北京郵電大學
Beijing University of Posts and Telecommunications

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 $\mathbf{f}_i \in \mathbb{R}^{n \times 1}, i \in 1, 2, ..., m$ stands for the data collected by the ith sensor and $\mathbf{F}$ is the spatially and temporally correlated data matrix.

MMV

$$\mathbf{Y} = \boldsymbol{\Xi} \mathbf{F} + \mathbf{V},$$

$$\mathbf{F} = \boldsymbol{\Psi} \mathbf{X}, \quad \boldsymbol{\Phi} = \boldsymbol{\Xi} \boldsymbol{\Psi}$$

$$\mathbf{Y} = \boldsymbol{\Phi} \mathbf{X} + \mathbf{V}, (1)$$

SMV

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}, (2)$$

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

$$\mathbf{x} = \mathrm{vec}[\mathbf{X}^\top] = [\mathbf{x}_1^\top, \cdots, \mathbf{x}_m^\top]^\top \in \mathbf{R}^{nm \times 1}$$

Gaussian likelihood:

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

$$p(\mathbf{x}; \gamma_i, \gamma_j, \mathbf{B}_{ij}, \forall i, j) \sim \mathcal{N}_x(\mathbf{0}, \mathbf{\Sigma}_0) \propto \exp[\mathbf{x}^\top \mathbf{\Sigma}_0^{-1} \mathbf{x}], (4)$$

$$\mathbf{\Sigma}_0 = \begin{bmatrix} \gamma_1 \gamma_1 \mathbf{B}_{11} & \gamma_1 \gamma_2 \mathbf{B}_{12} & \cdots & \gamma_1 \gamma_m \mathbf{B}_{1m} \\ \gamma_2 \gamma_1 \mathbf{B}_{21} & \gamma_2 \gamma_2 \mathbf{B}_{22} & \cdots & \gamma_2 \gamma_m \mathbf{B}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_m \gamma_1 \mathbf{B}_{m1} & \gamma_m \gamma_2 \mathbf{B}_{m2} & \cdots & \gamma_m \gamma_m \mathbf{B}_{mm} \end{bmatrix}, (5)$$

An example structure of the covariance matrix $\mathbf{\Sigma}_0$ of $\mathbf{x}$ with $m = 4, n = 6$.

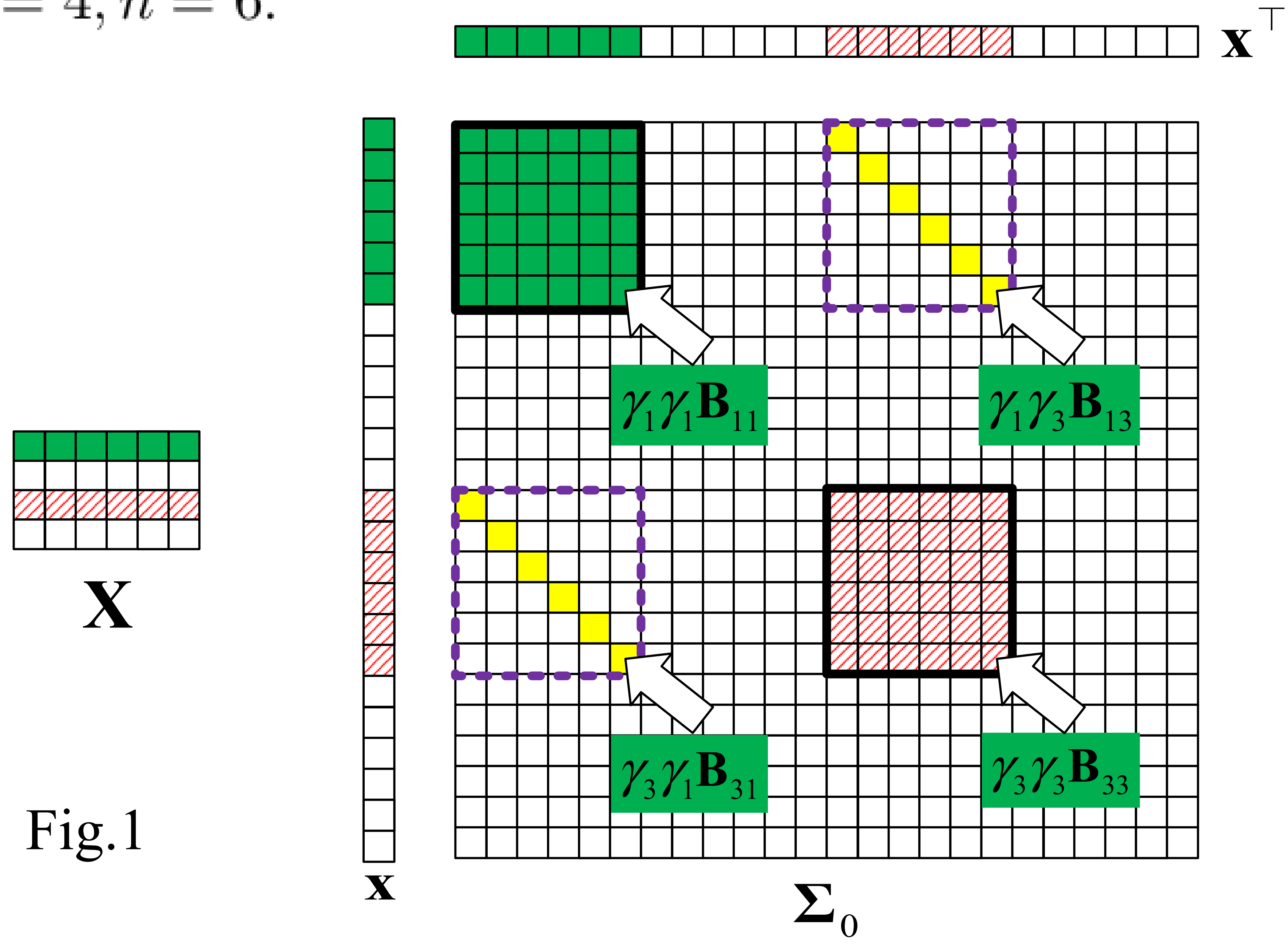

Fig.1

PROPOSED ALGORITHM

Using the Bayes rule, we have the posterior density of $\mathbf{x}$,

$$p(\mathbf{x} \mid \mathbf{y}; \lambda, \gamma_i, \gamma_j, \mathbf{B}_{ij}, \forall i, j) \sim \mathcal{N}_x(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x), (6)$$

$$\boldsymbol{\mu}_x = \frac{1}{\lambda} \boldsymbol{\Sigma}_x \mathbf{A}^\top \mathbf{y}, (7)$$

$$\boldsymbol{\Sigma}_x = (\boldsymbol{\Sigma}_0^{-1} + \frac{1}{\lambda} \mathbf{A}^\top \mathbf{A})^{-1} n$$

$$= \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_0 \mathbf{A}^\top (\lambda \mathbf{I} + \mathbf{A} \boldsymbol{\Sigma}_0 \mathbf{A}^\top)^{-1} \mathbf{A} \boldsymbol{\Sigma}_0, (8)$$

MAP estimation:

$$\hat{\mathbf{x}} = \text{vec}[\hat{\mathbf{X}}^\top] \overset{\Delta}{=} \boldsymbol{\mu}_x = (\lambda \boldsymbol{\Sigma}_0^{-1} + \mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A} \boldsymbol{\Sigma}_0$$

$$= \boldsymbol{\Sigma}_0 \mathbf{A}^\top (\lambda \mathbf{I} + \mathbf{A} \boldsymbol{\Sigma}_0 \mathbf{A}^\top)^{-1} \mathbf{y}, (9)$$

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

$$\boldsymbol{\Sigma}_0 = \boldsymbol{\Gamma} \otimes \mathbf{B}, (10)$$

where,

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1\gamma_1 & \gamma_1\gamma_2 & \cdots & \gamma_1\gamma_m \\ \gamma_2\gamma_1 & \gamma_2\gamma_2 & \cdots & \gamma_2\gamma_m \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_m\gamma_1 & \gamma_m\gamma_2 & \cdots & \gamma_m\gamma_m \end{bmatrix}. \quad (11)$$

Using Bayesian strategy

$$\max_{\mathbf{B}\in H^{+},\boldsymbol{\Gamma}\geq 0} \int p(\mathbf{y}\,|\,\mathbf{x};\mathbf{A},\lambda)\,p(\mathbf{x};\boldsymbol{\Gamma},\mathbf{B})d\mathbf{x}, \ (12)$$

which is equivalent to minimizing the cost function

$$\mathcal{L}(\boldsymbol{\Gamma},\mathbf{B},\lambda) = \mathbf{y}^{\top}\boldsymbol{\Sigma}_{y}^{-1}\mathbf{y} + \log|\boldsymbol{\Sigma}_{y}|, \ (13)$$

$$\boldsymbol{\Theta} = \{\boldsymbol{\Gamma},\mathbf{B},\lambda\}$$

$$\mathcal{L}(\boldsymbol{\Theta}) = \mathbf{y}^{\top}\boldsymbol{\Sigma}_{y}^{-1}\mathbf{y} + \log|\boldsymbol{\Sigma}_{y}|, \ (15)$$

$$\boldsymbol{\Sigma}_{y} = \mathbf{A}\boldsymbol{\Sigma}_{0}\mathbf{A}^{\top} + \lambda\mathbf{I}, \quad \boldsymbol{\Sigma}_{0} = \boldsymbol{\Gamma}\otimes\mathbf{B}. \ (14)$$

We first treat **x** as hidden variables in the EM formulation proceeding and then maximize

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\Theta}) &= E_{x|y;\boldsymbol{\Theta}^{(pre)}}[\log p(\mathbf{y},\mathbf{x};\boldsymbol{\Theta})] \\ &= E_{x|y;\boldsymbol{\Theta}^{(pre)}}[\log p(\mathbf{y}\,|\,\mathbf{x};\lambda)] \qquad ,(16) \\ &\quad + E_{x|y;\boldsymbol{\Theta}^{(pre)}}[\log p(\mathbf{x};\boldsymbol{\Gamma},\mathbf{B})]\end{aligned}$$

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

To estimate $\boldsymbol{\Gamma}$ and $\mathbf{B}$, we assume $\boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1^2, \cdots, \gamma_m^2)$ where $\mathrm{diag}(\cdot)$ denotes a diagonal matrix operator.

So, we can simplify the $\mathcal{Q}$ function (16) to

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

Then we have

$$\mathcal{Q}(\mathbf{\Gamma},\mathbf{B}) \propto -\frac{n}{2}\log(|\mathbf{\Gamma}|) - \frac{m}{2}\log(|\mathbf{B}|) - \frac{1}{2}\mathrm{tr}[(\mathbf{\Gamma}^{-1} \otimes \mathbf{B}^{-1})(\mathbf{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x^\top)]. \quad (18)$$

Then, we plug $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_x$ into (18). To estimate hyper-parameters $\boldsymbol{\Theta}$, we get the gradients of (18) over γ_i^2 and $\mathbf{B}$, respectively, and then we obtain $\gamma_i^{(pre)}$, $i = 1, \cdots, m$, and $\mathbf{B}^{(pre)}$. Thus, we will get $\boldsymbol{\Gamma}^{(pre)}$. Using the same way, we can get $\lambda^{(pre)}$. Finally, we get $\boldsymbol{\Theta}^{(pre)}$. Here, $\mathbf{A}^{(pre)}$ denotes a initial value of $\mathbf{A}$.

In order to get an exact result of $\boldsymbol{\Theta}$, 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 $\mathcal{L}(\boldsymbol{\Gamma}, \mathbf{B})$, we compute their bounds respectively.

Based on [9], for the first term in (13) we have

$$\mathbf{y}^\top \boldsymbol{\Sigma}_y^{-1}\mathbf{y} \le \frac{1}{\lambda} \| \mathbf{y} - \mathbf{A}\mathbf{x} \|_2^2 + \mathbf{x}^\top \boldsymbol{\Sigma}_0^{-1}\mathbf{x}, \quad (19)$$

For the second term,

$$\begin{aligned} \log|\boldsymbol{\Sigma}_y| &\equiv m\log|\mathbf{B}| + \log|\lambda \mathbf{A}^\top \mathbf{A} + \boldsymbol{\Sigma}_0^{-1}| \\ &\le m\log|\mathbf{B}| + \mathrm{tr}[\mathbf{B}^{-1}\nabla_{\mathbf{B}^{-1}}] + C, \qquad (20) \end{aligned}$$

where for the second term $\log|\lambda\mathbf{A}^\top\mathbf{A} + \boldsymbol{\Sigma}_0^{-1}|$, we use a first-order approximation with a bias term C to approximate it with equality whenever the gradient satisfies

$$\nabla_{\mathbf{B}^{-1}} = \sum_{i=1}^{m} \mathbf{B} - \mathbf{B}\mathbf{A}_i^{\top}(\mathbf{A}\boldsymbol{\Sigma}_0\mathbf{A}^{\top} + \lambda\mathbf{I})^{-1}\mathbf{A}_i\mathbf{B}, \quad (21)$$

where $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_m]$ and $\mathbf{A}_i \in \mathbb{R}^{p \times n}$. Finally using the upper bounds of (19), (20) and $\nabla_{\mathbf{B}^{-1}}$, we have the optimal $\mathbf{B}$ in closed form as

$$\mathbf{B}^{opt} = \arg\min_{\mathbf{X}} \mathrm{tr}[\mathbf{B}^{-1}(\mathbf{X}\mathbf{X}^{\top} + \nabla_{\mathbf{B}^{-1}})] + m \log |\mathbf{B}|$$

$$= \frac{1}{m}[\hat{\mathbf{X}}\hat{\mathbf{X}}^{\top} + \nabla_{\mathbf{B}^{-1}}]. \quad (22)$$

By starting with $\mathbf{B} = \mathbf{B}^{(pre)}$ and then iteratively computing (9), (21), and (22), we then have an estimate for $\mathbf{B}$, and a corresponding estimate for $\mathbf{x}$ 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: $\mathbf{y}, \mathbf{A}$;

Output: $\mathbf{X}$;

procedure

Initialize

$iters = 0,\ \delta = 10^{-6},$ **max iteration number** $= 500$;

assume $\mathbf{\Gamma} = \mathrm{diag}(\gamma_1^2, \cdots, \gamma_m^2)$;

compute $\mathbf{\Gamma}, \mathbf{B}, \lambda$ from (18);

$\mathbf{\Sigma}_0 \leftarrow \mathbf{\Gamma} \otimes \mathbf{B}$;

while $\|\mathbf{X} - \hat{\mathbf{X}}\|_2^2 \geq \delta$ **do**

compute $\hat{\mathbf{X}}$ from (9);

compute $\nabla_{\mathbf{B}^{-1}}$ from (21);

compute $\mathbf{B}^{opt}$ from (22);

$iters = iters + 1$;

if $iters \geq 500$ **STOP**; **end if**

end while

Get the best $\mathbf{B}^{opt}$ and $\mathbf{X}$.

end procedure

SIMULATION EXPERIMENTS

Synthetic Experiments

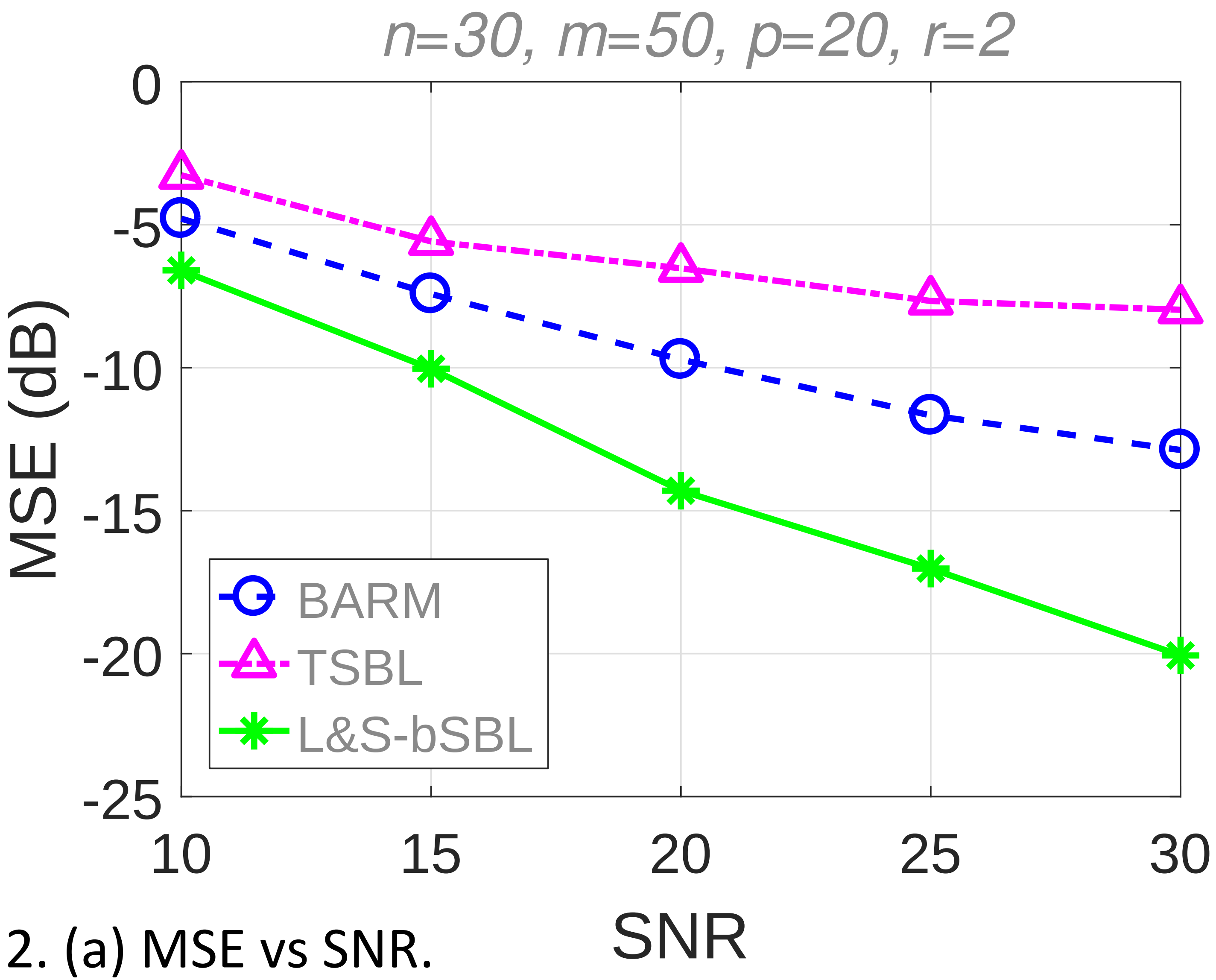

Fig. 2. (a) MSE vs SNR.

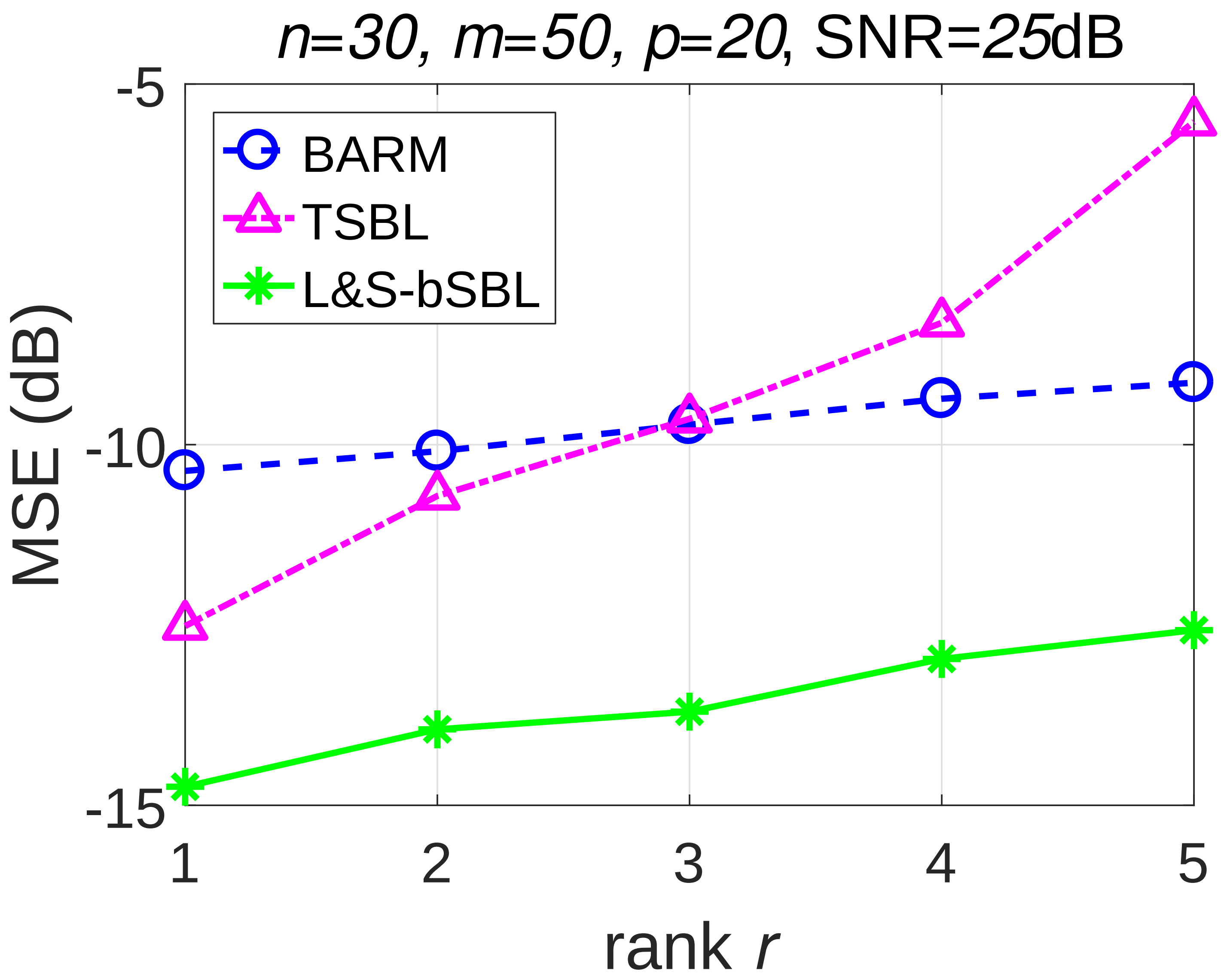

Fig. 2. (b) MSE vs Rank.

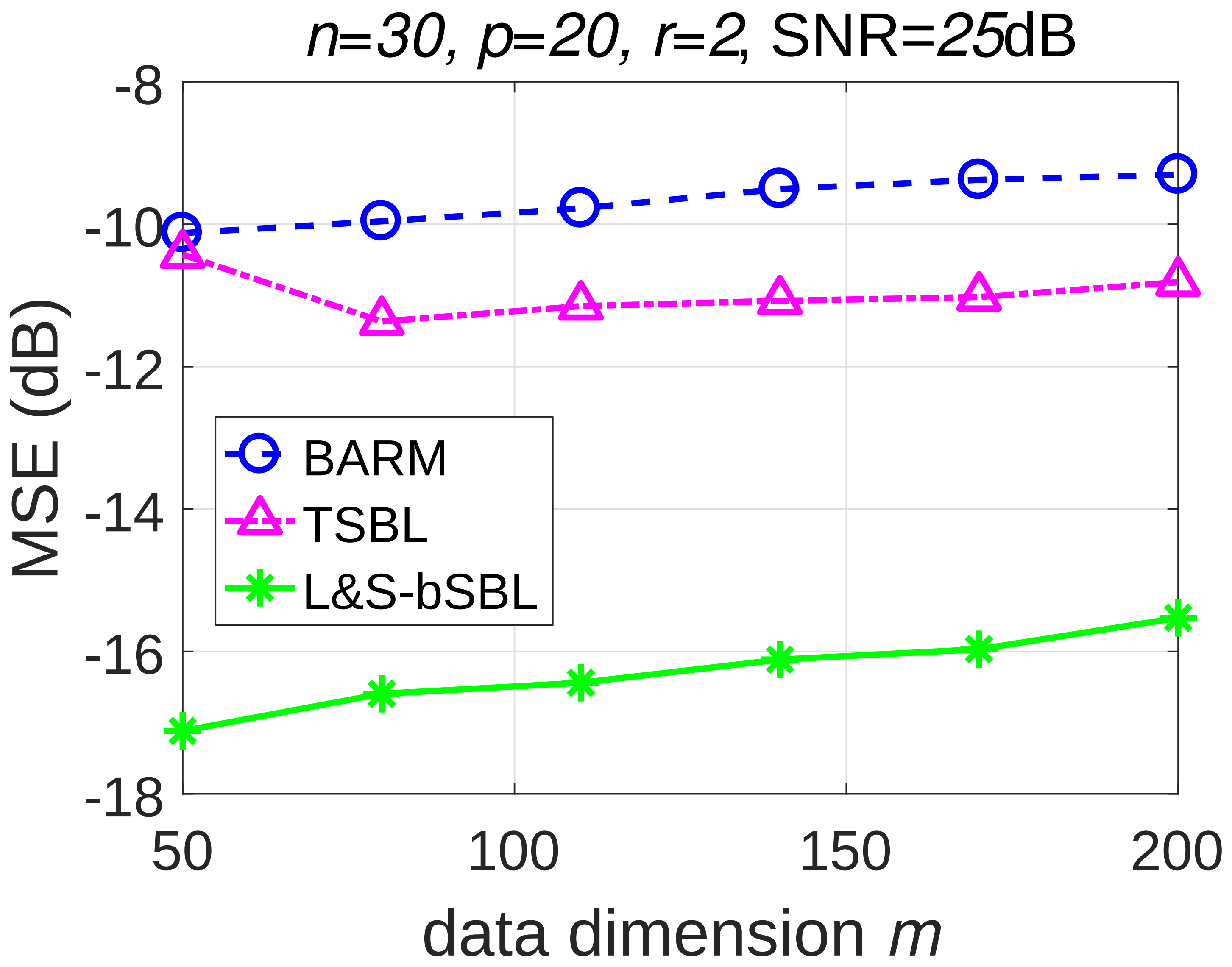

Fig. 3. (a) MSE vs m.

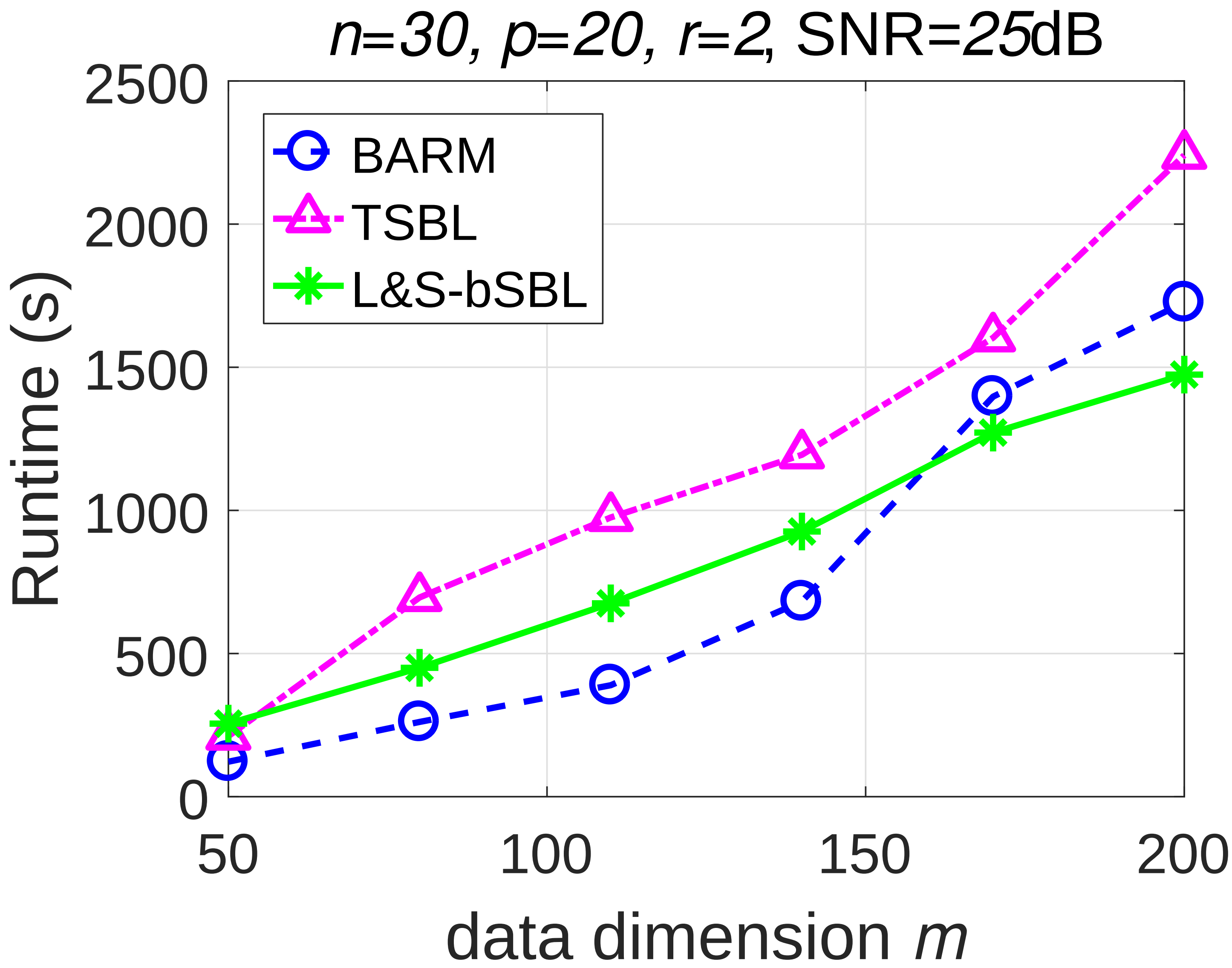

Fig. 3. (b) runtime vs m.

Experiments with Real Data —ECG data

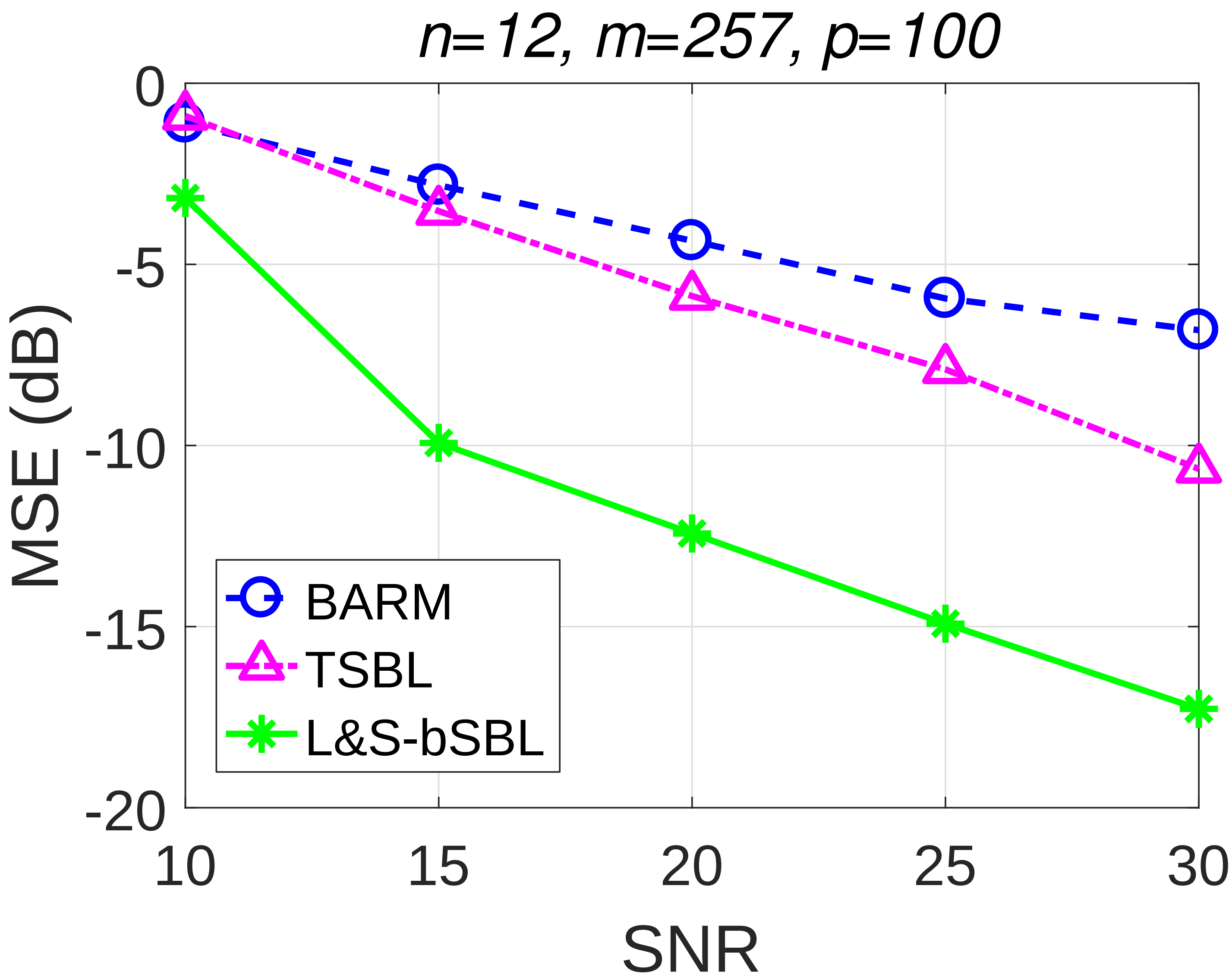

Fig. 4. MSE vs SNR.

http://physionet.org/physiobank/database/incartdb

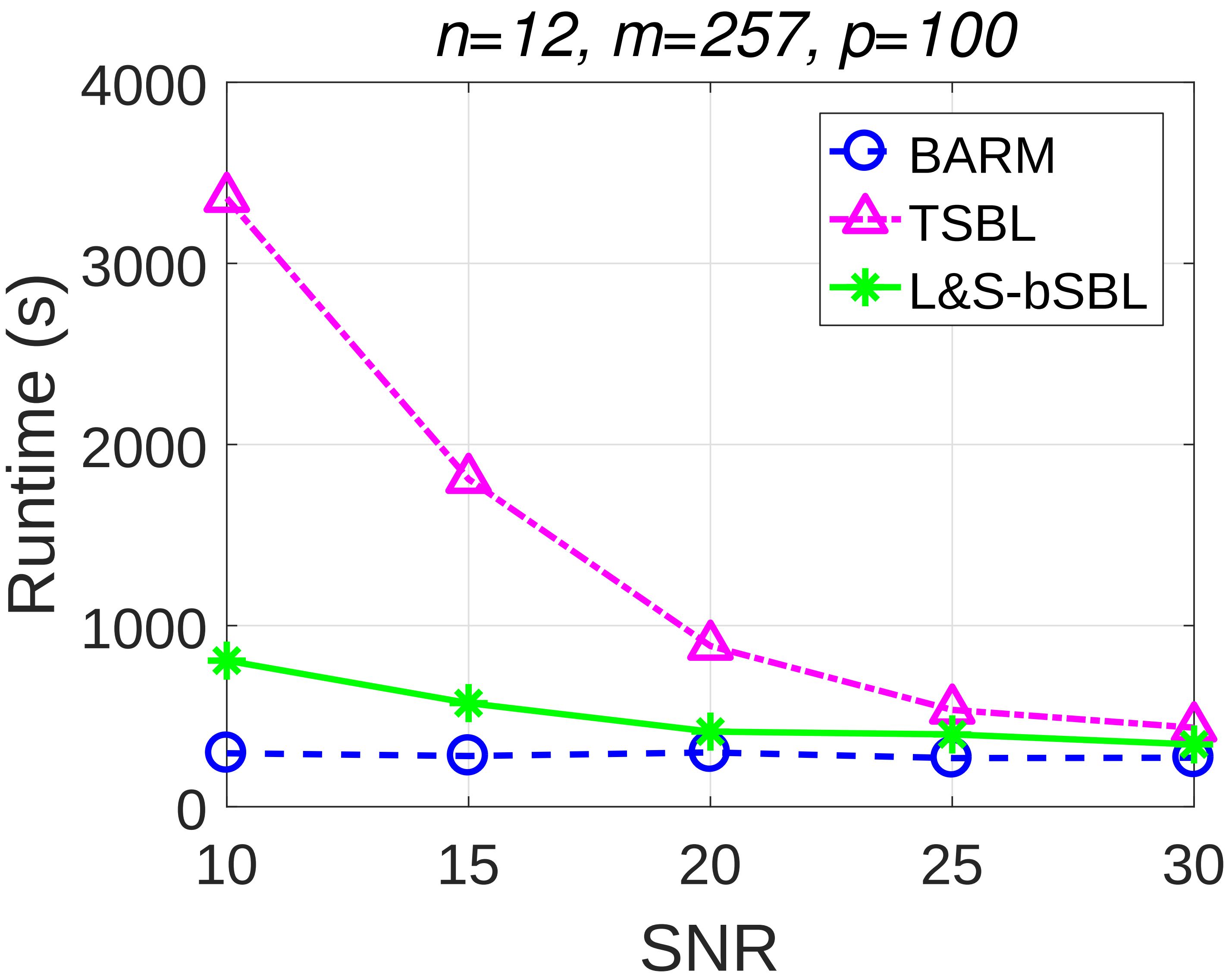

Fig. 4. runtime vs SNR.

CONCLUSION

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