Compressed Sensing, Compressed Classification and Joint Signal Recovery

Mohammad Emtiyaz Khan
Department of Computer Science
University of British Columbia
emtiyaz@cs.ubc.ca

Abstract

We review compressive sensing and its extension to classification and joint signal recovery. We present an overview of compressed sensing, followed by some simulation results on perfect reconstruction for sparse signals. We review previous work on compressed signal classification and discuss relations between the two earlier papers. Finally, we discuss joint signal reconstruction for compressed sensing.

1 Introduction

Compressive sensing involves talking randomized projection of a signal. It is now well known that the signal can be reconstructed back using a small number of measurements with optimization methods. Our motive in this report is to review some important results of compressive sensing. We are interested in the theory of compressive sensing and extensions to the classification and distributed estimation. There are quite a lot of literature available for the theory of compressive sensing. The work of Donoho et. al. and Candes et.al. is quite insightful in our view, and we review these two papers. There are some preliminary work on compressed classification of which we review the work of Davenport et. al. and Nowak et. al. In addition, we also review the work on distributed compressed sensing.

Contributions To summarize we make the following contributions: **(A)** We review the theory of compressed sensing discussed in [8], [5], [3]. We define compressed sensing, review methods, and show results on signal recovery. **(B)** We review compressed classification discussed in [9],[7]. We compare results of these two papers and show that there are some relations between these results. We verify results of [9]. **(C)** We review the joint signal reconstruction discussed in [11] and verify the simulations.

2 Compressive Sensing : Overview

In this section, we present an overview of compressed sensing based on [8], [5], [3].

Measure of Sparsity Let x be a vector in $\mathcal{R}^N$. Suppose there is an orthonormal basis $\{\Phi_i : i = 1, \ldots, N\}$ for $\mathcal{R}^N$. This basis can be an orthonormal wavelet basis, a Fourier basis or even euclidean basis vectors. A measure of sparsity called K-sparse is defined in [5]. A vector x is said to be K-sparse if it can be represented as a linear combination of K

vectors from the basis Φ. However note that this is a very restrictive definition, and doesn't include many real-wold signal. A better way to define sparsity would be to say that the signal is sparse if x is "well-approximated" by a linear combination of K vectors from the basis Φ. A more precise definition is given in [8]. Let $\theta_i = \langle x, \Phi_i \rangle$ are the projections of x on the basis Φ, then the vector x can be said to be sparse in basis Φ, if for some $0 < p < 2$ and for some $R > 0$,

$$||\theta||_p \equiv \Big(|\theta_i|^p\Big)^{1/p} \leq R \tag{1}$$

A trivial example is a "bounded" sparse signal in euclidean basis for which $||x||_1 \leq R$ where R is the bound on each value.

Information Operator and Reconstruction Algorithm Let X denote the set of vectors of interest. In compressive sensing, we are interested in designing an information operator $I_M : X \rightarrow \mathcal{R}^M$ that samples information from the vector. The sampler is a projection of the vector on $\mathcal{R}^n$: $I_M(x) = (\langle \zeta_1, x\rangle, \ldots, \langle \zeta_M, x\rangle)$. We are also interested in a reconstruction algorithm $A_n : \mathcal{R}^n \rightarrow R^m$ which reconstructs the signal back. Finally, we want a pair of optimal information operator and optimal recovery algorithm such that it minimizes a function of $||x - A_n(I_n(x))||_p$ for some p over all $x \in X$. In the existing literature, the object of interest are the sparse signals satisfying one of the sparsity measure. From now on, we will denote the set of K-sparse vectors as X_K.

Exact Recovery of K-Sparse Signals We now consider some results for compressed sensing of sparse signals. Although these results are first proposed in [6] and [5], the discussion in [3] are easier to understand, hence we follow the discussion in [3](although all the papers have almost identical results). We consider a K-sparse vector $x \in X_K$. It is shown that a matrix $\Phi \in \mathcal{R}^{M \times N}$, which follows *Restricted Isometry Property* (RIP), is an information operator, and given the measurements $y = \Phi x$ we can reconstruct the vector, with high probability, by solving the following linear program:

$$\min ||x||_1 \quad \text{subject to } \Phi x = y \tag{2}$$

RIP means that any two well-separated signals in $\mathcal{R}^N$ remains separated upon projection to $\mathcal{R}^M$. Formally Φ satisfies RIP of order K if for every $x \in X_K$, the following inequality is satisfied for some $\epsilon > 0$:

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{||\Phi x||_2}{||x||_2} \leq (1+\epsilon)\sqrt{\frac{M}{N}} \tag{3}$$

be a projection from $\mathcal{R}^N$ to $\mathcal{R}^M$. It is now known that various matrices satisfy this property if $M = O(K \log(N/K))$, for example, an i.i.d Gaussian matrix (see [5] more examples).

Exact Recovery of Noisy K-Sparse Signals If measurements are noisy, then a measure of sparsity can be defined in terms of the perturbation caused by noise, for e.g. $||e||_{l_2} \leq \epsilon$. In such cases, the signal can be reconstructed back from measurements $y = \Phi x + e$, by solving the following program:

$$\min ||x||_1 \quad \text{subject to } ||\Phi x - y|| \leq \epsilon \tag{4}$$

Note that both the programs given by Eq. (2) and (4) can be solved by reformulation as a linear program with linear/quadratic constraints. Various software packages are available to solve these optimization problems (for a list of software, see [1]). We used l_1-magic for our implementation [2].

Results We verify the results given in [5] for sparse signals. We used the l_1-magic optimization routines to get these plots. We consider both noisy and noiseless case. For signals without noise, we take $N = 1024$ with a sparsity $K = 50$ samples, and for noisy case we set $N = 512$ with $K = 20$ with noise $\sigma^2 = 0.05$. We consider Φ to be a matrix with i.i.d. Gaussian entries. Fig. 1(a) shows the original and reconstructed signals with $M = 250$

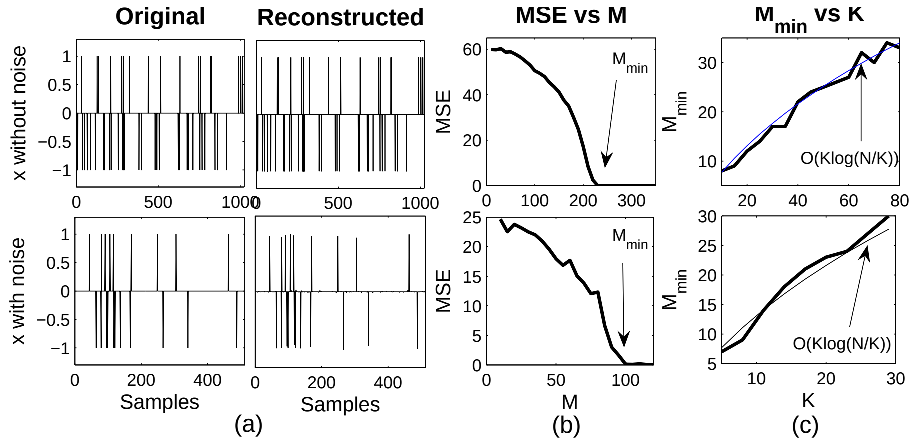

Figure 1: Top row contains results for signals without noise, while bottom row containes results for signal with noise. (a) Perfect reconstruction of a sparse signal (b) MSE vs number of measurements M (c) M_{min} vs K; thick line shows M_{min} while the thin line shows $O(K\log(N/K)$ scaled to match the starting point of M_{min} curve.

and $M = 120$ for signals without noise and with noise. We see that the reconstruction is perfect. Fig. 1(b) shows the plot of empirical mean-square error (MSE) computed with different measurements of length M. Because of the large computation time, we only did 50 simulations to obtain each point in MSE curve. We observe that the mean square error drops to zero at $M = 220$ for signals without noise, and at $M = 100$ for noisy signal. Next we find the minimum value of M at which we get perfect reconstruction as a function of K, the sparsity measure. For this purpose, we compute the MSE for various K and M. For each value of K, we visually find M_{min} at which MSE drops to zero. Fig. 1(c) shows the plot (thick line shows M_{min} while the thin line shows $O(K\log(N/K)$ scaled to match the starting point of M_{min} plot). We see that the minimum value of M follows $O(K\log(N/K)$, which verifies the results.

3 Compressive classification

We now review the work on application of compressive sensing for classification. We are interested finding out the effect of compressive sensing on classification, and obtain the bound on performance. There has been some work in the area, for example comparison between active learning and compressed sensing by Nowak et. al.. However most of the work is applied to special signals. We will discuss the results in [9] which is the most general case. We also find that a preliminary result published in [7] supports and justifies the results presented in [9]. Hence we review both [7] and [9], and verify the results shown in these papers.

Classical Detector We first discuss a classical detector and then described how a compressed detector can be derived using the same approach. Let us say that there are two hypotheses concerning the signal; that it is present in the measurements or it is not. The classical Neymon-Pearson (NP) detector involves a *likelihood ratio test* where the sufficient statistics $t \equiv \langle y, x \rangle$ is compared against a threshold γ. Here y are the measurements, x is the signal of interest and γ is set to achieve certain probability of false alarm rate $P_F \leq \alpha$ for some $0 \geq \alpha \leq 1$. It is easy to show that (for example, see [10]):

$$P_D(\alpha) = Q(Q^{-1}(\alpha) - \sqrt{\text{SNR}}) \tag{5}$$

where $Q(\cdot)$ is the flipped version of standard Gaussian cumulative distribution function.

Compressed Detector As described in [7], this theory can be easily extended to the case when the measurements are made using a compressed sampler. So we consider the following hypothesis:

$$\mathcal{H}_0 \quad : \quad y = \Phi n \tag{6}$$

$$\mathcal{H}_1 \quad : \quad y = \Phi(x + n) \tag{7}$$

where $n \sim \mathcal{N}(0, \sigma^2)$ is white Gaussian noise. It is straightforward to show that in this case the sufficient statistics is $\tilde{t} \equiv \langle y, \Phi x\rangle$. It is shown in [7] that for some $\epsilon > 0$, the probability of false rate is approximately given by the following equation:

$$P_D(\alpha) \approx Q(Q^{-1}(\alpha) - \sqrt{M/N}\sqrt{\text{SNR}}) \tag{8}$$

Comparing Eq. (8) and (5), we can see that the performance of the detector will be deteriorated with M decreasing (as expected), and the rate of performance degradation depends on SNR.

Compressed Classifier These results have been generalized for a multi-class classifier where we have a set $X = \{x_i\}$ of signals. It is straightforward to show that the detection rule is $\hat{x} = \min_{x_i \in X} ||y - \Phi x_i||_2$. There is an important point here. We discussed in Section 2 that the matrix Φ approximately preserves the distances in transformation. However for classification the distances are not preserved, rather *uniformly shrunken*. This can be easily seen in the following theorem proved in [7]: for any signal x, and typical values of ϵ, with probability at least $1 - \delta$, the following holds for all x_i,

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{||\Phi(x - x_i)||_2}{||(x - x_i)||_2} \leq (1+\epsilon)\sqrt{\frac{M}{N}} \tag{9}$$

Also because of the dependence on SNR, effect of noise can be amplified with the transformation.

An Upper Bound on Compressed Classification The following theorem proved in [9] gives a bound on the compressed classification: Let us say that the cardinality of set X is K, and there are M measurements, then the probability of error is upper bounded as follows,

$$P(\text{error}) \leq (K-1)\Big(1 + \frac{d_{min}}{4}\text{SNR}\Big)^{-M/2} \tag{10}$$

where $d_{min} = \min_{x_i, x_j} ||x_i - x_j||^2$. We see that the bound indeed depends on SNR and it has an exponential relation with M, which supports the results of [7]. Surprisingly, this relation has not been discussed in [9]. Also the classification rule $\hat{x}$ is called "empirical" in [9], while it is just a generalization of NP detectors and hence is theoretically justifiable.

Results We now show simulation results for compressed classification. We verify the simulations presented in [9]. We construct the set X by sampling a collection of linear chirp signal given as follows:

$$f_l(i) = \cos\Big(2\pi\frac{l \times 10^6}{2}\Big\lfloor\frac{i}{200 \times 10^6}\Big\rfloor\Big\lceil\frac{200 \times 10^6}{400}\Big\rceil\Big) \tag{11}$$

So we have a sampling rate of 200 MHz, with range 1 MHz to 100 MHz with 1 Hz increments. Hence N = 400, M = 100. We performed compressed detection with various number of measurements and compute the probability of error by running each simulation 10000 times. Results are shown in Fig. 2(left) for SNR = 2 (or 3dB), where we plot the log of probability of error versus M. The figures also show the theoretical bounds. We also plot the probability of error versus SNR (this simulation is not shown in [9]). The results are shown in Fig. 2(right). Note that the y-axis is logarithmic and hence we can see that the bound is very loose. However the results verifies that the performance is exponentially related to the number of measurements and SNR.

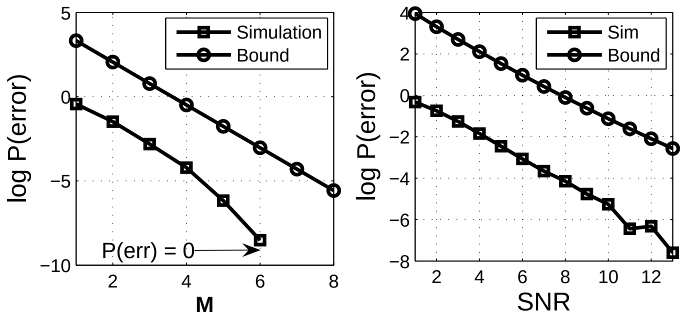

Figure 2: Compressed classification: Plot of log of Probability of error for various number of observations M (right) and SNR (left). Both simulated and the theoretical bounds are shown. Note y-axis is in logarithmic scale.

4 Joint Sparsity Model and Distributed Compressive Sensing

The notion of signal being sparse in some basis has been generalized to the notion of an ensemble of signals being *jointly sparse* in [4][11]. The main idea is that if some signals contain a common information, the number of measurements for perfect reconstruction can be decreased. In the model described in [4], the signals are assumed to be sum of two components with one component common to all the signals, i.e. for $i = 1, \dots, J$,

$$x_i = z_C + z_i \tag{12}$$

Here z_C is a $K-$sparse signal common to all the signals x_i, and z_i is a K_i-sparse innovation. A practical example could be a set of temperature sensors geographically separated, which measure a common temperature with a little variation in space. There are two other kinds of model described in the paper, however the basic idea is same and hence we don't discuss those here.

Joint Signal Recovery In this discussion we assume $J = 2$. The measurement model is $y_i = \Phi_i x_i$. We define,

$$z = \begin{bmatrix} z_C \\ z_1 \\ z_2 \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \Phi = \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix}, \tilde{\Psi} = \begin{bmatrix} \Psi & \Psi & 0 \\ \Psi & 0 & \Psi \end{bmatrix}$$

It is shown in [4] that by solving the following linear program, we can recover both the components:

$$\min ||z||_1 \quad \text{subject to } \Phi\tilde{\Psi}x = y \tag{13}$$

However in [4] the results are produced by using a modified penalty, $\gamma_z||z||_1 + \gamma_1||z_1||_1 + \gamma_2||z_2||_1$, where γ_i are found with optimization. We found that it is just because of the numerical issues in the solver, and for our implementation we could repeat the results without any modification of the penalty. We now describe the results.

Results We choose a signal length of $N = 50$ with different sparsity K,K_1 and K_2. We repeat the experiments 1000 times to calculate the probability of perfect reconstruction. Results are shown in Fig. 3. We note that as the value of K is decreased required measurements for perfect reconstruction with joint increases. Our results for joint reconstruction are exactly same as the results described in [4], however there seem to be a bug in our implementation of separate recovery. We can see that the separate recovery is giving much better results than joint reconstruction for the last case. The results of [4] suggests that both the curves should come closer and closer, which is intuitively correct.

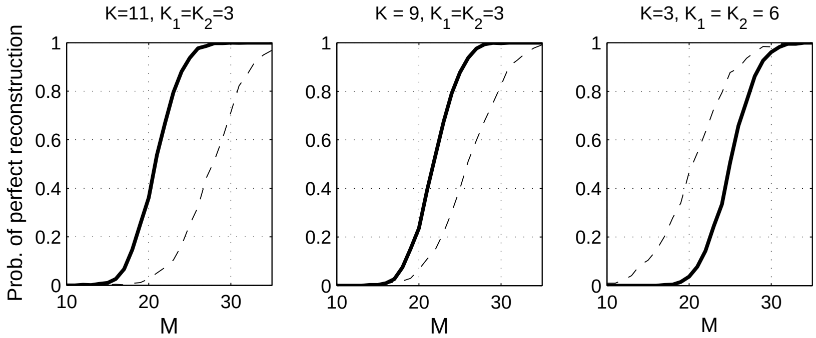

Figure 3: Comparison of joint reconstruction with separate reconstruction for various values of sparsity

5 Conclusions

In this paper, we reviewed and verified the results on compressed sensing, classification and joint signal recovery. We were able to repeat the results of the previous work and also conduct some experiment to gain new insight on the previous works. Our implementation on joint signal recovery still need some work.

References

[1] Compressed Sensing Resources. http://www.dsp.ece.rice.edu/cs/.

[2] L1-magic. http://www.acm.caltech.edu/l1magic/.

[3] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. The Johnson-Lindenstrauss Lemma Meets Compressed Sensing. *Preprint*, 100(1):0, 2006.

[4] D. Baron, M.F. Duarte, S. Sarvotham, M.B. Wakin, and R.G. Baraniuk. An information-theoretic approach to distributed compressed sensing. *Allerton Conf. Comm., Control, Comput*, 2005.

[5] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. on Pure and Applied Math*, 2005.

[6] EJ Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[7] M.A. Davenport, M.B. Wakin, and R.G. Baraniuk. Detection and Estimation with Compressive Measurements. Technical report, Department of Electrical and Computer Engineering, Rice University, 2006.

[8] DL Donoho. Compressed Sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[9] Jarvis Haupt, Rui Castro, Robert Nowak, Gerald Fudge, and Alex Yeh. Compressive sampling for signal classification. In *Proc. Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California, October*, 2006.

[10] S.M. Kay. *Fundamentals of statistical signal processing*. Prentice Hall, 1998.

[11] M.B. Wakin, M.F. Duarte, S. Sarvotham, D. Baron, and R.G. Baraniuk. Recovery of Jointly Sparse Signals from Few Random Projections. *Proc. Neural Information Processing Systems (NIPS)*, 2005.