Precision Matrix Estimation under the Horseshoe-like Prior–Penalty Dual

Ksheera Sagar K. N.**

Department of Statistics
Purdue University

collaborative work with Sayantan Banerjee (IIM-Indore), Jyotishka Datta (VTech) and Anindya Bhadra (Purdue). Preprint available at: **https://arxiv.org/abs/2104.10750

Outline

Introduction

Precision matrix estimation under $\mathcal{N}(0, \Omega_{p\times p})$ is a key for network estimation. It has major applications in:

- Gene Regulatory Networks (GRNs)
- Econometrics
- Neuroscience etc.

Say $X = (x_1, x_2, \ldots, x_p) \sim \mathcal{N}(0, \Omega_{p\times p})$, where $\Sigma^{-1} = \Omega = ((\omega_{ij}))$ then for any $i \in \{1, 2, \ldots, p\}$,

$$x_i | x_{-i} \sim \mathcal{N}\left(\sum_{j\in\{1,2,\ldots,p\}\setminus\{i\}} x_j \beta_{ij}, \quad \frac{1}{\omega_{ii}} \right)$$

$$\beta_{ij} = -\frac{\omega_{ij}}{\omega_{ii}}$$

If $\omega_{ij} = 0$ then $\beta_{ij} = 0$ and there is no 'edge' between $x_i,\ x_j$.

Introduction (cont.)

If we have n copies of $X = (x_1, x_2, \ldots, x_p)$ and $n << \binom{p}{2}$, we aim to construct the network with a 'penalized' estimate of Ω. Some important works in the area of penalized estimates are:

- Graphical LASSO (Friedman et al., 2008)
- Graphical SCAD (Fan et al., 2009)
- CLIME estimator (Cai et al., 2011)
- Bayesian Graphical LASSO (Wang, 2012)
- Spike and slab with double exponential priors (Gan et al., 2019)
- Graphical horseshoe (Li et al., 2019)
- ...

So why estimate Ω with one more prior (Horseshoe-like)?

Introduction (cont.)

Graphical horseshoe outperformed older methods like BGL, GSCAD and GLASSO etc. but the drawbacks were:

- posterior concentration properties were unknown and
- fast point estimation algorithms were unavailable, so it was not scalable.

Hence we resolve these two issues in the current work by proposing the *graphical horseshoe-like* prior, which is an extension of horseshoe-like prior (Bhadra et al., 2019) to Gaussian graphical models.

Horseshoe-like prior

Horseshoe prior has no closed form:

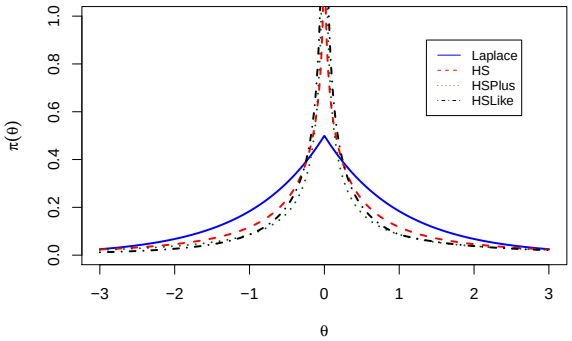
$$\omega_{ij} \mid \lambda_{ij}, \tau \sim \mathcal{N}\left(0, \lambda_{ij}^2 \tau^2\right), \ \pi(\lambda_{ij}^2) \sim \mathcal{C}^+(0,1), \ \pi(\tau^2) \sim \mathcal{C}^+(0,1).$$

Whereas horseshoe-like prior has, $\pi(\omega_{ij} \mid a) = (2\pi a^{1/2})^{-1} \log\left(1 + a/\omega_{ij}^2\right)$ and can be written with the hierarchy as follows:

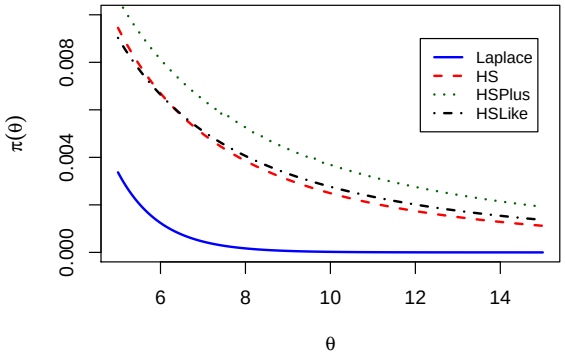
$$\omega_{ij} \mid \nu_{ij}, a \sim \mathcal{N}\left(0, \frac{a}{2\nu_{ij}}\right), \ \pi(\nu_{ij}) \sim \frac{1 - \exp(-\nu_{ij})}{2\pi^{1/2}\nu_{ij}^{3/2}}.$$

- It also closely approximates the horseshoe prior.
- With the hierarchy above, one can arrive at a point estimate of Ω using Expectation-Maximization type of algorithms.

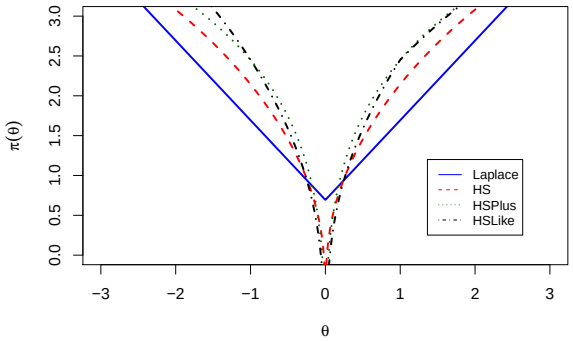
Horseshoe-like prior (cont.)



(a) Marginal prior densities near the origin



(b) Marginal prior densities in the tails



(c) Induced penalty functions

With the prior specification above, the log-posterior $\mathcal{L}$ thus becomes,

$$\mathcal{L} \propto \frac{n}{2}\log|\Omega| - \frac{n}{2}\text{tr}(S\Omega) + \sum_{i,j:i<j}\left\{\log\left(1-\exp(-\nu_{ij})\right) - \log\nu_{ij} - \frac{\nu_{ij}\omega_{ij}^2}{a}\right\}.$$

ECM and MCMC estimation

E Step: $\nu_{ij}^{(t)} = \mathbb{E}(\nu_{ij} \mid \omega_{ij}^{(t)}, a) = \left(\log\left(1 + \frac{a}{(\omega_{ij}^{(t)})^2}\right)\right)^{-1} \frac{a^2}{\left((\omega_{ij}^{(t)})^2 + a\right)\left((\omega_{ij}^{(t)})^2\right)}.$

CM Step: Use coordinate descent algorithm proposed by Wang (2014).

We compute a using the effective model size technique proposed by Piironen and Vehtari (2017) and keep it fixed throughout the ECM estimation.

MCMC Estimation: With a different hierarchy, $\omega_{ij} \mid t_{ij}, \tau \sim \mathcal{N}\left(0, \tau^2/t_{ij}^2\right)$, $\pi(t_{ij}, m_{ij}) = \frac{1}{2(2\pi)^{1/2}} \exp\left(\frac{-t_{ij}^2 m_{ij}}{2}\right) \mathcal{I}(0 < m_{ij} < 1),\ \tau^2 > 0,$ and following the remaining updates from the graphical horseshoe sampler of Li et al. (2019), the complete MCMC scheme for the graphical horseshoe-like is attained.

Theoretical properties

Assumption

(1) $p = n^b$, $b \in (0,1)$, *and* $(p+s)\log p/n = o(1)$.

Assumption

(2) The true precision matrix Ω_0 *belongs to the parameter space given by*

$$\mathcal{U}(\varepsilon_0, s) = \{\Omega \in \mathcal{M}_p^+ : \sum_{1 \le i < j \le p} \mathcal{I}(\omega_{ij} \neq 0) \le s,$$
$$0 < \varepsilon_0^{-1} \le \mathit{eig}_1(\Omega_0) \le \cdots \le \mathit{eig}_p(\Omega_0) \le \varepsilon_0 < \infty\}.$$

Assumption

(3) The bound $[L^{-1}, L]$ *on the eigenvalues of* Ω *satisfies* $L > \varepsilon_0$, *or, in other words,* $\varepsilon_0 = cL$, *for some* $c \in (0,1)$.

Theoretical properties (cont.)

Assumption

(4) The global shrinkage parameter a satisfies the condition, $a^{1/2} < n^{-1/2}p^{-b_1}$, for some constant $b_1 > 0$.

Theorem

Let $X^{(n)} = (X_1, \ldots, X_n)^T$ be a random sample from a p-dimensional normal distribution with mean 0 and covariance matrix $\Sigma_0 = \Omega_0^{-1}$, where $\Omega_0 \in \mathcal{U}(\varepsilon_0, s)$. Under prior specifications and assumptions on the prior as given in Assumptions (1)–(4), the posterior distribution of Ω satisfies

$$\mathbb{E}_0\left[P\{\|\Omega - \Omega_0\|_2 > M\epsilon_n \mid X^{(n)}\}\right] \to 0,$$

for $\epsilon_n = n^{-1/2}(p+s)^{1/2}(\log p)^{1/2}$ and a sufficiently large constant $M > 0$.

Theoretical properties (cont.)

Corollary

Under similar conditions as theorem above, the posterior distribution of Ω *has the posterior convergence rate* $\epsilon_n = n^{-1/2}(p+s)^{1/2}(\log p)^{1/2}$ *around* Ω_0 *with respect to the Frobenius norm under the graphical horseshoe prior.*

Theorem

Under similar conditions of theorem above, the MAP estimator of Ω*, given by* $\hat{\Omega}^{\mathrm{MAP}}$ *is consistent, in the sense that*

$$\|\hat{\Omega}^{\mathrm{MAP}} - \Omega_0\|_2 = O_P(\epsilon_n),$$

where $\epsilon_n = n^{-1/2}(p+s)^{1/2}(\log p)^{1/2}$.

Simulation results

Simulation Setting:

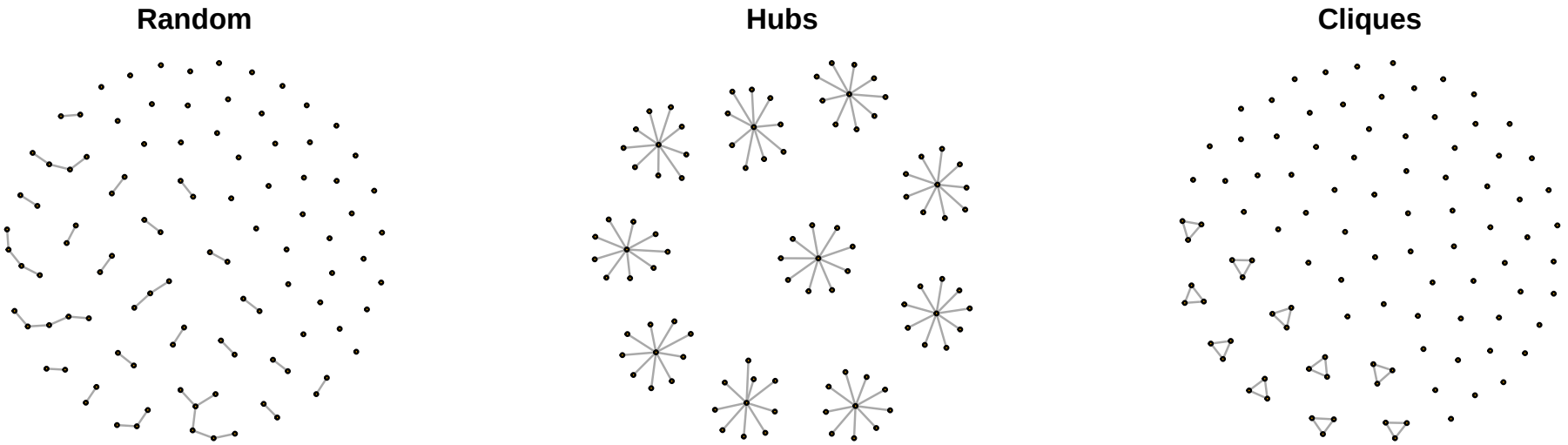


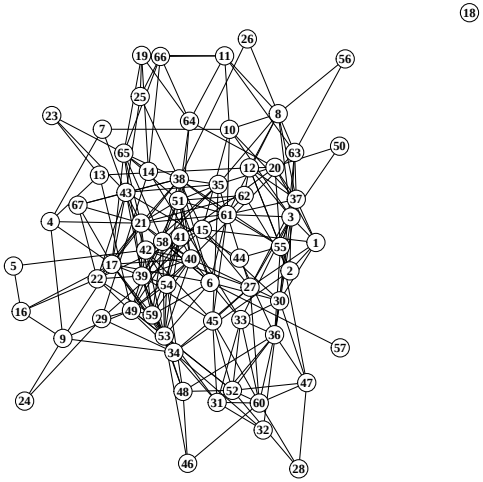
Figure: Visualization of true precision matrices under 3 different structures. Credits: Zhang et al. (2021).

We demonstrate results for $n = 120, p = 100$ for hub structure. Detailed simulation results for all other settings are available in our preprint: **https://arxiv.org/abs/2104.10750**

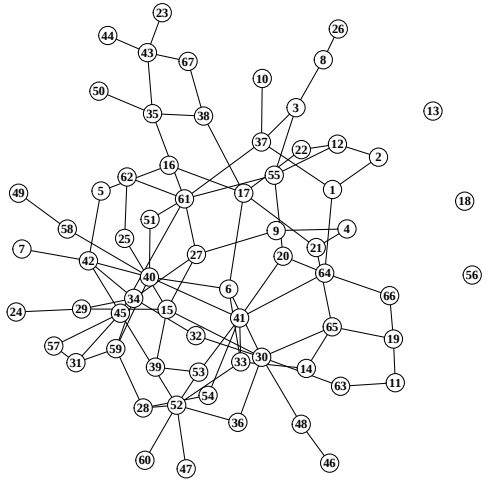
Simulation results (cont.)

	Hubs 90 nonzero pairs out of 4950 nonzero elements = 0.25						
	GL1	GL2	GSCAD	BGL	GHS	ECM	MCMC
Stein's loss	5.255	6.328	5.213	43.042	5.101	**4.22**	5.310
	(0.263)	(0.414)	(0.261)	(0.802)	(0.455)	(0.369)	(0.485)
F norm	3.018	3.432	3.003	4.295	2.544	**2.415**	2.687
	(0.091)	(0.112)	(0.093)	(0.156)	(0.126)	(0.103)	(0.141)
TPR	.995	.986	**.998**	.995	.872	0.985	0.754
	(.007)	(.017)	(.002)	(.008)	(.04)	(.014)	(0.004)
FPR	.101	.045	.983	.186	**.003**	.062	**0.003**
	(.016)	(.008)	(.012)	(.007)	(.001)	(0.005)	(0.001)
MCC	0.373	0.523	0.016	0.27	**0.85**	0.458	0.775
	(.027)	(.039)	(.006)	(.006)	(.027)	(.015)	(.033)
Avg CPU time	1.739	1.76	48.54	549.196	252.94	5.811	537.604

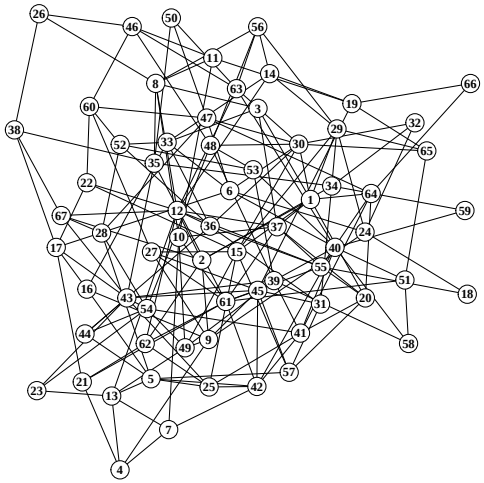
Next we analyze Reverse Phase Protein Array (RPPA) data of 33 patients with lymphoid neoplasm "Diffuse Large B-cell Lymphoma" to infer the protein interaction network.



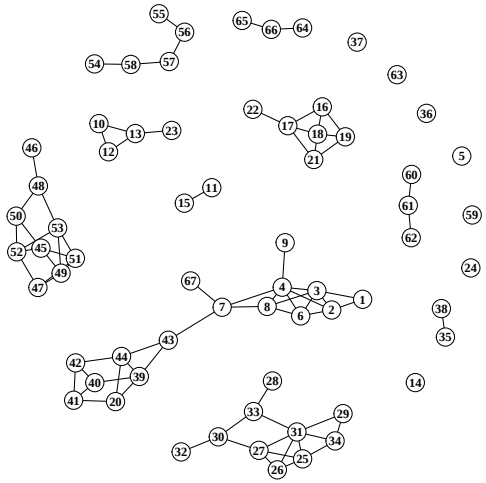
(a)



(b)



(c)



(d)

Figure: (a), (b), (c) and (d) correspond to RPPA networks for ECM, MCMC, GHS and PRECISE. The nodes are numbered from 1 to 67, which are proteins.

Conclusion and future scope

Contributions:

- Fully analytical prior-penalty dual termed the *graphical horseshoe-like*.
- First ever optimality results for both the frequentist point estimate as well as the fully Bayesian posterior for graphical horseshoe-like and for fully Bayesian posterior in graphical horseshoe.
- Simulation studies clearly establish that the family of horseshoe based priors perform the best among state-of-the-art competitors.

Future work:

- 2-means (Bhattacharya et al., 2015) or shrinkage factor thresholding (Tang et al., 2018) for variable selection in MCMC based methods.
- Establish Bayes risk under 0-1 loss.
- Extend to GLMs, e.g. graphical models with exponential families as node-conditional distributions (Yang et al., 2012).

References

Bhadra, A., Datta, J., Polson, N. G., and Willard, B. T. (2019). The horseshoe-like regularization for feature subset selection. *Sankhya B*, pages 1–30.

Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, 110(512):1479–1490.

Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Fan, J., Feng, Y., and Wu, Y. (2009). Network exploration via the adaptive lasso and SCAD penalties. *The Annals of Applied Statistics*, 3(2):521–541.

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

Gan, L., Narisetty, N. N., and Liang, F. (2019). Bayesian regularization for graphical models with unequal shrinkage. *Journal of the American Statistical Association*, 114(527):1218–1231.

Li, Y., Craig, B. A., and Bhadra, A. (2019). The graphical horseshoe estimator for inverse covariance matrices. *Journal of Computational and Graphical Statistics*, 28(3):747–757.

References (cont.)

Piironen, J. and Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2):5018–5051.

Tang, X., Xu, X., Ghosh, M., and Ghosh, P. (2018). Bayesian variable selection and estimation based on global-local shrinkage priors. *Sankhya A*, 80(2):215–246.

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

Wang, H. (2014). Coordinate descent algorithm for covariance graphical lasso. *Statistics and Computing*, 24(4):521–529.

Yang, E., Ravikumar, P., Allen, G. I., and Liu, Z. (2012). Graphical models via generalized linear models. In *NIPS*, volume 25, pages 1367–1375.

Zhang, R., Yao, Y., and Ghosh, M. (2021). Contraction of a quasi-bayesian model with shrinkage priors in precision matrix estimation.