BAGUS: Bayesian Regularization for Graphical Models with Unequal Shrinkage

Lingrui Gan, Naveen N. Narisetty, Feng Liang

Department of Statistics
University of Illinois at Urbana-Champaign

Table of Contents

Introduction

Problem Statement:

$$Y_1, \cdots, Y_n \overset{iid}{\sim} \mathsf{N}_p(0, \Theta^{-1}).$$

Denote $S = \frac{1}{n} \sum Y_i Y_i^t$ as the sample covariance matrix of the data, then the $\log$-likelihood is given by

$$l(\Theta) = \log L(\Theta) = \frac{n}{2}\Big(\log\det(\Theta) - \mathrm{tr}(S\Theta)\Big). \quad (1)$$

- Our target is to estimate Θ, the precision matrix.

To make this problem applicable under the high-dimensional scenario, assumptions need to be made.

Sparsity Assumption

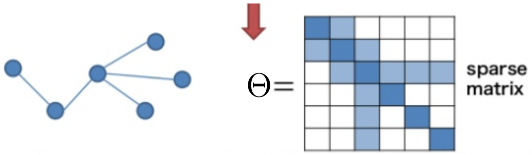
The sparsity assumption is the most common and practical useful one [Dempster, 1972]. It assumes that the majority of the entries are zero, while only a few entries in Θ are non-zero.

Gaussian Graphical Model

Well-known Fact:

- Consider Undirected graph G=(V,E) with V is the vertex set and E is the edge set
- No edge between $(Y_i, Y_j) \Leftrightarrow Y_i \perp\!\!\!\perp Y_j | Y_{[-i,-j]} \Leftrightarrow \Theta_{i,j} = 0$

$$Y \sim N_p(0, \Theta^{-1})$$



Literature Review

Penalized Likelihood

Minimize the negative log-likelihood function with an element-wise penalty on the off-diagonal entries of Θ, i.e.,

$$\arg\min_{\Theta} \left[-\frac{n}{2}\Big(\log\det(\Theta) - \text{tr}(S\Theta) \Big) + \lambda \sum_{i<j} \text{pen}(\theta_{ij}) \right].$$

- The penalty function $\text{pen}(\theta_{ij})$ is often taken to be L_1 [Yuan and Lin, 2007, Banerjee et al., 2008, Friedman et al., 2008],
- but SCAD is also been used [Fan et al., 2009].
- Asymptotic properties have been studied in [Rothman et al., 2008, Lam and Fan, 2009, Ravikumar et al., 2011]

Literature Review

Regression

- Sparse regression model is estimated separately in each column of Θ. Implicitly, they are modeling with under the likelihood $\prod_i P(\mathbf{Y}[i,]|\mathbf{Y}[-i,])$, instead of $P(\mathbf{Y})$.[a] [Meinshausen and Bühlmann, 2006, Peng et al., 2009]

[a] Denote $\mathbf{Y} = (Y_1, \cdots, Y_n)$.

- Other work: CLIME estimator [Cai et al., 2011];

Literature Review

Bayesian Regularization

- Several Bayesian approaches have also been proposed [Wang, 2012, Banerjee and Ghosal, 2015, Gan and Liang, 2016].
- However, Bayesian methods are not in wide use in this fields, because of the high computation cost of MCMC.

Our contributions

1. Propose a new approach for precision matrix estimation, named BAGUS. The adaptive shrinkage is due to the non-convex penalization from our Bayesian formulation.
2. With very mild conditions, the optimal estimation error rate is $O_p\left(\sqrt{\frac{\log p}{n}}\right)$ in the entrywise maximum norm under both exponential and polynomial tail distributions. Selection consistency is also proved.
3. A fast EM algorithm which produces the MAP estimate of the precision matrix and (approximate) posterior probabilities on all edges is proposed. The EM algorithm has computational complexity comparable to the state-of-the-art GLasso algorithm.

Table of Contents

Spike and Slab Prior

Double Exponential Spike and Slab Prior

The cornerstone of our Bayesian formulation is the following spike and slab prior [George and McCulloch, 1997, Ročková and George, 2014, Rocková, 2015] on the off diagonal entries θ_{ij} $(i < j)$:

$$\begin{cases} \theta_{ij} \mid r_{ij} = 0 & \sim & \text{DE}(0, v_0). \\ \theta_{ij} \mid r_{ij} = 1 & \sim & \text{DE}(0, v_1). \end{cases}$$

where $0 \leq v_0 < v_1$ and r_{ij} for all i,j, follows

$$r_{ij} \sim \text{Bern}(\eta).$$

Model Specification

Under the constraint $\|\Theta\|_2 \leq B$:

$$Y_1, \cdots, Y_n|\Theta \overset{iid}{\sim} \mathsf{N}_p(0, \Theta^{-1}).$$

$$\theta_{ij} \sim \eta \mathsf{DE}(0, v_1) + (1-\eta)\mathsf{DE}(0, v_0) \quad i < j$$

$$\theta_{ji} = \theta_{ji}$$

$$\theta_{ii} \sim \mathsf{Ex}(\tau)$$

Our target is the MAP estimate of Θ and the posterior inclusion probability of $r_{ij}|\cdot$.

Penalized Likelihood Perspective

This is equivalent to minimizing the following objective function under the **constraint** $\|\Theta\|_2 \leq B$ and $\Theta \succ 0$:

$$\begin{aligned} L(\Theta) &= -\log \pi(\Theta|Y_1, \cdots, Y_n) \\ &= -\ell(\Theta) - \sum_{i<j} \log \pi(\theta_{ij}|\eta) - \sum_{i} \log \pi(\theta_{ii}|\tau) + \text{Const.} \\ &= \frac{n}{2}\Big(\text{tr}(S\Theta) - \log\det(\Theta)\Big) + \sum_{i<j} \text{pen}_{SS}(\theta_{ij}) + \sum_{i} \text{pen}_1(\theta_{ii}) \end{aligned}$$

where $\text{pen}_1(\theta) = \tau|\theta|$.

Theorem

Local Convexity If $B \leq (2nv_0)^{\frac{1}{2}}$, *then* $\min_{\Theta\succ 0, \|\Theta\|_2 \leq B} L(\Theta)$ *is a strictly convex problem.*

Bayesian Regularization Function

The "signal" indicator r_{ij} can be treated as latent and integrate it out, then we get the Bayesian regularization function:

$$\begin{aligned} \text{pen}_{SS}(\theta_{ij}) &= -\log \int \pi(\theta_{ij}|r_{ij})\pi(r_{ij}|\eta)dr_{ij} \\ &= -\log\left[\left(\frac{\eta}{2v_1}\right)e^{-\frac{|\theta|}{v_1}} + \left(\frac{1-\eta}{2v_0}\right)e^{-\frac{|\theta|}{v_0}}\right] \end{aligned} \tag{2}$$

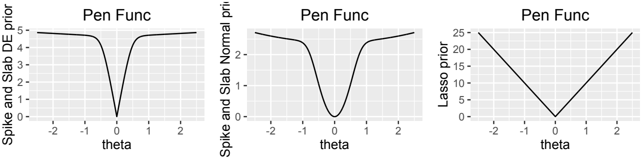


Table of Contents

Assumption

(A1) $\lambda_{\max}(\Theta^0) \leq 1/k_1 < \infty$ or equivalently $0 < k_1 \leq \lambda_{\min}(\Sigma^0)$.

(A2) The minimal "signal" entry satisfies $\min_{(i,j)\in S_g} |\theta_{ij}^0| \geq K_0\sqrt{\frac{\log p}{n}}$, where $K_0 > 0$ is a sufficiently large constant not depending on n.

Tail Conditions

(C1) Exponential tail condition: Suppose that there exists some $0 < \eta_1 < 1/4$ such that $\frac{\log p}{n} < \eta_1$ and

$$Ee^{tY^{(j)^2}} \leq K \text{ for all } |t| \leq \eta_1, \text{ for all } j = 1, \dots, p$$

where K is a bounded constant.

(C2) Polynomial tail condition: Suppose that for some γ, $c_1 > 0$, $p \leq c_1 n^\gamma$, and for some $\delta_0 > 0$,

$$E|Y^{(j)}|^{4\gamma+4+\delta_0} \leq K, \text{ for all } j = 1, \dots, p.$$

Rate of Convergence

Theorem (Estimation accuracy in entrywise ℓ_∞ norm)

Assume condition (A1) holds. For any pre-defined constants $C_3 > 0$, $\tau_0 > 0$, when the exponential tail (C1) or the polynomial tail (C2) condition holds. Assume that:

i) the prior hyper-parameters v_0, v_1, η, and τ are properly tuned;

ii) the spectral norm B satisfies $\frac{1}{k_1} + 2d(C_1 + C_3)K_{\Gamma^0}\sqrt{\frac{\log p}{n}} < B < (2nv_0)^{\frac{1}{2}}$,

iii) the sample size n satisfies $\sqrt{n} \geq M\sqrt{\log p}$,

where $M = \max\left\{2d(C_1 + C_3)K_{\Gamma^0}\max\left(3K_{\Sigma^0}, 3K_{\Gamma^0}K_{\Sigma}^{0\,3}, \frac{2}{k_1^2}\right), \frac{2C_3\varepsilon_1}{k_1^2}\right\}$.

Then, the MAP estimator $\tilde{\Theta}$ satisfies

$$\|\tilde{\Theta} - \Theta^0\|_\infty \leq 2(C_1 + C_3)K_{\Gamma^0}\sqrt{\frac{\log p}{n}}.$$

with probability greater than $1 - \delta_1$, where $\delta_1 = 2p^{-\tau_0}$ when condition (C1) holds, and $\delta_1 = O(n^{-\delta_0/8} + p^{-\tau_0/2})$ when condition (C2) holds.

Theorem (Selection consistency)

Assume the same conditions in Theorem 2 and condition (A2) with the following restriction:

$$\epsilon_0 < \frac{1}{\log p} \log\left(\frac{v_1(1-\eta)}{v_0\eta}\right) < (C_4 - C_3)\Big(K_0 - 2(C_1 + C_3)K_{\Gamma^0}\Big)$$

for some arbitrary small constant $\epsilon_0 > 0$. *Then, for any* T *such that* $0 < T < 1$, *we have*

$$\mathbb{P}\Big(\hat{S}_g = S_g\Big) \to 1.$$

Comparison with Existing Results

- **Graphical Lasso [Ravikumar et al., 2011]:**
  The irrepresentable condition, $|||\Gamma_{S_g^c S_g} \Gamma^{-1}_{S_g S_g}|||_\infty \leq 1 - \alpha$, is needed to establish the rate of convergence in entrywise ℓ_∞ norm. Such an assumption is quite restrictive, and is not needed for our results.
- Under the polynomial tail condition, the rate of convergence for Graphical Lasso is $O_p\left(\sqrt{\frac{p^c}{n}}\right)$, slower than our rate $O_p\left(\sqrt{\frac{\log p}{n}}\right)$.
- **CLIME [Cai et al., 2011] :**
  We assume boundedness of the largest eigenvalue of Θ^0, which is strictly weaker than the boundedness of $|||\Theta^0|||_\infty$ (the ℓ_∞/ℓ_∞ operator norm), the assumption imposed for CLIME.
- **Non-convex Penalties like SCAD, MCP [Loh and Wainwright, 2014] :**
  Beta-min condition (minimal signal strength) is needed for the rate of estimation accuracy established in [Loh and Wainwright, 2014]. In addition, their results are only available for sub-Gaussian distributions.

Table of Contents

EM Algorithm

- We treated R as latent and derive an EM algorithm to obtain a maximum a posterior (MAP) estimate of Θ in the M-step and the posterior distribution of R in the E-step.
- The updating scheme is in the similar fashion with [Friedman et al., 2008], i.e., updating one column and one row at a time. Without loss of generality, we describe the updating rule for the last column of Θ while fixing the others.

We list the following equalities from $W\Theta = \mathbf{I}_p$ which will be used in our algorithm:

$$\begin{bmatrix} W_{11} & w_{12} \\ \cdot & w_{22} \end{bmatrix} = \begin{bmatrix} \Theta_{11}^{-1} + \frac{\Theta_{11}^{-1}\theta_{12}\theta_{12}^T\Theta_{11}^{-1}}{\theta_{22}-\theta_{12}^T\Theta_{11}^{-1}\theta_{12}} & -\frac{\Theta_{11}^{-1}\theta_{12}}{\theta_{22}-\theta_{12}^T\Theta_{11}^{-1}\theta_{12}} \\ \cdot & \frac{1}{\theta_{22}-\theta_{12}^T\Theta_{11}^{-1}\theta_{12}} \end{bmatrix}. \tag{3}$$

M-step

Given Θ_{11}, to update the last column $(\theta_{12}, \theta_{22})$, we set the subgradient of Q with respect to $(\theta_{12}, \theta_{22})$ to zero. First take the subgradient of Q with respect to θ_{22}:

$$\frac{\partial Q}{\partial \theta_{22}} = \frac{n}{2} \frac{1}{\theta_{22} - \theta_{12}^T \Theta_{11}^{-1} \theta_{12}} - \frac{n}{2} (s_{22} + \tau) = 0. \qquad (4)$$

Due to Equations (3) and (4), we have

$$w_{22} = \frac{1}{\theta_{22} - \theta_{12}^T \Theta_{11}^{-1} \theta_{12}} = s_{22} + \frac{2}{n}\tau,$$

which leads to the following update for θ_{22}:

$$\theta_{22} \leftarrow \frac{1}{w_{22}} + \theta_{12}^T \Theta_{11}^{-1} \theta_{12}. \qquad (5)$$

M-step

Next take the subgradient of Q with respect to θ_{12}:

$$\begin{aligned}\frac{\partial Q}{\partial \theta_{12}} =& \frac{n}{2}\left(\frac{-2\Theta_{11}^{-1}\theta_{12}}{\theta_{22}-\theta_{12}^T\Theta_{11}^{-1}\theta_{12}} - 2s_{12}\right) - \left(\frac{1}{v_1}p_{12} + \frac{1}{v_0}(1-p_{12})\right) \odot \mathsf{sign}(\theta_{12}) \\ =& n(-\Theta_{11}^{-1}\theta_{12}w_{22} - s_{12}) - \left(\frac{1}{v_1}p_{12} + \frac{1}{v_0}(1-p_{12})\right) \odot \mathsf{sign}(\theta_{12}) = 0,\end{aligned} \tag{6}$$

where $A \odot B$ denotes the element-wise multiplication of two matrices. Here the second line of (6) is due to the identities in (3). To update θ_{12}, we then solve the following stationary equation for θ_{12} with coordinate descent, under the constraint $\|\Theta\|_2 \leq B$:

$$ns_{12} + nw_{22}\Theta_{11}^{-1}\theta_{12} + \left(\frac{1}{v_1}P_{12} + \frac{1}{v_0}(1-P_{12})\right) \odot \mathsf{sign}(\theta_{12}) = 0. \tag{7}$$

Algorithm 1 BAGUS

```
Initialize W = Θ = I
repeat
    Update P with each entry p_ij updated as log(p_ij/(1-p_ij)) ← ( log(v0/v1) + log(η/(1-η)) - |θ_ij^(t)|/v1 + |θ_ij^(t)|/v0 ).
    for j in 1 : p do
        Move the j-th column and j-th row to the end (implicitly), namely Θ11 := Θ_{-j-j}, θ12 := θ_{-jj}, θ22 := θ_jj
        Update w22 using w22 ← s22 + (2/n) τ
        Update θ12 by solving (7) with Coordinate Descent for θ12.
        Update θ22 using θ22 ← 1/w22 + θ12^T Θ11^{-1} θ12.
        Update W using (3)
    end for
until Converge
Return Θ, P
```

Where step 3 reads: Update P with each entry p_{ij} updated as $\log \frac{p_{ij}}{1-p_{ij}} \leftarrow \left(\log \frac{v_0}{v_1} + \log \frac{\eta}{1-\eta} - \frac{|\theta_{ij}^{(t)}|}{v_1} + \frac{|\theta_{ij}^{(t)}|}{v_0} \right)$.

Algorithm 2 Coordinate Descent for θ_{12}

```
Initialize θ12 from the previous iteration as the starting point.
repeat
    for j in 1 : (p - 1) do
        Solve the following equation for θ12_j:
    end for
until Converge or Max Iterations Reached.
If ||Θ||_2 > B : Return θ12 from the previous iteration
Else: Return θ12
```

The equation solved in step 4:

$$ns_{12j} + nw_{22}{\Theta_{11}^{-1}}_{j,\setminus j}\theta_{12\setminus j} + nw_{22}{\Theta_{11}^{-1}}_{j,j}\theta_{12j} + \left[\left(\frac{1}{v_1}P_{12} + \frac{1}{v_0}(1-P_{12})\right) \odot \text{sign}(\theta_{12})\right]_j = 0.$$

Our algorithm always ensures the symmetry and positive definiteness of the precision matrix estimation outputted.

Theorem (Positive Definiteness & Symmetry)

- *The estimate of Θ is always guaranteed to be symmetric.*
- *If $\Theta^{(0)} > 0$, i.e the initial estimate of precision matrix is positive definite,* $\Theta^{(t)} > 0$, $\forall t \geq 1$.

For the existing algorithms, the positive definiteness of the estimate usually doesn't hold [Mazumder and Hastie, 2012].

Table of Contents

Simulation Studies

1. Model 1: An $AR(1)$ model with $w_{ii} = 1$, $w_{i,i-1} = w_{i-1,i} = 0.5$
2. Model 2: An $AR(2)$ model $w_{ii} = 1$, $w_{i,i-1} = w_{i-1,i} = 0.5$ and $w_{i,i-2} = w_{i-2,i} = 0.25$.
3. Model 3: A circle model with $w_{ii} = 2$, $w_{i,i-1} = w_{i-1,i} = 1$, and $w_{1,p} = w_{p,1}$=0.9
4. Model 4: Random Select Model.

For each model, three scenarios will be considered: *Case* 1: $n = 100$, $p = 50$; *Case* 2: $n = 100, p = 100$; *Case* 3: $n = 100, p = 100$.

Metrics

Average Selection accuracy and L_2 distance between estimates and truths on 50 replications.

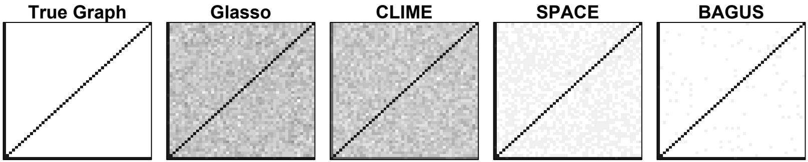


Figure: Average of the estimated precision matrices for the model with the star structure

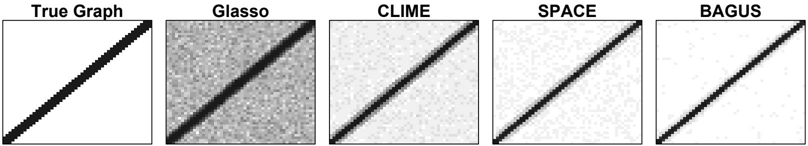


Figure: Average of the estimated precision matrices for the model with the AR(2) structure

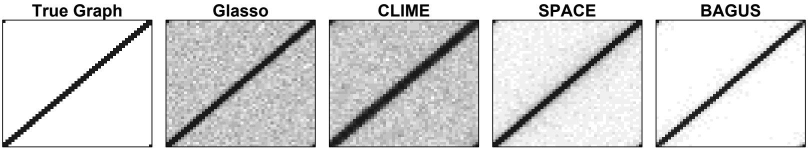


Figure: Average of the estimated precision matrices for the model with the **circle structure**

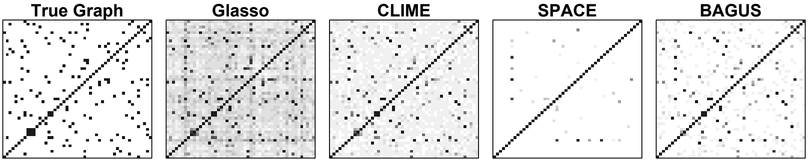


Figure: Average of the estimated precision matrices for the model with the **random structure**

Table: Model1 Star

	$n = 100, p = 50$			
	Fnorm	Specificity	Sensitivity	MCC
GLasso	2.301(0.126)	0.687(0.015)	0.998(0.004)	0.339(0.011)
CLIME	3.387(0.401)	0.452(0.051)	0.971(0.023)	0.168(0.021)
SPACE	2.978(0.244)	0.972(0.039)	1.000(0.003)	0.824(0.163)
BAGUS	**1.045(0.086)**	0.999(0.001)	1.000(0.003)	**0.989(0.012)**
	$n = 100, p = 100$			
	Fnorm	Specificity	Sensitivity	MCC
GLasso	4.219(0.118)	0.715(0.007)	0.989(0.008)	0.260(0.005)
CLIME	4.818(0.449)	0.998(0.004)	0.336(0.000)	0.131(0.067)
SPACE	3.207(0.311)	0.987(0.022)	0.996(0.024)	0.842(0.162)
BAGUS	**1.493(0.118)**	0.999(0.000)	0.938(0.026)	**0.954(0.017)**
	$n = 100, p = 200$			
	Fnorm	Specificity	Sensitivity	MCC
GLasso	3.028(0.068)	0.947(0.003)	0.999(0.002)	0.389(0.009)
CLIME	5.595(0.528)	0.978(0.018)	0.000(0.000)	-0.014(0.006)
SPACE	3.735(0.294)	0.985(0.007)	1.000(0.000)	0.656(0.138)
BAGUS	**2.009(0.092)**	1.000(0.000)	0.999(0.002)	**0.999(0.001)**

Table: Model 2: $AR(2)$

	$n = 100, p = 50$			
	Fnorm	Specificity	Sensitivity	MCC
GLasso	**3.361(0.240)**	0.479(0.056)	0.981(0.015)	0.251(0.028)
CLIME	3.758(0.381)	0.822(0.054)	0.906(0.039)	0.472(0.053)
SPACE	5.903(0.070)	0.982(0.004)	0.608(0.038)	0.656(0.029)
BAGUS	3.684(0.311)	0.997(0.002)	0.560(0.027)	**0.710(0.026)**
	$n = 100, p = 100$			
	Fnorm	Specificity	Sensitivity	MCC
GLasso	8.130(0.035)	0.901(0.007)	0.745(0.028)	0.382(0.017)
CLIME	5.595(1.578)	0.837(0.075)	0.821(0.191)	0.371(0.085)
SPACE	9.819(0.083)	0.991(0.002)	0.566(0.025)	0.625(0.021)
BAGUS	**5.501(0.080)**	0.998(0.001)	0.546(0.017)	**0.701(0.015)**
	$n = 100, p = 200$			
	Fnorm	Specificity	Sensitivity	MCC
GLasso	11.728(0.045)	0.990(0.001)	0.478(0.017)	0.481(0.014)
CLIME	11.552(0.382)	0.989(0.004)	0.580(0.031)	0.539(0.028)
SPACE	13.696(0.079)	0.995(0.000)	0.518(0.018)	0.588(0.013)
BAGUS	**8.026(0.084)**	0.999(0.000)	0.521(0.013)	**0.686(0.010)**

Table: Model 3: Circle

		$n = 100, p = 50$		
	Fnorm	Specificity	Sensitivity	MCC
GLasso	**4.319(0.174)**	0.492(0.064)	1.000(0.000)	0.196(0.024)
CLIME	5.785(0.440)	0.555(0.026)	1.000(0.000)	0.221(0.010)
SPACE	19.402(0.232)	0.930(0.006)	1.000(0.000)	0.595(0.019)
BAGUS	4.666(0.710)	0.991(0.004)	0.948(0.030)	**0.879(0.048)**
		$n = 100, p = 100$		
	Fnorm	Specificity	Sensitivity	MCC
GLasso	6.981(0.192)	0.647(0.005)	1.000(0.000)	0.189(0.002)
CLIME	19.282(2.802)	0.224(0.226)	0.995(0.015)	0.069(0.058)
SPACE	27.737(0.345)	0.975(0.010)	0.994(0.008)	0.674(0.062)
BAGUS	**5.954(0.582)**	0.996(0.001)	0.968(0.017)	**0.898(0.026)**
		$n = 100, p = 200$		
	Fnorm	Specificity	Sensitivity	MCC
GLasso	7.664(0.209)	0.752(0.003)	1.000(0.000)	0.172(0.001)
CLIME	33.009(0.535)	0.857(0.154)	0.769(0.167)	0.209(0.052)
SPACE	32.142(0.832)	0.981(0.012)	0.783(0.212)	0.485(0.129)
BAGUS	**7.103(0.953)**	0.997(0.00)	0.982(0.010)	**0.854(0.030)**

Table: Model4: Random Graph

		$n = 100, p = 50$		
	Fnorm	Specificity	Sensitivity	MCC
GLasso	7.017(0.256)	0.877(0.010)	0.766(0.039)	0.417(0.027)
CLIME	11.347(0.452)	0.971(0.012)	0.614(0.068)	0.572(0.042)
SPACE	12.278(0.183)	1.000(0.000)	0.073(0.031)	0.257(0.051)
BAGUS	**5.980(0.376)**	0.997(0.001)	0.451(0.051)	**0.629(0.040)**
		$n = 100, p = 100$		
	Fnorm	Specificity	Sensitivity	MCC
GLasso	11.851(0.900)	0.837(0.047)	0.720(0.049)	0.285(0.033)
CLIME	12.649(1.587)	0.735(0.153)	0.761(0.120)	0.243(0.123)
SPACE	17.706(0.203)	1.000(0.000)	0.068(0.015)	0.236(0.028)
BAGUS	**8.621(0.298)**	0.998(0.001)	0.382(0.032)	**0.565(0.029)**
		$n = 100, p = 200$		
	Fnorm	Specificity	Sensitivity	MCC
GLasso	15.054(0.356)	0.951(0.012)	0.633(0.029)	0.307(0.017)
CLIME	23.568(0.954)	0.993(0.004)	0.469(0.048)	0.492(0.038)
SPACE	24.997(0.213)	0.999(0.000)	0.090(0.014)	0.221(0.024)
BAGUS	**12.778(0.306)**	0.999(0.000)	0.392(0.017)	**0.583(0.018)**

Telephone call center arrival data prediction

- Forecast the call arrival pattern from one call center in a major U.S. northeastern financial organization.
- The training set contains data for the first 205 days. The remaining 34 days are used for testing.
- In the testing set, the first 51 intervals are assumed observed and we will predict the last 51 intervals, using the following relationship:

$$f(Y_{2i}|Y_{1i}) = \mathsf{N}(u_2 - \Theta_{22}^{-1}\Theta_{21}(Y_{1i} - u_1), \Theta_{22}^{-1})$$

Error Metric

To evaluate the prediction performance, we used the same criteria as [Fan et al., 2009], the average absolute forecast error (AAFE):

$$\mathsf{AAFE}_t = \frac{1}{34}\sum_{i=206}^{239} |\hat{y}_{it} - y_{it}|$$

where $\hat{y}_{it}$ and y_{it} are the predicted and observed values.

Telephone call center arrival data

From the results shown, our method has shown a significant improvement in prediction accuracy when compared with existing methods.

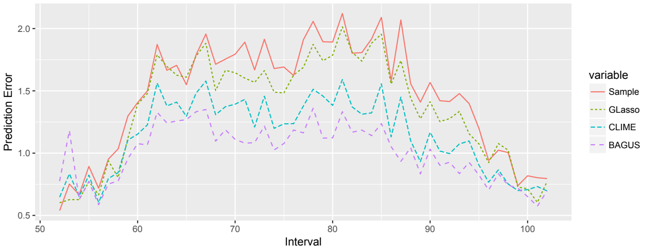


Figure: Prediction Error

Average Prediction Error						
	Sample	GLasso	Adaptive Lasso	SCAD	CLIME	BAGUS
Average AAFE	1.46	1.38	1.34	1.31	1.14	**1.00**

Summary

1. We propose a Bayesian model, using Spike and Slab Prior, for Gaussian Graphical Model.
2. An EM algorithm is derived to achieve the fast computation.
3. Simultaneous estimation and selection consistency of our method is proved.
4. Empirical Studies have shown promising results.

References I

Banerjee, O., El Ghaoui, L., and d'Aspremont, A. (2008).
Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data.
The Journal of Machine Learning Research, 9:485–516.

Banerjee, S. and Ghosal, S. (2015).
Bayesian structure learning in graphical models.
Journal of Multivariate Analysis, 136:147–162.

Cai, T., Liu, W., and Luo, X. (2011).
A constrained ? 1 minimization approach to sparse precision matrix estimation.
Journal of the American Statistical Association, 106(494):594–607.

Dempster, A. P. (1972).
Covariance selection.
Biometrics, pages 157–175.

Fan, J., Feng, Y., and Wu, Y. (2009).
Network exploration via the adaptive lasso and scad penalties.
The Annals of Applied statistics, 3(2):521.

Friedman, J., Hastie, T., and Tibshirani, R. (2008).
Sparse inverse covariance estimation with the graphical lasso.
Biostatistics, 9(3):432–441.

Gan, L. and Liang, F. (2016).
A bayesian em algorithm for graphical model selection.

References II

George, E. I. and McCulloch, R. E. (1997).
Approaches for Bayesian variable selection.
Statistica sinica, pages 339–373.

Lam, C. and Fan, J. (2009).
Sparsistency and rates of convergence in large covariance matrix estimation.
Annals of Statistics, 37(6B):4254.

Loh, P.-L. and Wainwright, M. J. (2014).
Support recovery without incoherence: A case for nonconvex regularization.
arXiv preprint arXiv:1412.5632.

Mazumder, R. and Hastie, T. (2012).
The graphical lasso: New insights and alternatives.
Electronic journal of statistics, 6:2125.

Meinshausen, N. and Bühlmann, P. (2006).
High-dimensional graphs and variable selection with the lasso.
The Annals of Statistics, pages 1436–1462.

Peng, J., Wang, P., Zhou, N., and Zhu, J. (2009).
Partial correlation estimation by joint sparse regression models.
Journal of the American Statistical Association, 104(486):735–746.

Ravikumar, P., Wainwright, M. J., Raskutti, G., Yu, B., et al. (2011).
High-dimensional covariance estimation by minimizing ?1-penalized log-determinant divergence.
Electronic Journal of Statistics, 5:935–980.

References III

Rocková, V. (2015).
Bayesian estimation of sparse signals with a continuous spike-and-slab prior.
Submitted manuscript, pages 1–34.

Ročková, V. and George, E. I. (2014).
Emvs: The em approach to bayesian variable selection.
Journal of the American Statistical Association, 109(506):828–846.

Rothman, A. J., Bickel, P. J., Levina, E., Zhu, J., et al. (2008).
Sparse permutation invariant covariance estimation.
Electronic Journal of Statistics, 2:494–515.

Wang, H. (2012).
Bayesian graphical lasso models and efficient posterior computation.
Bayesian Analysis, 7(4):867–886.

Yuan, M. and Lin, Y. (2007).
Model selection and estimation in the gaussian graphical model.
Biometrika, 94(1):19–35.