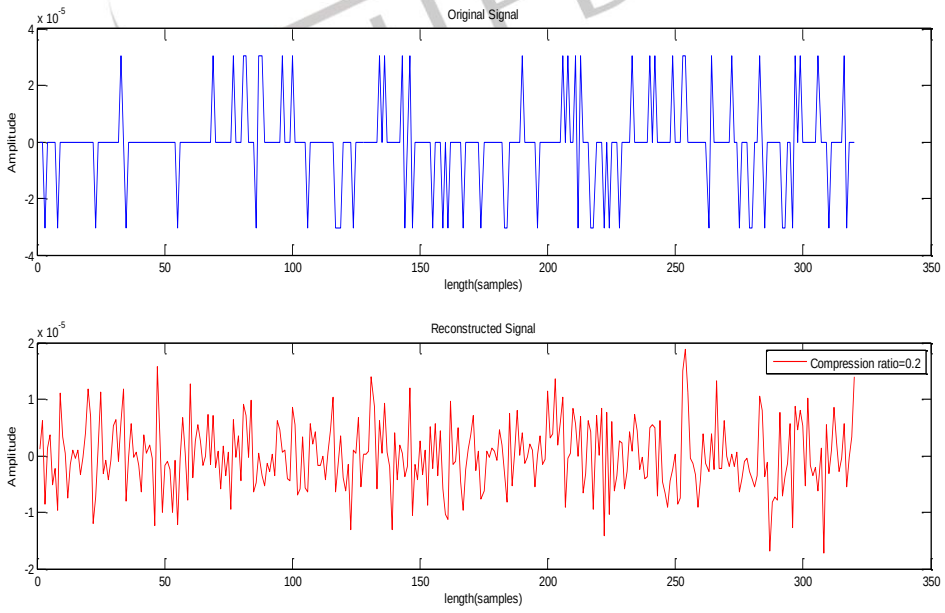


Figure 3 Original and Reconstructed signals for N=320, M=64, r=0.2

Compression Ratio r=0.4

Here the compression ratio r is equal to 0.4 which indicates that the number of projections M=128 for the frame of samples N = 320.The below figures 2 shows the time domain waveform of the original speech signal and adaptive CS reconstructed speech with compressed ratio of 0.4

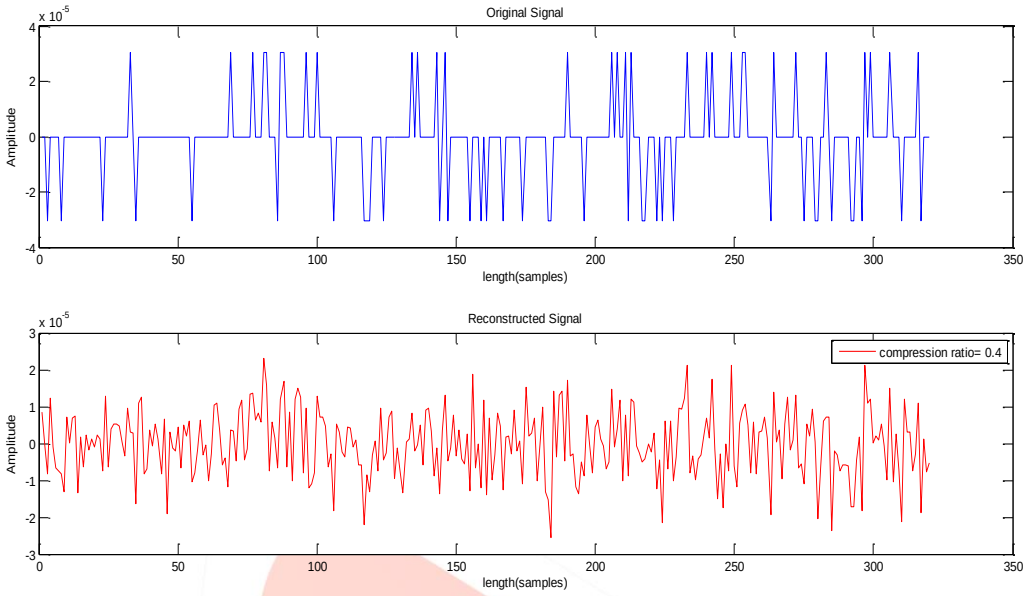


Figure 4 shows the time domain waveform of the original speech signal and adaptive CS reconstructed speech with compressed ratio of 0.4

IV. CONCLUSION AND FUTURE SCOPE

The adaptive projection matrix has been applied to the conventional compressed sensing and improved the average frame signal to noise ratio. It is also proved that the quality of the reconstructed signal increases as the compressed ratio increases. Thus the conventional non adaptive compressed sensing can be replaced by the adaptive compressed sensing to improve the efficiency of the system. During the design process, this module went through different tests and analysis in order to find the most adequate optimization technique to reconstruct the speech signal with few random measurements without losing the information. For simulation purposes, code was created in order to compress the speech signal below the Nyquist rate by taking only a few measurements of the signal.

The result shows that by keeping the length of the signal (L) and threshold window (Th) constant the desired compression of the signal can be achieved by making the signal sparse (K) to a certain amount which in turn increases the data rates. The speech signal was reconstructed without losing important information in order to achieve an increase in the data rates. After multiple simulations, it was found that the system worked as expected and the speech signal was reconstructed efficiently with a minimum error. Performance of compressive sensing is better when compared to wavelet compression as there is a minimum error with same compression rate using different parameters.

In this research the design of a new signal acquisition system using adaptive compressive sensing as been implemented. The proposed system should fulfill the accurate reconstruction of the speech signal. Different transformations need to be tested in order to find the most efficient one for this application design and measurement matrix that will be optimum for speech signals.

REFERENCES

[1] Tingting Xu, Zhen Yang and Xi Shao. IEEE *Adaptive Compressed Sensing of speech signal [J], Communications*, APCC 2009

[2] D. Donoho, *"Compressed sensing" IEEE Trans. on Information Theory [J]*, vol. 52(4), 2006, pp. 1289-1306.

[3] Donoho D,Tsaig Y. *Extensions of compressed sensing [J], Signal Processing*, 2006,86(3):533-548.

[4] J. Tropp and A. Gilbert, *"Signal recovery from random measurements via orthogonal matching pursuit", IEEE Trans. Inform. Theory [J]*, vol. 53(12), 2007, pp. 4655-4666.