SUPPLEMENT TO ADJUSTED REGULARIZATION IN LATENT GRAPHICAL MODELS: APPLICATION TO MULTIPLE-NEURON SPIKE COUNT DATA

By Giuseppe Vinci[*], Valérie Ventura[†,§]
Matthew A. Smith[‡,§], and Robert E. Kass[†,§]

Rice University[*], *Carnegie Mellon University*[†],
University of Pittsburgh[‡], *and Center for the Neural Basis of Cognition*[§]

APPENDIX A: MORE ON THE GENERAL SPIKE AND SLAB PRIOR

A.1. Special cases. Our general spike and slab prior in Equation 2.3 admits common models as special cases and several other special cases not explored in the literature. Assume that $\pi_{ij} = 1$ for all (i, j) unless specified otherwise. Then:

1. If G is a point-mass at $v = 1$, $p(\omega)$ in Equation 2.4 is the Gaussian distribution with mean m and variance $1/\lambda^2$, which produces ridge-type regularization (Hoerl and Kennard, 1970).
2. If $V \sim G$ is such that $V^2 \sim \Gamma(1, 1)$, then $p(\omega) \propto \exp\left\{-\sqrt{2}\lambda|\omega - m|\right\}$ is the Laplace distribution with location parameter m, which produces lasso-type regularization when $m = 0$ (Tibshirani, 1996; Yuan and Lin, 2007; Friedman, Hastie, and Tibshirani, 2008; Rothman et al., 2008; Ravikumar et al., 2011; Mazumder and Hastie, 2012; Wang, 2012; Vinci et al., 2018a).
3. If V is as in 2, and the matrix $M = [m_{ij}]$ in Equation 2.4 satisfies $-M = B \succeq 0$ with prior density $h(B) \propto \exp\{-\xi \mathrm{tr}(B)\} I(B \succeq 0)$, then we obtain the latent variable graphical model (Chandrasekaran et al., 2012; Yuan, 2012; Giraud and Tsybakov, 2012), which assumes $\Omega = \mathrm{Sparse} - \mathrm{LowRank}$, where the component $\mathrm{LowRank}$ quantifies the effect of unobserved variables (e.g. unrecorded neurons) of the network on the observed ones (see also Section 5 and end of Appendix C.1). To be specific, let Σ be the covariance matrix of $d+q$ variables, with $q < d$, and let $\Psi = \Sigma^{-1}$ be sparse. Thus, the conditional dependence between pairs of the first d variables given all variables is encoded by the submatrix Ψ_d. We have that $\Sigma_d^{-1} = \Psi_d - \Psi_{dq}\Psi_q^{-1}\Psi_{qd}$, where the second component has rank $\leq q < d$ (i.e. low-rank). Hence, the decomposition $\Sigma_d^{-1} = \mathrm{Sparse} - \mathrm{LowRank}$.
4. If $V^2 = (Z - 1)/(2bZ)$ with Z distributed as $\Gamma(1/2, a^2/4b)$ and truncated at $Z > 1$, then $p(\omega) \propto \exp\{-a|\omega| - b\omega^2\}$ (Lemma 4, Appendix B), which produces the elastic net (Zou and Hastie, 2005). Note that $a \to 0^+$ yields

the Gaussian prior with variance $(2b)^{-1}$, and $b \to 0^+$ the Laplace prior with variance $2/a^2$. See Figure 9.

5. If V is as in 2, and we let $\pi \in (0,1)$, then we obtain the spike-and-slab regularization (Banerjee and Ghosal, 2015; Wang, 2015).
6. A generalization of case 2 is given by $V = U^\gamma$, where $\gamma > 0$ and $U \sim \Gamma(\alpha,\beta)$, or equivalently $V \sim \text{Gen}\Gamma(a = \beta^{-\gamma}, b = \alpha\gamma^{-1}, c = \gamma^{-1})$ with p.d.f.

$$g_{ij}(v) = \frac{ca^{-b}}{\Gamma\left(\frac{b}{c}\right)} v^{b-1} e^{-\left(\frac{v}{a}\right)^c} \tag{A.1}$$

Case 2 corresponds to $\gamma = 1/2, \alpha = 1, \beta = 1$. If $\beta = (\Gamma(\alpha+2\gamma)/\Gamma(\alpha))^{\frac{1}{2\gamma}}$, then $\mathbb{E}[V^2] = 1$, and in Figure 9 we show the resulting Gaussian mixture for $\alpha = 1$ and $\gamma \in (0, 1.5]$: a larger γ increases the concentration of $p(\omega)$ in Equation 2.4 about the mean m.

7. If $V^2 \sim \text{Inv}\Gamma(\nu/2, \nu/2)$, then $p(\omega)$ is student-t with ν degrees of freedom (Gelman et al., 2004).
8. If G is a Stable distribution, then $p(\omega) \propto \exp\left\{-|\omega|^b\right\}$ for $b \in [1,2]$ (West, 1987).
9. If $V \sim \text{Cauchy}_+(0,1)$ (half-Cauchy), then $p(\omega)$ is the horseshoe distribution.

A.2. Utility of the general prior. To investigate the benefit of using the general prior on the estimation Ω, we simulated data (sample size $n = 500$; 20 repeats) from a 50-dimensional Gaussian distribution $N(0, \Omega^{-1})$ whose partial correlations varied with an auxiliary quantity $W \sim \text{Unif}(0,5)$, as depicted in Figure 10A. That is, partial correlations decrease with W on average, they concentrate further around their means as W increases, and they equal zero when $W \in (2,3)$.

We estimated the graphs using the following prior configurations:

- Model "g": $f(w) = 0$, $g(w)$ step function, and $\eta(w) = 1$;
- Model "(f,g)": $f(w)$ and $g(w)$ step functions, and $\eta(w) = 1$;
- Model "(g,η)": $f(w)$ constant, and $g(w)$ and $\eta(w)$ step functions;
- Model "(f,g,η)": $f(w)$, $g(w)$, and $\eta(w)$ step functions,

where the functions f, g, and η (Equations 2.5, 2.6, and 2.7) were taken to be step functions with five steps at the empirical quantiles of W. The models that fit a non-constant f should perform best because the average partial correlation decreases with W. We fitted these models, taking the Gaussian mixture component in Equation 2.3 to be either Gaussian, Laplace, Elastic net ($a = \sqrt{2}$ and $b = 1/2$, special case 4, Appendix A.1), or Generalized Gamma mixtures (Equation A.1, with

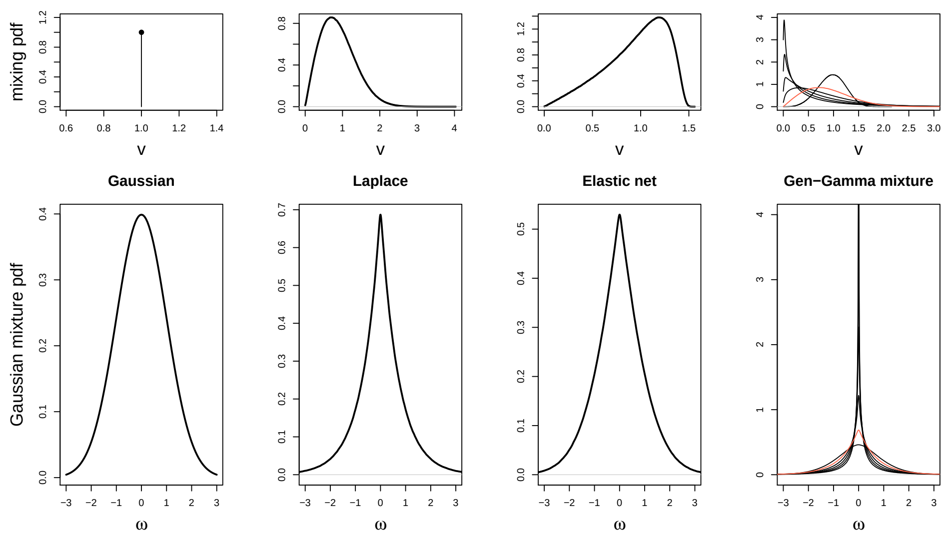

FIG 9. *Examples of Gaussian scale mixtures with mean $m = 0$. Top: mixing densities. Bottom: resulting mixture densities. The red density in the Generalized-Gamma case is the Laplace density.*

$\alpha = 1$, $\gamma = 1$, and $\beta = (\Gamma(\alpha + 2\gamma)/\Gamma(\alpha))^{\frac{1}{2\gamma}}$); see Figure 9. Figure 10A shows example posterior means of f, g, and η for these models, obtained by using a Gaussian prior p.d.f. component in Equation 2.3, and Figure 10B shows the ROC curves of unsigned posterior (p, δ) graphs with AUC and rescaled VUS (Section 3.5) in parenthesis. All Gaussian mixtures (Gaussian, Laplace, elastic net, and generalized Gamma) produced AUC and VUS within ± 0.03 of the Gaussian prior case, so we show only curves for the latter for clarity. We plan to compare the properties of the different Gaussian mixtures more fully in the future. We also fitted two models that do not incorporate auxiliary information:

- Model "Bglasso" (Bayesian glasso): $f(w) = 0$, $g(w)$ constant, and $\eta(w) = 1$;
- Model "Sp-Sl" (spike-and-slab): $f(w) = 0$, $g(w)$ constant, and $\eta(w)$ constant.

All estimators that used auxiliary information outperformed the others, with more precise signed edge detection (higher VUS) in the models that fitted f, as we expected.

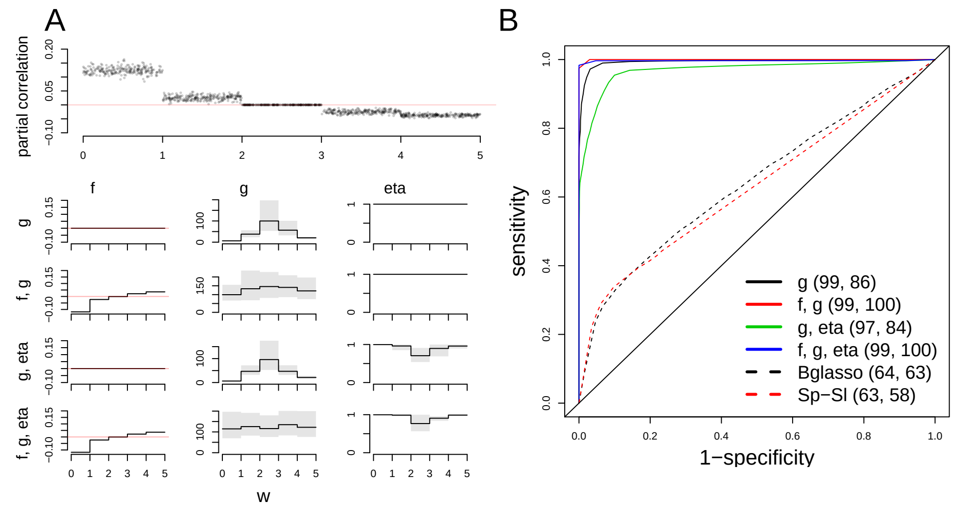

FIG 10. **(A)** *Top: partial correlations of the ground truth model. Bottom: posterior means of f, g, η (in columns) of models that use auxiliary information (in rows); e.g. the first row is model "g" with $f(w) = 0$, $g(w)$ a step function, and $\eta(w) = 1$.* **(B)** *ROC curves of the fitted models with average AUC and rescaled VUS in parenthesis.*

APPENDIX B: LEMMAS

LEMMA 1. *Let*

$$p(\omega) = \int_0^\infty \frac{\lambda}{v}\phi\left(\frac{\lambda}{v}(\omega - m)\right) dG(v),$$

where ϕ is the p.d.f. of $N(0,1)$ and $V \sim G$ is a positive random variable. Then, $p(\omega)$ exists if and only if $\mathbb{E}[V^{-1}] < \infty$, with $\mathbb{E}[\omega] = m$ and $\mathrm{Var}(\omega) = \mathbb{E}[V^2]/\lambda^2$.

PROOF. For $\lambda > 0$ and $\forall \omega \in \mathbb{R}$

$$\begin{aligned} p(\omega) &\leq p(m) \\ &= \frac{\lambda}{\sqrt{2\pi}} \int_0^\infty v^{-1}\, dG(v) \\ &= \frac{\lambda}{\sqrt{2\pi}} \mathbb{E}[V^{-1}] \end{aligned}$$

Finally note that $\omega \overset{D}{=} m + ZV/\lambda$, where $Z \sim N(0,1)$, so that

$$\begin{aligned} \mathbb{E}[\omega] &= m + \mathbb{E}\left\{\mathbb{E}\left[ZV/\lambda \mid V\right]\right\} \\ &= m + \mathbb{E}\left\{\mathbb{E}\left[Z \mid V\right] V/\lambda\right\} \\ &= m \end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}(\omega) &= \mathbb{E}\left[\mathrm{Var}\left(ZV/\lambda \mid V\right)\right] + \mathrm{Var}\left(\mathbb{E}\left[ZV/\lambda \mid V\right]\right) \\ &= \mathbb{E}\left[V^2/\lambda^2\right] + \mathrm{Var}\left(\mathbb{E}\left[Z \mid V\right] V/\lambda\right) \\ &= \mathbb{E}[V^2]/\lambda^2\end{aligned}$$

□

LEMMA 2. *In Equations 3.6 and 3.7,* $1 < A(\theta, W, Y, a) < \infty$ *and* $1 < B(\theta, W, Y, V, a) < \infty$.

PROOF. We have

$$\begin{aligned}B(\theta, W, Y, V, a)^{-1} &= \int_{\Omega \succ 0} \prod_{i \le j} \tau_{ij}\phi\left(\tau_{ij}(\omega_{ij} - Y_{ij}m_{ij})\right)\, d\Omega \\ &< \int_{\mathcal{S}} \prod_{i \le j} \tau_{ij}\phi\left(\tau_{ij}(\omega_{ij} - Y_{ij}m_{ij})\right)\, d\Omega \\ &= \prod_{i<j} \int_{\mathbb{R}} \tau_{ij}\phi\left(\tau_{ij}(\omega_{ij} - Y_{ij}m_{ij})\right)\, d\omega_{ij} \\ &\quad \times \prod_{i=1}^{d} \int_{\mathbb{R}_+} \tau_{ii}\phi\left(\tau_{ii}(\omega_{ii} - m_{ii})\right)\, d\omega_{ii} \\ &\le 1\end{aligned}$$

where $\mathcal{S} = \{\Omega \in \mathbb{R}^{d\times d} : \omega_{ij} = \omega_{ji},\ \omega_{ii} > 0\}$ is the set of symmetric matrices with positive diagonals. Thus, $B(\theta, W, Y, V, a) > 1$. The positive definite cone $\{\Omega \in \mathbb{R}^{d\times d} : \Omega \succ 0\} \subset \mathcal{S}$ is a non-empty convex subset of $\mathbb{R}^{d\times d}$, and therefore the integral over this set is strictly positive, i.e. $B(\theta, W, Y, V, a) < \infty$. Finally, note that

$$\begin{aligned}A(\theta, W, Y, a)^{-1} &= \int_{\mathbb{R}_+^{d(d+1)/2}} B(\theta, W, Y, V, a)^{-1} \prod_{i\le j} g_{ij}(v_{ij}) dv \\ &= \mathbb{E}_V[B(\theta, W, Y, V, a)^{-1}]\end{aligned}$$

where $v = \{v_{ij}\}_{i\le j}$ and $V \sim \prod_{i\le j} g_{ij}(v_{ij})$. Therefore $1 < A(\theta, W, Y, a) < \infty$, $\forall a > 0$.

□

LEMMA 3. *Equation 3.4 is satisfied by Equation 3.5.*

PROOF.

$$
\begin{aligned}
&\int h_a(\Omega \mid \theta, W, V)\, h_a(V \mid \theta, W, Y)\, h_a(\theta \mid W)\, dV \\
= &\prod_{i\le j}\left\{Y_{ij}\int_0^\infty \tfrac{\lambda_{ij}}{v}\phi\left(\tfrac{\lambda_{ij}}{v}(\omega_{ij}-m_{ij})\right) g_{ij}(v)\,dv + \ (1-Y_{ij})a^{-1}\phi\left(a^{-1}\omega_{ij}\right)\right\} \\
&\times I(\Omega \succ 0)\, h^*(\theta \mid W)
\end{aligned}
$$

where $Y_{ii}=1$ for all i, and

$$
\lim_{a\to 0^+} a^{-1}\phi\left(a^{-1}\omega_{ij}\right) = \delta(\omega_{ij}).
$$

Taking the expectation with respect to $Y \sim h(Y \mid \theta, W)$ in Equation 3.8 completes the proof.

□

LEMMA 4. *Elastic net. We have*

$$
\exp\{-a|x| - bx^2\} \propto \mathbb{E}\left[V^{-1}\phi\left(V^{-1}x\right)\right],
$$

where V has p.d.f.

$$
p(v) = C^{-1}v(1-2bv^2)^{-3/2}\exp\left\{-\tfrac{a^2/4b}{1-2bv^2}\right\} I(0<v<(2b)^{-1/2}) \tag{B.1}
$$

with $C = \frac{\sqrt{\pi}}{2a\sqrt{b}}[1-\mathrm{erf}(a/(2\sqrt{b}))]$ *and* $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$.

PROOF.

$$
\begin{aligned}
\exp\{-a|x| - bx^2\} &\propto \exp\{-a|x|\}\exp\{-bx^2\} \\
&\propto \mathbb{E}\left[\tfrac{1}{\sqrt{U}}\phi\left(\tfrac{x}{\sqrt{U}}\right)\right]\ \exp\{-bx^2\}
\end{aligned}
$$

where $U \sim \Gamma(1, a^2/2)$. Thus

$$
\begin{aligned}
\exp\{-a|x| - bx^2\} &\propto \mathbb{E}\left[\tfrac{1}{\sqrt{U}}\exp\left\{-\tfrac{x^2}{2U} - bx^2\right\}\right] \\
&\propto \mathbb{E}\left[\tfrac{1}{\sqrt{U}}\exp\left\{-\tfrac{x^2}{2U/(1+2bU)}\right\}\right] \\
&\propto \mathbb{E}\left[\tfrac{\sqrt{U/(1+2bU)}}{\sqrt{U}}\tfrac{1}{\sqrt{U/(1+2bU)}}\exp\left\{-\tfrac{x^2}{2U/(1+2bU)}\right\}\right] \\
&\propto \mathbb{E}\left[(1+2bU)^{-1/2}\tfrac{1}{\sqrt{U/(1+2bU)}}\phi\left(\tfrac{x}{\sqrt{U/(1+2bU)}}\right)\right] \\
&\propto \mathbb{E}\left[\tfrac{1}{V}\phi\left(\tfrac{x}{V}\right)\right],
\end{aligned}
$$

where $V = \sqrt{\frac{Z-1}{2bZ}}$ with $Z \sim \Gamma_1(1/2, a^2/4b)$, and V has distribution in Equation B.1.

□

LEMMA 5. *Consider Equations 3.18, 3.20, 3.21, 3.23, 3.24, and 3.26. We have* $\mathrm{FDP} \leq \mathrm{FDP}^*$, *which implies* $\mathrm{FDR} \leq \mathrm{FDR}^*$ *and* $\mathrm{FDR}_\Pi \leq \mathrm{FDR}^*_\Pi$.

PROOF.

$$\begin{aligned}
\mathrm{FDP}^* &\geq \frac{\sum_{i<j}\left\{I(\widehat{E}_{ij}>0)I(E_{ij}(\delta)=0)+I(\widehat{E}_{ij}<0)I(E_{ij}(\delta)=0)\right\}}{\sum_{i<j}|\widehat{E}_{ij}|} \\
&= \frac{\sum_{i<j} I(\widehat{E}_{ij}\neq 0)I(E_{ij}(\delta)=0)}{\sum_{i<j}|\widehat{E}_{ij}|} \\
&= \mathrm{FDP}
\end{aligned}$$

Applying monotonicity of expectations completes the proof. □

LEMMA 6. *Let* $f : \mathbb{R}^d \times \{0,1\} \to \mathbb{R}$, *and let* $(X, Y) \in \mathbb{R}^d \times \{0,1\}$ *have joint distribution* P. *Then*

$$f(x,y) = yf(x,1) + (1-y)f(x,0)$$

and

$$\mathbb{E}[f(X,Y)] = P(Y=1)\mathbb{E}[f(X,1)] + P(Y=0)\mathbb{E}[f(X,0)]$$

APPENDIX C: ALGORITHMS

Assume $V_{ii}^2 \sim \Gamma(1, 1/2)$ for all $i = 1, ..., d$, and integrate $(V_{11}, ..., V_{dd})$ out of Equation 3.3. Further assume $m_{ii} \leq 0, \forall i = 1, ..., d$; see comments in Section 3.1.

C.1. GGM estimation: Full Bayes.

In this section we describe the Gibbs sampling of $(\mu, \Omega, V, Y, \theta) \mid \mathbb{X}_n, W$ (Equation 3.3) with step functions f, g, and η (Equation 3.10).

C.1.1. $\mathbf{V} \mid$ **rest.** The conditional distribution of $V_{ij} \mid \mathrm{rest}$ is

$$h_a(V_{ij} \mid \mathrm{rest}) \propto \begin{cases} g_{ij}(V_{ij}), & \text{if } Y_{ij}=0 \\ g_{ij}(V_{ij})\frac{\lambda_{ij}}{V_{ij}}\phi\left(\frac{\lambda_{ij}}{V_{ij}}(\omega_{ij}-m_{ij})\right), & \text{if } Y_{ij}=1 \end{cases} \tag{C.1}$$

If it is easy to sample from g_{ij}, then the following rejection sampling algorithm can always be implemented for the case $Y_{ij} = 1$:

$$\text{(C.2)} \quad \begin{cases} 1.\ \text{Draw } \widetilde{V} \sim g_{ij} \\ 2.\ \text{Draw } \widetilde{U} \sim \text{Uniform}(0,1) \\ 3.\ \text{If } f(\widetilde{V})/f(v^*) > \widetilde{U},\ \text{then keep } \widetilde{V},\ \text{otherwise repeat } 1-2. \end{cases}$$

where $v^* = \lambda_{ij}|\omega_{ij} - m_{ij}|$ is the maximum point of $f(v) = \frac{\lambda_{ij}}{v}\phi\left(\frac{\lambda_{ij}}{v}(\omega_{ij} - m_{ij})\right)$ so that $f(\widetilde{V})/f(v^*) \leq 1$.

However, there exist some special cases where the conditional distribution of $V_{ij} \mid \text{rest}$ belongs to a known family of distributions allowing for a faster sampling.

Special case: Laplace prior. Assume $V_{ij}^2 = U$ where $U \sim \Gamma(1,1)$. Then

$$p(u \mid \text{rest}) \propto u^{-1/2} \exp\left\{-\frac{\lambda_{ij}^2}{2u}(\omega_{ij} - m_{ij})^2 - u\right\}$$

Now let $z = u^{-1} \Rightarrow du = z^{-2}dz$ so that

$$p(z \mid \text{rest}) \propto \left(\frac{1}{z^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{a(z-b)^2}{2b^2 z}\right\},$$

where $a = 2$ and $b = \sqrt{2}(\lambda_{ij}|\omega_{ij} - m_{ij}|)^{-1}$. Therefore, $V_{ij} \mid \text{rest} \overset{D}{=} Z^{-1/2}$, where $Z \sim \text{InvGaussian}(\sqrt{2}(\lambda_{ij}|\omega_{ij} - m_{ij}|)^{-1}, 2)$.

C.1.2. $\mathbf{Y} \mid \textbf{rest}$. We have

$$P(Y_{ij} = 1 \mid \text{rest}) = \pi_{ij}\frac{\lambda_{ij}}{V_{ij}}\phi\left(\frac{\lambda_{ij}}{V_{ij}}(\omega_{ij} - m_{ij})\right)/D,$$

where $D = \pi_{ij}\frac{\lambda_{ij}}{V_{ij}}\phi\left(\frac{\lambda_{ij}}{V_{ij}}(\omega_{ij} - m_{ij})\right) + (1 - \pi_{ij})a^{-1}\phi\left(a^{-1}\omega_{ij}\right)$.

C.1.3. $\mu \mid \textbf{rest}$. For simplicity we assume $h(\mu) \propto 1$, so that $\mu \mid \text{rest} \sim N(\bar{X}_n, \Omega^{-1}/n)$.

C.1.4. $\boldsymbol{\Omega} \mid \textbf{rest}$. The assumption $m_{ii} \leq 0$ implies $|\omega_{ii} - \nu_{ii}| = \omega_{ii} - \nu_{ii}$. Thus, distribution of $\Omega \mid \text{rest}$ is given by

$$\begin{aligned} h_a(\Omega \mid \text{rest}) \quad &\propto \quad (\det\Omega)^{\frac{n}{2}} \exp\{-\text{tr}(S\Omega)/2\} \\ &\times \prod_{i<j} \tau_{ij} e^{-\frac{\tau_{ij}^2}{2}(\omega_{ij} - \nu_{ij})^2} \times \prod_{i=1}^{d} \tau_{ii} e^{-\tau_{ii}\omega_{ii}} \times I(\Omega \succ 0) \end{aligned}$$

where $S = \sum_{r=1}^{n}(X_r - \mu)(X_r - \mu)'$, $\tau_{ij} = Y_{ij}\lambda_{ij}/V_{ij} + (1 - Y_{ij})/a$ for $i < j$ and $\tau_{ii} = \lambda_{ii}$ for $i = 1, ..., d$, and $\nu_{ij} = Y_{ij}m_{ij}$. We sample $\Omega \mid \text{rest}$ column-by-column guaranteeing its positive definiteness at each iteration (Wang, 2012; Vinci et al., 2018a). Partition the precision matrix as

$$\Omega = \begin{bmatrix} \Omega_{11} & \omega_{12} \\ \omega_{12}' & \omega_{22} \end{bmatrix}$$

where ω_{22} is a scalar, and partition the matrices S, $N = [\nu_{ij}]$ and $\mathcal{T} = [\tau_{ij}]$ similarly. By applying the Schur complement decomposition we obtain $\det(\Omega) = \det(\Omega_{11})(\omega_{22} - \omega_{12}'\Omega_{11}^{-1}\omega_{12})$, where $\det(\Omega_{11}) > 0$ and $(\omega_{22} - \omega_{12}'\Omega_{11}^{-1}\omega_{12}) > 0$ whenever $\Omega \succ 0$. Further define the diagonal matrix

$$D_{\mathcal{T}} = \text{diag}(\tau_{1d}^2, ..., \tau_{(i-1)d}^2, \tau_{(i+1)d}^2, ..., \tau_{dd}^2).$$

The conditional distribution of $\omega_{12}, \omega_{22} \mid \text{rest}$ is given by

$$\begin{aligned} & h(\omega_{12}, \omega_{22} \mid \text{rest}) \\ \propto \; & (\omega_{22} - \omega_{12}'\Omega_{11}^{-1}\omega_{12})^{\frac{n}{2}} \times I(\omega_{22} - \omega_{12}'\Omega_{11}^{-1}\omega_{12} > 0) \\ & \times \exp\left\{ -\frac{1}{2}\left[\omega_{12}' D_{\mathcal{T}}\omega_{12} + 2(s_{12}' - \nu_{12}' D_{\mathcal{T}})\omega_{12} + (s_{22} + 2\tau_{22})\omega_{22}\right] \right\} \end{aligned}$$

Let $\psi = \omega_{12}$ and $\zeta = \omega_{22} - \omega_{12}'\Omega_{11}^{-1}\omega_{12}$. The joint distribution of $(\psi, \zeta) \mid \text{rest}$ is

$$\begin{aligned} & h(\psi, \zeta \mid \text{rest}) \\ \propto \; & \zeta^{\frac{n}{2}} \exp\left\{ -\frac{1}{2}\left[\psi' D_{\mathcal{T}}\psi + 2(s_{12}' - m_{12}' D_{\mathcal{T}})\psi + (s_{22} + 2\tau_{22})(\zeta + \psi'\Omega_{11}^{-1}\psi)\right] \right\} \\ & \times I(\zeta > 0) \\ \propto \; & \zeta^{\frac{n}{2}} \exp\left\{ -\frac{s_{22} + 2\tau_{22}}{2}\zeta \right\} I(\zeta > 0) \times \exp\left\{ -\frac{1}{2}(\psi - \widetilde{\mu})'\widetilde{\Sigma}^{-1}(\psi - \widetilde{\mu}) \right\} \end{aligned}$$

that is $(\zeta, \psi) \mid \text{rest} \sim \Gamma\left(\frac{n}{2} + 1, \frac{s_{22} + 2\tau_{22}}{2}\right) N\left(\widetilde{\mu}, \widetilde{\Sigma}\right)$, where $\widetilde{\Sigma} = (D_{\mathcal{T}} + (s_{22} + 2\tau_{22})\Omega_{11}^{-1})^{-1}$ and $\widetilde{\mu} = -\widetilde{\Sigma}[s_{12}' - \nu_{12}' D_{\mathcal{T}}]$. Therefore, we update $\Omega \mid \text{rest}$ as follows: for $i = 1, ..., d$,

$$\left\{ \begin{array}{l} \text{1. Sample } \zeta \sim \Gamma\left(\frac{n}{2} + 1, \frac{S_{ii} + 2\lambda_{ii}}{2}\right) \text{ and } \psi \sim N\left(\widetilde{\mu}, \widetilde{\Sigma}\right), \\ \quad \text{where } \widetilde{\Sigma} = (D_{\mathcal{T}} + (S_{ii} + 2\lambda_{ii})\Omega_{-i-i}^{-1})^{-1}, \;\; \widetilde{\mu} = -\widetilde{\Sigma}[S_{-ii}' - N_{-ii}' D_{\mathcal{T}}], \\ \quad \text{and } \; D_{\mathcal{T}} = \text{diag}(\tau_{1i}^2, ..., \tau_{(i-1)i}^2, \tau_{(i+1)i}^2, ..., \tau_{di}^2). \\ \text{2. Set } \Omega_{-ii} = \Omega_{i-i}' := \psi \text{ and } \Omega_{ii} := \zeta + \psi'\Omega_{-i-i}^{-1}\psi. \end{array} \right.$$

C.1.5. $\theta \mid$ **rest**. Assume Equation 3.10 so that $\theta = \{\alpha, m, \lambda, \gamma, \beta, \chi\}$, and in Equation 3.1 assume

$$
\begin{aligned}
h^*(\theta \mid W) \quad \propto \quad & \prod_{i=1}^{d} \alpha_i^{r-1} e^{-s\alpha_i^2/2} \times \prod_{k=1}^{K_g} \beta_k^{r'-1} e^{-s'\beta_k^2} \times \prod_{k=1}^{K_\eta} \chi_k^{-0.5} (1-\chi_k)^{-0.5} \\
& \times \lambda_{ii}^{r''-1} e^{-s''\lambda_{ii}} \times h(m, \gamma \mid \alpha, W),
\end{aligned}
$$

where $h(m, \gamma \mid \alpha, W)$ has support such that $m_{ii} \leq 0$, for all $i = 1, ..., d$, as in Appendix C.1.4, and r, r', r'', s, s', s'' are hyperparameters. Thus, starting from Equation 3.5 we obtain

$$
\begin{aligned}
& h_a(\theta \mid \text{rest}) \quad \propto \quad h_a(\Omega, \theta, Y, V, \mid W) \\
\propto \quad & h^*(\theta \mid W) \times \prod_{ij:Y_{ij}=1} \alpha_i \alpha_j \left(\sum_{k=1}^{K_g} \beta_k I_{B_k}(W_{ij}) \right) V_{ij}^{-1} \\
& \times \prod_{ij:Y_{ij}=1} \phi \left(\alpha_i \alpha_j \left(\sum_{k=1}^{K_g} \beta_k I_{B_k}(W_{ij}) \right) V_{ij}^{-1} \left(\omega_{ij} - \frac{\sum_{k=1}^{K_f} \gamma_k I_{A_k}(W_{ij})}{\alpha_i \alpha_j} \right) \right) \\
& \times \prod_{i<j} \left(\sum_{k=1}^{K_\eta} \chi_k I_{C_k}(W_{ij}) \right)^{Y_{ij}} \left(1 - \sum_{k=1}^{K_\eta} \chi_k I_{C_k}(W_{ij}) \right)^{1-Y_{ij}} \times \prod_{i=1}^{d} \lambda_{ii} e^{-\lambda_{ii}\omega_{ii}}
\end{aligned}
$$

For notational convenience, let $g_{ij} = g(W_{ij})$, $f_{ij} = f(W_{ij})$, and $\eta_{ij} = \eta(W_{ij})$. We obtain the following conditional distributions:

1.

$$
h(\alpha_i \mid \text{rest}) \propto \alpha_i^{A_i - 1} e^{-\frac{B_i}{2}(\alpha_i - C_i)^2} I(\alpha_i > 0) h(m, \gamma \mid \alpha, W) \tag{C.3}
$$

where

$$
A_i = r + \sum_{j \neq i} Y_{ij}, \qquad B_i = s + \sum_{j \neq i} Y_{ij} V_{ij}^{-2} \alpha_j^2 \omega_{ij}^2 g_{ij}^2,
$$

and

$$
C_i \quad = \quad B_i^{-1} \sum_{j \neq i} Y_{ij} V_{ij}^{-2} \alpha_j \omega_{ij} g_{ij}^2 f_{ij}
$$

If $h(m, \gamma \mid \alpha, W)$ does not depend on α, we use rejection sampling to sample $\alpha_i \mid$ rest with proposal distribution $\alpha_i \sim N(C_i + D/B_i, B_i^{-1})$ truncated at $\alpha_i > 0$ and acceptance rule $\gamma(x; A_i, D)/\max_w \gamma(w, A_i, D) > U$, where

$U \sim \mathrm{Uniform}(0,1)$, $D > 0$ is a parameter chosen to minimize the distance between the Gaussian mean $C_i + D_i/B_i$ and the mode of the Gamma p.d.f. $\gamma(x; A_i, D) = \Gamma(A_i)^{-1} D^{A_i} x^{A_i - 1} e^{-Dx}$, ultimately increasing the acceptance probability.

2.

$$h(\lambda_{ii} \mid \text{rest}) \propto \lambda_{ii}^{r''} e^{-(s''+\omega_{ii})\lambda_{ii}} \tag{C.4}$$

that is, $\lambda_{ii} \mid \text{rest} \sim \Gamma(r''+1, s''+\omega_{ii})$.

3.

$$h(\beta_k \mid \text{rest}) \propto \beta_k^{A-1} e^{-B\beta_k^2} I(\beta_k > 0) \tag{C.5}$$

where

$$A = r'' + \sum_{i<j} I_{B_k}(W_{ij}) Y_{ij}$$

and

$$B = s'' + \sum_{i<j} I_{B_k}(W_{ij}) Y_{ij} V_{ij}^{-2} \left(\alpha_i \alpha_j \omega_{ij} - f_{ij}\right)^2 /2,$$

that is $\beta_k^2 \mid \text{rest} \sim \Gamma(A/2, B)$.

4.

$$h(\gamma_k \mid \text{rest}) \quad \propto \quad \exp\left(-\zeta_k (\gamma_k - \xi_k)^2/2\right) h(m, \gamma \mid \alpha, W) \tag{C.6}$$

where

$$\zeta_k = \sum_{i<j} I_{A_k}(W_{ij}) Y_{ij} g_{ij}^2 V_{ij}^{-2}$$

and

$$\xi_k = \zeta_k^{-1} \beta_k^2 \sum_{i<j} I_{A_k}(W_{ij}) Y_{ij} V_{ij}^{-2} \alpha_i \alpha_j \omega_{ij}$$

5.

$$\chi_k \mid \text{rest} \sim \mathrm{Beta}(0.5 + S_k, 0.5 + N_k - S_k) \tag{C.7}$$

where

$$S_k = \sum_{i<j} I_{C_k}(W_{ij}) Y_{ij}$$

and

$$N_k = \sum_{i<j} I_{C_k}(W_{ij})$$

6.

$$h(m \mid \text{rest}) \propto h(m, \gamma \mid \alpha, W) \tag{C.8}$$

Special cases.

- If $h(m, \gamma \mid \alpha, W) \propto \prod_{i=1}^{d} \delta(m_{ii}) \prod_{k=1}^{K_f} \phi(b_k(\gamma_k - a_k))$, $\lambda_{ii} \equiv \alpha_i^2$, $r'' = 1$, and $s'' = 0$, then A_i, B_i, and C_i in Equation C.3 are replaced by $\widetilde{A}_i = A_i + 2$, $\widetilde{B}_i = B_i + 2\omega_{ii}$, and $\widetilde{C}_i = \widetilde{B}_i^{-1} B_i C_i$, respectively, and Equation C.4 is not implemented; $\gamma_k \mid \text{rest} \sim N((\xi_k \zeta_k + a_k b_k)/(\zeta_k + b_k), (\zeta_k + b_k)^{-1})$, where $a_k = 0$ may be a standard choice; this is the configuration used in our data analyses.
- If $h(m, \gamma \mid \alpha, W) \propto \exp\{\kappa \sum_{i=1}^{d} m_{ii}\} I(-M \succeq 0)$, where we specify $m_{ij} \equiv (\alpha_i \alpha_j)^{-1} \sum_{k=1}^{K_f} \gamma_k I_{A_k}(W_{ij})$ for $i \neq j$, then we obtain a covariate dependent Bayesian sparse-low rank model.

C.2. GGM estimation: Empirical Bayes. We want to maximize

$$h(\theta \mid \mathbb{X}_n, W) = \int \int \int \sum_{Y \in \mathcal{Y}} h_a(\mu, \Omega, \theta, Y, V \mid \mathbb{X}_n, W) \, dV \, d\mu \, d\Omega$$

with respect to θ, where $h_a(\mu, \Omega, \theta, Y, V \mid \mathbb{X}_n, W)$ is defined in Equation 3.3. We use the following EM algorithm (Dempster, 1977).

E-STEP. Given the current estimate θ^{old}, we compute the expectation

$$\mathbb{E}\left[\log h_a(\mu, \Omega, \theta, Y, V \mid \mathbb{X}_n, W) \;\mid\; \mathbb{X}_n, W, \theta^{old}\right] \tag{C.9}$$

with respect to $(\mu, \Omega, Y, V) \mid \mathbb{X}_n, W, \theta^{old}$ and θ fixed. By applying Lemma 6 and simple algebra, we obtain that maximizing Equation C.9 with respect to θ is equivalent to maximize the function

$$\begin{aligned} \mathcal{Q}(\theta \mid \theta^{old}) \;&=\; \sum_{i<j} D_{ij} \left\{\log(\alpha_i \alpha_j g(W_{ij})) - \alpha_i^2 \alpha_j^2 g(W_{ij})^2 A_{ij}\right\} \\ &+ \sum_{i<j} D_{ij} \left\{-g(W_{ij})^2 f(W_{ij})^2 B_{ij} + \alpha_i \alpha_j g(W_{ij})^2 f(W_{ij}) C_{ij}\right\} \\ &+ \sum_{i<j} \left\{D_{ij} \log\left(\tfrac{\eta(W_{ij})}{1-\eta(W_{ij})}\right) + \log\left(1 - \eta(W_{ij})\right)\right\} \\ &+ \sum_{i=1}^{d} \{\log \lambda_{ii} - \lambda_{ii} E_i\} + \log h^*(\theta \mid W), \end{aligned} \tag{C.10}$$

where

$$A_{ij} = \mathbb{E}\left[\tfrac{\omega_{ij}^2}{2V_{ij}^2} \mid \mathbb{X}_n, W, \theta^{old}\right], \qquad B_{ij} = \mathbb{E}\left[\tfrac{1}{2V_{ij}^2} \mid \mathbb{X}_n, W, \theta^{old}\right],$$

$$C_{ij} = \mathbb{E}\left[\frac{\omega_{ij}}{V_{ij}^2} \mid \mathbb{X}_n, W, \theta^{old}\right], \qquad D_{ij} = \mathbb{E}\left[Y_{ij} \mid \mathbb{X}_n, W, \theta^{old}\right],$$

and $E_i = \mathbb{E}\left[\omega_{ii} \mid \mathbb{X}_n, W, \theta^{old}\right]$, which we approximate with the Gibbs sampler.

M-STEP. The function $\mathcal{Q}(\theta \mid \theta^{old})$ in Equation C.10 can be maximized by circularly optimizing it with respect to each component of θ, that is maximizing the following functions:

$$\mathcal{Q}_g(g) = \sum_{i<j} D_{ij}\left\{\log g(W_{ij}) - g(W_{ij})^2 Z_{ij}\right\} + \log h^*(\theta \mid W) \tag{C.11}$$

where $Z_{ij} = \alpha_i^2\alpha_j^2 A_{ij} + f(W_{ij})^2 B_{ij} - \alpha_i\alpha_j f(W_{ij}) C_{ij}$;

$$\mathcal{Q}_f(f) = -\sum_{i<j} R_{ij}(T_{ij} - f(W_{ij}))^2 + \log h^*(\theta \mid W) \tag{C.12}$$

where $R_{ij} = D_{ij} g(W_{ij})^2 B_{ij}$ and $T_{ij} = \frac{\alpha_i\alpha_j C_{ij}}{2B_{ij}}$;

$$\mathcal{Q}_\eta(\eta) = \sum_{i<j}\left\{D_{ij}\log\left(\frac{\eta(W_{ij})}{1-\eta(W_{ij})}\right) + \log(1-\eta(W_{ij}))\right\} + \log h^*(\theta \mid W), \tag{C.13}$$

$$\mathcal{Q}_m(m) = \log h^*(\theta \mid W) \tag{C.14}$$

$$\mathcal{Q}_\lambda(\lambda) = \sum_{i=1}^{d}\{\log\lambda_{ii} - \lambda_{ii}E_{ii}\} + \log h^*(\theta \mid W) \tag{C.15}$$

$$\begin{aligned}\mathcal{Q}_\alpha(\alpha) \;=\; & \sum_{i<j} D_{ij}\left\{\log(\alpha_i\alpha_j) - \alpha_i^2\alpha_j^2 g(W_{ij})^2 A_{ij} + \alpha_i\alpha_j g(W_{ij})^2 f(W_{ij}) C_{ij}\right\} \\ & + \log h^*(\theta \mid W) \qquad \text{(C.16)}\end{aligned}$$

However, if we further assume $\lambda_{ii} \equiv \alpha_i^2$, $\forall i$, then the optimization of Equations C.15 and C.16 is replaced by the maximization of

$$\begin{aligned}\mathcal{Q}^*_\alpha(\alpha) \;=\; & \sum_{i<j} D_{ij}\left\{\log(\alpha_i\alpha_j) - \alpha_i^2\alpha_j^2 g(W_{ij})^2 A_{ij} + \alpha_i\alpha_j g(W_{ij})^2 f(W_{ij}) C_{ij}\right\} \\ & + \sum_{i=1}^{d}\left\{2\log\alpha_i - \alpha_i^2 E_i\right\} + \log h^*(\theta \mid W) \qquad \text{(C.17)}\end{aligned}$$

The functions f, g, and η can all be fitted either as parametric or nonparametric regressions: Equation C.11 is a Gamma regression problem, Equation C.12 is a weighted least squares regression, and Equation C.13 is a logistic regression with proportions. If $h^*(\theta \mid W)$ does not depend on α, then the maximum point of Equation C.17 can be obtained by iteratively updating $\alpha_i := (b_i + \sqrt{b_i^2 + 4a_i c_i})(2a_i)^{-1}$ for $i = 1, ..., d$, where

$$a_i = 2E_i + 2\sum_{j\neq i} D_{ij}\alpha_j^2 g(W_{ij})^2 A_{ij}, \quad b_i = \sum_{j\neq i} D_{ij}\alpha_j g(W_{ij})^2 f(W_{ij}) C_{ij},$$

and $c_i = 2 + \sum_{j\neq i} D_{ij}$.

C.3. Poisson-lognormal estimation. The p.m.f. of a d-variate Poisson-lognormal random vector $Z^{(r)} = (Z_1^{(r)}, ..., Z_d^{(r)})$ in Equation 2.1 with parameter (μ, Ω) is given by

$$p_\theta(z) = \int_{\mathbb{R}^d} p(z \mid x)\phi(x \mid \mu, \Omega)dx$$

where

$$p(z \mid x) = \exp\left\{\sum_{i=1}^d (x_i z_i - e^{x_i})\right\} / \prod_{i=1}^d z_i!$$

and

$$\phi(x \mid \mu, \Omega) = \sqrt{\det(\Omega)/2\pi}\exp\left\{-\frac{1}{2}(x-\mu)'\Omega(x-\mu)\right\}.$$

Given n samples of count vectors $\mathbb{Z}_n = \{Z^{(1)}, ..., Z^{(n)}\}$, let $p(\mathbb{Z}_n \mid \mathbb{X}_n) = \prod_{r=1}^n p(Z^{(r)} \mid X^{(r)})$. We use the Gibbs sampler in Algorithm 1 to sample from the full joint posterior distribution

$$h(\mu, \Omega, \theta, Y, V, \mathbb{X}_n, \mid \mathbb{Z}_n, W) \propto h_a(\Omega, \theta, Y, V \mid W)L(\mu, \Omega; \mathbb{X}_n)p(\mathbb{Z}_n \mid \mathbb{X}_n) \tag{C.18}$$

where $h_a(\Omega, \theta, Y, V \mid W)$ is defined in Equation 3.5. We use method of moments estimates as starting values (Vinci et al., 2016). An Empirical Bayes approach is also possible (Algorithm 3).

To sample $X_{ir} \mid$ rest in step 1 of Algorithm 1, we use the rejection sampling scheme in Algorithm 2 to draw $R_{ir} = e^{X_{ir}} \mid$ rest whose distribution can be written as

$$p(R_{ir} \mid \text{rest}) \propto \gamma(R_{ir}; K+1, 1) \times LN(R_{ir}; \nu + (Z_{ir} - K)\omega_{ii}^{-1}, \omega_{ii}^{-1}), \tag{C.19}$$

for any $K \geq 0$. The parameter K is chosen to make the expectations of the components Gamma ($\gamma(x; a, b) = \Gamma(a)^{-1}b^a x^{a-1}e^{-bx}$) and log-normal ($LN(x; a, b) = e^{-(\log x - a)^2/(2b)}/(x\sqrt{b2\pi})$) in Equation C.19 close together, substantially increasing the acceptance probability.

Algorithm 1 *Poisson-lognormal - Full Bayes*

Input: Data $\mathbb{Z}_n$, starting values of $\mu, \Omega, \mathbb{X}_n$, and θ.
For $b = 1, ..., B$:
1. $X_{ir} \mid \text{rest} \sim p(x) \propto \exp\{Z_{ir}x - e^x\} \exp\{-\frac{\omega_{ii}}{2}(x-\nu)^2\}$,
where $\nu = \mu_i - \sum_{j\neq i}(X_{jr} - \mu_j)\omega_{ij}/\omega_{ii}$, for $i = 1, ..., d$, and $r = 1, ..., n$ via Algorithm 2.
2. Update θ, Y, V, μ, and Ω, given $\mathbb{X}_n$ according to Appendix C.1.
3. Set $\mu^{(b)} = \mu,\ \Omega^{(b)} = \Omega,\ \ \mathbb{X}_n^{(b)} = \mathbb{X}_n$, and $\theta^{(b)} = \theta$.

Output: Sequences $\{\mu^{(1)}, \Omega^{(1)}, \mathbb{X}_n^{(1)}, \theta^{(1)}\}, ..., \{\mu^{(B)}, \Omega^{(B)}, \mathbb{X}_n^{(B)}, \theta^{(B)}\}$.

Algorithm 2 *Rejection sampling for step 1 of Algorithm 1*

Input: ν, Z_{ir}, ω_{ii}.
1. $K^* = \underset{K\geq 0}{\arg\min} \left|\exp\{\nu + (Z_{ir} - K)\omega_{ii}^{-1} + 0.5\omega_{ii}^{-1}\} - (K+1)\right|$.
2. $\widetilde{R} \sim \log N(\nu + (Z_{ir} - K^*)\omega_{ii}^{-1}, \omega_{ii}^{-1})$
3. $U \sim \text{Uniform}(0, 1)$
4. If $\gamma(\widetilde{R}; K^*+1, 1)/\gamma(K^*; K^*+1, 1) > U$, then $\widetilde{R}$ is retained, otherwise repeat steps 2–3.

Output: $\widetilde{X}_{ir} = \log \widetilde{R}$.

C.4. Computational time. The Full Bayes algorithm (3500 iterations including 1000 burn-in period) for Gaussian data takes about 1.5 and 13 minutes for $d = 50$ and $d = 100$, respectively; correspondingly, the Empirical Bayes algorithm takes about 10 and 60 minutes to run 30 EM iterations to achieve convergence. For Poisson-lognormal data times have to be multiplied by about a factor of 1.5. The Empirical Bayes algorithm could be made more efficient by using some alternate faster approximation of the expectations in Equation C.10 in place of the Gibbs sampler. Computations were implemented using the programming language R, CPU Quad-core 2.6 GHz Intel Core i7, and RAM 16 GB 2133 MHz DDR4.

APPENDIX D: MODEL EDA AND DIAGNOSTICS

D.1. Empirical data analysis. Figure 5 shows the functions f, g, and η fitted to the data in condition $(\vartheta, \zeta) = (135, 1)$ – fitted models in the other conditions were similar: f and η are close to 0 and 1, respectively, but g increases with W, which means that the partial correlations were penalized more for neurons that are further apart. These findings are consistent with the plot of sample partial correlations of square-rooted spike counts, $\widetilde{\rho}_{ij}$, versus W in Figure 11: although this plot is very variable and the $\widetilde{\rho}_{ij}$ are not estimates of the log firing rate partial correlations, the values appear centered at zero and their spread seems to decrease with W. This is confirmed by the regression of $\widetilde{\rho}_{ij}$ on W, which is not significantly different than the zero constant, and by the Gamma regression of $\widetilde{\rho}_{ij}^2$ on W, which has a significantly negative slope ($p < 2\times 10^{-16}$). Because the $\widetilde{\rho}_{ij}$ are variable and

Algorithm 3 *Poisson-lognormal - Empirical Bayes*

Input: Data $\mathbb{Z}_n$, starting values of $\mu, \Omega, \mathbb{X}_n$, and θ.

1. E-STEP: Approximate $\mathcal{Q}(\theta \mid \theta^{\text{old}})$ in Equation C.10 by Gibbs sampler based on steps 1-2 of Algorithm 1 but with θ fixed equal to θ^{old}.
2. M-STEP: Maximize $\mathcal{Q}(\theta \mid \theta^{\text{old}})$ based on Equations C.11-C.17.
3. Iterate 1-2 until convergence.

Output: Estimate of θ.

thus do not equal zero even if $\pi_{ij} = 0$, the plot does not help determine if η, which modulates the true proportion of zero entries in Ω, varies with W.

Testing independence between V4 and PFC. For each experimental condition, we tested whether the square-rooted spike count partial correlations between V4 and PFC neurons were all zero, which is equivalent to testing whether their covariances are all equal to zero thanks to properties of block matrix inversion. We used the zero mean Gaussian Kullback-Leibler divergence (Cover and Thomas, 2012) as the test statistic:

$$KL(\widehat{\Sigma}_0 \| \widehat{\Sigma}) = \frac{1}{2}\{\text{tr}(\widehat{\Sigma}^{-1}\widehat{\Sigma}_0) - d + \log\det\widehat{\Sigma} - \log\det\widehat{\Sigma}_0\} \tag{D.1}$$

where $\widehat{\Sigma}$ is the sample covariance matrix of square-rooted spike counts and $\widehat{\Sigma}_0 = \widehat{\Sigma}$ but with all covariances between pairs of V4-PFC neurons set to zero. We approximated the null distribution of KL by bootstrapping trials for each neural area independently. The null hypothesis was rejected in all four conditions ($p < 0.001$), suggesting that V4 and PFC are not independent. This conclusion was confirmed by parametric bootstrap and asymptotic likelihood ratio test under the assumption that the square-rooted spike counts are Gaussian.

D.2. Model checking. In Figure 12 we compare the parametric (PLN) estimates of the spike-count means, variances, and correlations with their empirical counterparts. The good agreement between sample and model estimated quantities suggests an adequate model fit.

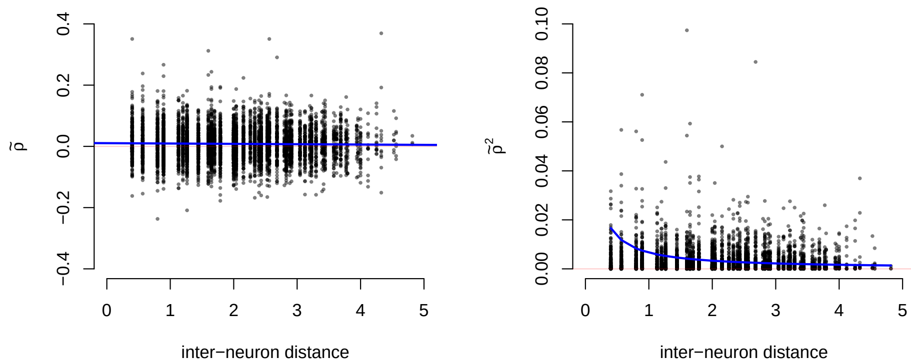

FIG 11. *Empirical data analysis: square root spike count sample partial correlation vs inter-neuron distance in one experimental condition. Values appear centered at zero and their spread seems to decrease with W, which suggests that $f \approx 0$ and g increases with W.*

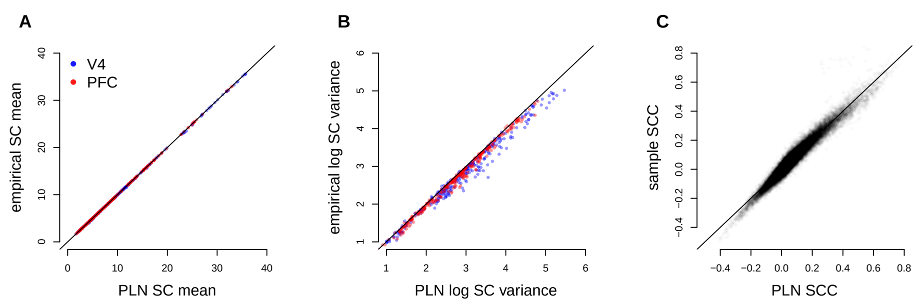

FIG 12. *Diagnostics. The PLN model fit implied spike count (SC) means in (A), variances in (B), and spike count correlations (SCC) in (C) agree with their empirical counterparts, suggesting an adequate model fit.*

D.3. Results for Section 4.1.

	SAM		Glasso		Aglasso	
	attend in	attend out	attend in	attend out	attend in	attend out
135° only	66	129	2255	1966	1822	2058
45° only	114	207	2309	2678	2066	2134
intersection	36	80	2010	2665	1137	1586
graph similarity	17%	19%	31%	36%	23%	27%

TABLE 2
Same as Table 1 but using sample partial correlations, Glasso, and Aglasso.

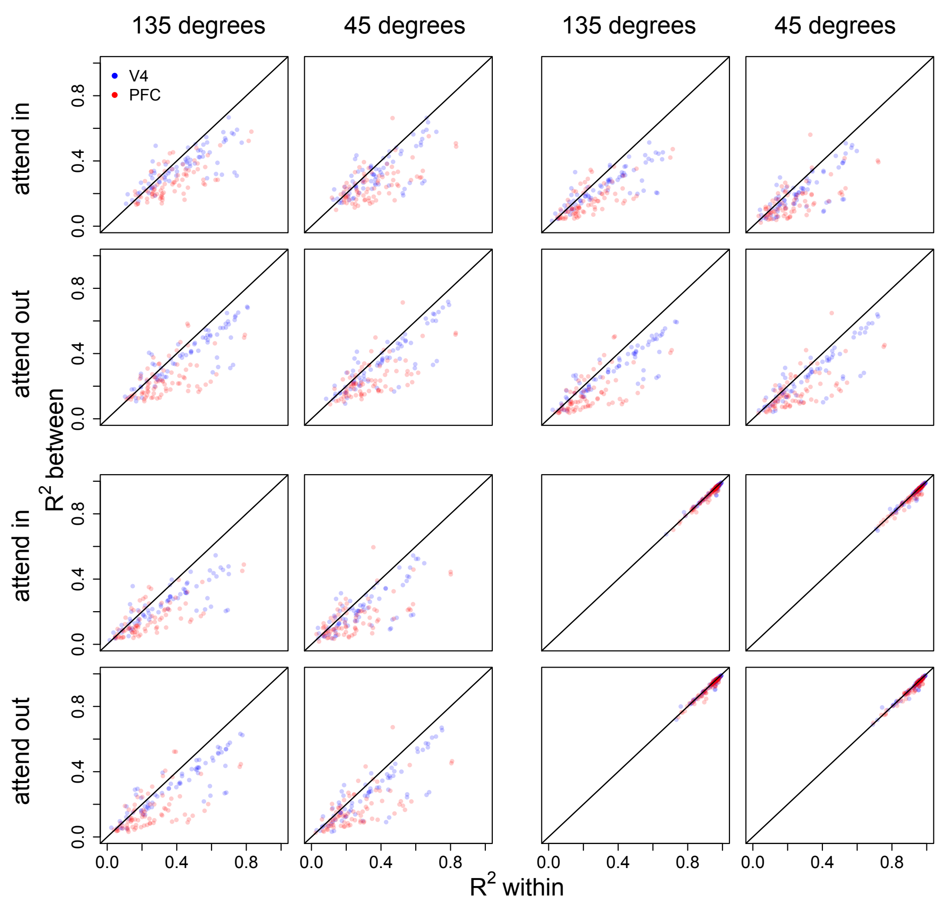

FIG 13. *Same analyses of Figure 7, but using sample covariances (top left), Glasso (top right), Aglasso (bottom left), and factor analysis (bottom right).*

REFERENCES

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

Chandrasekaran, V., Parrilo, P. A., and Willsky, A. S. (2012). Latent variable graphical model selection via convex optimization. Annals of Statics. Volume 40, Number 4 (2012), 1935-1967.

Cover, T. M., and Thomas, J. A. (2012). Elements of information theory. John Wiley and Sons.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society. Series B (methodological), 1-38.

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

Gelman, A., Carlin, J., Stern, H. S., and Rubin, D. B. (2004). Bayesian Data Analysis CRC Press. New York.

Giraud, C., and Tsybakov, A. (2012). Discussion: Latent variable graphical model selection via convex optimization. Annals of Statistics, 40(4), 1984–1988.

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

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

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. Electronic Journal of Statistics, 5, 935-980.

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.

Vinci, G., Ventura, V., Smith, M. A., and Kass, R. E. (2016). Separating spike count correlation from firing rate correlation. Neural computation.

Vinci, G., Ventura, V., Smith, M. A., and Kass, R. E. (2018a). Adjusted regularization of cortical covariance. *To appear in* Journal of Computational Neuroscience.

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

Wang, H. (2015). Scaling it up: Stochastic search structure learning in graphical models. Bayesian Analysis, 10(2), 351-377.

West, M. (1987). On scale mixtures of normal distributions. Biometrika, 1, 646–8.

Yuan, M., and Lin, Y. (2007). Model selection and estimation in the Gaussian graphical model. Biometrika, 19-35.

Yuan, M. (2012). Discussion: Latent variable graphical model selection via convex optimization. Annals of Statistics, 40(4), 1968–1972.

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.

RICE UNIVERSITY
DEPARTMENT OF STATISTICS
DUNCAN HALL
6100 MAIN ST
HOUSTON, 77005, TX, USA
E-MAIL: gv9@rice.edu

CARNEGIE MELLON UNIVERSITY
DEPARTMENT OF STATISTICS
BAKER HALL 132
5000 FORBES AVENUE
PITTSBURGH, 15203, PA, USA
E-MAIL: vventura@stat.cmu.edu
kass@stat.cmu.edu

University of Pittsburgh
Department of Ophthalmology
Eye and Ear Institute, Room 914
203 Lothrop St.
Pittsburgh, PA 15213, USA
E-mail: smithma@pitt.edu