Nonparametric Sparse Representation

Mahmoud Ramezani Mayiami  
ECE Department, Shahed University  
Tehran, Iran  
ma.ramezani@shahed.ac.ir  

Babak Seyf  
ECE Department, Shahed University  
Tehran, Iran

Abstract—This paper suggests a nonparametric scheme to find the sparse solution of the underdetermined system of linear equations in the presence of unknown impulsive or non-Gaussian noise. This approach is robust against any variations of the noise model and its parameters. It is based on minimization of rank pseudo norm of the residual signal and $l_1 - norm$ of the signal of interest, simultaneously. We use the steepest descent method to find the sparse solution via an iterative algorithm. Simulation results show that our proposed method outperforms the existing methods like OMP, BP, Lasso, and BCS whenever the observation vector is contaminated with measurement or environmental non-Gaussian noise with unknown parameters. Furthermore, for low SNR condition, the proposed method has better performance in the presence of Gaussian noise.

Index Terms—Non-Gaussian noise, rank pseudo norm, sparse representation, sparse signal.

I. INTRODUCTION

The approaches for over complete sparse decomposition aim at representing the signals as the combination of a few numbers of columns, namely atoms, which are chosen from an over complete dictionary. Recently, this problem has been applied in a number of fields, like Compressed Sensing or Compressive Sampling (CS) [1], [2] and Blind Source Separation (BSS) [3].

The main assumption in this framework is that the signal has a sparse representation (SR). It means that the total number of nonzero elements, $K$ in the signal coefficient vector is far less than the vector dimension $N$. Roughly speaking, the main problem is to find the sparsest possible $s \in \mathbb{R}^N$ from signal vector $r = \Phi s + n$ when $r$, $n$ are $M \times 1$ signal and noise vectors respectively and $\Phi$ is $M \times N$ dictionary where $M < N$. A standard formulation for the SR problem is given by

$$\hat{s} = \arg \min \|s\|_0 \text{ subject to } \|r - \Phi s\|_2^2 \leq \delta.$$  

(1)

where $\delta$ is the threshold of error and $\|\cdot\|_0$ counts the nonzero elements. In general case, $\|s\|_p = \left(\sum_{i=1}^{N} |s_i|^p \right)^{1/p}$ for $p \geq 1$.

A number of approaches have been proposed to approximate $s$ based on (1) by using pursuit algorithms. For example, Mallat and Zhang [4] have used the maximum correlation between the residual signal $(r - \Phi \hat{s})$ and the atoms of the dictionary to find one of the nonzero elements in each iterations, namely Matching Pursuit (MP). Also, there are other iterative algorithms which used MP with some modifications to find the active coefficients of $s$ and their amplitudes, like Orthogonal MP (OMP) [5] and Stagewise Orthogonal MP (StOMP) [6].

Since finding the sparsest solution via (1) is an NP-hard combinatorial problem [7], another solution for reconstructing $s$ from $r$ is given by

$$\hat{s} = \arg \min \|s\|_1 \text{ subject to } \|r - \Phi s\|_2^2 \leq \delta.$$  

(2)

or equivalently

$$\hat{s} = \arg \min \|r - \Phi s\|_2^2 + \lambda \|s\|_1.$$  

(3)

here $\lambda$ is the Lagrangian multiplier and it controls the effect of the sparsity terms $\|s\|_1$ in the solution. Basis Pursuit (BP) is one of the most familiar approach which solve (3) by using linear programming [8]. Moreover, researchers applied Bayesian framework to solve (3), like sparse Bayesian learning and the relative vector machine (RVM) [9], and Bayesian Compressive Sensing (BCS) [10].

In all of the conventional approaches, researchers have assumed that $n$ has Gaussian model. However, Middleton [11] showed that human activities in urban area may produce some noises with non-Gaussian distribution. Furthermore, non-Gaussian model for noise were used in some applications [12]-[14]. In general, the parameters of the impulsive or non-Gaussian model of noise were supposed to be unknown. Moreover, the non-Gaussian noise for sparse representation problem were considered; for example, Wright and his colleagues [15], found the sparse solution in the presence of impulsive noise via the following optimization problem

$$\hat{s} = \arg \min \|s\|_1 + \|n\|_1 \text{ subject to } r = \Phi s + n.$$  

(4)

Since they used some labeled training images for recognizing the class of the sparse images, the proposed approach in [15] was restricted to apply for object recognition and computer vision problem. Therefore, we cannot apply this method for solving SR problem in the presence of general non-Gaussian model for environmental or measurement noise while it arises in many situations in signal processing or communication area. More precisely, suppose a wireless communication noisy channel with non-Gaussian model of noise. The transmitter sends the sparse signal $s$ with encoded form $\Phi s$ through this channel and the receiver observes the signal $r = \Phi s + n$ (like compressive sensing measurement vector) while the model and the parameters of the noise vector $n$ are unknown for the receiver. The receiver aims to find the sparse coefficient vector $s$ without any knowledge but the elements of $r$ and $\Phi$. In this paper, we investigate SR problem in the presence...
of ambient impulsive noise which is non-Gaussian with unknown parameters.

Seyfe and Sharafat [16] used the properties of pseudo norm to design nonparametric multiuser detector that does not need any a priori information about the ambient noise model.

At the same way, in this paper, a new optimization algorithm for recovering sparse signals based on minimization of rank pseudo norm of \( (r - \Phi s) \) is proposed while we minimize \( l_1\) -norm of the signal of interest \( s \), simultaneously. Our proposed approach is robust in the sense of the variations in the noise model and parameters since we do not use any a priori knowledge about the ambient noise.

The reminder of this paper is organized as follows. Some preliminaries about pseudo norm, especially rank pseudo norm are defined in section II. In section III, our main idea is proposed to find the sparse solution in the presence of non-Gaussian noise. The section IV and V reserved for simulation and conclusion, respectively.

II. CHARACTERISTIC OF PSEUDO NORM

As is well known, the norm function \( \| \cdot \| \) is defined on \( \mathbb{R}^N \) and has four characteristics for \( l \) and \( m \in \mathbb{R}^N \) as following

(a) \( \| l + m \| \leq \| l \| + \| m \| \),
(b) \( \| a l \| = |a| \| l \| \) for \( a \in \mathbb{R} \),
(c) \( \| l \| \geq 0 \) for all \( l \in \mathbb{R}^N \),
(d) \( \| l \| = 0 \) if and only if \( l = 0_N \).

Where \( 0_N \) is a \( N \times 1 \) all zero components vector. The pseudo norm satisfies the first three properties of the norm function, but it has a looser constraint instead of (d). These properties are as follows

(a') \( \| l + m \| \leq \| l \| + \| m \| + \| l \| + \| m \| 

(b') \( \| a l \| = |a| \| l \| \) for \( a \in \mathbb{R} \),
(c') \( \| l \| \geq 0 \) for all \( l \in \mathbb{R}^N \),
(d') \( \| l \| = 0 \) if and only if \( l_i = l_2 = ... = l_N \).

where \( \| \cdot \| \) denotes to the pseudonorm operator and \( l_i, i = 1, 2, ..., N \) are the components of the vector \( l \).

Now, we define the rank pseudo norm by

\[ \| l \|_\rho = \sum_{i=1}^N a(R(l_i))l_i. \]

Here \( R(l_i) \) is the rank of \( l_i, i = 1, 2, ..., N \) among \( N \) elements of the input vector \( l \). Also, \( a(\cdot) \) is a score function which satisfies \( a(1) \leq a(2) \ldots \leq a(N) \) and \( \sum a(i) = 0 \) [17]. It was proved that the rank pseudo norm with \( a(i) = -a(N + 1 - i) \) is a kind of pseudo norm [16].

The general rank score was proposed [17] as \( a_\rho(l) = \varphi \left( \frac{l}{N+1} \right) \) whenever \( \varphi(\cdot) \) satisfies the following constraints

\[ \int_0^1 \varphi(x)dx = 0, \quad \int_0^1 \varphi^2(x)dx = 1. \]

The Wilcoxon function \( \varphi(x) = \sqrt{\pi}Z(x - 0.5) \) and the sign function \( \varphi(x) = \text{sgn}(2x - 1) \) are two kinds of score function \( a(\cdot) \) [16], [17].

III. MAIN IDEA

In the majority of works in the SR research domain, the observation vector is contaminated with Gaussian noise. Hence, the optimum norm function which minimizes the noise power is Euclidean norm. Thus (1), (2) or (3) is an appropriate formulation for SR problem in the presence of known Gaussian noise. However, our simulations show that these optimization approaches fail when the sparse linear model is considered with unknown impulsive or non-Gaussian model of noise.

On the other hand, the characteristics of the rank pseudo norm cause this function to be an efficient tool for overcoming the effect of non-Gaussian noise in the sparse linear regression model [18]. Hence, we suggest our main optimization formulations as follows

\[ \hat{s} = \arg \min_s \| s \|_1 \text{ subject to } \| r - \Phi s \|_\rho \leq \zeta, \]

or equivalently

\[ \hat{s} = \arg \min_s \| r - \Phi s \|_\rho + \lambda \| s \|_1. \]

Since \( \| r - \Phi s \|_\rho \) is a continuous and convex function of \( s \) [19], we can find the sparse vector \( \hat{s} \) through solving \( \nabla f(s) = 0_N \), where \( f(s) = \| r - \Phi s \|_\rho + \lambda \| s \|_1 \) as following

\[ -s^T a(R(r - \Phi s)) + \lambda u(s) = 0. \]

where \( (\cdot)^T \) denotes to the vector transpose. Since \( \| s \|_1 \) is a non-differentiable function, we use the subgradient \( u(s) \) instead of the gradient (see the Appendix). Also we have

\[ a(R(r - \Phi s)) = a(R(r_1 - \Phi_1^T s)), a(R(r_2 - \Phi_2^T s)), \ldots, a(R(r_M - \Phi_M^T s)). \]

Here, \( \Phi_i, i = 1, 2, ..., M \) is the \( i \)-th row of the dictionary. There is not a closed form for the solution of (11), since both of its terms are nonlinear with respect to \( s \). Therefore, we have investigated the iterative solution for recovering \( s \) via (11). But we have achieved a sparse solution with relatively good performance compared with the commonplace reconstruction methods. Instead, we minimize the optimization problem (10) directly through the iterative algorithms. We applied the steepest descent method with using the \( l_1 - \text{norm} \) to find the descent step direction \( \Delta s_{sd} \) and the exact line search for calculating the step length \( t \) in each iteration (see [20] for more details).

All things considered, our proposed algorithm for reconstructing \( s \) from the optimization formulation (10) via steepest descent method is written in Algorithm 1. We named this algorithm Non Parametric Sparse Representation or NPSR.

In this algorithm, first of all, since we know that our solution is sparse, a zero vector is considered to initialize the sparse vector \( s \). Then, in each iteration loops, one element will be added to the sparse estimated vector \( \hat{s} \). It causes the \( \| \hat{s} \|_1 \) to be increased while the reconstruction error \( \| r - \Phi s \|_\rho \) is decreasing. This loop must be repeated unless the reconstruction error becomes less than the threshold \( \xi \).
In this section, the performance of NPSR is studied and compared with the previous ones like OMP, BCS, BP, and Lasso [21]. The Matlab code for BCS is available online at http://www.ece.duke.edu/~shiji/BCS.html. Also, we use SparseLab toolbox for simulating other methods, available from http://www.sparselab.stanford.edu.

The $N \times 1$ ($N = 1000$) signal $s$ is generated with $K = 20$ nonzero elements drawn from an i.i.d Gaussian pdf with zero mean and unit variance. The positions of nonzero coefficients are chosen uniformly at random. The elements of the dictionary $\Phi$ are i.i.d realization of a Gaussian pdf with zero mean and variance $\frac{1}{M}$ ($M = 300$). In each trial, a noise vector is generated and added to $\Phi s$. The total number of trials is set to 1000 and the estimations of the signal through different methods for each trial are found. The reconstruction error is calculated by $E \frac{\|s - \hat{s}\|_2}{\|s\|_2}$, where $\hat{s}$ denotes to one of the reconstruction methods like the OMP, BCS, BP, Lasso, and NPSR.

The program is run for 10 different values of noise variance and the results for three kinds of noise are shown in Figs. 1, 2, and 3. Fig. 1 and Fig. 2 confirm that our proposed method for sparse representation has a better performance compared with the previous schemes. Furthermore, Fig. 3 shows that the proposed approach has a lower reconstruction error in the presence of the Gaussian noise for low SNR

where $SNR = \frac{\|s\|_2^2}{\|\hat{s} - s\|_2^2}$

V. CONCLUSION

In this paper, we proposed an optimization approach for finding the sparse solution of linear regression model $r = \Phi s + n$ in the presence of non-Gaussian and impulsive noise. We have called this method as Non Parametric SR because we do not use any a priori information about noise parameters and model to calculate the sparse solution analytically.

Furthermore, we found that our proposed algorithm outperforms the existing greedy methods (like MP, OMP, and StOMP) and Bayesian approaches.

APPENDIX

Here, we reviewed the concept of subgradient for easy reference (more details are available in [22] and [23]). Roughly speaking, the subgradient is a generalization of gradient for a convex non differentiable function.

Definition: let $f : \chi \rightarrow \mathbb{R}$ be a convex function where $\chi \subseteq \mathbb{R}^N$ is a convex set. The vector $g \in \mathbb{R}^N$ is a subgradient of non differentiable function $f$ at point $x^* \in \chi$ if $g$ satisfies the following condition

$$f(x) \geq f(x^*) + g^T (x - x^*), \forall x \in \chi.$$  

There can be more than one subgradient at $x^*$ and the set of all subgradients is named subdifferential at $x^*$ and is denoted by $\partial f(x^*)$. If $f$ is differentiable, $\partial f(x) = \{g\}$ where $g = \nabla f(x)$ and $\{g\}$ denotes to the set of subgradient vectors. The subgradient has scaling and addition properties which have given by (14) and (15), respectively

$$\partial (af(x)) = a \partial f(x)$$  \hspace{1cm} (14)

$$\partial (f_1(x) + f_2(x)) = \partial f_1(x) + \partial f_2(x).$$  \hspace{1cm} (15)

As is well known, if $x^*$ is the minimize of convex differentiable function $f$, $\nabla f(x^*) = 0$ must be satisfied at the same way. For a convex non differentiable $f$, the optimality condition generalizes as following

$$f(x^*) = \min_{x \in \mathbb{R}^N} f(x) \Leftrightarrow 0 \in \partial f(x^*).$$  \hspace{1cm} (16)

which follows from the definition of the subgradient

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*), x \in \mathbb{R}^N.$$  \hspace{1cm} (17)

For example, when the function of interest is $\|s\|_1$, the subgradient is given by

$$\frac{\partial f(x)}{\partial x} = u(s), u_i = \{ \text{sgn}(s_i), s_i \neq 0 \}$$

REFERENCES

[1] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. on Information Theory, vol. 52, no. 2, pp. 489–509, 2006.

[2] D. Donoho, “Compressed sensing,” IEEE Trans. on Information Theory, vol. 52, no. 4, pp. 5406–5425, 2006.

[3] R. Gribonval and S. Lesage, “A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges,” ESSAN’06, pp. 323–330, 2006.
Fig. 1. NPSR versus $l_2$-norm based approaches for sparse representation in the presence of double-exponential noise.

Fig. 2. NPSR versus $l_2$-norm based approaches for sparse representation in the presence of Middelton’s class A model for non-Gaussian noise with $\kappa = 1000$ and $\varepsilon = 0.01$.

Fig. 3. NPSR versus $l_2$-norm based approaches for sparse representation in the presence of Gaussian noise.

[4] S. Mallat, and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” IEEE Trans. Signal Processing, vol. 41, no.12, pp. 3397–3415, 1993.

[5] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in Proc. 27th Ann. silmar Conf. Signals, Systems, and Computers, 1993.

[6] D. L. Donoho, Y. Tsaig, I. Drori, and J.-C. Starck, “Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit,” Tech. Rep. 2006-02, Dept. of Statistics, Stanford University, Stanford, CA, 2006. (software available at http://sparselab.stanford.edu/).

[7] B. K. Natarajan, “Sparse approximate solutions to linear systems,” SIAM Journal on Computing, vol. 24, pp. 227–234, 1995.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” SIAM Journal on Scientific Computing, vol. 20, no. 1, pp. 33–61, 1998.

[9] SM. E. Tipping, “Sparse bayesian learning and the relevance vector machine,” Journal of Machine Learning Research, vol. 1, pp. 211–244, 2001.

[10] S. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” IEEE Trans. Signal. Proc, vol. 56, pp. 2346–2356, June 2008.