ATLANTIS PRESS
Advances in Computer Science Research, volume 92
International Conference on Mathematics, Big Data Analysis and Simulation and Modeling (MBDASM 2019)

Sparse Signal Recovery Based On A Mixture Distribution

Hongjie Wan[1,*], Haiyun Zhang[2]
[1]Information Engineering Dept, Beijing University of Chemical Technology, Beijing, China
[2]College of Information Science and Engineering, Fujian University of Technology, Fuzhou, China
[*]Corresponding author

***Abstract*—Based on the sparse characteristic of most signals under some transformation, compressive sampling has been brought out to replace the traditional Nyquist theory based sampling. This paper presents a Bayesian method to recovery the original signal from compressed measurements. A hierarchical Bayesian model is built to model the relation between the measurement and the underlying sparse coefficients. To model the sparse property of the signal, a mixture distribution is placed over the coefficient, which enforces the sparsity of the coefficient. The Variational Bayesian theory is applied to the model, and the estimation is obtained. To demonstrate the performance of the algorithm, experiments are carried out on both synthetic sparse signal and image signal.**

Keywords—sparse signal; variational bayesian; compressive sensing

I. INTRODUCTION

The Nyquist sampling theory has been existed for a long history for data acquisition, until recent decade the compressive sampling has been put forward [1]. In signal processing, most of the signals are sparse under a certain base transformation, which can be expressed by matrix form as $\mathbf{y} = \mathbf{\Phi\theta}$. In this equation, $\mathbf{y}$ represents the original signal, $\mathbf{\Phi}$ is the transformation matrix and $\mathbf{\theta}$ is the transformed coefficient. The transformation matrix is the key to analyze the characteristic of the signal. The typical transformations include discreet cosine transform (DCT), wavelet transform [2], Fourier transform and other dictionary based methods that can transform the original signal into a sparse domain [3]. In compressive sensing, the signal is measured by a measurement matrix $\boldsymbol{\psi}$, and the measurement may be corrupted by noise or there will exist measurement noise. Taking the above into consideration, the relation between the measurement and the spare coefficient can be expressed by the following typical form

$$\mathbf{x} = \mathbf{L\theta} + \boldsymbol{\varepsilon} \tag{1}$$

Where $\mathbf{x}$ denotes the measurement vector, $\mathbf{L}$ is the product of $\mathbf{\Phi}$ and $\boldsymbol{\psi}$, $\boldsymbol{\varepsilon}$ represents the measurement noise. The size of the measurement is always smaller or much smaller than that of the original signal, thus the compression aim of signal processing can be achieved. It is also obvious that the signal can be recovered after estimating the sparse coefficient $\mathbf{\theta}$.

There have been many algorithms proposed to recover the sparse coefficient, and the algorithms are classified according to their properties by some authors. The most typical and the earliest class of the algorithms are greedy searching algorithms which iteratively search the item which will result the best recovery result, including the OMP method, the CoSamp [4] and ROMP [5] methods. Another class of the methods is l_1 based methods, which aims to pursuit a minimization of l_1 norm on the coefficient. In this class, the lasso method [6] for regression is very famous, and this fact can be easily believed from the form of Eq. (1) as this is the same form for linear regression.

In this paper, a Bayesian model is built for the compressive measurement processing as in Eq. (1). From Eq. (1), it can be seen that the unknown parameters are sparse coefficient $\mathbf{\theta}$ and the parameters of the measurement noise $\boldsymbol{\varepsilon}$. A mixture distribution is placed over each element of the coefficient $\mathbf{\theta}$ as a prior, this type of prior can enforce sparsity and will cut off the zero elements and keep the non zeros. The principle of this prior will be discussed in the following contents. The noise is assumed to be Gaussian distributed with a zero mean and a variance parameter. Variational Bayesian method is applied to the model and the parameters are estimated iteratively. Experiments are carried out on synthetic sparse signal and image signal and demonstrate the performance of the algorithm.

The structure of the rest of paper is organized as following. The 2nd section gives the model building for compressive sensing measurement principle. In section III, the parameter estimation based on Variational Bayesian is described. Experiments on synthetic sparse signal and image signal are given in section IV. Finally, section V concludes the paper.

II. MODELING OF COMPRESSIVE SENSING

In Eq. (1), the unknown parameters are $\mathbf{\theta}$ and the parameters of $\boldsymbol{\varepsilon}$. In Eq. (1), the size of $\mathbf{x}$ is $N\times1$, the size of L is $N\times M$, the size of $\mathbf{\theta}$ is $M\times1$, the size of $\boldsymbol{\varepsilon}$ is $N\times1$. Firstly, a Gaussian distribution with zero mean and variance δ_i is set on each item of $\mathbf{\theta}$ as its prior

$$p(\theta_i|\delta_i) = \mathrm{N}(\theta_i \mid 0, \delta_i), i = 1, 2, \ldots M \tag{2}$$

It can be deduced from the above equation that when the variance approaches zero, then the distribution will concentrate at zero and thus cut off the zero element. Secondly, an exponential distribution with rate parameter $\lambda_i/2$ is set on each of the variance parameter as its prior

$$p(\delta_i \mid \lambda_i) = \lambda_i/2 \times \exp\left(-\lambda_i\delta_i/2\right), i = 1,2,...M \tag{3}$$

Finally, in order to have hierarchical prior, a gamma distribution is also set on each of the parameter λ_i

$$p(\lambda_i) = \mathcal{G}(\lambda_i \mid a_\lambda, b_\lambda), i = 1,2,...M \tag{4}$$

The shape and rate parameters of all the gamma distributions are the same as the same prior is set on all the parameters λ_i. In addition, to obtain uninformative prior, a_λ and b_λ are set to be a very tiny values, e.g. 1e-6.

After the above setting, the joint probability of the three parameters θ_i, δ_i and λ_i can be derived as following

$$p(\theta_i, \delta_i, \lambda_i) = p(\theta_i \mid \delta_i) p(\delta_i \mid \lambda_i) \mathrm{p}(\lambda_i) \tag{5}$$

Integrating out λ_i, the joint probability of θ_i, δ_i can be achieved

$$p(\theta_i, \delta_i) \propto \delta_i^{-1/2} \exp\left\{-\theta_i^2/(2\delta_i)\right\}\left(b_\lambda + \delta_i/2\right)^{-a_\lambda - 1} \tag{6}$$

It should be noted that very tiny values have been set on a_λ and b_λ, so assuming the extreme case that zeros are set on b_λ, and while θ_i tends to zero, the following relation can be obtained

$$p(\theta_i) \propto \int \delta_i^{-1/2}\left(b_\lambda + \delta_i/2\right)^{-a_\lambda - 1} d\delta_i \tag{7}$$

From this equation, it can be sure that while θ_i tends to zero, the distribution will also tends to infinity, thus the concentration property of the each element of the coefficient can be guaranteed. This is also how the title "mixture distribution " has come, which come from a mixture of Gaussian distribution, an exponential distribution and a gamma distribution.

Now it turns to the noise parameters. A Gaussian distribution with zero mean and an inverse variance parameter is set on each item of the noise vector $\boldsymbol{\varepsilon}$ following

$$p(\varepsilon_n) = \mathcal{N}(0, \gamma^{-1}), n = 1,2,...N \tag{8}$$

It should be noted that the inverse variance parameter is the same for each item of the noise vector. In the same way as Eq. (4), a gamma distribution with shape a_γ and rate b_γ is set on γ

$$p(\gamma) = \mathcal{G}(\gamma \mid a_\gamma, b_\gamma) \tag{9}$$

Similarly, very tiny values are also set on a_γ and b_γ.

After the above hierarchical priors setting on $\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$, the Bayesian model is built for compressive sensing. All the parameters involved in the model building can be written into a complete parameter set $\Theta = \{\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\lambda}, \gamma\}$, where each vector contains all the items of the same type.

III. Parameter Estimation Based on Variational Bayesian

First of all, the full joint probability of the model can be calculated as following

$$\mathrm{p}(\mathbf{x}, \Theta) = \mathrm{p}(\mathbf{x} \mid \boldsymbol{\theta}, \boldsymbol{\delta}, \gamma)\, \mathrm{p}(\boldsymbol{\theta} | \boldsymbol{\delta})\, \mathrm{p}(\boldsymbol{\delta} \mid \boldsymbol{\lambda})\, \mathrm{p}(\boldsymbol{\lambda})\, \mathrm{p}(\gamma) \tag{10}$$

In Variational Bayesian method, an auxiliary distribution $q(\Theta)$ is used to approach the full joint the distribution. The auxiliary distribution $q(\Theta)$ can be factorized as product of auxiliary distributions on each parameter vector in Θ

$$q(\Theta) = q(\boldsymbol{\theta}) q(\boldsymbol{\delta}) q(\boldsymbol{\lambda}) q(\gamma) \tag{11}$$

Each auxiliary distribution can be the estimation of the parameters. The estimation of the parameters of the auxiliary distributions is the aim of the Variational Bayesian method. This is achieved by minimizing the Kullback–Leibler (KL) distance between the full probability in Eq. (10) and the auxiliary distribution in Eq. (11). After minimization, the distribution form of each distribution in Eq. (11) can be the following general form, the first item on $\boldsymbol{\theta}$ is taken as an example [7]

$$q(\boldsymbol{\theta}) \propto \exp\left\langle \log\left[p(\mathbf{x} \mid \boldsymbol{\theta}, \gamma) p(\boldsymbol{\theta} \mid \boldsymbol{\delta})\right]\right\rangle_{q(\gamma)q(\boldsymbol{\delta})} \tag{12}$$

Where the sign $\langle \bullet \rangle_q$ denotes the expectation computation with respect to the distribution q, the subscript with respect to will be omitted in the following contents to make the description concise. After careful calculation, the form of $\boldsymbol{\theta}$ is obtained as a multivariate Gaussian distribution with variance $\boldsymbol{\Sigma}$ and mean $\boldsymbol{\mu}$ as following

$$\boldsymbol{\Sigma} = \left(\langle \gamma \rangle \mathbf{L}^T\mathbf{L} + diag\left\langle \boldsymbol{\delta}^{-1} \right\rangle\right)^{-1} \tag{13}$$

$$\boldsymbol{\mu} = \langle \gamma \rangle \boldsymbol{\Sigma} \mathbf{L}^T \mathbf{x} \tag{14}$$

Where the operator $diag(\bullet)$ has the ability to expand the column vector into a matrix, and the diagonal of the matrix is the column vector.

Then it turns to the $\boldsymbol{\delta}$ parameters, which is not the same as in Eq. (12) as the form is hard to lead to a tractable form.

$$q(\boldsymbol{\delta}) \propto \exp\left[-1/2\sum_{i=1}^{M}\left(\log\delta_i + \langle\theta_i^2\rangle/\delta_i + \langle\lambda\rangle\delta_i\right)\right] \tag{15}$$

In order to get the expectations required for computation, the MCMC method is used, and the Metropolis–Hastings (MH) sampler [8] is adopted to get samples and the normal distribution is adopted. The mean of $\langle\boldsymbol{\delta}^{-1}\rangle$ and $\langle\boldsymbol{\delta}\rangle$ are computed after sampling. It must be noticed that the two quantities do not have inverse relationships.

Now the elements of the parameter $\boldsymbol{\lambda}$, using the similar process as in Eq. (12), each element has the form of gamma distribution with shape and rate as

$$\alpha_\lambda = a_\lambda + 1 \tag{16}$$

$$\beta_{\lambda_i} = b_{\lambda_i} + 1/2 \times \langle\delta_i\rangle \tag{17}$$

It can be seen that all the shape parameters are the same.

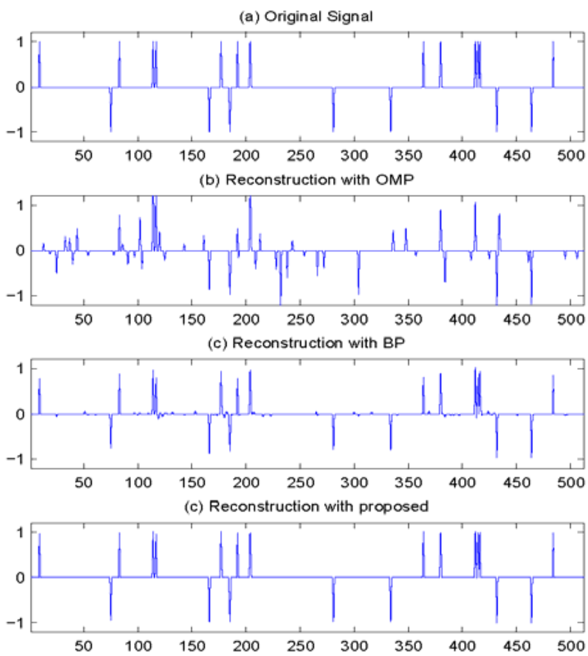

FIGURE I. THE ORIGINAL SIGNAL, RECOVERY OF OMP, BP AND THE PROPOSED METHOD.

Finally, using the similar manner, the auxiliary distribution of the noise precision can be obtained as a gamma distribution with shape and rate parameters as follows

$$\alpha_\gamma = a_\gamma + N/2 \tag{18}$$

$$\beta_\gamma = b_\gamma + 1/2 \times \left\langle(\mathbf{x}-\mathbf{L}\boldsymbol{\theta})^T(\mathbf{x}-\mathbf{L}\boldsymbol{\theta})\right\rangle \tag{19}$$

After estimating all the parameters of the auxiliary distributions, the parameters of the Bayesian model can be achieved and the coefficient of the signal can be estimated.

IV. EXPERIMENTS

In the experiments, the OMP and BP [9] methods are compared with the proposed algorithm. In the first experiment, synthetic sparse signals are used. The size of the sparse coefficient is set to be 512, and only a few items are randomly selected as non zero values. The non zero values are randomly set to be +1 or -1. The measurement matrix is built by randomly selected values from a standard normal distribution, and the norm of each column is regularized to be 1. In figure 1, results of the recovery for 100 measurement and 20 nonzero items. From the results, it can be seen that the proposed algorithm got the best result.

TABLE I. THE RECOVERY ON DIFFERENT NONZERO NUMBER

Non zeros	*OMP*	*BP*	*Proposed*
10	0.14	0.15	0.06
15	0.14	0.20	0.07
20	4.41	0.74	0.11

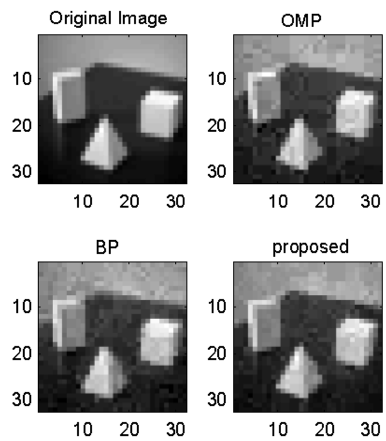

FIGURE II. THE ORIGINAL IMAGE, RECOVERY OF OMP, BP AND PROPOSED METHOD

Then the non zero number is varied to demonstrate the performance of the algorithm. The numbers of 10, 15, and 20 are tested and the norms of the reconstruction error are displayed in Table I. The results show that the errors of the proposed algorithm are always smaller than that of OMP and BP. It can also be seen that the error decreases as the nonzero number decreases.

In the second experiment, an image signal is used to be the original signal, and the size of the image is 32×32. The image is decomposed by wavelet method, and the wavelet type is chosen as db1 with the decomposition level 3. The measurement number is set to be 600. The reconstruction error of the proposed method is 253, while that of OMP and BP are all 316.

V. CONCLUSION

In this paper, a Bayesian model is built to model the process of compressive sensing. A mixture model is placed on each item of the sparse coefficient. Variational Bayesian method is applied to the model and the parameter estimation is given out. Experiments on synthetic sparse signal and image signal are carried out and the performance of the algorithm has been demonstrated.

REFERENCES

[1] D. L. Donoho, “Compressed sensing, ” IEEE Transactions on Information Theory vol.52(4), pp:1289-1306, 2006.

[2] S. Mallat, A Wavelet Tour of Signal Processing, 2nd ed. Academic: New York, 1998.

[3] E.Sakhaee, A. Entezari, “Joint Inverse Problems for Signal Reconstruction via Dictionary Splitting, ” IEEE Signal Processing Letters, vol. 24(8), pp.1203-1207, 2017.

[4] D. Needle, J.A. Tropp, “CoSaMP: iterative signal recovery from incomplete and inaccurate samples. ”Appl. Comput. Harmon. Anal. Vol. 26(3), pp. 301–321, 2009.

[5] D. Needell, R. Vershynin, “Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. ” IEEE J. Sel. Top. Signal Process. , vol. 4(2), pp. 310–316, 2010.

[6] R. Tibshirani, “Regression shrinkage and selection via the lasso. ” J. R. Stat. Soc. B, vol. 58, pp.267–288, 1996.

[7] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[8] Gilks, W. R., Richardson, S., & Spiegelhalter, D. J., Markov Chain Monte Carlo in Practice, London: Chapman and Hall, 1996.

[9] S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” SIAM J. Sci. Comput., vol. 20(1), pp. 33–61, 1999.