

# Quantize And CS : Report ID

---

Liu Benyuan

LL Corporation

· Xi'an ·

# 第 1 章 The Algorithm

## 1.1 The Overview of BSBL Framework

A block sparse signal  $\mathbf{x} \in \mathbb{C}^{N \times 1}$  has the following structure:

$$\mathbf{x} = [\underbrace{x_1, \dots, x_{d_1}}_{\mathbf{x}_1^H}, \dots, \underbrace{x_1, \dots, x_{d_g}}_{\mathbf{x}_g^H}]^H, \quad (1.1)$$

which means  $\mathbf{x}$  has  $g$  blocks, and only a few blocks are nonzero. Here  $d_i$  is the block size for the  $i$ th block. To model the intra-block correlation in the  $i$ -th block, the BSBL framework suggests to use the parameterized Gaussian distribution:

$$p(\mathbf{x}_i; \gamma_i, \mathbf{B}_i) = \mathcal{CN}(\mathbf{x}_i; \mathbf{0}, \gamma_i \mathbf{B}_i). \quad (1.2)$$

with unknown deterministic parameters  $\gamma_i$  ( $\gamma \geq 0$ ) and  $\mathbf{B}_i$ .

The observation  $\mathbf{y} \in \mathbb{C}^{M \times 1}$  is obtained by

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}, \quad (1.3)$$

where  $\Phi$  is the sensing matrix and  $\mathbf{n}$  is the observation noise. The observation noise is assumed to be independent and Complex Gaussian with zero mean and variance equal to  $\beta^{-1} \mathbf{I}$ .  $\beta$  is also unknown. Thus the likelihood is given by

$$p(\mathbf{y}|\mathbf{x}; \beta) = \mathcal{CN}(\mathbf{y}; \Phi \mathbf{x}, \beta^{-1} \mathbf{I}). \quad (1.4)$$

The hierarchical Bayesian model can be illustrated using a probabilistic graph, see fig 1.1. During the learning procedure most  $\gamma_i$  tend to be zero, due to the mechanism of automatic relevance determination[1]. Thus sparsity in the block level is encouraged[2].  $\mathbf{B}_i \in \mathbb{C}^{d_i \times d_i}$  is a positive definite matrix capturing the correlation structure of the  $i$ -th block. Although blocks have intra-block correlation, the framework assumes that blocks are mutually independent.

The posterior  $p(\mathbf{x}|\mathbf{y}; \{\gamma_i, \mathbf{B}_i\}_i, \beta)$  and the likelihood  $p(\mathbf{y}|\{\gamma_i, \mathbf{B}_i\}_i, \beta)$  can be calculated as

$$p(\mathbf{x}|\mathbf{y}; \{\gamma_i, \mathbf{B}_i\}_i, \beta) = \frac{p(\mathbf{x}_i; \gamma_i, \mathbf{B}_i) p(\mathbf{y}|\mathbf{x}; \beta)}{p(\mathbf{y}|\{\gamma_i, \mathbf{B}_i\}_i, \beta)}, \quad (1.5)$$

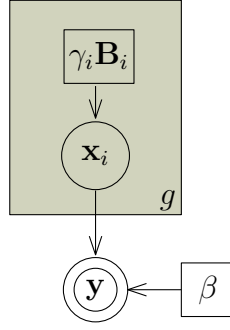


图 1.1 The block structural model. The variable  $\mathbf{y}$  in the double circled is denoted as observations. The  $\mathbf{x}_i$  in the single circled is variables while the value  $\gamma_i \mathbf{B}_i$  and  $\beta$  in square box are parameters. Shaded box means that each prior  $\gamma_i \mathbf{B}_i$  is employed on  $\mathbf{x}_i$  for the range  $i \in [1, \dots, g]$ .

given the data model (1.2) and the observation model (1.4), we apply the Gaussian Identity[3],

$$p(\mathbf{x}|\mathbf{y}; \{\gamma_i, \mathbf{B}_i\}_i, \beta) = \mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (1.6)$$

$$p(\mathbf{y}|\{\gamma_i, \mathbf{B}_i\}_i, \beta) = \mathcal{CN}(\mathbf{y}; \mathbf{0}, \mathbf{C}), \quad (1.7)$$

where  $\boldsymbol{\Sigma} \triangleq (\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Phi}^H \beta \boldsymbol{\Phi})^{-1}$ ,  $\boldsymbol{\mu} \triangleq \boldsymbol{\Sigma} \boldsymbol{\Phi}^H \beta \mathbf{y}$  and  $\mathbf{C} \triangleq \beta^{-1} \mathbf{I} + \boldsymbol{\Phi} \boldsymbol{\Gamma} \boldsymbol{\Phi}^H$  where  $\boldsymbol{\Gamma}$  denotes a block diagonal matrix with the  $i$ th principal block given by  $\gamma_i \mathbf{B}_i$ .

Once all the parameters, namely  $\{\gamma_i, \mathbf{B}_i\}_i, \beta$ , are estimated, the MAP estimate of the signal  $\mathbf{x}$  can be directly obtained from the mean of the posterior, i.e.,

$$\mathbf{x} = \boldsymbol{\Sigma} \boldsymbol{\Phi}^H \beta \mathbf{y}. \quad (1.8)$$

To estimate the parameters, the following cost function is generally used, which is derived from the Type II maximum likelihood [4]:

$$\mathcal{L}(\{\gamma_i, \mathbf{B}_i\}_i, \beta) = \log |\mathbf{C}| + \mathbf{y}^H \mathbf{C}^{-1} \mathbf{y}, \quad (1.9)$$

There are several methods to minimize the cost function [4]. In the following we consider to use the marginalized likelihood maximization method, which was used by Tipping et al. [5] for their basic SBL algorithm and later was used by Ji et al. [6] for their Bayesian compressive sensing algorithm.

## 1.2 The Proposed Block SBL Algorithm

### 1.2.1 The Mainbody of the Algorithm

The cost function (1.9) can be optimized in a block way. We denote by  $\Phi_i$  the  $i$ th block in  $\Phi$  with the column indexes corresponding to the  $i$ th block of the signal  $\mathbf{x}$ . Then  $\mathbf{C}$  can be rewritten as:

$$\mathbf{C} = \beta^{-1} \mathbf{I} + \sum_{m \neq i} \Phi_m \gamma_m \mathbf{B}_m \Phi_m^H + \Phi_i \gamma_i \mathbf{B}_i \Phi_i^H, \quad (1.10)$$

$$= \mathbf{C}_{-i} + \Phi_i \gamma_i \mathbf{B}_i \Phi_i^H, \quad (1.11)$$

where  $\mathbf{C}_{-i} \triangleq \beta^{-1} \mathbf{I} + \sum_{m \neq i} \Phi_m \gamma_m \mathbf{B}_m \Phi_m^H$ . Using the Woodbury Identity,

$$|\mathbf{C}| = |\gamma_i \mathbf{B}_i| |\mathbf{C}_{-i}| |A_i^{-1} + s_i|, \quad (1.12)$$

$$\mathbf{C}^{-1} = \mathbf{C}_{-i}^{-1} - \mathbf{C}_{-i}^{-1} \Phi_i (A_i^{-1} + s_i)^{-1} \Phi_i^H \mathbf{C}_{-i}^{-1}, \quad (1.13)$$

where  $A_i \triangleq \gamma_i \mathbf{B}_i$ ,  $s_i \triangleq \Phi_i^H \mathbf{C}_{-i}^{-1} \Phi_i$ , and  $\mathbf{q}_i \triangleq \Phi_i^H \mathbf{C}_{-i}^{-1} \mathbf{y}$ , the Equation (1.9) can be rewritten as:

$$\begin{aligned} \mathcal{L} &= \log |\mathbf{C}_{-i}| + \mathbf{y}^H \mathbf{C}_{-i}^{-1} \mathbf{y} \\ &\quad + \log |\mathbf{I}_{d_i} + A_i s_i| - \mathbf{q}_i^H (A_i^{-1} + s_i)^{-1} \mathbf{q}_i, \end{aligned} \quad (1.14)$$

$$= \mathcal{L}(-i) + \mathcal{L}(i), \quad (1.15)$$

where  $\mathcal{L}(-i) \triangleq \log |\mathbf{C}_{-i}| + \mathbf{y}^H \mathbf{C}_{-i}^{-1} \mathbf{y}$ , and

$$\mathcal{L}(i) = \log |\mathbf{I}_{d_i} + A_i s_i| - \mathbf{q}_i^H (A_i^{-1} + s_i)^{-1} \mathbf{q}_i, \quad (1.16)$$

which only depends on  $\gamma_i$  and  $\mathbf{B}_i$ .

Setting  $\frac{\partial \mathcal{L}(i)}{\partial \gamma_i} = 0$ , we have the updating rule

$$\gamma_i = \frac{1}{d_i} \text{Tr} \left[ \mathbf{B}_i^{-1} s_i^{-1} (\mathbf{q}_i \mathbf{q}_i^H - s_i) s_i^{-1} \right]. \quad (1.17)$$

Setting  $\frac{\partial \mathcal{L}(i)}{\partial \mathbf{B}_i} = \mathbf{0}$ , we have the updating rule

$$\mathbf{B}_i = \frac{s_i^{-1} (\mathbf{q}_i \mathbf{q}_i^H - s_i) s_i^{-1}}{\gamma_i}. \quad (1.18)$$

**Remark I:** The update equation (1.17) for  $\gamma_i$  guaranteed that  $\gamma_i \in \mathbb{R}$ . The matrix  $\mathbf{C}$  (see eq(1.11)) is a Hermite matrix with that  $\mathbf{C}^H = \mathbf{C}$ , and consequently  $s_i =$

$\Phi_i^H C_{-i}^{-1} \Phi_i$  is also a Hermite matrix. We let  $\mathbf{a} = \text{diag}[s_i^{-1}(\mathbf{q}_i \mathbf{q}_i^H - s_i) s_i^{-1}]$ , and the following property holds:  $\mathbf{a} \in \mathbb{R}^{d_i \times 1}$  and  $\gamma_i = \text{mean}(\mathbf{a})$ .

**Remark II:** The algorithm start with  $\mathbf{B}_i = \mathbf{I}_{d_i}, \forall i$ , iteratively update  $\gamma_i$  using (1.17) and then calculating the intra-correlation  $\mathbf{B}_i$  using (1.18).

### 1.2.2 The Stationary Point of Block SBL

The stationary solution of block BCS can be analyzed by inspecting the second partial derivation (see [5, 7]) of (1.16). In the Fast Marginalized Block SBL algorithm, the analysis of the stationary point is put in the Appendix for not affecting the reader. We only give the simple conclusion in this section.

The minimization of  $\mathcal{L}(\boldsymbol{\gamma})$  can be iteratively updated using,

$$\tilde{\gamma}_i = \begin{cases} \gamma_i, & \gamma_i > 0 \\ 0, & \gamma_i \leq 0 \end{cases}, \quad (1.19)$$

where  $\gamma_i$  is calculated using (1.17) and  $\tilde{\gamma}_i$  denotes the updated value of  $\gamma_i$ .

### 1.2.3 The Change of The Log-Likelihood of Fast Marginalized block SBL

By calculating all  $\rho_i$ , the potential operation on the basis is illustrated in fig 1.2. The change of the log-likelihood (see eq (1.16)) is calculated as:

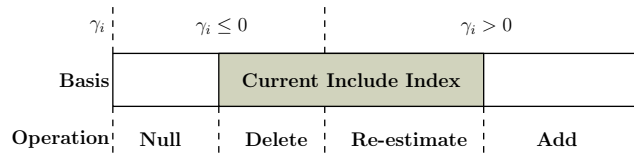


图 1.2 The operation on the basis w.r.t  $\gamma_i$ . The currently included block basis at iteration  $k$  is denoted as the shaded area. Those indices correspond to  $\gamma_i \leq 0$  will be pruned out at iteration  $k + 1$  while the one with  $\gamma_i > 0$  will be add in or re-estimated.

$$\Delta \mathcal{L}(\gamma_i) = \mathcal{L}(\tilde{\gamma}_i) - \mathcal{L}(\gamma_i). \quad (1.20)$$

By calculating  $\Delta \mathcal{L}(\gamma_i), \forall i$ , the minimal value of  $\Delta \mathcal{L}(\gamma_i)$  and the correspond basis is selected to boost the convergence speed in saughting the minimum of the cost function  $\mathcal{L}(\boldsymbol{\gamma})$ .

### 1.2.4 Remarks on $\beta$

The parameter  $\beta^{-1}$  is the noise variance in our model. It can be estimated by a number of methods. However, the resulting updating rule is generally not robust and requires some regularization [2, 4]. In practice, people treat it as a regularizer and assign some specific values to it<sup>1</sup>. Similar to [6], we select  $\beta = 10^{-6}$  in noiseless simulations,  $\beta = 0.1\|y\|_2^2$  in general noisy scenarios (e.g. SNR < 20 dB), and  $\beta = 0.01\|y\|_2^2$  in high SNR scenarios (e.g. SNR  $\geq$  20 dB).

### 1.2.5 The Fast Marginalized block SBL Algorithm

The Fast Marginalized block SBL algorithm (**BSBL-FM**) is given in Fig. 1.1, In the remaining of this paper, we denote **BSBL-FM(0)** as the algorithm ignores the intra-block correlation (i.e.  $\mathbf{B}_i = \mathbf{I}_{d_i}$ ), **BSBL-FM(1)** as the one iteratively update the intra-block correlation  $\mathbf{B}_i$ .

---

**Algorithm 1.1** The BSBL-FM Algorithm.

---

```

1: outputs:  $\mathbf{w}, \Sigma, \gamma$ .
2: while convergence criterion not met do
3:   Calculate  $\gamma_i, \mathbf{B}$  using (1.17) and (1.18);
4:   Calculate  $\Delta\mathcal{L}(\gamma_i)$ ;
5:   Select block  $i$  with  $\Delta\mathcal{L}(\gamma_i) = \min \Delta\mathcal{L}(\gamma_i)$ ;
6:   if add then
7:     Process add on quantities  $\mu, \Sigma, s, q$ .
8:   else if re-estimate then
9:     Process re-estimate on  $\mu, \Sigma, s, q$ .
10:  else if delete then
11:    Process delete on quantities  $\mu, \Sigma, s, q$ .
12:  end if
13:  Re-calculate the convergence criterion.
14: end while
```

---



---

<sup>1</sup>For example, one can see this by examining the published codes of the algorithms in [6, 8].

### 1.3 Reformulated ARD and Incremental Block ARD Schemes

The objective function  $\mathcal{L}(\boldsymbol{\gamma})$  can be up bounded as

$$\mathcal{L}(\mathbf{z}, \boldsymbol{\gamma}) = \mathbf{z}^T \boldsymbol{\gamma} - h^*(\mathbf{z}) + \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}, \quad (1.21)$$

where  $\mathbf{z} = [\mathbf{z}_1, \dots, \mathbf{z}_g]^T$  and  $h^*(\mathbf{z})$  is the concave conjugate of  $\log |\mathbf{C}|$  defined by the duality relationship  $h^*(\mathbf{z}) = \min_{\boldsymbol{\gamma}} \mathbf{z}^T \boldsymbol{\gamma} - \log |\mathbf{C}|$ . Then for  $\boldsymbol{\gamma}$  fixed at  $\hat{\boldsymbol{\gamma}}$ , the tightest upper bound of  $\boldsymbol{\gamma}$  is obtained by minimizing  $\mathcal{L}(\boldsymbol{\gamma}, \mathbf{z})|_{\boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}} = \mathcal{L}(\mathbf{z}; \hat{\boldsymbol{\gamma}})$  over all  $\mathbf{z}$ , the corresponding minimizer can be found in closed form:

$$\mathbf{z} = \text{diag}(\boldsymbol{\Phi}^T \mathbf{C}^{-1} \boldsymbol{\Phi}), \quad (1.22)$$

now be fix  $\mathbf{z}$  at  $\hat{\mathbf{z}}$ , the  $\mathcal{L}(\boldsymbol{\gamma}; \hat{\mathbf{z}})$  is minimized by solving

$$\boldsymbol{\gamma} = \arg \min_{\boldsymbol{\gamma}} \hat{\mathbf{z}}^T \boldsymbol{\gamma} + \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y}, \quad (1.23)$$

which again can be partitioned as,

$$\mathcal{L}(\boldsymbol{\gamma}; \hat{\mathbf{z}}) = \sum_{k \neq i} \mathbf{z}_k^T \boldsymbol{\gamma}_k + \mathbf{y}^T \mathbf{C}_{-i}^{-1} \mathbf{y} \quad (1.24)$$

$$+ \mathbf{z}_i^T \boldsymbol{\gamma}_i - \mathbf{q}_i^T (\mathbf{A}_i^{-1} + \mathbf{s}_i)^{-1} \mathbf{q}_i, \quad (1.25)$$

### 1.4 Structured Regularized Block-SBL Method

**biblio blaze: structured matrices, Geometric methods for estimation of structured covariances, structure regularized learning,**

#### 1.4.1 Regularization to $\mathbf{B}_i$

As noted in [4], regularization to  $\mathbf{B}_i$  is required due to limited data. It has been shown [2] that in noiseless cases the regularization does not affect the global minimum of the cost function (1.9), i.e., the global minimum still corresponds to the true solution; the regularization only affects the probability of the algorithm to converge to the local minima. A good regularization can largely reduce the probability of local convergence. Although theories on regularization strategies are lacked, some empirical methods [2, 4] were presented in literature.

It should be noted that the information of the signal structure within each block can be well described by the covariance matrices  $\mathbf{B}_i$ . Exploiting such information can

potentially improve the recovery performance. In the following, we firstly adopting the empirical approximation method proposed in [4]; and then proposed a geometric methods for estimating the underlying Toeplitz matrix.

#### 1.4.2 Regularized using AR(1) Model[9]

Our algorithm can conveniently utilizing the regularization used in [4], which model the entries in each block as a first-order Auto-Regressive (AR) process with the AR coefficient  $r_i$ . As a result,  $\mathbf{B}_i$  has the following form

$$\mathbf{B}_i = \text{Toeplitz}([1, r_i, \dots, r_i^{d_i-1}]). \quad (1.26)$$

where  $\text{Toeplitz}(\cdot)$  is a Matlab command expanding a real vector into a symmetric Toeplitz matrix. Thus the correlation level of the intra-block correlation is reflected by the value of  $r_i$ .  $r_i$  can be estimated from the cost function (1.9) directly, or can be empirically roughly calculated from the estimated  $\mathbf{B}_i$  in (1.18) as shown in [4]. We used the latter method, since it provides satisfactory results and saves lots of computation. According to [4],  $r_i$  is calculated by  $r_i \triangleq \frac{m_1^i}{m_0^i}$ , where  $m_0^i$  (res.  $m_1^i$ ) is the average of entries along the main diagonal (res. the main sub-diagonal) of the matrix  $\mathbf{B}_i$ . This calculation cannot ensure  $r_i$  has a feasible value, i.e.  $|r_i| < 0.99$ . Thus in practice, we calculate  $r_i$  by <sup>1</sup>

$$r_i = \text{sign}\left(\frac{m_1^i}{m_0^i}\right) \min\left\{\left|\frac{m_1^i}{m_0^i}\right|, 0.99\right\}. \quad (1.27)$$

Our algorithm using this regularization (1.26)-(1.27) is denoted by **BSBL-FM(1)**.

In many real-world applications, the intra-block correlation in each block of a signal tends to be positive and high together. Thus, one can further constrain that all the intra-block correlation values of blocks have the same AR coefficient  $r$  [4], where  $r$  is the average of all the  $r_i$ ,

$$r = \frac{1}{g} \sum_{i=1}^g r_i. \quad (1.28)$$

Then,  $\mathbf{B}_i$  is reconstructed as

$$\mathbf{B}_i = \text{Toeplitz}([1, r, \dots, r^{d_i-1}]). \quad (1.29)$$

---

<sup>1</sup>this constrain guaranteed that the determinant of  $\mathbf{B}_i$ , which is  $(1 - r_i^2)^{d_i-1}$ , is positive



Our algorithm using this regularization (1.28)-(1.29) is denoted by **BSBL-FM(2)**.

### 1.4.3 Regularized using Geometric method with Toeplitz Constrains

The update equation for  $\mathbf{B}_i$  (see (1.18)) can not be directly used, due to the fact that the matrix obtained using (1.18) may not be toeplitz due to statistical errors, and is sometimes badly conditioned, which will cause numerical errors during the iteration process.

Many other forms of constrains, such as Temporal Smooth Matrix[10, 11] or Band-Toeplitz Matrix can also be used in a similar manner.

We focus on the problem of estimating the Toeplitz covariance from finite observations. The Toeplitz matrix is sought as the one closest to  $\mathbf{B}_i$  in a suitable geometry.

#### 1.4.3.1 Geometry Viewpoint

Given  $\mathbf{B}_i$ , we consider the problem to minimize

$$\min_{\mathbf{T} \in \mathcal{T}} d(\mathbf{T}, \mathbf{B}), \quad (1.30)$$

over the class of admissible matrices

$$\mathcal{T} \triangleq \{\mathbf{T} : \mathbf{T} \succeq 0, \mathbf{T} \text{ being Toeplitz}\}. \quad (1.31)$$

## 参考文献

- [1] Tipping M E. Sparse Bayesian Learning and the Relevance Vector Machine [J]. Journal of Machine Learning Research. 2001, 1: 211–244.
- [2] Zhang Z, Rao B D. Sparse Signal Recovery with Temporally Correlated Source Vectors Using Sparse Bayesian Learning [J]. IEEE Journal of Selected Topics in Signal Processing. 2011, 5 (5): 912–926.
- [3] Petersen K B, Pedersen M S. The Matrix Cookbook [EB/OL]. 2008. <http://matrixcookbook.com>.
- [4] Zhang Z, Rao B D. Extension of SBL Algorithms for the Recovery of Block Sparse Signals with Intra-Block Correlation [J]. to appear in IEEE Transactions on Signal Processing. 2012.
- [5] Faul A C, Tipping M E. Analysis of Sparse Bayesian Learning [J]. Neural Inform. Process. Syst. 2002, 14: 383–389.
- [6] Ji S, Xue Y, Carin L. Bayesian Compressive Sensing [J]. IEEE Transactions on Signal Processing. 2008, 56: 2346–2356.
- [7] Tipping M E, Faul A C. Fast Marginal Likelihood Maximisation for Sparse Bayesian Models [C] // Bishop C M, Frey B J. In Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics. Key West, FL, 2003: 3–6.
- [8] Babacan S D, Nakajima S, Do M N. Bayesian Group-Sparse Modeling and Variational Inference [J]. 2012, Sep: 1–20.
- [9] Zhang Z, Rao B D. Clarify Some Issues on the Sparse Bayesian Learning for Sparse Signal Recovery [R]. 2012.
- [10] Chen Z, Cichocki A. Nonnegative Matrix Factorization with Temporal Smoothness and/or Spatial Decorrelation Constrains [J]. NIPS. 2007, 1: 1–10.

- [11] Zhou J, Yuan L, Liu J, et al. A Multi-Task Learning Formulation for Predicting Disease Progression [J]. KDD. 2011, 1: 1–9.