Derivation of Error Probability of Compressive Sensing Methods Based on Information Theoretic Concepts

Sepideh Almasi*, Hamid Soltanian-Zadeh**
*Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran 14395-515, Iran.
, almasi@ut.ac.ir
**Image Analysis Laboratory, Radiology Department, Henry Ford Health System, Detroit, MI 48202, USA.
Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran 14395-515, Iran.
, hszadeh@ut.ac.ir, hamids@rad.hfh.edu

Abstract: *A methodical definition of the probability of error for each of compressive sensing methods, a performance evaluating tool that has not been defined yet, is the theme of this work. To this end, an efficient information theoretic method is proposed in order to achieve a reliable measure. With making a replication with the communication channels, the probability of error is set as a function of the Mutual Information (MI) with the Compressive sensing process. Other methods of performance assessment including different kinds of error such as Mean Square Error (MSE) are in use prevalently; however, fail in some cases because of the lack in inclusion of the coherency between structure of the result and that of the original signal. MI helps to overcome this deficiency in the performance analysis of the compressive sensing methods.*

Keywords: compressive sensing, mutual information, error probability.

1. Introduction

Nearly most of the natural and artificial signals employed by human have a sparse representation in a specific space called sparse space. This means that they have very few nonzero elements in the sparse space. Owing to this specification, one can utilize appropriate algorithms to work with sparse signals more efficiently with less energy. In recent years, compressive sensing has been considered as a powerful tool for manipulating this common area of signal processing and information theory. It has found wide range of applications in different sciences and fields of study such as data networks, medical technologies, image processing, radar and sonar, communication systems, and optical devices. In spite of this applicability, insufficient amount of work has been directed toward establishing concrete formulations regarding its performance evaluation. **[1]**, **[2]** are two cases investigating this subject from an error point of view. They set bounds on the *MSE* under exclusive conditions. However, a unified measure of performance has never been studied for probability of error.

By a comparison between communication channels and compressive sensing, a fundamental similarity in their structure is detectable **[3]**. They both encode a message and finally decode the received noisy version of it with some nonlinear methods. Employing this similarity, one can take advantages of analyses presented for the communication channels in the hope of establishing a performance measure for the compressive sensing methods. There exists a proven connection between the error probability and mutual information for data transmission over the communication channels **[4]**. After some modifications, it is applicable to the compressive sensing context. Experimental results in Section 5 verify this claim.

In this paper, evaluation for the probability of error is done for two methods of Basis Pursuit and Orthogonal Matching Pursuit **[5]**, representatives of two main groups of sparse reconstruction methods. This task is accomplished through connecting probability of error to the mutual information specified to each of these algorithms. Being able to represent non-linear statistical dependence between random variables makes MI an ideal choice consistent with the inherent non-linearity of CS. It takes its highest amount when the original signal and the reconstructed signal are exactly the same so that given its value, one can say how correct the answer is. Therefore, it has a close and unitary connection with the error probability in concept.

The materials in the paper are presented in the following order. First, a mathematical description of compressive sensing and two methods of basis pursuit and orthogonal matching pursuit are provided in Section 2. Next, in Section 3, definition of mutual information is given. The error probability formulation approach is also thoroughly presented in Section 4, and the experimental results and conclusion are provided in Section 5 and 6, respectively.

2. Description of The Compressive Sensing Methods

In this section a brief review of the CS methods is presented. There are two separate parts for a CS process: sampling and reconstruction. The general sampling process is shown in Equation (1).

$$y_{m\times1} = A_{m\times n}.x_{n\times1} + n_{m\times1} \tag{1}$$

where x is the original sparse signal, A is the measurement matrix, and n is the additional noise. In order to have most informative samples, matrix A must have statistically independent entries. Sparse, Fourier, Hadamard, and Gaussian matrices are qualified instances for this end. There are miscellaneous algorithms capable of reconstructing the original signal with high precision under specific conditions. From these methods, Basis Pursuit and Orthogonal Matching Pursuit are explained in the following sections.

2.1 Basis Pursuit

The most precise CS methods are those based on convex optimization methods such as l_p minimizing ($0 < p \leq 1$) and different modifications of this method: Basis Pursuit (BP), defined in Equation (2), is in use more than the others.

$$\min \|x\|_1 \; s.t. \;\; y = A.x \tag{2}$$

Due to computational complexity, this method has a low rate of response, and is suitable only for cases that precision plays the key role. When measurements are noisy, Equation (3) will be used instead of (2).

$$\min \|x\|_1 \; s.t. \;\; \|y - A.x\|_2 \leq \varepsilon \tag{3}$$

where $\varepsilon > 0$ is a noise-related parameter.

2.2 Orthogonal Matching Pursuit

In addition to the above methods, there is also another group of algorithms that find a solution based on iterative calculations. Less computational complexity of this set has made them very fast. This property makes it advantageous over the former group of algorithms in the cases which prefer low computational complexity relative to precision. Orthogonal Matching Pursuit (OMP), Stage-wise OMP (St-OMP), and Regularized OMP (ROMP) are three examples of this group of methods. In this group, OMP is the most common tool for signal reconstruction. The way by which OMP reconstructs sparse signals in seven steps is as follows **[5]**:

1. Initialization: residual $r_0 = v$, index set $\Lambda_0 = \emptyset$, and counter t=1.
2. Finding the index λ_t that solves the easy optimization problem:

$$\lambda_t = argmax_{j=1,\dots,d} \langle r_{t-1}, \varphi_j \rangle$$

3. Augmenting the index set and the matrix of chosen atoms:

$$\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\} and\ \Phi_t = [\Phi_{t-1}\ \varphi_{\lambda_t}]$$

4. Solving a least squares problem to obtain a new signal estimate:

$$x_t = argmin_x \|v - \Phi_t x\|_2$$

5. Calculating new approximation of the data and new residual

$$\alpha_t = \Phi_t\ x_t$$
$$r_t = v - \alpha_t$$

6. Incrementing t and returning to step 2 if $t < m$.
7. The estimate $\hat{s}$ for the ideal signal has nonzero indices at the components listed in Λ_m. The value of the estimate $\hat{s}$ in component λ_j equals the jth component of x_t.

3. Mutual Information

Mutual information is a methodical measure of the dependence between random variables **[4]**. Being always non-negative, dimensionless, in units of bits, and zero only in the cases that the variables are statistically independent are its key properties. In brief, the mutual information functions to determine the degree of structural dependence between random variables; In other words, mutual information is a kind of tool from some other ones that determines how much one random variable tells us about another; it can be thought of as the reduction in uncertainty about one random variable given characteristics of another. The formula by which MI can be calculated is shown in Equation (4).

$$I(X;Y) = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} = E_{P_{XY}} \log \frac{P_{XY}}{P_X P_Y} \tag{4}$$

Also, another definition for the mutual information I between m (scalar) random variables y_i, $i=1...m$, is as follows.

$$I(y_1, y_2, \dots, y_m) = \sum_i H(y_i) - H(y) \tag{5}$$

where H denotes differential entropy.

These definitions require the probability distribution of both of the variables with their joint probability distribution. This causes computational complexities and difficulties regarding numerical integration of the functions in Equation (4) and Equation (5). To use the definition of entropy, we have to estimate the density functions. This problem has made restrictions on the use of mutual information in some applications **[6]**. In some research works, an approximation of the mutual information is made based on polynomial density expansions **[6,7]** or on the cumulative residuals **[8]**. By means of histograms, calculating MI between random variables is also viable **[9]**. In this work, the later approach is selected to achieve proper results.

4. Error Probability Verification

One can summarize a message transmission through a communication channel as shown in Fig. 1.

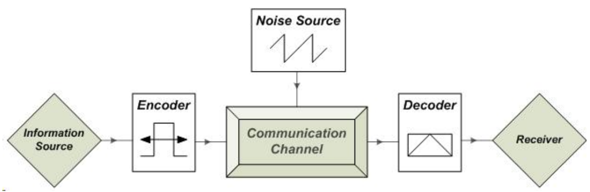

Fig. 1: Block diagram of a communication channel.

The message *x* would be encoded to a codeword from a collection of *M* possible alternatives according to the model presented in Fig. 2. Here, the length of the codeword is considered *m* which equals to $\log_2 M$. After this stage, channel will add noise to the messages and gives the y vector. In most applications, an additive white Gaussian noise (AWGN) is considered for the above model. By decoding of this vector, recovered message $\hat{x}$ is obtained.

$$x \xrightarrow{Encoding} s_1 s_2 \dots s_m \rightarrow [noisy\ channel{:}\ P_{Y|S}(y|s)] \rightarrow y_1 y_2 \dots y_m \xrightarrow{Decoding} \hat{x}$$

Fig. 2: Mathematical model of the communication channel.

The final objective is that $\hat{x}$ is most similar to *x*. Though, presence of noise makes it open to not having a definite recovery, and information loss is inevitable. According to the noisy channel coding theorem, when transmission rate, $R = \frac{\log(M)}{n}$, exceeds the channel capacity, probability of error increases rapidly. In addition, the highest number of messages that can be transmitted almost error-free is a function of the mutual information between *x* and $\hat{x}$. In this way, mutual information and communication channels will be linked. From large deviation theory **[4]**, probability of error for each message (P_n), when *n* is large enough, is connected with the mutual information as follows.

$$P_n = 2^{-nI\ (x;\hat{x})} \tag{6}$$

We can relate the error probability to the P_n as shown in equation (7). An approximation is also made here for more simplicity in computations when *M* is large.

$$P_e = 1 - (1 - P_n)^M \approx 1 - \exp(-MP_n) \tag{7}$$

Simply, we can apply this model to the CS process as it has been done before in other scenarios **[10-12]**, **[3]**. To this end, some adaptations are made here. First, the *M* value is chosen as the number of possible ways that one can construct a binary k-sparse signal with *n* dimension so that $M = \binom{n}{k}$. When the measurement matrix has a Gaussian random distribution with each column normalized to unity and the original signal is binary, according to the Central Limit Theory (CLT), it is possible to consider information loss as an additive Gaussian noise:

$$P_{n,CS} = 2^{-\frac{I'(x;\hat{x})}{R}} = 2^{-n\frac{I'(x;\hat{x})}{k}} \tag{8}$$

where $I'(x;\hat{x})$ is the MI between *x* and $\hat{x}$ normalized by the *x* entropy:

$$I'(x;\hat{x}) = \frac{I(x;\hat{x})}{H(x)} \tag{9}$$

By substituting *M* and $\mathrm{P_n}$ in (7), the final expression for the error probability will be:

$$P_e \approx 1 - \exp\left(-\binom{n}{k} 2^{-n\frac{I'(x;\hat{x})}{k}}\right) \tag{10}$$

5. Experimental Results

For the purpose of validating the established relationship, a series of experiments are done. The experimental results are compared to the theoretical ones. The settings considered here for the experiments are as follows: measurement noise has Gaussian distribution and the Signal to Noise Ratio (*SNR*) is 100. The length of the original signals (*n*) equals 80 and number of measurements (*m*) is 30. Original binary signal has a uniform distribution of non-zero elements so that they are highly unpredictable from an information theoretic point of view. Measurement matrix has randomly selected entries from a Gaussian distribution with zero mean and unit variance. Columns of this matrix are normalized to one as stated in the previous section.

The criterion by which the practical probability of error is computed is the *MSE* defined in (11).

$$MSE = E\{\|x - \hat{x}\|^2\} \tag{11}$$

Since we do not have the joint probability of x, because of the high computational complexity in its computation, the expectation operator is substituted with the averaging over the vector elements. When the *MSE* is greater than 0.01, a failure in reconstruction can be detected. Over a hundred trials, the number of failures divided by 100 equals the practical error probability.

The first part of experiments is dedicated to the normalized MI determination for two methods. The results of training based computation of the normalized MI are shown in Fig. 3. Through several examinations, it is established that these amounts of mutual information are constant for a specific setting; for several settings, a complete collection of mutual information amounts for different methods is obtainable and applicable.

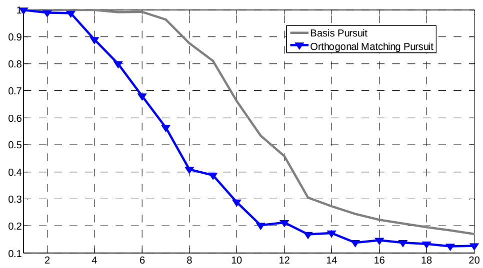

Fig. 3: Normalized mutual information for BP and OMP methods vs. sparsity of the original signal

Results of practical error calculation with the theoretical amounts are shown in fig. 4 and fig. 5 for BP and OMP methods, respectively. It is apparent that the estimations are highly precise and almost similar to the real results.

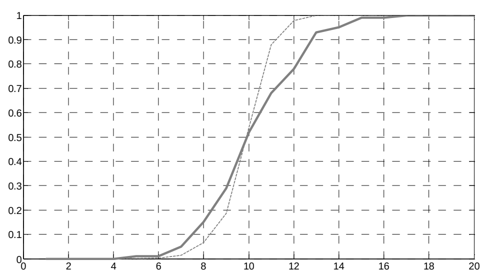

Fig. 4: Theoretical (dashed curve) and experimental values (solid curve) of error probability for BP method vs. sparsity of the original signal.

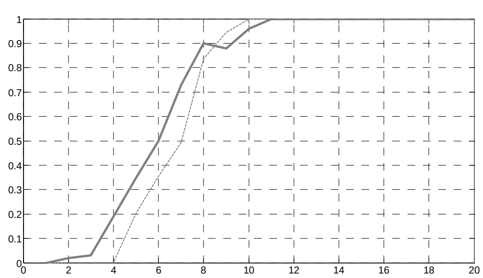

Fig. 5: Theoretical (dashed curve) and experimental values (solid curve) of error probability for OMP method vs. sparsity of the original signal.

5. Conclusion

In this work, an estimation of error probability for two basic methods of Basis Pursuit and Orthogonal Matching Pursuit is worked out. This formula can be generalized to other methods of compressive sensing as well. Making a scientific comparison between communication channels and compressive sensing has helped us to drive this approximation. By this mean, probability of error is linked to the mutual information, a coherency detection measure between two random variables. Experimental results suggest that this measure is always constant for a specific compressive sensing process with predefined settings. Based on this verity, one can obtain it for each method with a training practice, and substitute it in the proposed formula in order to obtain the related probability of error. As the test results show, this is a reliable approximation.

References

[1] M. Babaie-Zadeh, H. Mohimani, and C. Jutten, "An upper bound on the estimation error of the sparsest solution of underdetermined linear systems," in *Proc. Sparse2009,* Saint-Malo, France, Apr. 2009.

[2] B. Babadi, N. Kalouptsidis, and V. Taokh, "Asymptotic achievability of the Cramer-Rao bound for noisy compressive sampling," *IEEE Trans. on Signal Processing,* vol. 57, no. 13, pp. 1233-1236, Mar. 2009.

[3] S. Sarvotham, D. Baron, and R. Baraniuk, "Measurement vs. bits: Compressed sensing meets information theory," in *Proc. Allerton Conf. Communication, Control, and Computing,* Monticello, IL, pp. 11-14, Sep. 2006.

[4] T.M. Cover and J.A. Thomas, *Elements of Information Theory,* 2nd Edition, Wiley- Interscience, New Jersey, 2006.

[5] J.A. Tropp and A.C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Information Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[6] J.W. Williams and Y. Li, *Estimation of Mutual Information: A Survey*, Springer, Berlin Heidelberg, 2009.

[7] K.E. Hild, D. Erdogmus, and J. Principe, "Blind source separation using Renyi'a mutual information," *IEEE letters Signal Processing,* vol. 8, no. 6, pp. 174-176, June 2001.

[8] F. Wang, B.C. Vemuri, M. Rao, and Y. Chen, "Cumulative residual entropy: a new measure of information," *IEEE Trans. on Information Theory,* vol. 50, no. 6, pp. 1220-1228, June 2004.

[9] H.H. Yang and J. Moody, "Feature selection based on joint mutual information," in *Int. Con. Computer Science*, Rochester, NY, June. 2000.

[10] A.K. Fletcher, S. Rangan, and V.K. Goyal, "On the rate-distortion performance of compressed sensing," in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing (ICASSP),* Honolulu, Hawaii, pp. 885-888, Apr. 2007.

[11] A.K. Fletcher, S. Rangan, and V.K. Goyal, "Rate-distortion bounds for sparse approximation," in *Proc. IEEE Int. Workshop on Statistical Signal Processing (SSP),* Madison, Wisc., pp. 254-258, Aug. 2007.

[12] M. Akcakaya and V. Tarokh, "A frame construction and a universal distortion bound for sparse representations," *IEEE Trans. on Signal Processing*, vol. 56, no. 6, pp. 2443-2450, June 2008.