Compressive Sensing

team:No.6, leader:XiaofanZhu, members:HaoZhang, ZhaoqinMa, ChenxianZhu

March 29, 2009

Contents

1 Introduction

Conventional approaches to sampling signals or images follow Shannons celebrated theorem: the sampling rate must be at least twice the maximum frequency present in the signal (the Nyquist rate). In fact, this principle underlies nearly all signal acquisition protocols used in consumer audio and visual electronics, medical imaging devices, radio receivers, and so on. However, it has been proved that this theorem can be substituted when the problem deals with sparse signals. Consider the signal below:

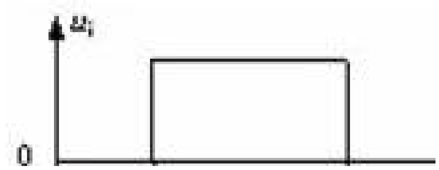

Figure 1

Due to Shannons theorem, we need an extremely high sampling rate although the shape of this signal is simple and most values of the samples are zero. However, recent breakthroughs in compressed sensing have shown that merely M samples (M $\ll$ required number of sampling points according to Shannons theory) can reconstruct the origin signal successfully. Compressive Sampling (CS), also known as Compressed Sensing, is a generalization of conventional point sampling where observations are inner products between an unknown signal and a set of user-defined test vectors. Recent theoretical results show that, for certain ensembles of test vectors, CS projections provide an effective method of encoding the salient information in any sparse (or nearly sparse) signal. Further, these projection samples can be used to obtain a consistent estimate of the unknown signal even in the presence of noise. These results are remarkable because the number of samples required for low-distortion reconstruction is on the order of the number of relevant signal coefficients, which is often far fewer than the ambient dimension in which the signal is observed. This huge reduction in sampling makes CS a practical and viable option in many resource constrained applications. The whole process is illustrated by figure 2 and figure 3:

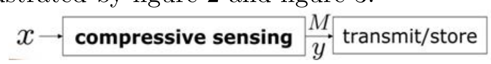

Figure 2

Figure 3

1.1 Sensing of signals

Next we will discuss some main characteristics of compressive sensing in detail.

1.1.1 Sparsity

Definition: If X_i is all zero but K entries, the vector is called k-sparse.

Consider a general linear measurement process that computes $M < N$ inner products between x and a collection of vectors $\{\phi_j\}_{j=1}^{M}$ j=1 as in $y_i =< x, \phi_j >$ Arrange the measurements y_j in an $M \times 1$ vector y and the measurement vectors ϕ_j^T as rows in an $M \times N$ matrix ϕ .It can be proved that when data is sparse, we can directly acquire a condensed representation with no/little information loss through dimensionality reduction: $y = \phi * x$,where $k < M \ll N$, to a more precise degree, M =O(K logN).

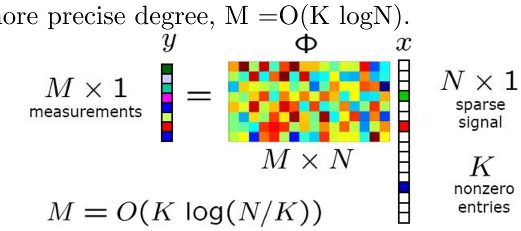

Figure 4: Compressive Data Acquisition. If x is an $N \times 1$ sparse signal with only K nonzero entries, then it can be projected by an $M \times N$ matrix, to form an $M \times 1$

vector. In addition, M is a little larger than K and much smaller than N. It means although x seems to need a lot of samples, its sparsity indicate that it can be measured just with M measurements. P.S. a random projection will work quite well.

Further studies have extended $y = \phi x$ to non-sparse signals. Suppose the observed signal x is not sparse, but instead a suitably transformed version of it is. That is, if T is a transformation matrix then $\alpha = \Psi^{-1}x$ is sparse. The CS observations can be written as $y = \phi\Psi\alpha$ .This property highlights the universality of compressive sensing.

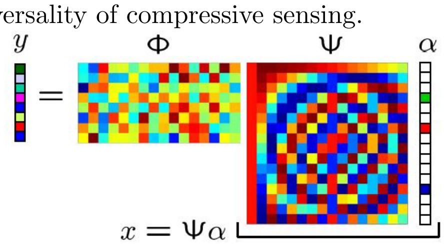

Figure 5: Assuming that x is not sparse itself, but it can still be represented by a sparse signal in certain basis, compressive sensing work appropriately as well.

1.1.2 Incoherence

Suppose we are given a pair (ϕ, Ψ) of orthobases of R^n .The first basis ϕ is used for sensing the object and the second is used to represent f. The restriction to pairs of orthobases is not essential and will merely simplify our treatment.

Definition: the coherence between the sensing basis ϕ and the representation basis Ψ is

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \le k, j \le n} |\langle \varphi_k, \psi_j \rangle|$$

The coherence measures the largest correlation between any two elements of Ψ and Φ ; If Φ and Ψ contain correlated elements, the coherence is large. Otherwise, it is small. As for how large and how small, it follows from linear algebra that $(\Psi, \phi) \in [1, \sqrt{n}]$.

Compressive sampling is mainly concerned with low coherence pairs, the reason for which we will explain below.

Due to the theorem presented by E. Cands and J. Romberg, if we fix $f \in R^n$ and suppose that the coefficient sequence x of f in the basis Ψ is S-sparse. Select m measurements in the ϕ domain uniformly at random. Then if

$m \geq C \cdot \mu^2(\Phi, \Psi) \cdot S \cdot logn$

for some positive constant C, the solution to reconstruct x is exact with overwhelming probability. Considering what we have said above, it is apparently that the smaller the coherence, the fewer samples are needed, hence our emphasis on low coherence systems in the previous section.

Random matrices are largely incoherent with any fixed basis Ψ. Select an orthobasis ϕ uniformly at random, which can be done by orthonormalizing n vectors sampled independently and uniformly on the unit sphere. Then with high probability, the coherence between ϕ and Ψ is about $\sqrt{2logn}$. By extension, random waveforms with independent identically distributed entries, e.g., Gaussian or 1 binary entries, will also exhibit a very low coherence with any fixed representation ϕ. If sensing with incoherent systems is good, then efficient mechanisms ought to acquire correlations with random waveforms, e.g., white noise.

1.2 Reconstruction of signals

Sparse Recovery It is necessary to introduce **Restricted Isometry Property** in advance, known as RIP.

Let δ_k be the smallest number such that:

$(1 - \delta_k)\|X\|_2^2 \geq \|\phi_x\|_2^2 \geq (1 + \delta_k)\|X\|_2^2$

for all k-sparse vectors x/inR^n where $\phi = [\phi_1 \dots \phi_n] \in R^{m \times n}$

The theorem presented by E. J. Cands tells us that:

If $\delta_{2k} < \sqrt{2} - 1$,then for all k-sparse vectors x such that $\phi_x = b$, the solution of (l_1) is equal to the solution of (l_0).

Here,$l_1 = min\|x\|_1 : \phi_x = b, x \in R^n$

$l_0 = min\|x\|_0 : \phi_x = b, x \in R^n$

The same compressed data could be generated by many n-dimensional vectors, but we have to find the sparsest one, i.e. the vector whose number of nonzero data is smallest. This might seem to require that any reconstruction algorithm must exhaustively search over all sparse vectors. However, this procedure is impossible, just to provide a way to measure whether the vector we find is appropriate or not. But fortunately, applying the RIP we have just discussed above, we can use the l_1 norm as a proxy for sparsity instead of l_0 norm so that the process is much more tractable. Given a vector of (noise-free) observations $y = \theta x$, the unknown k-sparse signal x can recovered exactly as the unique solution to

$min\|x\|_1$ subject to $y = \phi x$

which is known as l_1 minimization.

Of course, there are other effective recovery techniques for CS, such as matching pursuit, iterative thresh holding and total variation minimization, but the coverage of them is beyond this article. Let us look at the l_1 minimization from a geometrical point of view. The line denotes all the x vectors which satisfy the equation $\phi_x = b$, so that they all have the possibility to be reconstructed. The diamond represents $\|x\|_1$, all the points on the edge of this diamond have an equal $\|x\|_1$, and the points in the inner space of this diamond have a smaller $\|x\|_1$,and vice versa.

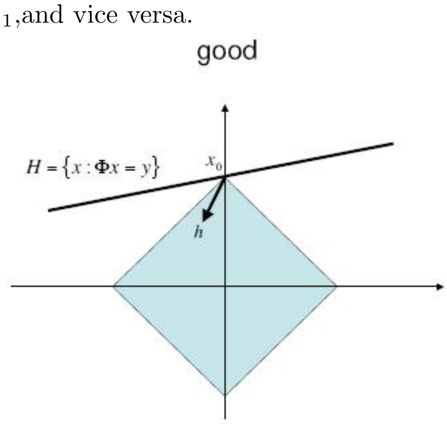

Figure 6: This picture obeys RIP, meaning that finding x_0 equals to finding the vector which obeys l_1 minimization. So that it is good for applying l_1 minimization to reconstruct x, since the line has only one point of intersection with the diamond, which determines the uniqueness of reconstruction. The other vectors on the line all have a larger l_1.

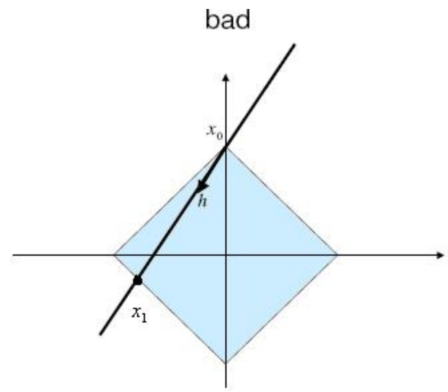

Figure 7: This picture does not obey RIP. It is bad for applying l_1 minimization to reconstruct x, since the line has more than one point of intersection with the diamond. Some other vectors rather than x_0 on the line lie in the inner space of the diamond, which have a smaller l_1. So that l_1 minimization will find x_1 in stead of x_0.

In addition, RIP also acts as a stable embedding, which means that if x_1 is close to x_2 in R^n , then when they are projected by ϕ ,ϕ_{x_1} is close to ϕ_{x_2} as well. It can be proved by the inequation:

$(1 - \delta_{2k}) \geq \frac{\|\phi_{x_1} - \phi_{x_2}\|_2^2}{\|x_1 - x_2\|_2^2} \geq (1 + \delta_{2k})$

Further more, if δ_{2k} is less than 0.41, then tractable recovery, robust recovery and stable recovery are ensured.

Compressed sensing remains quite effective even when the samples are corrupted by additive noise, which is important from a practical point of view since any real system will be subjected to measurement inaccuracies. We present noisy measurement as:

$y = \phi' \alpha_0 + e$, $\|e\|_2 \geq \varepsilon$

A variety of reconstruction methods have been proposed to recover (an approximation of) x when observations are corrupted by noise. The fundamental solution is to relax the recovery program, i.e. solve

$min\|\alpha\|_{l_1}$ subject to $\|\phi'\alpha - y\| \geq \varepsilon$

Out of doubt, since we have relaxed the condition, there exist some errors generated by this relaxation. However, the recovery error obeys:

$\|\alpha_0 - \alpha^*\|_2 \geq \sqrt{\frac{N}{M}} \cdot \epsilon + \frac{\|\alpha_0 - \alpha_{0,k}\|_{l_1}}{\sqrt{K}}$

The first part of the right is called measurement error, for it correlates to the number of measurements M. Moreover, we can see easily that the larger M is, which means we acquire more samples, the smaller the error becomes. The second part is called approximation and α_0, K stands for best K-term approximation.

2 Application: Exploiting CS in WSNs

2.1 Introduction of WSN

Due to recent technological advances, the manufacturing of small and low cost sensors became technically and economically feasible. The sensing electronics measure ambient conditions related to the environment surrounding the sensor and transform them into an electric signal. Processing such a signal reveals some properties about objects located and/or events happening in the vicinity of the sensor. A large number of these disposable sensors can be networked in many applications that require unattended operations, which later develops to the wireless sensor networks. Nowadays, wireless sensor networks (WSNs) have been used for numerous applications including military surveillance, facility monitoring and environmental monitoring. Typically WSNs have a large number of sensor nodes with the ability to communicate among themselves and also to an external sink or a base-station. The sensors could be scattered randomly in harsh environments such as a battlefield or placed at specified locations. The sensors coordinate among themselves to form a communication network such as a single multi-hop network or a hierarchical organization with several clusters and cluster heads. The sensors periodically sense the data, process it and transmit it to the base station.

2.2 CS for WSNs

A typical wireless sensor network, consists of a large number of wireless sensor nodes, spatially distributed over a region of interest, that can sense (and potentially actuate) the physical environment in a variety of modalities, including acoustic, seismic, thermal, and infrared. A wide range of applications of sensor networks are being envisioned in a number of areas, including geographical monitoring, inventory management, homeland security, and health care. The essential task in many applications of sensor networks is to extract some relevant information from distributed data and wirelessly deliver it to a distant destination (the sink node). While this task can be accomplished in a number of ways, one particularly attractive technique leverages the theory of CS and corresponds to delivering random projections of the sensor network data to the sink. In contrast to classical approaches, where the data is first compressed and then transmitted to a given destination, with CS the compression phase can be jointly executed with data transmission. This is important for WSNs as compressing the data before the transmission to the data gathering point (hereafter called the sink) requires to know in advance the correlation properties of the input signal over the entire network (or over a large part of it) and this implies high transmission costs. With CS, the content of packets can be mixed as they are routed towards the sink. Under certain conditions, CS allows to reconstruct all sensor readings of the network using much fewer transmissions than routing or aggregation schemes. When we utilize CS at the sink node, we obtain more valuable information, and the received values are linear random combinations of the sensor nodes. Nevertheless, there still remain certain problems critical for evaluating the performance of CS for WSNs. (1) How to choose two matrices ϕ and Ψ in the data gathering protocol since the sparsity requirements and the incoherence ought to be met. (2) The energy consumption of transmitting the random combined data of the sensor nodes in the process of CS should be took into consideration due to the energy limitation of WSNs. (3) The robustness of CS in WSNs is also a problem because of the node fails and the harsh environment.

3 Conclusion

3.1 Current achievements

We have explained carefully about the sensing and recovery procedure of data applying compressive sensing. The key points are almost covered in this paper, such as the sparsity of signals, how to deal with it when the signal is not sparse at all, what incoherence means to the application of compressive sensing. Also, we take a quite deep look at the recovery, since this is the most vital process to possess the data we would like to get. During the procedure of explaining sparse recovery, RIP is interpreted, basic methods and some definition of the symbols that we are not familiar with are covered, moreover, we view the recovery method from the angle of geometry. Further more, we present how to function properly when noise is added and analyze the error occurred during sparse recovery.

3.2 Our obstacles

Our first touch with compressive sensing is accidental, but we are totally attracted by this revolution in data sampling and acquisition. Before we are informed of the conception of compressive sensing, we never knew that we have the capacity to reconstruct any signal by sampling at a rate less than its highest frequency. Our group soon developed a great interest in the field of CS and thus changed the project topic from "routing protocols in WSNs" to "compressive sensing". Although we have done much work on our previous topic, we finally made up our minds to exploit further the field of compressive sensing. Accordingly, the time left for us to read papers is rather limited. Due to the time constraint and a lack of certain basic knowledge in mathematic, it is rather a tough task for us to understand the general ideas of compressive sensing. And there is also no doubt that we are only able to develop several preliminary ideas on CS, but we will stick to optimizing the opinion.

3.3 Further work

Through less than a week's study, it seems to us that compressive sensing is such an innovative idea that it almost overturned the beliefs Shannon has built in our mind. Moreover, compressive sensing was built upon mathematic base rather than certain experience, which determines it a theory that can be utilized to a myriad of aspects. Because we have also read a lot of papers in the area of WSNs, our further work may focus on the application of CS in WSNs and develop a data gathering protocol for CS. At last, we will make a evaluation on this application. To achieve our goals, we will not satisfy only about the application of compressive sensing, but to pursue a general comprehension of the whole procedure, i.e. we will apprehend not only how this new technique can be applied but also the foundation it lies on, such as the intrinsic meaning of mathematical theorems and so forth. In the next few weeks, we will keep reading papers in this area and get in touch with the CS coding in order to attain a better understand of the thought.

4 References

References

[1] Richard Baraniuk, Justin Romberg, Michael Wakin, "Tutorial on Compressive Sensing", Information Theory and Applications Workshop, 2008

[2] Olga V. Holtz, "An Introduction to Compressive Sensing", UC Berkeley and TU Berlin, Stanford, January 2009

[3] Emmanuel J. Candes and Michael B. Wakin, "An Introduction To Compressive Sampling", IEEE Signal Processing Magazine, March 2008 Workshop, 2008

[4] Jarvis Haupt, Waheed U. Bajwa, Michael Rabbat, and Robert Nowak, "Compressed Sensing for Networked Data," December 5, 2007

[5] Antonio Ortega, "Novel Distributed Wavelet Transforms and Routing Algorithms for Efficient Data Gathering," Jet Propulsion Laboratory, NASA

[6] Jarvis Haupt and Robert Nowak, "A Generalized Restricted Isometry Property," University of Wisconsin Technical Report ECE-07-1 May 2007, Revised December 2007

[7] Emmanuel J. Cands, Justin Romberg, Member, IEEE, and Terence Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information," IEEE Transactions on Information Theory, VOL. 52, NO. 2, February 2006

[8] GiorgioQuer, Riccardo Masiero, Daniele Munaretto, "On the Interplay Between Routing and Signal Representation for Compressive Sensing in Wireless Sensor Networks," DEI, University of Padova, via Gradenigo 6/B - 35131, Padova, Italy.

[9] Yair Weiss, Hyun Sung Chang and William T. Freeman, "Learning Compressed Sensing", Allerton 2007

[10] Emmanuel Cand‘es? and Justin Romberg, "Sparsity and Incoherence in Compressive Sampling", Applied and Computational Mathematics, Caltech, Pasadena, CA 91125, Electrical and Computer Engineering, Georgia Tech, Atlanta, GA 90332, November 2006

[11] David L. Donoho, "Compressed Sensing." IEEE Transactions on Information Theory, VOL. 52, NO. 4, APRIL 2006

[12] Justin Romberg and Michael Wakin, "Compressed Sensing: A Tutorial", IEEE Statistical Signal Processing Workshop, Madison, Wisconsin, August 26, 2007

[13] Emmanuel J. Cands, "The Restricted Isometry Property and Its Implications for Compressed Sensing," Applied Computational Mathematics, California Institute of Technology, Pasadena, CA 91125-5000, 2008

[14] Emmanuel Cand‘es and Justin Romberg, "l1-magic: Recovery of Sparse Signals via Convex Programming", October 2005

[15] Richard G. Baraniuk, "Compressive Sensing," IEEE Signal Processing Magazine July 2007

[16] Jarvis Haupt and Rob Nowak, "Signal reconstruction from noisy random projections", IEEE Trans. on Information Theory, 52(9), pp. 4036-4048, September 2006

[17] R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, and M. J. Strauss, "Combining geometry and combinatorics: A

unified approach to sparse signal recovery", 2008

[18] Gotz Pfander and Holger Rauhut, "Sparsity in time-frequency representations", 2007

[19] I.F. Akyildiz et al., Wireless sensor networks: a survey, Computer Networks 38 (4) (2002) 393C422.

[20] S. Tilak et al., A Taxonomy of Wireless Microsensor Network Models, in ACM Mobile Computing and Communications Review (MC2R), June 2002.