Asymptotic Analysis of Marginal Posterior PDF via Belief Propagation in Noisy Sparse Estimation

Jaewook Kang, Heung-No Lee and Kiseon Kim
Gwangju Institute of Science and Technology, Republic of Korea
Email:{jwkkang,heungno,kskim}@gist.ac.kr

1. Introduction

Recently, belief propagation (BP)-based sparse recovery algorithms have received significant attention due to the availavility of low complexity [1]. The aim of such algorithms is to estimate a sparse vector $\mathbf{X} \in \mathbb{R}^N$ from noisy measurements $\mathbf{Y} \in \mathbb{R}^M$ which is a linear projection of an underdetermined system $(M < N)$, represented as

$$\mathbf{Y} = \mathbf{\Phi X} + \mathbf{W} \quad (1)$$

where the measurements are corrupted by an additive Gaussian noise vector $\mathbf{W} \in \mathbb{R}^M$.

In particular, we are interested in the BP-based signal recovery in conjuction with sparse measurement matrices $\mathbf{\Phi} \in \{0,-1,1\}^{M\times N}$ which is tree-like structured in factor graph representation. Such BP algorithms iteratively approximate the marginal posterior corresponding to each element X_i by exchanging probability messages over the factor graph associated with $\mathbf{\Phi}$. The iterative behavior of BP have been well investigated in applications to channel coding [3], but still remains elusive in the sparse estimation problem.

In this paper, we asymptotically analyze such BP behavior in the sparse estimation problem using the *decoupling principle* [3-5]. According to this principle, the vector measurement channel using BP, shown in Fig.1-(a), can be aymptotically decoupled to a sequence of scalar Gaussian channels, shown in Fig.1-(b), if the system is a *large-sparse-system* (LSS) satisfying two conditions below:

1. *Large-system-limit*: The system size is very large, *i.e.*, $(N,M \to \infty)$ where M/N remains a fixed constant.
2. *No-short-cycle*: The bipartite graph of the sparse measurement matrix $\mathbf{\Phi}$ does not include cycles shorter than the number of the BP-iterations denoted by l.

In this work, we consider spike-and-slab PDF as the prior for the sparse signal estimation. Our result shows that the marginal posterior PDF takes the form of a spike-and-slab PDF, being a function of the signal magnitude and noise level.

2 Main Result

We aim to derive an expression of the marginal posterior PDF $f_{X_i}(x\,|\,Z_i^{(l)},\mathbf{\Phi})$ from the scalar Gaussian channel using the concept of the relaxed BP [3]. Let $V_{X_i\to Y_j}^{(l)}$ and $U_{Y_j\to X_i}^{(l)}$ be RVs representing the BP-messages passed from X_i to Y_j and from Y_j to X_i, respectively, for all pairs of $(i,j):\phi_{ji}\neq 0$ at the l-th iteration. Our derivation starts from the update rule of $U_{Y_j\to X_i}^{(l)}$, expressed as

$$\begin{aligned} U_{Y_j\to X_i}^{(l)} &:= \phi_{ji}\left(Y_j - \sum_{k\neq i}\phi_{jk}E\{X_k\,|\,V_{X_k\to Y_j}^{(l)},\mathbf{\Phi}\}\right) \\ &\overset{(a)}{=}\phi_{ji}^2 X_i + \mathrm{CEI}_{ji}^{(l)} + \phi_{ji}W_j, \\ &\overset{(b)}{=}X_i + \mathrm{CEI}_{ji}^{(l)} + W_j, \end{aligned} \quad (2)$$

where (a) holds since we know $Y_j = \sum_k \phi_{jk}X_k$ and define the cross-element-interference (CEI) term passed through Y_j to X_i as

$$\mathrm{CEI}_{ji}^{(l)} := \phi_{ji}\sum_{k\neq i}\phi_{jk}\left(X_k - E\{X_k\,|\,V_{X_k\to Y_j}^{(l)},\mathbf{\Phi}\}\right), \quad (3)$$

and (b) also holds since we know $\phi_{ji}^2 = 1$ and scaling the zero-mean Gaussian RV W_j by $\phi_{ji}\in\{1,-1\}$ does not change its statistics. By the central limit theorem (CLT), as a sum of *i.i.d.* RVs, $\mathrm{CEI}_{ji}^{(l)}$ is asymptotically Gaussian distributed with zero-mean and the variance

$$\begin{aligned} (\sigma^2_{\mathrm{CEI}_{ji}})^{(l)} &= (\sigma^2_{\mathrm{CEI}_j})^{(l)} \\ &\overset{(a)}{\approx}(R-1)\mathbf{Var}\left\{X_k\,|\,V_{X_k\to Y_j}^{(l)},\mathbf{\Phi}\right\}, \end{aligned} \quad (4)$$

under the LSS setup, where R is the number of nonzero in a row of the matrix $\mathbf{\Phi}$, and the approximation of (a) is vaild by the weak law of large numbers. Therefore, from (2)-(4), it turns out that the message $U_{Y_j\to X_i}^{(l)}$ is Gaussian distributed with the mean X_i and the variance $\eta_j^{(l)} := \sigma_W^2 + (\sigma^2_{\mathrm{CEI}_j})^{(l)}$, *i.e.*,

$$f_{U_{Y_j\to X_i}}^{(l)}(x\,|\,X_i,\mathbf{Y},\mathbf{\Phi}) \overset{\mathrm{CLT}}{\to} \mathcal{N}(x;X_i,\eta_j^{(l)}). \quad (5)$$

The marginal posterior of X_i is obtained after a certain number of the BP-iterations l^*. Let $\mathcal{U}_i^{(l)}$ denote a set of the messages from the measurement side with respect to X_i, *i.e.*, $\mathcal{U}_i^{(l)} := \{U_{Y_j\to X_i}^{(l)}\,|\,\phi_{ji}\neq 0\}$. Then, using the Bayesian rule (Posterior $\propto$ Prior $\times$ Likelihood), the

This work was supported by Haek-Sim Research Program (NO. 2012-047744) and Do-Yak Research Program (NO.2013-035295) through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology (MEST)

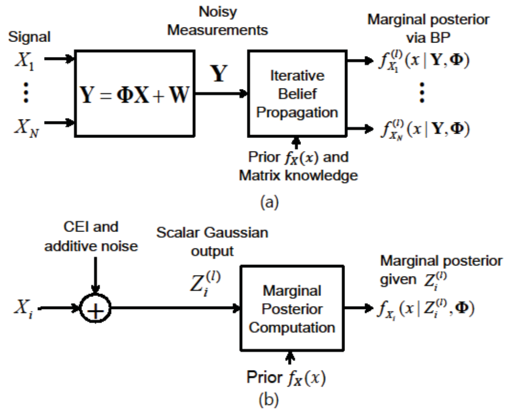

Fig. 1 (a) Vector measurement channel with BP, (b) The asymptotically equivalent scalar Gaussian channel

marginal posterior is given as

$$f_{X_i}^{(l^*)}(x|\mathbf{Y},\mathbf{\Phi}) \overset{\text{BP}}{\propto} f_X(x)\times f_{\mathcal{U}_i}^{(l^*)}(x|X_i,\mathbf{Y},\mathbf{\Phi}) \quad (6)$$
$$= f_X(x)\times \prod_{k:\phi_{ji}\neq 0} f_{U_{Y_j\to X_i}}^{(l^*)}(x|X_i,\mathbf{Y},\mathbf{\Phi}),$$

where the likelihood PDF $f_{\mathcal{U}_i}^{(l)}(x|X_i,\mathbf{Y},\mathbf{\Phi})$ can be represented by the product of the message PDFs in $\mathcal{U}_i^{(l)}$ since the *no-short-cycle* property ensures the statistical independency among $\mathcal{U}_i^{(l)}$. By applying the fact that the product of Gaussian PDFs results in a scaled Gaussian PDF, the likehood PDF is given as

$$f_{\mathcal{U}_i}^{(l)}(x|X_i,\mathbf{Y},\mathbf{\Phi}) = \mathcal{N}(x;X_i,(\sum_j \frac{1}{\eta_j^{(l)}})^{-1}). \quad (7)$$

It is important to note from (6) that the PDF of the scalar channel output $Z_i^{(l)}$ corresponds to the likelihood PDF $f_{\mathcal{U}_i}^{(l)}(x|X_i,\mathbf{Y},\mathbf{\Phi})$ in the *decoupling principle* [3-5], *i.e.,*

$$f_{\mathcal{U}_i}^{(l)}(x|X_i,\mathbf{Y},\mathbf{\Phi}) \overset{\text{LSS}}{\to} f_{Z_i}^{(l)}(x|X_i,\mathbf{\Phi}), \quad (8)$$

in probability. Hence, the scalar output $Z_i^{(l)}$ is Gaussian distributed with the mean X_i and the variance $\eta^{(l)}$ following

$$\left(\sum_j \frac{1}{\eta_j^{(l)}}\right)^{-1} \overset{\text{LSS}}{\to} \eta^{(l)}. \quad (9)$$

In addition, we can describe the iterative behavior of $\eta^{(l)}$ by an fixed point equation resulting from (4) and (9), given as

$$\eta^{(l)} = \frac{\sigma_W^2}{L} + \frac{R}{L}\mathbf{Var}\left\{X_k \,|\, Z_i^{(l-1)},\eta^{(l-1)},\mathbf{\Phi}\right\}, \quad (10)$$

where the number of nonzeros L in a column of the matrix $\mathbf{\Phi}$ should be properly chosen to satisfy the weak law of large numbers while holding the *no-short-cycle* property. The variance $\eta^{(l)}$ consists of two factors: the additive noise and CEI. The CEI factor is represented as the variance of X_k at the previous iteration, obtained from

$$\mathbf{Var}\left\{X_k \,|\, Z_i^{(l-1)},\eta^{(l-1)},\mathbf{\Phi}\right\}$$
$$= \int\left(x-\mathbf{E}\left\{X_k \,|\, Z_i^{(l-1)},\eta^{(l-1)},\mathbf{\Phi}\right\}\right)^2 f_{X_i}(x|Z_i^{(l)},\mathbf{\Phi})dx, \quad (11)$$

and it will be continuously approches to zero as $l\to\infty$ if the SNR level is sufficiently high [5]. Such a fixed point equation in (10) was originally discussed by Guo and Verdu [4] , and has widely investigated to analyze performance of the BP-based algorithms.

In the sparse estimation, the marginal posterior is computed by imposing a sparsifying prior PDF. In this work, we consider the spike-and-slab PDF as the prior, given as

$$f_X(x) := q\mathcal{N}(x;0,\sigma_X^2)+(1-q)\delta_0. \quad (12)$$

In addition, it is obvious from (8) that

$$f_{X_i}^{(l)}(x|\mathbf{Y},\mathbf{\Phi}) \overset{\text{LSS}}{\to} f_{X_i}(x|Z_i^{(l)},\mathbf{\Phi}) \quad (13)$$

for every i and x, in probability under the LSS setup. Then , the marginal posterior PDF is expressed as

$$f_{X_i}^{(l)}(x|Z_i^{(l)},\mathbf{\Phi}) \quad (14)$$
$$\propto qc_2\mathcal{N}(x;\frac{X_i\sigma_X^2}{\sigma_X^2+\eta^{(l)}},\frac{\sigma_X^2\eta^{(l)}}{\sigma_X^2+\eta^{(l)}})+(1-q)c_1\delta_0,$$

where c_1, c_2 are some constants such that the marginal posterior is valid as a PDF, *i.e.*, $\int f_{X_i}^{(l)}(x|Z_i^{(l)},\mathbf{\Phi})dx=1$.

3 Conclusion

From (14), it turns out that the marginal posterior PDF is a spike-and-slab PDF whose parameters are functions of the signal value X_i and $\eta^{(l)}$ which includes the noise variance. For the calculation of (11), we refer to the paper by Krzakala *et al.* [6]. This analysis result will be very useful to evaluate the BP performance in the sparse estimation problems.

References

[1] D. Baron, *et al*., "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269-280, Jan. 2010.

[2] T. Richardson *et al*., " The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599-618, Feb. 2001.

[3] D. Guo *et al.,* "Asymptotic mean-square optimality of belief propagation for sparse linear systems," *Proc. in IEEE Inform. Theory Workshop (ITW)*, pp. 194-198, Chengdu, China, Oct. 2006.

[4] D. Guo *et al.*, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1983-2010, Jun. 2005.

[5] G. Guo *et al.,*, "Multiuser detection of sparsely spread CDMA," *IEEE J. Sel. Areas Comm.*, vol. 26, no. 3, pp. 421-431, Mar. 2008.

[6] F. Krzakala *et al.* "Probabilistic reconstruction in compressed sensing: algorithm, phase diagrams, and threshold achieving matrices," , *J. Stat. Mech.,* P08009, 2012.