LOW-RANK AND JOINT-SPARSE SIGNAL RECOVERY FOR SPATIALLY AND TEMPORALLY CORRELATED DATA USING SPARSE BAYESIAN LEARNING

Yanbin Zhang, Yangqing Li, Changchuan Yin, and Kesen He
{iyanbin4, liyq, ccyin,kesenhe}@bupt.edu.cn
Beijing Laboratory of Advanced Information Networks,
Beijing Key Laboratory of Network System Architecture and Convergence,
Beijing University of Posts and Telecommunications, Beijing 100876, P. R. China
INTRODUCTION

Motivation: The data in WBAN prevalently has both spatial and temporal correlations at the same time. And it will obtain a superior performance when we consider the structured spatial and temporal correlations jointly by assuming the spatio-temporal correlated data satisfies simultaneous low-rank and joint-sparse (L&S) structure.

The proposed method:
- We first formulate our problem and transform it into a block single measurement problem.
- The structure of the covariance matrix of the L&S data is given.
- The inference problem is split into two steps: Firstly, we get initial values of hyperparameters. Secondly, we get the optimal reconstructed data.
**PROBLEM FORMULATION AND SIGNAL MODEL**

We consider a typical WBAN scenario in which there are $m$ sensors to collect data $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_m]^\top \in \mathbb{R}^{m \times n}$ in time synchronization, where $\mathbf{f}_i \in \mathbb{R}^{n \times 1}$, $i \in 1, 2, \ldots, m$ stands for the data collected by the $i$th sensor and $\mathbf{F}$ is the spatially and temporally correlated data matrix.

\[
\begin{align*}
\mathbf{Y} &= \Xi \mathbf{F} + \mathbf{V}, \\
\mathbf{F} &= \Psi \mathbf{X}, \quad \Phi = \Xi \Psi
\end{align*}
\]

\[\mathbf{y} = \mathbf{Ax} + \mathbf{v}, \quad (2)\]

where $\mathbf{y} = \text{vec}[\mathbf{Y}^\top] \in \mathbb{R}^{np \times 1}$, $\mathbf{A} = \Phi \otimes \mathbf{I}_n \in \mathbb{R}^{np \times nm}$,

\[
\mathbf{x} = \text{vec}[\mathbf{X}^\top] = [\mathbf{x}_1^\top, \cdots, \mathbf{x}_m^\top]^\top \in \mathbb{R}^{nm \times 1}
\]
Gaussian likelihood:

\[
p(y \mid x; A, \lambda) \sim \mathcal{N}_{y\mid x}(Ax, \lambda I) \propto \exp[-\frac{1}{2\lambda} \| Ax - y \|_2^2], \quad (3)
\]

\[
p(x; \gamma_i, \gamma_j, B_{ij}, \forall i, j) \sim \mathcal{N}_x(0, \Sigma_0) \propto \exp[x^\top \Sigma_0^{-1} x], \quad (4)
\]

\[
\Sigma_0 = \begin{bmatrix}
\gamma_1 \gamma_1 B_{11} & \gamma_1 \gamma_2 B_{12} & \cdots & \gamma_1 \gamma_m B_{1m} \\
\gamma_2 \gamma_1 B_{21} & \gamma_2 \gamma_2 B_{22} & \cdots & \gamma_2 \gamma_m B_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_m \gamma_1 B_{m1} & \gamma_m \gamma_2 B_{m2} & \cdots & \gamma_m \gamma_m B_{mm}
\end{bmatrix}, \quad (5)
\]
An example structure of the covariance matrix $\Sigma_0$ of $\mathbf{x}$ with $m = 4, n = 6$. 

Fig. 1
Using the Bayes rule, we have the posterior density of \( \mathbf{x} \),

\[
p(\mathbf{x} | \mathbf{y}; \lambda, \gamma_i, \gamma_j, \mathbf{B}_{ij}, \forall i, j) \sim \mathcal{N}_x (\mu_x, \Sigma_x),
\]

\[
\mu_x = \frac{1}{\lambda} \sum_x \mathbf{A}^\top \mathbf{y},
\]

\[
\Sigma_x = (\Sigma_0^{-1} + \frac{1}{\lambda} \mathbf{A}^\top \mathbf{A})^{-1} n
\]

\[
= \Sigma_0 - \Sigma_0 \mathbf{A}^\top (\lambda \mathbf{I} + \mathbf{A} \Sigma_0 \mathbf{A}^\top)^{-1} \mathbf{A} \Sigma_0,
\]
MAP estimation:

\[ \hat{x} = \text{vec}[\hat{X}^\top] = \mu_x = (\lambda \Sigma_0^{-1} + A^\top A)^{-1} A \Sigma_0 \]

\[ = \Sigma_0 A^\top (\lambda I + A \Sigma_0 A^\top)^{-1} y, \quad (9) \]

Using a common positive definite matrix \( \mathcal{B} \) to model all the covariance matrices \( \mathcal{B}_{ij} \), so, (5) turns into

\[ \Sigma_0 = \Gamma \otimes \mathcal{B}, \quad (10) \]
where,

\[
\Gamma = \begin{bmatrix}
\gamma_1 \gamma_1 & \gamma_1 \gamma_2 & \cdots & \gamma_1 \gamma_m \\
\gamma_2 \gamma_1 & \gamma_2 \gamma_2 & \cdots & \gamma_2 \gamma_m \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_m \gamma_1 & \gamma_m \gamma_2 & \cdots & \gamma_m \gamma_m
\end{bmatrix} \quad (11)
\]
Using Bayesian strategy

\[
\max_{\mathcal{B} \in \mathcal{H}^+, \Gamma \geq 0} \int p(y | x; A, \lambda) p(x; \Gamma, B) dx, \quad (12)
\]

which is equivalent to minimizing the cost function

\[
\mathcal{L}(\Gamma, B, \lambda) = y^\top \Sigma_y^{-1} y + \log | \Sigma_y |, \quad (13)
\]

\[
\Theta = \{ \Gamma, B, \lambda \}
\]

\[
\mathcal{L}(\Theta) = y^\top \Sigma_y^{-1} y + \log | \Sigma_y |, \quad (15)
\]

\[
\Sigma_y = A \Sigma_0 A^\top + \lambda I, \quad \Sigma_0 = \Gamma \otimes B. \quad (14)
\]
We first treat \( x \) as hidden variables in the EM formulation proceeding and then maximize

\[
Q(\Theta) = E_{x \mid y; \Theta^{(pre)}} \left[ \log p(y, x; \Theta) \right]
\]

\[
= E_{x \mid y; \Theta^{(pre)}} \left[ \log p(y \mid x; \lambda) \right]
\]

\[
+ E_{x \mid y; \Theta^{(pre)}} \left[ \log p(x; \Gamma, B) \right]
\]

where \( \Theta^{(pre)} \) denotes the hyperparameters which have been estimated in the previous iteration.

To estimate \( \Gamma \) and \( B \), we assume \( \Gamma = \text{diag}(\gamma_1^2, \cdots, \gamma_m^2) \) where \( \text{diag}(\cdot) \) denotes a diagonal matrix operator.
So, we can simplify the $Q$ function (16) to

$$Q(\Gamma, B) = E_{x|y;\Theta^{(pre)}}[\log p(x; \Gamma, B)], \quad (17)$$

Then we have

$$Q(\Gamma, B) \propto -\frac{n}{2} \log(|\Gamma|) - \frac{m}{2} \log(|B|)$$

$$- \frac{1}{2} \text{tr}[(\Gamma^{-1} \otimes B^{-1})(\Sigma_x + \mu_x \mu_x^\top)]. \quad (18)$$
Then, we plug $\mu_x$ and $\Sigma_x$ into (18). To estimate hyperparameters $\Theta$, we get the gradients of (18) over $\gamma_i^2$ and $B$, respectively, and then we obtain $\gamma_i^{(pre)}$, $i = 1, \cdots, m$, and $B^{(pre)}$. Thus, we will get $\Gamma^{(pre)}$. Using the same way, we can get $\lambda^{(pre)}$. Finally, we get $\Theta^{(pre)}$. Here, $A^{(pre)}$ denotes a initial value of $A$.

In order to get an exact result of $\Theta$, we employ standard upper bounds for solving (13) which known as a non-convex optimization problem leading to an EM-like algorithm. For the first and second terms of $L(\Gamma, B)$, we compute their bounds respectively.
Based on [9], for the first term in (13) we have

\[ y^\top \Sigma_y^{-1} y \leq \frac{1}{\lambda} \| y - Ax \|_2^2 + x^\top \Sigma_0^{-1} x, \quad (19) \]

For the second term,

\[ \log | \Sigma_y | \equiv m \log | B | + \log | \lambda A^\top A + \Sigma_0^{-1} | \]

\[ \leq m \log | B | + \text{tr}[B^{-1} \nabla B^{-1}] + C, \quad (20) \]

where for the second term \( \log | \lambda A^\top A + \Sigma_0^{-1} | \), we use a first-order approximation with a bias term \( C \) to approximate it with equality whenever the gradient satisfies
\[
\n\nabla_{B^{-1}} = \sum_{i=1}^{m} B - BA_i^\top \left( A\Sigma_0 A^\top + \lambda I \right)^{-1} A_i B, \quad (21)
\]

where \( A = [A_1, \ldots, A_m] \) and \( A_i \in \mathbb{R}^{p \times n} \). Finally using the upper bounds of (19), (20) and \( \nabla_{B^{-1}} \), we have the optimal \( B \) in closed form as

\[

B^{opt} = \arg \min_x \text{tr}[B^{-1} (XX^\top + \nabla_{B^{-1}})] + m \log |B|

= \frac{1}{m} \left[ \hat{X} \hat{X}^\top + \nabla_{B^{-1}} \right]. \quad (22)
\]
By starting with $B = B^{(pre)}$ and then iteratively computing (9), (21), and (22), we then have an estimate for $B$, and a corresponding estimate for $x$ given by (9).

We refer to this approach as L&S-bSBL algorithm which is outlined in Algorithm 1.
Algorithm 1 L&S-bSBL

Input: \( y, A; \)
Output: \( X; \)

procedure
  Initialize
  \( \text{iters} = 0, \, \delta = 10^{-6}, \, \text{max iteration number} = 500; \)
  assume \( \Gamma = \text{diag}(\gamma_1^2, \ldots, \gamma_m^2); \)
  compute \( \Gamma, B, \lambda \) from (18);
  \( \Sigma_0 \leftarrow \Gamma \otimes B; \)
  while \( \|X - \hat{X}\|_2^2 \geq \delta \) do
    compute \( \hat{X} \) from (9);
    compute \( \nabla_{B^{-1}} \) from (21);
    compute \( B^{opt} \) from (22);
    \( \text{iters} = \text{iters} + 1; \)
    if \( \text{iters} \geq 500 \) STOP; end if
  end while
  Get the best \( B^{opt} \) and \( X. \)
end procedure
Fig. 2. (a) MSE vs SNR.

- BARM
- TSBL
- L&S-bSBL

n=30, m=50, p=20, r=2

Synthetic Experiments
$n=30$, $m=50$, $p=20$, SNR=$25$ dB

Fig. 2. (b) MSE vs Rank.
Fig. 3. (a) MSE vs m.

$n=30$, $p=20$, $r=2$, SNR=25dB

MSE (dB)

data dimension $m$
Fig. 3. (b) runtime vs m.

$n=30$, $p=20$, $r=2$, SNR=25dB

- BARM
- TSBL
- L&S-bSBL
Fig. 4. MSE vs SNR.

http://physionet.org/physiobank/database/incartdb
Fig. 4. runtime vs SNR.
In this paper, we studied joint sparse reconstruction of spatially and temporally correlated data in WBAN, assuming that the signal matrix satisfies the L&S model. We proposed an algorithm based L&S structure to recover data using a bSBL-based algorithm. The proposed approach presented a better performance than other two methods through numerical results.
Thank you