Laplace Prior Based Distributed Compressive Sensing

Liang Tang, Zheng Zhou, Lei Shi, Haipeng Yao, Jing Zhang, Yabin Ye☆

Key Lab of Universal Wireless Communications, MOE

Wireless Network Lab, Beijing University of Posts and Telecommunications

Beijing, China

E-mail: tangliangbupt@gmail.com, zzhou@bupt.edu.cn

☆ European Research Center，Huawei Technologies Duesseldorf GmbH

Abstract—**Bayesian compressive sensing (BCS) utilizes the prior distribution of signal coefficients to reconstruct the original signal. The widely used prior is Laplace and Gaussian distributed. In this paper, we use the scene of L sets of signal sparse coefficients which are statistically related and take advantage of Laplace prior and statistically interrelationship among signals to propose the Laplace prior based distributed Bayesian compressive sensing. We provide the experiment result to demonstrating that the proposed method is an effective reconstruction algorithm and has a good performance.**

Keywords: Bayesian compressive sensing (BCS); Laplace prior; statistically interrelationship; distributed Bayesian compressive sensing.

I. Introduction

Compressive sensing has received a lot of attentions in recent years due to its active theoretical characteristic and its practical use in many application areas. Let $\mathbf{x} \in \mathbb{R}^N$ be the original signal which has a sparse representation in one basis. Most of its coefficients are zero for its sparsity. Consider the measure system $\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}$, where $M \times 1$ linear measurement $\mathbf{y}$ of original signal $\mathbf{x}$ is taken with the $M \times N$ projection matrix $\mathbf{\Phi} = [\mathbf{\varphi}_1, \mathbf{\varphi}_2, \cdots, \mathbf{\varphi}_N]$ and $\mathbf{n}$ represents the measure noise. The mapping from $\mathbf{y}$ to the approximation of $\mathbf{x}$ is under-determined, since $M \leq N$. According to the theory of compressive sensing[1-2], with the certain condition and reconstruction algorithm, the original signal $\mathbf{x}$ can be reconstructed accurately even if the number of measurement is highly smaller than the number of coefficients($M \leq N$).

The common reconstruction algorithms utilize the sparsity of the signal $\mathbf{x}$ and constrain the l_p norm of $\mathbf{x}$. The formulation is [2]:

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{\Phi x}\|_p^2 + \varepsilon \|\mathbf{x}\|_p \right\}. \qquad (1)$$

With the case of $p=0$, $\|\mathbf{x}\|_0$ is the number of nonzero

coefficients of $\mathbf{x}$, and it is a NP-hard problem. Thus, the common value of p is 1. A number of inversion algorithms have been proposed to solve the problem based on (1), including linear programming [3] and greedy algorithms [4] [5]. Unlike the point estimate of the coefficient in [3-5], the CS reconstruction algorithm can also been formulated in a Bayesian framework [6] [7]. In [6], the prior of sparse coefficients and relevance vector machine (RVM) [8] are used to resolve the CS problem. The Laplace prior of sparse coefficients is utilized in [7]. The experiment result in [7] shows that the reconstruction of Bayesian compressive sensing based on Laplace prior is better than BCS proposed in [6].

In this paper, we formulate the reconstruction of L sets of statistically related signal coefficients using the Laplace prior. As will be shown, because of utilizing the statistically relationship among the signals, our formulation includes the formulation in [7] as a special case (L=1), and result in the less number of total measurement needed by the reconstruction of all signals. This statistically related framework has been proposed in [8]. But in the literature [8], the algorithm bases on the BCS. We will demonstrate with the experiment result that the proposed algorithm provides the better reconstruction performance than the method in [8] and the separate Bayesian compressive sensing based on Laplace prior in [7].

The rest of this paper is organized as follow: in section II, we analyze the distribution hierarchical Bayesian modeling based on Laplace prior and propose a reconstruction algorithm for this problem. We present the experiment result and analyze it in section III. The conclusion is drawn in section IV.

II. Distribution Bayesian CS Modeling

In Bayesian framework, all unknowns are modeled as the stochastic quantities with the some probability distribution. The signal $\mathbf{x}$ is assigned the distribution $p\{\mathbf{x} \,|\, \gamma\}$, and the measurement $\mathbf{y}$ is also a random quantity with the distribution $p\{\mathbf{y} \,|\, \mathbf{x}\}$. The parameter γ decides the distribution, and it is called hyperparameter. The reconstruction of signal $\mathbf{x}$ depends on the determination of the hyperparameter. This model is also called hierarchical Bayesian model.

Assume that the L sets of measurements are performed, and they are statistically related. The L sets of measurements are represented as $\{\mathbf{y}_i\}_{i=1,\cdots,L}$, where $\mathbf{y}_i = \mathbf{\Phi}_i \mathbf{x}_i + \mathbf{n}_i$. In general every measurement vector $\mathbf{y}_i \in \mathbb{R}^{M_i}$ employs a

different random projection matrix $\mathbf{\Phi}_i \in \mathbb{R}^{M_i \times N}$, where M_i represents the number of ith set of measurements. $\mathbf{n}_i$ is modeled as M_i i.i.d. draws of a zero-mean Gaussian random variable with a same unknown precision β (variance $1/\beta$).

For simplicity but without loss of generality, we assume that the sparse basis $\mathbf{\Psi} = \mathbf{I}_N$, and we can obtain $\mathbf{x}_i = \mathbf{\Psi}_i \mathbf{\theta}_i = \mathbf{\theta}_i$, then $\mathbf{y}_i = \mathbf{\Phi}_i \mathbf{\theta}_i + \mathbf{n}_i$. Thus the probability distribution of $\mathbf{y}_i$ is:

$$p\left(\mathbf{y}_i \mid \mathbf{\theta}_i, \beta\right) = \mathcal{N}\left(\mathbf{y}_i \mid \mathbf{\Phi}_i \mathbf{\theta}_i, \beta^{-1}\right), \tag{2}$$

where a Gamma prior can be given to the precise β. However, the estimation of β is not performed well with the greedy algorithm and it would make the reconstruction performance bad. So we put β a fix value $\beta = 0.01 * \|\mathbf{y}\|_2^2$ in the beginning of algorithm like in [6].

The coefficients $\mathbf{\theta}_i$ is assigned the Laplace prior with the parameter λ [7]. This Laplace distribution is shared by all signals, and the L signals are statistically related in this sense. That is:

$$p\left(\mathbf{\theta}_i \mid \lambda\right) = \left(\lambda/2\right)^N \exp\left(-\lambda|\mathbf{\theta}_i|\right) \tag{3}$$

However, this model of Laplace prior cannot be used directly, since it is not conjugate to the Gaussian distribution in (2). To solve this problem, the hierarchical prior is employed. Letting the $\theta_{i,j}$ represent the jth sparse coefficient of the ith signal, we first assign the zero-mean Gaussian random distribution with the non-zero precise $\mathbf{r} = \left(r_1, \cdots, r_N\right)$ to $\mathbf{\theta}_i$:

$$p\left(\mathbf{\theta}_i \mid \mathbf{r}\right) = \prod_{j=1}^{N} \mathcal{N}\left(\theta_{i,j} \mid 0, r_j\right) \tag{4}$$

After that, secondly we use the following hyperprior on r_j:

$$p\left(r_j \mid \lambda\right) = \Gamma\left(r_j \mid 1, \lambda/2\right) = \left(\lambda/2\right)\exp\left(-\frac{\lambda r_j}{2}\right), r_j \geq 0, \lambda \geq 0 \tag{5}$$

We have known that the L signals share the Laplace distribution in the (3) and the parameter λ. Consequently, in the hierarchical model we assume that the Gaussian distribution on $\mathbf{\theta}_i$ in (4) is shared by the L signals. It is important to note that the hyperparameters $\mathbf{r} = \left(r_1, \cdots, r_N\right)$ and λ are shared among the all signals. We use the data from all L sets of measurements $\left\{\mathbf{y}_i\right\}_{i=1,\cdots,L}$ to learn the common hyperparameters $\mathbf{r}$ and λ. This makes it possible to reduce the number of total measurements.

In the final, we have $p\left(\lambda\right) \propto 1/\lambda$ and based on the above, we can get

$$p\left(\mathbf{\theta}_i \mid \lambda\right) = \int_0^{+\infty} p\left(\mathbf{\theta}_i \mid \mathbf{r}\right) p\left(\mathbf{r} \mid \lambda\right) d\mathbf{r} = \frac{\lambda^{N/2}}{2^N}\exp\left(-\sqrt{\lambda}\sum_j |\theta_{i,j}|\right) \tag{6}$$

This hierarchical model includes three stages. The hyperprior in (4) and (5) result in the Laplace distribution $p\left(\mathbf{\theta}_i \mid \lambda\right)$ in (6).

We have completed the formation of the signal model. Having defined the prior, we carry out the Bayesian inference by computing the posterior $p\left(\mathbf{\theta}_i, \mathbf{r}, \lambda, \beta \mid \mathbf{y}_i\right)$. The type-II maximum likelihood approach (or evidence maximization) is used to perform the Bayesian inference in this paper. In order to complete our inference procedure, we decompose the posterior $p\left(\mathbf{\theta}_i, \mathbf{r}, \lambda, \beta \mid \mathbf{y}_i\right)$ as

$$p\left(\mathbf{\theta}_i, \mathbf{r}, \lambda, \beta \mid \mathbf{y}_i\right) = p\left(\mathbf{\theta}_i \mid \mathbf{r}, \lambda, \beta, \mathbf{y}_i\right) p\left(\mathbf{r}, \beta, \lambda \mid \mathbf{y}_i\right) \tag{7}$$

The $p\left(\mathbf{\theta}_i \mid \mathbf{r}, \lambda, \beta, \mathbf{y}_i\right)$ is a multivariate Gaussian distribution $\mathcal{N}\left(\mathbf{\theta}_i \mid \mu_i, \Sigma_i\right)$ with the parameters like:

$$\mathbf{\mu}_i = \beta \Sigma_i \mathbf{\Phi}_i^T \mathbf{y}_i \tag{8}$$

$$\Sigma_i = \left(\Lambda + \beta \mathbf{\Phi}_i^T \mathbf{\Phi}_i\right)^{-1}, \tag{9}$$

where $\Lambda = diag\left(1/r_1, \cdots, 1/r_N\right)$.

Now we utilize the posterior $p\left(\mathbf{r}, \beta, \lambda \mid \mathbf{y}_i\right)$ to compute the hyperparameters. Exploiting

$$p\left(\mathbf{r}, \lambda, \beta \mid \mathbf{y}_i\right) \propto p\left(\mathbf{y}_i \mid \mathbf{r}, \lambda, \beta\right) p\left(\mathbf{r}\right) p\left(\lambda\right),$$

we estimate the hyperparameters by compute the maximum of the joint distribution $p\left(\mathbf{y}_i \mid \mathbf{r}, \lambda, \beta\right) p\left(\mathbf{r}\right) p\left(\lambda\right)$, or equivalently its logarithm

$$\mathcal{L}\left(\mathbf{r}, \lambda\right) = \sum_{i=1}^{L} \log \int p\left(\mathbf{y}_i \mid \mathbf{\theta}_i, \beta\right) p\left(\mathbf{\theta}_i \mid \mathbf{r}\right) p\left(\mathbf{r} \mid \lambda\right) p\left(\lambda\right) d\mathbf{\theta}_i$$

$$= \sum_{i=1}^{L}\left(-\frac{1}{2}\log|\mathbf{C}_i| - \frac{1}{2}\mathbf{y}_i^t \mathbf{C}_i^{-1}\mathbf{y}_i + N\log\frac{\lambda}{2} - \frac{\lambda}{2}\sum_j r_j - \log\lambda\right) \tag{10}$$

with $\mathbf{C}_i = \beta^{-1}\mathbf{I} + \mathbf{\Phi}_i \Lambda^{-1} \mathbf{\Phi}_i^T$. There are two approaches to maximize $\mathcal{L}\left(\mathbf{r}, \lambda\right)$ with respect to $\mathbf{r}$ and λ.

1) Iterative Solution: We respectively differentiate (10) with respect to $\mathbf{r}$ and λ, set the results to zero and have

$$r_j^{new} = \frac{-L + \sqrt{L^2 + 4\lambda L \sum_{i=1}^{L}\left(\mu_{i,j}^2 + \Sigma_{i,jj}\right)}}{2\lambda L} \quad , \quad (11)$$

$$\lambda^{new} = \frac{2(N-1)}{\sum_j r_j} \quad , \quad (12)$$

where $\Sigma_{i,jj}$ is the jth diagonal component of the Σ of the ith signal and $\mu_{i,j}$ is the jth component of μ of the ith signal. We note $\mathbf{r}^{new}$ is the function of the $\{\mu_i\}_{i=1\cdots N}$ and $\{\Sigma_i\}_{i=1\cdots N}$. At the same time, $\{\mu_i\}_{i=1\cdots N}$ and $\{\Sigma_i\}_{i=1\cdots N}$ are the functions of $\mathbf{r}^{new}$. An iterative algorithm is performed among (8), (9) and (11). λ^{new} is get from $\mathbf{r}^{new}$ in (12). The all hyperparameters are estimated through the above iterative procedure.

However, the iterative solution has a limitation that it needs a $o(N^3)$ operation. It makes that the approach is very slow when it is applied to a large-scale problem. So a fast algorithm, similar to [9], is utilized in the following.

2) Fast Algorithm: Considering that $\mathcal{L}(\mathbf{r}, \lambda)$ dependents on the hyperparameter r_j, $j \in \{1, \cdots, N\}$, we can get

$$\mathbf{C}_i = \beta^{-1}\mathbf{I} + \sum_{k \neq j} r_k \Phi_{i,k}\Phi_{i,k}^T + r_j \Phi_{i,j}\Phi_{i,j}^T = \mathbf{C}_{i,-j} + r_j \Phi_{i,j}\Phi_{i,j}^T$$
$$\quad , (13)$$

where $\Phi_{i,j}$ is the jth column of Φ_i, and $\mathbf{C}_{i,-j}$ is the contribution to $\mathbf{C}_i$ without $\Phi_{i,j}$. Applying the above determination, the correspond parts in (12) can be written as

$$\mathbf{C}_i^{-1} = \mathbf{C}_{i,-j}^{-1} - \frac{\mathbf{C}_{i,-j}^{-1}\Phi_{i,j}\Phi_{i,j}^T\mathbf{C}_{i,-j}^{-1}}{1/r_j + \Phi_{i,j}^T\mathbf{C}_{i,-j}^{-1}\Phi_{i,j}}, \quad (14)$$

$$\left|\mathbf{C}_i\right| = \left|\mathbf{C}_{i,-j}\right|\left|1 + r_j \Phi_{i,j}^T\mathbf{C}_{i,-j}^{-1}\Phi_{i,j}\right|. \quad (15)$$

Using the equation (14) and (15) and regarding $\mathcal{L}(\mathbf{r}, \lambda)$ as the function of r_j, we can get

$$\mathcal{L}(\mathbf{r}) = -\frac{1}{2}\sum_{i=1}^{L}\left(\log\left|\mathbf{C}_{i,-j}\right| + \mathbf{y}_i^t\mathbf{C}_{i,-j}^{-1}\mathbf{y}_i + \frac{\lambda}{2}\sum_{k \neq j}r_k\right)$$
$$+ \frac{1}{2}\sum_{i=1}^{L}\left(\log\left(\frac{1}{1+r_js_{i,j}}\right) + \frac{q_{i,j}^2 r_j}{1+r_js_{i,j}} - \lambda r_j\right)$$
$$= \mathcal{L}(\mathbf{r}_{-j}) + l(r_j) \quad (16)$$

where $l(r_j)$ is $\frac{1}{2}\sum_{i=1}^{L}\left(\log\left(\frac{1}{1+r_js_{i,j}}\right) + \frac{q_{i,j}^2 r_j}{1+r_js_{i,j}} - \lambda r_j\right)$ and $\mathbf{r}_{-j}$ is $\mathbf{r}$ without the jth component. $s_{i,j}$ and $q_{i,j}$ are defined as

$$s_{i,j} = \Phi_{i,j}^T\mathbf{C}_{i,-j}^{-1}\Phi_{i,j}, \quad q_{i,j} = \Phi_{i,j}^T\mathbf{C}_{i,-j}^{-1}\mathbf{y}_i. \quad (17)$$

The maximum of $\mathcal{L}(\mathbf{r})$ with respect to r_j is the same as the maximum of $l(r_j)$. Differentiating $l(r_j)$ with respect to r_j, we obtain

$$\frac{dl(r_j)}{dr_j} = \frac{1}{2}\sum_{i=1}^{L}\left[-\frac{s_{i,j}}{1+r_js_{i,j}} + \frac{q_{i,j}^2}{\left(1+r_js_{i,j}\right)^2} - \lambda\right]. \quad (18)$$

In order to simplify the zero-finding procedure in (18), an approximation that $r_js_{i,j} \gg 1$ is used in here. This approximation has been found to be valid numerically, e.g., typically $s_{i,j} > 20*1/r_j$ [9]. Therefore, setting the equation (18) to be zero and getting the approximate result as

$$r_j \approx \frac{-\left[L+2\lambda\sum_{i=1}^{L}(1/s_{i,j})\right] + \sqrt{\left[L+2\lambda\sum_{i=1}^{L}(1/s_{i,j})\right]^2 - 4L\lambda\sum_{i=1}^{L}\left[\left(\lambda+s_{i,j}-q_{i,j}^2\right)/s_{i,j}^2\right]}}{2L\lambda}$$

$$\text{if } \sum_{i=1}^{L}\left[\left(\lambda+s_{i,j}-q_{i,j}^2\right)/s_{i,j}^2\right] < 0 \quad (19)$$

$$r_j = 0 \quad \text{otherwise.} \quad (20)$$

The data from all the signals are used to update r_j, and other processes of this fast algorithm, which is respectively performed on each signal, are same as those in literature [7].

III. Experiment Result

In this section, we present the experiment results with $L = 2$ signals of length $N = 512$. Each signal has 20 spikes in 20 random locations. The spike denotes the coefficient which is non-zero and realized of the unit variance Laplace distribution. The non-zero coefficients of the two signals have the same locations and random amplitudes consistent with identical Laplace distribution. The measurement matrix Φ_i is constructed by first obtaining a $M_i \times N$ matrix with i.i.d. draws of a zero-mean Gaussian distribution $\mathcal{N}(0,1)$, and then the rows of matrix Φ_i are normalized to unit. We add a zero-mean white Gaussian noise with standard deviation 0.005 to the measurement. In the Fig.1, the number of measurements is set $M_1 = M_2 = 50$, and in the Fig.2, the number of measurements changes from 40 to 70 in step of 1.
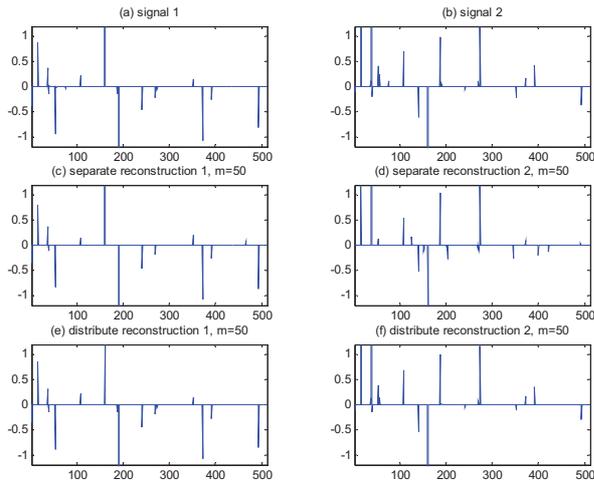
Fig.1. Reconstruction of the signals with Laplace prior. (a),(b)Signals. (c),(d)Reconstruction by separate Laplace based CS. (e),(f) Reconstruction by distributed Laplace based CS.

Fig.1 (a) and (b) is the two original signals with the Laplace prior. The reconstruction result with the BCS based on Laplace prior is presented in Fig.1(c) and (d). Because of the number of the measurement is smaller than quantity needed for precise reconstruction, the reconstructed signals are heavily noisy. Since the two signals are not statistically independent, the distributed Laplace based CS can utilize this interrelationship and obtain the almost perfect reconstructions [see Fig.1(e)(f)].

To compare the reconstruction performance of distributed Laplace based CS with MT-BCS[8] and Laplace based CS[7], in the second experiment we use the same dataset as Fig.1 and represent the probability of successful reconstruction as a function of the number of measurements (average 1000 trails). It is clear that our proposed method outperforms other methods. The separate Laplace based CS do not utilize the statistically relationship between the signals, so it gets the most poor result. MT-BCS and our method both take advantage of the statistically relationship of the original signals. However, our proposed method utilizes the Laplace prior of the signals and MT-BCS bases on the Gaussian prior. Thus our method represents the best reconstruction performance.

IV. Conclusions

In this paper, we formulate the reconstruction of L sets of statistically related signal coefficients using the Laplace prior. Because of utilizing the statistically relationship among the signals, our formulation includes the formulation in separate Laplace prior based CS as a special case (L=1), and result in the less number of total measurements needed by the precise reconstruction of all signals. Our proposed method takes advantage of the Laplace prior, and performs better than another distributed Bayesian compressive sensing MT-BCS based on the Gaussian prior. We demonstrate that the proposed method outperforms the method in [7] [8].
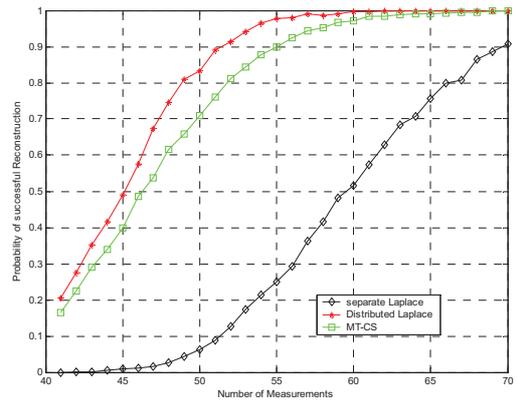

Fig.2. Reconstruction with different methods as a function of M.

V. References

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

[2] D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol. 52, no. 4, pp: 1289–1306, Apr. 2006.

[3] 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, 1999.

[4] D. L. Donoho and Y. Tsaig, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," Preprint, March 2006.

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

[6] S. Ji, Y. Xue and L. Carin, "Bayesian compressive sensing," IEEE Trans. Signal Processing, vol. 56, no. 6, pp: 2346–2356, June 2008.

[7] S. D. Babacan, R. Molina and A. K. Katsaggelos, "Bayesian Compressive Sensing Using Laplace Priors", Image Processing. IEEE Transactions on, Volume 19, Issue 1, pp: 53 - 63, Jan. 2010

[8] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," Journal of Machine Learning Research, pp: 211–244, 2001.

[9] S. Ji, D. Dunson and L. Carin, "Multitask Compressive Sensing," Signal Processing. IEEE Transactions on，Volume 57, Issue 1, pp: 92 – 106, Jan. 2009.

[10] M. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, C. M. Bishop and B. J. Frey, Eds., 2003.