PARTITIONING NEURAL CO-VARIABILITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Individual sensory neurons signal precise stimulus information despite high trial-to-trial variability.. Much of this variability is overdispersed — spike-count variance exceeds the mean — and can be partitioned into a stimulus-driven component and a gain that fluctuates across repeated presentations of the same stimulus. While the structure of neural activity, i.e., the stimulus-driven portion of the spike rate, has been explored widely in modeling population-level activity, the role of how modulatory gain factors are shared between neurons or changes over time has not been as thoroughly explored. Recovering the latent structure of the gains would offer a way to separate the stimulus drive from the shared gain that modulates a population over time. Most current overdispersion models thus treat each neuron's gain as independent across neurons and time. To capture the network-level statistics of the population, we thus present the Poisson matrix-normal latent variable model (PMNLV), which extends single-neuron overdispersion to neural populations and across trials, allowing the stimulus drive and the shared population gain to be estimated jointly rather than assumed independent. The model places a matrix-normal prior on the latent gain, and binned spike counts are drawn as Poisson from a rate that sums a per-neuron tuning term and the latent gain passed through a soft-rectifying link; its only structural assumption is a Kronecker factorization of the gain covariance into a neuron factor and a time factor. We derive a variational expectation-maximization algorithm that recovers both factors together with the tuning curves. On synthetic data the algorithm recovers the tuning curves and the covariance factors, and remains accurate at firing rates as low as 0.3 Hz. Compared against related population and overdispersion methods, PMNLV compares favorably in recovering the underlying covariance structure. Applied to Neuropixels recordings from mouse visual cortex, we find that the population's activity covariance and its estimated gain covariance are spectrally misaligned, indicating that the drivers of gain covariance may operate separately from the activity covariance. The PMNLV framework applies to any simultaneously recorded population in which structured gain covariance is of scientific interest, providing a tool for basic neuroscience that resolves population-level structure without compromising single-neuron statistics.

1 INTRODUCTION

Neural coding, or how neural responses change as a function of external stimuli or behavior, forms a fundamental language for how neuroscience quantifies information processing in the brain. A classical observation has been that beyond firing pattern changes in response to different stimuli (Zohary et al., 1994), neurons also exhibit significant variability between presentations of the same stimuli (Carandini, 2004). Modeling this trial-to-trial variability has proven important for creating better models of neural activity that appropriately capture the distribution of response observations (Keeley et al., 2020a; Churchland et al., 2011; Dichter et al., 2016; Lakshminarasimhan et al., 2023). Due its ubiquity, various approaches to quantifying population level changes range from identifying simple linear pairwise correlations (Ecker et al., 2010; Dichter et al., 2016) to flexible nonlinear manifold geometries (Koay et al., 2022; De & Chaudhuri, 2023) has become a centerpiece of data-driven neuroscience.

Current approaches that model neural response and their variability include the Fano factor (ratio of the response variance to the mean), doubly stochastic models, state-space models, and renewal processes that characterize variability via inter-spike-interval (ISI) statistics Goris et al. (2014); Charles

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

et al. (2018); Macke et al. (2011); Koyama (2015); Cunningham et al. (2007); Adams et al. (2009); Pillow et al. (2008); Rupasinghe et al. (2025); Gao et al. (2015); Cunningham et al. (2007); Yu et al. (2008). Within this body of work, special attention has been paid to modeling overdispersion, i.e., modeling additional sources of variability in a Poisson distribution resulting a greater variance to mean relationship Gur et al. (1997); Barberini et al. (2001); Friedenberger & Naud (2023). Prior work has shown that such models can capture changing response statistics within and across trials Goris et al. (2014); Charles et al. (2018). However, many of these models, impose an independence assumption between neurons and ignore population-level coordination in trial-to-trial gain fluctuations or identify a temporally restricted, low dimensional noise subspace without explicitly analyzing or relating this to spiking overdispersion or the population code Keeley et al. (2020a;b). As a result, while informative, such models are limited in their insight as to how these gain factors are shared throughout a network.

To bridge this gap, we propose a Poisson matrix-normal latent variable (PMNLV) model that relaxes the typical assumption taken in single-neuron overdispersion models where the residual gain fluctuations are statistically independent across neurons. In our model, spike counts are Poisson with a rate driven by the sum of a stimulus tuning term and a latent gain drawn from a matrix-normal prior whose Kronecker-factored covariance $\boldsymbol{\Sigma} = \boldsymbol{V} \otimes \boldsymbol{U}$ jointly encodes inter-neuron structure (row factor $\boldsymbol{U}$) and temporal structure (column factor $\boldsymbol{V}$). However, it is known that this prior leads to an intractable marginal likelihood (Charles et al., 2018; Wu et al., 2017; 2018; Keeley et al., 2020a;b). Although the covariance separability cannot express shared gain and correlation between neurons, it is readily interpretable and standard for estimating large datasets.

We demonstrate that by fitting our model with a modified variational expectation-maximization algorithm, we can better infer gain-level correlations between neurons and over time in controlled simulations. To demonstrate the applicability of our model to real data, we then apply VEM to Neuropixel recordings across four cortical regions. On this data, we first show that we can replicate previously published results that marginal single-neuron variability is near flat across the visual hierarchy. Next, we reveal that shared population co-variability peaks in primary visual area (VISp) and subsequently declines in higher visual areas. Finally, we find that the population's activity covariance and its estimated gain covariance are spectrally misaligned, indicating that the drivers of gain covariance may operate separately from the more typically studied activity covariance. These results illustrate how VEM can detect patterns in large neural data sets that are invisible to current models.

In summary, we present the following contributions:

- Extend prior single-unit overdispersion models to the population level.
- Derive a consistent and accurate fitting procedure based on variational expectation maximization (VEM).
- Demonstrate that our approach better identifies the gain statistics in a variety of simulated settings when compared to the current best models.
- Characterize gain-vs.-activity population statistics across the visual system.

2 Related Methods

Accurately modeling correlated gains is important, as they reflect large-scale activity in the population (Carandini, 2004), providing insight into neural processing beyond the recorded spikes. One related approach is to modify the classic Poisson spike count model by introducing a neuron-specific latent gain in addition to network drive which modulates the firing rate (Goris et al., 2014; Charles et al., 2018; Rupasinghe et al., 2025). Another related approach is to permit gain to come from a fixed kernel and explain the trajectory by a low-dimensional representation of neural trajectories (Lawrence, 2005; Santhanam et al., 2009; Yu et al., 2008; Keeley et al., 2020a; Wu et al., 2018) including population models such as GPFA and GP-LVM (Lawrence, 2005; Yu et al., 2008; Wu et al., 2018). While capturing smooth, shared variability across neuronal activity through a low-dimensional latent trajectory, such observation models are often Gaussian or assume independent noise. Other models do extend the latent activity across neurons and either time or conditions but require an independence assumption (Wu et al., 2018; Keeley et al., 2020a). To our knowledge,

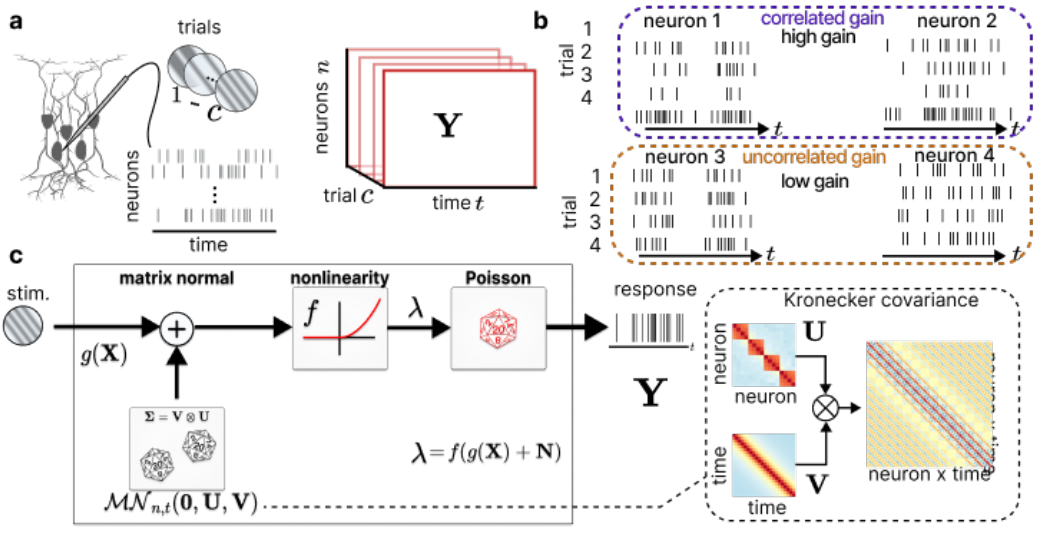


Figure 1: Poisson matrix-normal latent variable model. (a) (left) Data collection from N simultaneously recorded neurons, C experimental conditions, and T time steps yields a data tensor $\mathcal{Y}$. (right) A block-structured $NT \times NT$ Kronecker covariance matrix (right) decomposes into a neuron factor $\boldsymbol{U}$ (top left) capturing inter-neuron correlations and a time factor $\boldsymbol{V}$ (bottom left) capturing temporal covariance, which serves as the prior for the generative model. (b) Conceptualization of Kronecker covariance for two neurons across four example trials. (c) The stimulus-driven tuning term $f(\boldsymbol{X})$ and a zero-mean matrix-normal gain $\mathcal{N} \sim \mathcal{MN}(\boldsymbol{0}, \boldsymbol{U}, \boldsymbol{V})$ are summed and passed through the quadratic soft-rectifying link $g(\cdot)$ before Poisson spike generation.

PMNLV is the first population model of overdispersion to learn unrestricted gain covariances over both neurons and time,

3 Poisson Matrix Normal Latent Variable Model

We first specify the generative model and a solution using variational expectation maximization (VEM). We then study an optimal solution and modify our VEM algorithm resulting in two complementary estimation algorithms in Sections 3.2

Generative model: We consider simultaneously measured neuronal electrophysiological recordings from a population of neurons $n \in \{1, \ldots, N\}$ with binned times $t \in \{1, \ldots, T\}$ across multiple trials or conditions $c \in \{1, \ldots, C\}$. Each condition indexes an observed data matrix $\boldsymbol{Y}_c \in \mathbb{R}^{N \times T}$, which stacked across conditions produces a tensor $\mathcal{Y}$ of dimensions $N \times T \times C$. Following the single-neuron framework (Charles et al., 2018), we model each count as Poisson modulated additively by latent population gain $\boldsymbol{N}_c$

$$\boldsymbol{Y}_c \mid \boldsymbol{N}_c \sim \text{Poisson}\Big(g\big(f(\boldsymbol{X})_c + \boldsymbol{N}_c\big)\Big), \tag{3.1}$$

where $f(\boldsymbol{X})_c \in \mathbb{R}^{N \times T}$ is the matrix of stimulus tuning values for condition c and $g(\cdot) = \text{softplus}(\cdot)$ is a soft-rectifying link applied element-wise. The key departure from the single-neuron model is that the additive latent gain $\boldsymbol{N}_c \in \mathbb{R}^{N \times T}$ is no longer independent across neurons which allows for its co-variance structure to encode population co-variability across neurons and time. In particular, we model each $\boldsymbol{N}_c$ as drawn from a mean 0 matrix-normal distribution with row variance $\boldsymbol{U}$ and column variance V, i.e., $\boldsymbol{N}_c \sim \mathcal{MN}(\boldsymbol{0}, \boldsymbol{U}, \boldsymbol{V})$, separating neuronal and temporal structure into interpretable factors. Each latent gain $\boldsymbol{N}_c$ is drawn independently across conditions.

Marginal likelihood: Since the gain and tuning curves are unobserved, we estimate them by maximizing the marginal likelihood over $\Theta = \{f(\boldsymbol{X}), \boldsymbol{U}, \boldsymbol{V}\}$,

$$\text{p}(\boldsymbol{Y} \mid \Theta) = \int \text{p}(\boldsymbol{Y} \mid \boldsymbol{N}, \Theta)\ \text{p}(\boldsymbol{N} \mid \boldsymbol{U}, \boldsymbol{V})\ d\boldsymbol{N}. \tag{3.2}$$

The nonlinear link $g(\cdot)$ and the non-conjugacy between the Poisson likelihood and the matrix-normal prior, however, render this integral intractable. To resolve this, we first present an unconstrained VEM framework (§ 3.2) to recover factor estimates for $\boldsymbol{U}$ and $\boldsymbol{V}$.

3.1 MARGINALIZATION

As direct marginalization over $\boldsymbol{N}$ in equation 3.2 is intractable, we instead begin by maximizing a variational lower bound on $\log \mathrm{p}(\boldsymbol{Y} \mid \Theta)$. We approximate the posterior of each slice by $\mathrm{q}_\Psi(\boldsymbol{N}_c) = \mathcal{MN}_{N\times T}(\boldsymbol{M}_c, \boldsymbol{\Psi}_U, \boldsymbol{\Psi}_V)$, with variational parameters consisting of the means $\boldsymbol{M} = \{\boldsymbol{M}_c\}_{c=1}^C$ and row and column covariances $\boldsymbol{\Psi}_U \in \mathbb{R}^{N\times N}$ and $\boldsymbol{\Psi}_V \in \mathbb{R}^{T\times T}$, both positive definite and shared across slices.

$$\mathcal{L} = \mathbb{E}_{\mathrm{q}}\big[\log \mathrm{p}(\boldsymbol{Y} \mid \boldsymbol{N}, \boldsymbol{X})\big] - \mathrm{KL}\big(\mathrm{q}_\Psi(\boldsymbol{N}) \,\|\, \mathcal{MN}(\boldsymbol{0}, \boldsymbol{U}, \boldsymbol{V})\big) \tag{3.3}$$

$$= \mathbb{E}_{\mathrm{q}}\big[\log \mathrm{p}(\boldsymbol{Y} \mid \boldsymbol{N}, \boldsymbol{X})\big] - \frac{C}{2}\,\mathrm{tr}\Big(\boldsymbol{U}^{-1}\boldsymbol{\Psi}_U\Big)\,\mathrm{tr}\Big(\boldsymbol{V}^{-1}\boldsymbol{\Psi}_V\Big) + \frac{CN}{2}\log|\boldsymbol{\Psi}_V| + \text{const}, \tag{3.4}$$

where "const" refers to values that do not depend on the variatoinal parmaeter $\Psi_{\boldsymbol{V}}$ and under the matrix normal family. Because the Poisson likelihood factorizes over entries (Aitchison & Ho, 1989), it contributes only to the diagonal term of the posterior precision and all off-diagonal posterior structure is explained by the prior. However, a standard mean-field approximation, treating each variational observation as independent, discards exactly this structure (Blei et al., 2017). And since the ELBO penalizes $\mathrm{KL}(\mathrm{q}_\Psi \parallel \mathrm{p})$, the M-step is biased toward a diagonal matrix.

3.2 VARIATIONAL EXPECTATION-MAXIMIZATION (VEM)

We first approximated the posterior of each slice with a matrix-normal family whose row and column covariances are shared across slices,

$$q(\boldsymbol{Z}_c) = \mathcal{MN}_{N\times T}\big(\boldsymbol{M}_c, \boldsymbol{\Psi}_U, \boldsymbol{\Psi}_V\big), \tag{3.5}$$

so that $\mathrm{Cov}_q[\mathrm{vec}\,\boldsymbol{Z}_c] = \boldsymbol{\Psi}_V \otimes \boldsymbol{\Psi}_U$ inherits the Kronecker structure of the prior. Under this family the expected log-likelihood separates over entries with marginal variances $[\boldsymbol{\Psi}_U]_{nn}[\boldsymbol{\Psi}_V]_{tt}$. We evaluated the liklihood in the E-step by Gauss–Hermite quadrature while the M-step has the closed form

$$\widehat{\boldsymbol{U}} = \frac{1}{CT}\sum_{c=1}^{C} \boldsymbol{M}_c \boldsymbol{V}^{-1} \boldsymbol{M}_c^\top + \frac{\mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{\Psi}_V)}{T}\,\boldsymbol{\Psi}_U, \tag{3.6}$$

with the symmetric update for $\boldsymbol{V}$ to prevent collapse during maximization (Chiquet et al., 2019). Alternating these steps recovered the tuning curves and the neural factor well, however the estimated temporal factor was interestingly more spectrally peaked than the truth. To determine whether this was caused by the optimizer or the variational family, we examined the ELBO's dependence on $\boldsymbol{\Psi}_V$.

In equation 3.4, the expected log-likelihood depends on $\boldsymbol{\Psi}_V$ only through its diagonal, since each entry of $\boldsymbol{N}_c$ has marginal variance $[\boldsymbol{\Psi}_U]_{nn}[\boldsymbol{\Psi}_V]_{tt}$. For a fixed diagonal, $\log|\boldsymbol{\Psi}_V|$ is maximized when $\boldsymbol{\Psi}_V$ is diagonal by Hadamard's inequality Horn & Johnson (1985), yet the trace term is the only connection between $\boldsymbol{\Psi}_V$ and $\boldsymbol{V}$. As a consequence, the free variational posterior $\boldsymbol{\Psi}_V$ off-diagonal entries must be absorbed during the M-step. This raises the question of what the posterior covariance should be to prevent this absorption.

3.3 PRECISION CORRECTED VARIATIONAL EXPECTATION-MAXIMIZATION (VEM)

By characterizing the Gaussian posterior we can offer a correction to the objective to improve estimation of the prior factors. We proceed by first analyzing the critical points of the objective and then show that for the Poisson likelihood with a matrix-normal prior and any link satisfying Lemma 2, the resulting target covariance has Kronecker rank greater than one (Loan & Pitsianis, 1992).

We begin by characterizing the fixed points of the objective function in order to characterize the optimal solution posterior. As equation 3.2 has no closed-form, neither does the ELBO; however, the form of its optimal Gaussian posterior can be characterized. Proofs for the lemmas and theorem below are provided in the appendix D.

Algorithm 1 Free VEM

Require: $\{\boldsymbol{Y}_c\}_{c=1}^{C}, \boldsymbol{X}, \{\beta, \boldsymbol{M}, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\Psi}_U, \boldsymbol{\Psi}_V\}$
while not converged **do**

E-step
$\boldsymbol{\Sigma}_q \leftarrow \boldsymbol{\Psi}_V \otimes \boldsymbol{\Psi}_U$
$\mathrm{Var}[\eta_{nt}] \leftarrow [\boldsymbol{\Sigma}_q]_{(nt),(nt)}$
$(\boldsymbol{M}, \boldsymbol{\Psi}_U, \boldsymbol{\Psi}_V) \leftarrow \textsc{OptStep}(\mathcal{L})$

M-step
$\widehat{\boldsymbol{U}} \leftarrow \frac{1}{TC}\sum_c \boldsymbol{M}_c \boldsymbol{V}^{-1} \boldsymbol{M}_c^\top + \frac{1}{T}\sum_{tt'} [\boldsymbol{V}^{-1}]_{tt'} [\boldsymbol{\Sigma}_q]_{tt'}$
$\widehat{\boldsymbol{V}} \leftarrow \frac{1}{NC}\sum_c \boldsymbol{M}_c^\top \widehat{\boldsymbol{U}}^{-1} \boldsymbol{M}_c + \frac{1}{N}\sum_{nn'} [\boldsymbol{U}^{-1}]_{nn'} [\boldsymbol{\Sigma}_q]^{nn'}$
$s \leftarrow \mathrm{tr}(\widehat{\boldsymbol{U}})/N; (\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}) \leftarrow (\widehat{\boldsymbol{U}}/s, s\widehat{\boldsymbol{V}})$
end while

Algorithm 2 Precision Correction VEM

Require: $\{\boldsymbol{Y}_c\}_{c=1}^{C}, \boldsymbol{X}, \{\beta, \boldsymbol{M}, \boldsymbol{U}, \boldsymbol{V}\}$
while not converged **do**

E-step
$\boldsymbol{\Sigma}_q \leftarrow (\boldsymbol{V}^{-1} \otimes \boldsymbol{U}^{-1} + \gamma \boldsymbol{I})^{-1}$
$\mathrm{Var}[\eta_{nt}] \leftarrow [\boldsymbol{\Sigma}_q]_{(nt),(nt)}$
$(\boldsymbol{M}, \gamma) \leftarrow \textsc{OptStep}(\mathcal{L})$

M-step
$\widehat{\boldsymbol{U}} \leftarrow \frac{1}{TC}\sum_c \boldsymbol{M}_c \boldsymbol{V}^{-1} \boldsymbol{M}_c^\top + \frac{1}{T}\sum_{tt'} [\boldsymbol{V}^{-1}]_{tt'} [\boldsymbol{\Sigma}_q]_{tt'}$
$\widehat{\boldsymbol{V}} \leftarrow \frac{1}{NC}\sum_c \boldsymbol{M}_c^\top \widehat{\boldsymbol{U}}^{-1} \boldsymbol{M}_c + \frac{1}{N}\sum_{nn'} [\boldsymbol{U}^{-1}]_{nn'} [\boldsymbol{\Sigma}_q]^{nn'}$
$s \leftarrow \mathrm{tr}(\widehat{\boldsymbol{U}})/N; (\widehat{\boldsymbol{U}}, \widehat{\boldsymbol{V}}) \leftarrow (\widehat{\boldsymbol{U}}/s, s\widehat{\boldsymbol{V}})$
end while

Lemma 1 (Gaussian Posterior Critical Points)**.** *Let* $\mathrm{Diag}(\boldsymbol{D})$ *denote the* $NT \times NT$ *diagonal matrix with diagonal* $\mathrm{vec}(\boldsymbol{D})$*. Suppose* $\mathrm{vec}(\boldsymbol{N}_c) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\Sigma} = \boldsymbol{V} \otimes \boldsymbol{U} \succ 0$*, and* $p(\boldsymbol{Y} \mid \boldsymbol{\eta}) = \prod_{c,n,t} p(y_{ntc} \mid \eta_{ntc})$ *for* $\boldsymbol{\eta}_c = f(\boldsymbol{X}) + \boldsymbol{N}_c$*. For variational factors* $q_c = \mathcal{N}(\boldsymbol{m}_c, \boldsymbol{\Sigma}_q)$ *sharing* $\boldsymbol{\Sigma}_q \succ 0$*, every critical point of the ELBO satisfies*

$$\boldsymbol{\Sigma}_q = \left(\boldsymbol{\Sigma}^{-1} + \mathrm{Diag}(\boldsymbol{D})\right)^{-1}, \qquad D_{nt} = \frac{1}{C}\sum_{c=1}^{C} \mathbb{E}_{q_c}\left[-\partial_\eta^2 \log p(y_{ntc} \mid \eta_{ntc})\right]. \tag{3.7}$$

The optimal posterior precision is therefore the prior precision plus a diagonal. The next lemma guarantees that this diagonal is positive.

Lemma 2 (Positivity of the observed curvature)**.** *Let* $g : \mathbb{R} \to (0, \infty)$ *be at least* C^3*, log-concave, and satisfy* $g''(\eta) > 0$ *for all* $\eta \in \mathbb{R}$*. For Poisson observations with rate* $g(\eta)$*,*

$$\mathbf{J}(y, \eta) := -\partial_\eta^2 \log p(y \mid \eta) = -y(\log g)''(\eta) + g''(\eta) > 0, \qquad y \geq 0,\ \eta \in \mathbb{R}. \tag{3.8}$$

In particular, the softplus link satisfies these conditions. Even when $\boldsymbol{D}$ is constant, however, the target lies outside the Kronecker family.

Theorem 1 (Target not in family)**.** *Let* $\boldsymbol{U}, \boldsymbol{V} \succ 0$ *and both* $\boldsymbol{U}, \boldsymbol{V} \neq \alpha \mathbf{I}$*, for some* $\alpha \in \mathbb{R}^+$*. Let the link satisfy Lemma 2. Replace* $\boldsymbol{D}$ *in Lemma 1 by its average* $\gamma \mathbf{1}\mathbf{1}^\top$*, where* $\gamma = \frac{1}{NT}\sum_{n,t} D_{nt}$*. Then* $\gamma > 0$*, and* $(\boldsymbol{\Sigma}^{-1} + \gamma \boldsymbol{I})^{-1} \neq \boldsymbol{\Psi}_V \otimes \boldsymbol{\Psi}_U$ *for all* $\boldsymbol{\Psi}_U, \boldsymbol{\Psi}_V \succ 0$*.*

The averaged target $(\boldsymbol{\Sigma}^{-1} + \gamma \boldsymbol{I})^{-1}$ is therefore unreachable by the Kronecker family, and we adopt it directly as the variational family, treating γ as a free parameter of the ELBO. Algorithm 2 shows how to leverage our theoretical results into a modified estimation procedure that computes VEM with the derived precision correction (change shown in purple).

Remark. We arrive at Lemma 1 through the observed Fisher information of our objective, and it coincides with the fixed-point characterization of variational Gaussian approximations for Gaussian priors and factorizing likelihoods (Opper & Archambeau, 2009).

Initialization: The Kronecker factors are initialized by the square root transform of the spike counts after centering the tensor over the condition axis $\widetilde{\boldsymbol{Y}}_c \in \mathbb{R}^{N \times T}$. The C slices are then contracted over columns for $\boldsymbol{U}$ and over rows for $\boldsymbol{V}$,

$$\boldsymbol{U}^{(0)} = \frac{1}{CT}\sum_{c=1}^{C} \widetilde{\boldsymbol{Y}}_c \widetilde{\boldsymbol{Y}}_c^\top, \qquad \boldsymbol{V}^{(0)} = \frac{1}{CN}\sum_{c=1}^{C} \widetilde{\boldsymbol{Y}}_c^\top \widetilde{\boldsymbol{Y}}_c, \tag{3.9}$$

Similarly, we initialize the tuning curves by regressing $\widetilde{\boldsymbol{Y}}$ onto a four-term Fourier design matrix $\boldsymbol{X}$, and the variational mean $\boldsymbol{M}$ to the centered residual $\widetilde{\boldsymbol{Y}}_c$.

Table 1: VEM and PC-VEM Comparison

Scenario $(\mathbf{U}, \mathbf{V})$		**relative error** ε		**correlation** ρ	
		VEM	PC-VEM	VEM	PC-VEM
(LR,M)	$\widehat{\mathbf{U}}$	0.434 ± 0.005	$\mathbf{0.320 \pm 0.004}$	0.951 ± 0.002	$\mathbf{0.960 \pm 0.002}$
	$\widehat{\mathbf{V}}$	0.468 ± 0.006	$\mathbf{0.247 \pm 0.016}$	0.949 ± 0.005	$\mathbf{0.960 \pm 0.004}$
	$\hat{f}$	$\mathbf{0.065 \pm 0.001}$	0.067 ± 0.002	0.989 ± 0.000	0.989 ± 0.000

E-step: The Kronecker structure $\boldsymbol{V} \otimes \boldsymbol{U}$ of the prior is maintained, but the target posterior takes the form derived by theorem 1. Eigendecomposition of $\boldsymbol{U} = \boldsymbol{Q}_U \boldsymbol{\Lambda}_U \boldsymbol{Q}_U^\top$ and $\boldsymbol{V} = \boldsymbol{Q}_V \boldsymbol{\Lambda}_V \boldsymbol{Q}_V^\top$ diagonalizes the posterior covariance in the joint basis $\boldsymbol{Q}_V \otimes \boldsymbol{Q}_U$,

$$(\boldsymbol{Q}_V \otimes \boldsymbol{Q}_U)^\top \boldsymbol{\Sigma}_q (\boldsymbol{Q}_V \otimes \boldsymbol{Q}_U) \;=\; \boldsymbol{\Lambda}_\Sigma, \qquad [\boldsymbol{\Lambda}_\Sigma]_{ij} \;=\; \left(\frac{1}{\lambda_i^U \lambda_j^V} + \gamma \right)^{-1}, \tag{3.10}$$

so the posterior covariance reduces to a $N \times T$ array $\boldsymbol{\Lambda}_\Sigma$. It follows that the marginal variance of each η_{nt} is then,

$$\mathrm{Var}_{\mathrm{q}}[\eta_{nt}] \;=\; \sum_{i,j} [\boldsymbol{\Lambda}_\Sigma]_{ij} \, [\boldsymbol{Q}_U]_{ni}^2 \, [\boldsymbol{Q}_V]_{tj}^2, \tag{3.11}$$

the squared eigenvectors, without forming the $NT \times NT$ Kronecker product. Next, γ is computed by maximizing the ELBO, and the posterior mean is updated at that value.

M-step: Following the E-step, $\widehat{\boldsymbol{U}}$ and $\widehat{\boldsymbol{V}}$ are updated via flip-flop (Dutilleul, 1999), holding each Kronecker factor fixed while updating the other. Each factor requires the posterior covariance summed over the axis it does not describe. From the E-step and definition of trace, $\boldsymbol{\Lambda}_\Sigma$ provides the residual sum in closed form. By taking the sum over j against $\boldsymbol{\Lambda}_V^{-1}$ leaves the neuron axis, and similarly over i against $\boldsymbol{\Lambda}_U^{-1}$ leaves the time axis. Resulting in

$$\widehat{\boldsymbol{U}} \;=\; \frac{1}{TC} \sum_{c=1}^{C} \boldsymbol{M}_c \boldsymbol{V}^{-1} \boldsymbol{M}_c^\top \;+\; \frac{1}{T} \boldsymbol{Q}_U \, \mathrm{diag}\Big(\textstyle\sum_j [\boldsymbol{\Lambda}_\Sigma]_{ij} / \lambda_j^V \Big) \boldsymbol{Q}_U^\top, \tag{3.12}$$

with a symmetric update for $\widehat{\boldsymbol{V}}$ by swapping $1/T$ for $1/N$ and using the updated $\widehat{\boldsymbol{U}}^{-1}$ in the first term and similarly swapping the index from j to i sum in the second term.

4 Results on Synthetic data

We first assessed recovery of the Kronecker factors on simulated data sampled from the generative model with dimensions matched to real data N=45, T=12, C=75 (§5). The observations are first variance stabilized by the standard square-root transform (Yu et al., 2008; Bartlett, 1936). We assessed recovery by relative Frobenius error $\varepsilon_F(\hat{\boldsymbol{U}}) = \|\widehat{\boldsymbol{U}} - \boldsymbol{U}\|_F / \|\boldsymbol{U}\|_F$ as the primary metric and off-diagonal correlation $\rho(\widehat{\boldsymbol{U}})$ as a secondary metric of structural fidelity with hardware details in the appendix B.

4.1 Ground truth recovery

We evaluated VEM and PC-VEM in the four covariance regimes formed by independently assigning $\boldsymbol{U}$ and $\boldsymbol{V}$ a low-rank (LR) or Matérn (M) ground truth (Table 1). The two algorithms share data, initialization, and M-step, and differ only in the variational covariance with the all regimine comparisons in appendix E.1.2). We found that both VEM and its precision corrected version recovered the tuning curves accurately ($\varepsilon_F(\widehat{f}) = 6.0$–$6.4\%$, $\rho(\widehat{f}) \geq 0.986$ in every regime). What differentiates them is that PC-VEM reduced $\varepsilon_F(\widehat{\boldsymbol{U}})$ by 12–29% and $\varepsilon_F(\widehat{\boldsymbol{V}})$ by 39–65% in comparison to VEM and raised ρ for both factors in every regime. Because cortical populations are often low-dimensional, we show the (LR, M) regime in Fig. 2. There PC-VEM recovered $\boldsymbol{U}$ with $\varepsilon_F = 31.9\%$ and $\rho = 0.964$, and $\boldsymbol{V}$ with $\varepsilon_F = 26.1\%$ and $\rho = 0.972$. Recovery of $\boldsymbol{U}$ was harder whenever $\boldsymbol{V}$ was Matérn, with $\varepsilon_F(\widehat{\boldsymbol{U}})$ of 0.319 and 0.314 against 0.176 and 0.193 when $\boldsymbol{V}$ was low-rank.

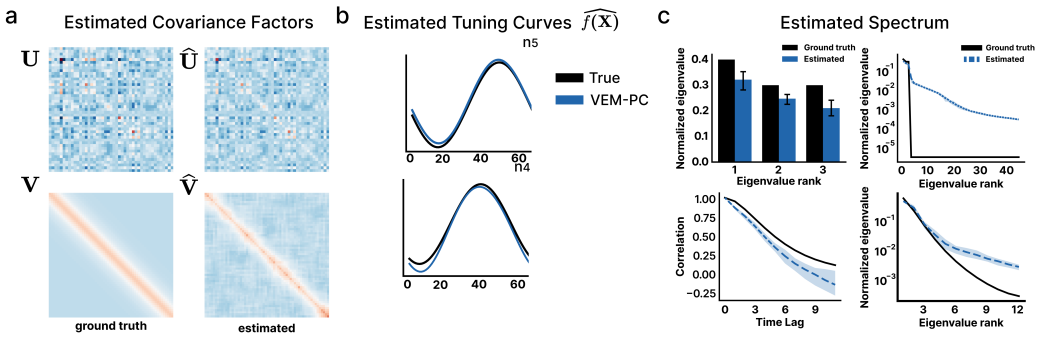


Figure 2: Covariance and tuning curve recovery in the (LR, M) regime. (a) Heatmaps of ground truth and VEM estimates of the neural factor $\boldsymbol{U}$ (top) and temporal factor $\boldsymbol{V}$ (bottom). b) Estimated tuning curves $\widehat{f}(\boldsymbol{X})$ for four example neurons in no order. (c) Low rank (LR) factor (top), Matérn factor (bottom). LR spectrum eigenvalues(top left, blue) eigenvalues of LR ground truth closely recovered black, LR estimated spectrum (top right, blue) falls off at at same truncation as ground truth (black), estimated lag correlation (bottom left) closely recovers ground truth lag (blue), Matérn spectrum (bottom right) shown trace normalized mean over 12 seeds; range shaded.

Table 2: Co-smoothing prediction

scenario	**predictive performance**					
	oracle	PC-VEM	GPFA	P-GPFA	SNP-GPFA	GP-LVM
(LR, LR)	+0.139	**+0.134 (0.96)**	+0.120 (0.87)	+0.107 (0.77)	+0.120 (0.86)	+0.126 (0.91)
(LR, M)	+0.149	**+0.144 (0.96)**	+0.125 (0.83)	+0.109 (0.73)	+0.140 (0.94)	+0.132 (0.88)
(M, M)	+0.154	**+0.147 (0.95)**	+0.118 (0.77)	+0.101 (0.66)	+0.138 (0.90)	+0.123 (0.80)

4.2 Model Comparison

Covariance recovery can only be measured against a known ground truth, and the competing models parameterize covariance differently. With the goal comparing different models, we use the co-smoothing metric to provide an empirical comparison across different models (Pei et al., 2022). The comparison proceeds by holding or leaving out a neuron and seeing how other neuronal activity is predicts the activity of the held out neuron. This predictive cross validation has been used perviously to compare these models (Keeley et al., 2020a; Wu et al., 2017; Macke et al., 2011). We simulated data with N=45, T=60, C=75 from the three matrix-normal regimes. Each with five fold hold outs of 20% of the condition slices, and on each held-out slice a random 25% of neurons. All parameters are learned on the training slices with every neuron observed and are not refit at test time. Performance is the held-out Poisson log-likelihood in bits per spike relative to a per-neuron mean-rate baseline, and we also report it as a fraction of an oracle that applies the same predictor at the true parameters.

PC-VEM attained 0.95–0.96 of the oracle across all regimes and scored higher than each baseline on each of the 15 seed–fold splits shared by all models (Table 2). The closest competitors were SNP-GPFA (0.86–0.94) and GP-LVM (0.80–0.91). While synthetic data generated from our model establishes that PC-VEM reliably recovers ground truth, this advantage also generalizes out-of-distribution: on data simulated from the signal-plus-noise framework of Keeley et al. (2020a), PC-VEM retained 0.91 of oracle performance compared to 0.89 for SNP-GPFA, outperforming it on 14 of 15 splits (Appendix E.1.2). As our model assumes separable covariance of the prior, we additionally performed the same cross validation test with data generated from PLDS (Macke et al., 2011). Our results show that PC-VEM still performed the other models but performing with a score of 0.97 relative to the oracle baseline. A complete table is available in the appendix E.1.2.

5 Results on Visual Coding Dataset

To demonstrate the utility of PMNLV for studying neural population dynamics, we next applied PC-VEM to Neuropixels recordings from the Allen Institute Visual Coding project. This dataset contains simultaneous recordings of single unit activity from distinct regions of mouse visual processing (Siegle et al., 2021). During recording, mice were head-fixed and free to run on a wheel while visual stimuli were presented. We focus on four consecutive regions along the visual hierar-

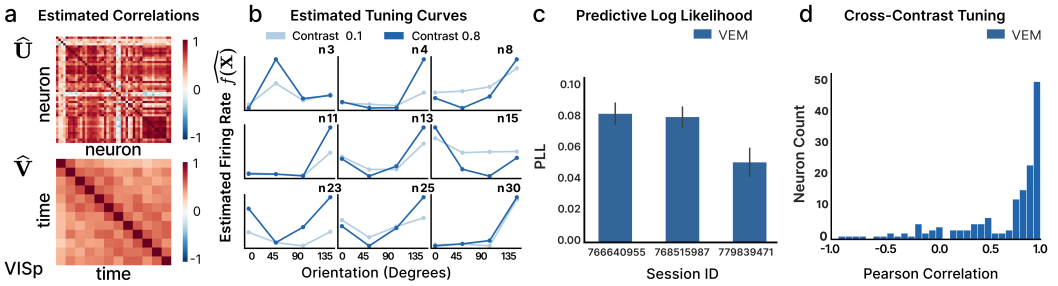


Figure 3: VEM applied to the Allen Institute Visual Coding dataset from VISp at contrast 0.8 and 90° orientation. (a) Estimated latent covariance matrices normalized to show correlation of neuronal and temporal dimensions. (b) histogram of cross-contrast tuning stability in VISp. (c) Example estimated tuning curves. (d) PLL of VEM.

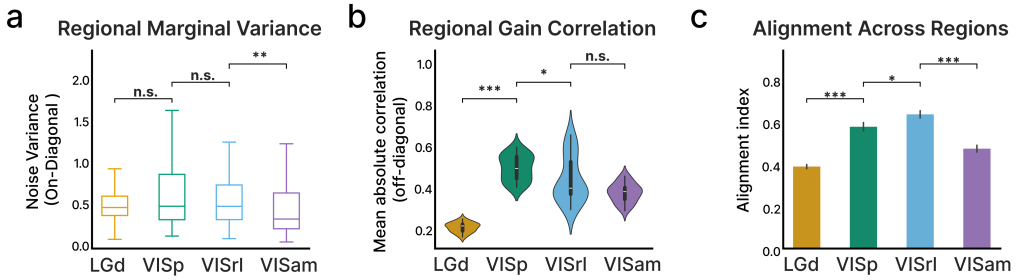


Figure 4: PC-VEM applied to the Allen Institute Visual Coding dataset across four connected brain regions. (a) Marginal variance across regions with no significant difference between LGd and VISp and VISp and VISrl. (b) Off-diagonal correlation significant differences from LGd to VISp and VISp to VISrl (c) Alignment between neural activity and $\hat{U}$ ***, p ¡ 0.001, *, p ¡ 0.05, computed by Wilcoxon rank-sum test.

chy: dorsal lateral geniculate nucleus (LGd), primary visual area (VISp), rostrolateral visual area (VISrl), and anteromedial visual area (VISam). We isolate recordings taken during the presentation of drifting grating stimuli of 4 orientations (0°, 45°, 90°, 135°) and 2 contrast levels (0.1, 0.8) presented for 2 s and repeated 75 times. To account for differences in visual processing latency, we excluded the first 200 ms of stimulus presentation and binned the subsequent 1.8 s of spike counts into 150 ms windows. We strictly included units that passed standardized spike sorting quality control metrics (ISI violations $< .5$, presence ratio $> .95$, amplitude cutoff $< .1$) and had mean firing rate > 2.0 Hz. We selected the three sessions with the most isolated units that had ≥ 20 units per brain region.

We first validated PMNLV by evaluating its held-out predictive log-likelihood (PLL) against a baseline mean-firing rate model with a 70-30 data split for each session/stimulus (Fig. 3). We then subtracted the log-likelihood of a baseline Poisson model, parameterized by the empirical mean firing rate of each neuron in each time-bin, from our model's log-likelihood. We found that PC-VEM consistently outperform the baseline by $.053 - .086$ per observation (Fig. 3d). Moreover, PC-VEM recovers plausible orientation-selective tuning curves whose baseline preferences remain stable across contrast conditions (Fig. 3c). Together, these results indicate that our model is both generalizable and captures realistic orientation selective tuning curves in VISp phenomena (Niell & Stryker, 2008).

We next investigated the question: how does population variability change across the visual hierarchy? Consistent with past work (Akella et al., 2025), we found that the marginal variances of our model, which represent single-neuron variability, change minimally along the hierarchy (Fig. 4a). However, we are also uniquely positioned to explore how population covariance evolves along the hierarchy. To do so, we calculated the mean absolute value across the off-diagonal terms of the neuron gain correlation matrices (Fig. 4b). We find that, in contrast to single-neuron variability, shared co-variability is highly region dependent, with gain correlation initially low in the thalamus (LGd), peaking in VISp, and then systematically decreasing into higher-order visual areas (VISrl, VISam). This pattern suggests that while individual neurons remain similarly variable, ascending

the cortical hierarchy decorrelates the population level co-variability (Siegle et al., 2021; Jia et al., 2022; Montjin et al., 2016).

Finally, we explored the difference between co-variability in neural activity and neural gain (see appendix C.1). Specifically, we computed the alignment index (Elsayed et al., 2016) between the covariance of the neural activity and the first 10 principal components of the recovered neuron gain covariance matrices. As with the total neural gain co-variability, alignment begins low in LGd, peaks in early stages of visual cortex (VISp and VISrl), and then declines in higher order cortex (VISam). We note that, generally, the subspaces defining the neural activity and neuron noise covariance are substantially misaligned (mean alignment index < 0.65). This discrepancy indicates that the drivers of neural gain covariance may operate separately from the activity covariance.

6 Conclusion

We introduce here the Poisson matrix-normal latent variable (PMNLV) model for population-level overdispersion. PMNLV is a generative framework that simultaneously respects the observed Poisson spiking statistics and the correlated gain fluctuations across neurons and time. We derive an estimation algorithm that improves upon VEM for the non-conjugate prior by targeting the precision of the Gaussian prior. We show that as a non-conjugate posterior, the Kronecker-rank moves the optimization target outside of the Kronecker family. By targeting the correct precision through PC-VEM, the covariance and correlation of the neuronal factors is largely improved. We show on synthetic data that both the PC-VEM outperform standard methods in recovering the Kronecker factors.

By applying our method to simulated data reflective of empirical neural recordings, we demonstrate that it can accurately recover ground truth covariance matrices for the matrix normal prior. VEM recovers both the tuning curves and covariance matrices of the population. Both methods only require that the data be simultaneously recorded across neurons and time-bins without assumptions on the stimulus design.

By applying the PMNLV framework to mouse visual data, we demonstrate that our proposed method capture realistic single-neuron tuning curves as well as exhibit generalizability through PLL validation. We also show, through the estimated correlation, that the within brain region variance tends to show little change between VISp and VISrl regions and decreases across the hierarchy. Although the change between VISp and VISrl is not significant, we found that the noise co-variability in spiking data surprisingly increases after LGd. Finally, by applying these methods to study the geometry of gain covariance, we also show that, unexpectedly, the noise covariance is misaligned with the covariance of the neural activity. In sum, the PMNLV model extends findings from studies on single-neuron variability to a broader context. Future work will extend both methods to quantify uncertainty over the recovered Kronecker factors.

Limitations: The Kronecker factorization $\mathbf{\Sigma} = \boldsymbol{V} \otimes \boldsymbol{U}$ is a standard assumption for matrix-variate data (Werner et al., 2007; Tsiligkaridis & Hero, 2013), and it is what allows both factors to be recovered without parametric constraints from a modest number of trials. It also gives each factor a direct interpretation: $\boldsymbol{U}$ describes how neurons covary and $\boldsymbol{V}$ how that covariation unfolds in time. The cost of this interpretability is that all pairs of neurons share the same temporal profile $\boldsymbol{V}$ and differ only in the magnitude $U_{nn'}$. Rather than isolating slow gain modulations from fast shared responses across distinct neural subpopulations, the recovered factors characterize the integrated effect of both processes. Sums of Kronecker terms, $\mathbf{\Sigma} = \sum_k \boldsymbol{V}_k \otimes \boldsymbol{U}_k$, assign each pattern of neuronal covariation its own timescale, and extending our variational family to this setting is a natural next step.

AI use statement

In this work, we used generative AI tools for: editing software code, editing sentences in writing, and checking code for errors. We have not used generative AI tools for: establishing the conceptual framework, deriving or writing proofs, identify relevant literature, formatting references, analyzing existing literature, writing the manuscript (outside of minor grammatical editing), or creating figures

or images (outside of creating software to plot results). We have reviewed all AI-assisted work. We manually checked LLM-generated code for correctness and verified it by 2 authors.

Reproducibility statement

All code is available freely online at https://anonymous.4open.science/r/iclr-0E0B/. The full Allen Institute Visual Coding dataset is publicly available at https://portal.brain-map.org/explore/circuits/visual-coding-neuropixels.

References

Ryan Prescott Adams, Iain Murray, and David J. C. MacKay. Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 9–16, Montreal Quebec Canada, June 2009. ACM. ISBN 978-1-60558-516-1. doi: 10.1145/1553374.1553376. URL https://dl.acm.org/doi/10.1145/1553374.1553376.

J. Aitchison and C. H. Ho. The Multivariate Poisson-Log Normal Distribution. *Biometrika*, 76(4): 643–653, 1989. ISSN 0006-3444. doi: 10.2307/2336624. URL https://www.jstor.org/stable/2336624.

Shailaja Akella, Peter Ledochowitsch, Joshua H. Siegle, Hannah Belski, Daniel D. Denman, Michael A. Buice, Severine Durand, Christof Koch, Shawn R. Olsen, and Xiaoxuan Jia. Deciphering neuronal variability across states reveals dynamic sensory encoding. *Nature Communications*, 16(1):1768, February 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-56733-w. URL https://www.nature.com/articles/s41467-025-56733-w.

Crista L. Barberini, Gregory D. Horwitz, and William T. Newsome. A Comparison of Spiking Statistics in Motion Sensing Neurones of Flies and Monkeys. In Johannes M. Zanker and Jochen Zeil (eds.), *Motion Vision: Computational, Neural, and Ecological Constraints*, pp. 307–320. Springer, Berlin, Heidelberg, 2001. ISBN 978-3-642-56550-2. doi: 10.1007/978-3-642-56550-2_16. URL https://doi.org/10.1007/978-3-642-56550-2_16.

M. S. Bartlett. The Square Root Transformation in Analysis of Variance. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 3(1):68–78, January 1936. ISSN 1369-7412, 1467-9868. doi: 10.2307/2983678. URL https://academic.oup.com/jrsssb/article/3/1/68/7026428.

Ted Bergstrom and Mark Bagnoli. Log-concave probability and its applications. *Economic Theory*, 26:445–469, 2005. doi: 10.1007/s00199-004-0514-4. URL https://escholarship.org/uc/item/0xt8k6sd.

David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518):859–877, April 2017. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2017.1285773. URL http://arxiv.org/abs/1601.00670. arXiv:1601.00670 [stat].

Matteo Carandini. Amplification of Trial-to-Trial Response Variability by Neurons in Visual Cortex: e264. *PLoS Biology*, 2(9):E264, September 2004. ISSN 15449173. doi: 10.1371/journal.pbio.0020264. URL https://www.proquest.com/docview/1291077381/abstract/EF3BE215479C4F99PQ/1. Num Pages: E264 Place: San Francisco, United States.

Adam S. Charles, Mijung Park, J. Patrick Weller, Gregory D. Horwitz, and Jonathan W. Pillow. Dethroning the Fano Factor: A Flexible, Model-Based Approach to Partitioning Neural Variability. *Neural Computation*, 30(4):1012–1045, April 2018. ISSN 0899-7667, 1530-888X. doi: 10.1162/neco_a_01062. URL https://direct.mit.edu/neco/article/30/4/1012-1045/8368.

Julien Chiquet, Stephane Robin, and Mahendra Mariadassou. Variational Inference for sparse network reconstruction from count data. In *Proceedings of the 36th International Conference on*

Machine Learning, pp. 1162–1171. PMLR, May 2019. URL https://proceedings.mlr.press/v97/chiquet19a.html.

Anne K. Churchland, R. Kiani, R. Chaudhuri, Xiao-Jing Wang, Alexandre Pouget, and M. N. Shadlen. Variance as a Signature of Neural Computations during Decision Making. *Neuron*, 69(4):818–831, February 2011. ISSN 0896-6273. doi: 10.1016/j.neuron.2010.12.037. URL https://www.cell.com/neuron/abstract/S0896-6273(10)01087-1.

John P Cunningham, Byron M Yu, Krishna V Shenoy, and Maneesh Sahani. Inferring Neural Firing Rates from Spike Trains Using Gaussian Processes. In *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007. URL https://papers.nips.cc/paper_files/paper/2007/hash/5ef698cd9fe650923ea331c15af3b160-Abstract.html.

Anandita De and Rishidev Chaudhuri. Common population codes produce extremely nonlinear neural manifolds. *Proceedings of the National Academy of Sciences*, 120(39):e2305853120, September 2023. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2305853120. URL https://www.pnas.org/doi/10.1073/pnas.2305853120.

Benjamin K. Dichter, Kristofer E. Bouchard, and Edward F. Chang. Dynamic Structure of Neural Variability in the Cortical Representation of Speech Sounds. *Journal of Neuroscience*, 36(28):7453–7463, July 2016. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.0156-16.2016. URL https://www.jneurosci.org/content/36/28/7453.

Pierre Dutilleul. The mle algorithm for the matrix normal distribution. *Journal of Statistical Computation and Simulation*, 64(2):105–123, September 1999. ISSN 0094-9655. doi: 10.1080/00949659908811970. URL https://doi.org/10.1080/00949659908811970._eprint: https://doi.org/10.1080/00949659908811970.

Alexander S. Ecker, Philipp Berens, Georgios A. Keliris, Matthias Bethge, Nikos K. Logothetis, and Andreas S. Tolias. Decorrelated Neuronal Firing in Cortical Microcircuits. *Science*, 327(5965):584–587, January 2010. doi: 10.1126/science.1179867. URL https://www.science.org/doi/full/10.1126/science.1179867.

Gamaleldin F. Elsayed, Antonio H. Lara, Matthew T. Kaufman, Mark M. Churchland, and John P. Cunningham. Reorganization between preparatory and movement population responses in motor cortex. *Nature Communications*, 7(1):13239, October 2016. ISSN 2041-1723. doi: 10.1038/ncomms13239. URL https://www.nature.com/articles/ncomms13239.

Zachary Friedenberger and Richard Naud. Dendritic excitability controls overdispersion, June 2023. URL https://www.biorxiv.org/content/10.1101/2022.11.18.517108v2. Pages: 2022.11.18.517108 Section: New Results.

Yuanjun Gao, Lars Busing, Krishna V Shenoy, and John P Cunningham. High-dimensional neural spike train analysis with generalized count linear dynamical systems. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/hash/9996535e07258a7bbfd8b132435c5962-Abstract.html.

Robbe L. T. Goris, J. Anthony Movshon, and Eero P. Simoncelli. Partitioning neuronal variability. *Nature Neuroscience*, 17(6):858–865, June 2014. ISSN 1546-1726. doi: 10.1038/nn.3711. URL https://www.nature.com/articles/nn.3711.

Moshe Gur, Alexander Beylin, and D. Max Snodderly. Response Variability of Neurons in Primary Visual Cortex (V1) of Alert Monkeys. *The Journal of Neuroscience*, 17(8):2914–2920, April 1997. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.17-08-02914.1997. URL https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.17-08-02914.1997.

Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1 edition, December 1985. ISBN 978-0-521-38632-6 978-0-521-30586-0 978-0-511-81081-7. doi: 10.1017/CBO9780511810817. URL https://www.cambridge.org/core/product/identifier/9780511810817/type/book.

Xiaoxuan Jia, Joshua H Siegle, Séverine Durand, Greggory Heller, Tamina K Ramirez, Christof Koch, and Shawn R. Olsen. Multi-regional module-based signal transmission in mouse visual cortex. *Neuron*, 2022.

Stephen Keeley, Mikio Aoi, Yiyi Yu, Spencer Smith, and Jonathan W Pillow. Identifying signal and noise structure in neural population activity with Gaussian process factor models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 13795–13805. Curran Associates, Inc., 2020a. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/9eed867b73ab1eab60583c9d4a789b1b-Abstract.html.

Stephen Keeley, David Zoltowski, Yiyi Yu, Spencer Smith, and Jonathan Pillow. Efficient non-conjugate gaussian process factor models for spike count data using polynomial approximations. In *International conference on machine learning*, pp. 5177–5186. PMLR, 2020b.

Sue Ann Koay, Adam S. Charles, Stephan Y. Thiberge, Carlos D. Brody, and David W. Tank. Sequential and efficient neural-population coding of complex task information. *Neuron*, 110 (2):328–349.e11, January 2022. ISSN 0896-6273. doi: 10.1016/j.neuron.2021.10.020. URL https://www.cell.com/neuron/abstract/S0896-6273(21)00835-7.

Shinsuke Koyama. On the Spike Train Variability Characterized by Variance-to-Mean Power Relationship. *Neural Computation*, 27(7):1530–1548, July 2015. ISSN 0899-7667. doi: 10.1162/NECO_a_00748. URL https://doi.org/10.1162/NECO_a_00748.

Kaushik J. Lakshminarasimhan, Eric Avila, Xaq Pitkow, and Dora E. Angelaki. Dynamical latent state computation in the male macaque posterior parietal cortex. *Nature Communications*, 14 (1):1832, April 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-37400-4. URL https://www.nature.com/articles/s41467-023-37400-4.

Neil Lawrence. Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models. *Journal of Machine Learning Research*, 6(60):1783–1816, 2005. ISSN 1533-7928. URL http://jmlr.org/papers/v6/lawrence05a.html.

Charles Van Loan and Nikos Pitsianis. Approximation with Kronecker Products. Technical report, Cornell University, November 1992. URL https://hdl.handle.net/1813/5484.

Jakob H Macke, Lars Buesing, John Cunningham, Byron M Yu, Krishna V Shenoy, and Maneesh Sahani. Empirical models of spiking in neural populations. In *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper_files/paper/2011/hash/7143d7fbadfa4693b9eec507d9d37443-Abstract.html.

Jorrit S Montjin, Guido T Meijer, Carien S Lansink, and Cyriel M A Pennartz. Population-level neural codes are robust to single-neuron variability from a multidimensional coding perspective. *Cell Reports*, 16(9):2486–2498, 2016.

Cristopher M. Niell and Michael P. Stryker. Highly Selective Receptive Fields in Mouse Visual Cortex. *The Journal of Neuroscience*, 28(30):7520–7536, July 2008. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.0623-08.2008. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3040721/.

Manfred Opper and Cédric Archambeau. The Variational Gaussian Approximation Revisited. *Neural Computation*, 21(3):786–792, March 2009. ISSN 0899-7667. doi: 10.1162/neco.2008.08-07-592. URL https://doi.org/10.1162/neco.2008.08-07-592.

Felix Pei, Joel Ye, David Zoltowski, Anqi Wu, Raeed H. Chowdhury, Hansem Sohn, Joseph E. O'Doherty, Krishna V. Shenoy, Matthew T. Kaufman, Mark Churchland, Mehrdad Jazayeri, Lee E. Miller, Jonathan Pillow, Il Memming Park, Eva L. Dyer, and Chethan Pandarinath. Neural Latents Benchmark '21: Evaluating latent variable models of neural population activity, January 2022. URL http://arxiv.org/abs/2109.04463. arXiv:2109.04463 [cs].

Jonathan W. Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M. Litke, E. J. Chichilnisky, and Eero P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, August 2008. ISSN 1476-4687. doi: 10.1038/nature07140. URL https://www.nature.com/articles/nature07140.

Anuththara Rupasinghe, Adam S. Charles, and Jonathan W. Pillow. Continuous partitioning of neuronal variability. *bioRxiv*, pp. 2025.07.23.666404, July 2025. ISSN 2692-8205. doi: 10.1101/2025.07.23.666404. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12330597/.

Gopal Santhanam, Byron M. Yu, Vikash Gilja, Stephen I. Ryu, Afsheen Afshar, Maneesh Sahani, and Krishna V. Shenoy. Factor-Analysis Methods for Higher-Performance Neural Prostheses. *Journal of Neurophysiology*, 102(2):1315–1330, August 2009. ISSN 0022-3077. doi: 10.1152/jn.00097.2009. URL https://journals.physiology.org/doi/full/10.1152/jn.00097.2009.

Joshua H. Siegle, Xiaoxuan Jia, Séverine Durand, Sam Gale, Corbett Bennett, Nile Graddis, Greggory Heller, Tamina K. Ramirez, Hannah Choi, Jennifer A. Luviano, Peter A. Groblewski, Ruweida Ahmed, Anton Arkhipov, Amy Bernard, Yazan N. Billeh, Dillan Brown, Michael A. Buice, Nicolas Cain, Shiella Caldejon, Linzy Casal, Andrew Cho, Maggie Chvilicek, Timothy C. Cox, Kael Dai, Daniel J. Denman, Saskia E. J. de Vries, Roald Dietzman, Luke Esposito, Colin Farrell, David Feng, John Galbraith, Marina Garrett, Emily C. Gelfand, Nicole Hancock, Julie A. Harris, Robert Howard, Brian Hu, Ross Hytnen, Ramakrishnan Iyer, Erika Jessett, Katelyn Johnson, India Kato, Justin Kiggins, Sophie Lambert, Jerome Lecoq, Peter Ledochowitsch, Jung Hoon Lee, Arielle Leon, Yang Li, Elizabeth Liang, Fuhui Long, Kyla Mace, Jose Melchior, Daniel Millman, Tyler Mollenkopf, Chelsea Nayan, Lydia Ng, Kiet Ngo, Thuyahn Nguyen, Philip R. Nicovich, Kat North, Gabriel Koch Ocker, Doug Ollerenshaw, Michael Oliver, Marius Pachitariu, Jed Perkins, Melissa Reding, David Reid, Miranda Robertson, Kara Ronellenfitch, Sam Seid, Cliff Slaughterbeck, Michelle Stoecklin, David Sullivan, Ben Sutton, Jackie Swapp, Carol Thompson, Kristen Turner, Wayne Wakeman, Jennifer D. Whitesell, Deric Williams, Ali Williford, Rob Young, Hongkui Zeng, Sarah Naylor, John W. Phillips, R. Clay Reid, Stefan Mihalas, Shawn R. Olsen, and Christof Koch. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852):86–92, April 2021. ISSN 1476-4687. doi: 10.1038/s41586-020-03171-x. URL https://www.nature.com/articles/s41586-020-03171-x.

Theodoros Tsiligkaridis and Alfred O. Hero. Covariance Estimation in High Dimensions Via Kronecker Product Expansions. *IEEE Transactions on Signal Processing*, 61(21):5347–5360, November 2013. ISSN 1941-0476. doi: 10.1109/TSP.2013.2279355. URL https://ieeexplore.ieee.org/abstract/document/6584826.

Karl Werner, Magnus Jansson, and Petre Stoica. Kronecker Structured Covariance Matrix Estimation. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 3, pp. III–825–III–828, April 2007. doi: 10.1109/ICASSP.2007.366807. URL https://ieeexplore.ieee.org/abstract/document/4217837.

Anqi Wu, Nicholas A. Roy, Stephen Keeley, and Jonathan W Pillow. Gaussian process based nonlinear latent structure discovery in multivariate spike train data. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/b3b4d2dbedc99fe843fd3dedb02f086f-Abstract.html.

Anqi Wu, Stan Pashkovski, Sandeep R Datta, and Jonathan W Pillow. Learning a latent manifold of odor representations from neural responses in piriform cortex. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/hash/17b3c7061788dbe82de5abe9f6fe22b3-Abstract.html.

Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen Ryu, Krishna V Shenoy, and Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. In *Advances in Neural Information Processing Systems*, volume 21. Curran Associates, Inc., 2008. URL https://papers.nips.cc/paper_files/paper/2008/hash/ad972f10e0800b49d76fed33a21f6698-Abstract.html.

Ehud Zohary, Michael N. Shadlen, and William T. Newsome. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370(6485):140–143, July 1994.

ISSN 1476-4687. doi: 10.1038/370140a0. URL https://www.nature.com/articles/370140a0.

A Appendix

B Machine Information

All experiments were run in Python 3.12 on a Windows 11 workstation equipped with an Intel Core i7 CPU and an NVIDIA RTX 4070 SUPER GPU in `float64` precision. Per-iteration cost of VEM was approximately 1 s on GPU (Quadro RTX 5000) and 2 s on CPU (2× Xeon Gold 6248R, 48 threads), float64. Mean total wall-clock time across the four regimes was 475 s on GPU and 974 s on CPU; three of the four runs stopped at the 500-iteration cap, and only the Matérn–low-rank regime met the convergence tolerance (iteration 447).

C Experimental

C.1 Subspace Alignment

We computed the subspace alignment between the neural activity and the latent gain by using the projection method from Elsayed et al. (2016).

$$\frac{\mathrm{Tr}(\boldsymbol{E}^T \boldsymbol{C} \boldsymbol{E})}{\sum_{i=1}^{10} \lambda_i(\boldsymbol{C})}, \tag{A1}$$

where E is a matrix of the first 10 eigenvectors of $\boldsymbol{U}$, $\boldsymbol{C}$ is the neural activity covariance matrix, and $\lambda_i(\boldsymbol{C})$ is the i$^{\text{th}}$ eigenvalue value of $\boldsymbol{C}$.

D Derivations

Below we provide proofs of the lemmas in the main paper.

D.1 Proof of Lemma 1

Proof. Let $d = NT$ and write $\boldsymbol{S} = \boldsymbol{\Sigma}_q$. For slice c, let

$$q_c(\mathrm{vec}\, \boldsymbol{N}_c) = \mathcal{N}(\boldsymbol{m}_c, \boldsymbol{S}), \qquad \boldsymbol{\mu}_c = \mathrm{vec}\, f(\boldsymbol{X}) + \boldsymbol{m}_c.$$

Index entries of $\boldsymbol{\eta}_c$ by $i = 1, \ldots, d$, equivalently by pairs (n, t). Since the likelihood factorizes entrywise, the ELBO terms depending on $\boldsymbol{S}$ are

$$\mathcal{L}(\boldsymbol{S}) = \sum_{c=1}^{C} \sum_{i=1}^{d} \mathbb{E}_{q_c}\left[\ell(y_{ic}, \eta_{ic})\right] - \sum_{c=1}^{C} \mathrm{KL}\big(\mathcal{N}(\boldsymbol{m}_c, \boldsymbol{S}) \,\|\, \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})\big)\,,$$

where $\ell(y, \eta) = \log p(y \mid \eta)$.

We first compute the covariance derivative of the expected log-likelihood. If $Z \sim \mathcal{N}(\mu, s)$ and ϕ is twice differentiable with sufficient integrability, then CITE

$$\frac{\partial}{\partial s} \mathbb{E}[\phi(Z)] = \frac{1}{2}\, \mathbb{E}[\phi''(Z)]. \tag{A1}$$

Indeed, the Gaussian density p_s satisfies $\partial_s p_s = \frac{1}{2} \partial_z^2 p_s$, and integration by parts twice gives

$$\frac{\partial}{\partial s} \int \phi(z) p_s(z)\, dz = \frac{1}{2} \int \phi(z) \partial_z^2 p_s(z)\, dz = \frac{1}{2} \int \phi''(z) p_s(z)\, dz.$$

For fixed i, c, the expectation $\mathbb{E}_{q_c}[\ell(y_{ic}, \eta_{ic})]$ depends on $\boldsymbol{S}$ only through the marginal variance S_{ii}. Applying equation A1,

$$d\left[\sum_{c=1}^{C}\sum_{i=1}^{d}\mathbb{E}_{q_c}\ell(y_{ic},\eta_{ic})\right] = \frac{1}{2}\sum_{c=1}^{C}\sum_{i=1}^{d}\mathbb{E}_{q_c}\left[\partial_\eta^2\ell(y_{ic},\eta_{ic})\right]\, dS_{ii}.$$

Since $\mathbf{J}(y,\eta) = -\partial_\eta^2\ell(y,\eta)$, this is

$$d\left[\sum_{c,i}\mathbb{E}_{q_c}\ell(y_{ic},\eta_{ic})\right] = -\frac{C}{2}\,\operatorname{tr}\big(\operatorname{Diag}(\boldsymbol{D})\,d\boldsymbol{S}\big)\,,$$

where

$$D_{nt} = \frac{1}{C}\sum_{c=1}^{C}\mathbb{E}_{q_c}\big[\mathbf{J}(y_{ntc},\eta_{ntc})\big].$$

Hence

$$\nabla_{\boldsymbol{S}}\sum_{c,i}\mathbb{E}_{q_c}\ell(y_{ic},\eta_{ic}) = -\frac{C}{2}\operatorname{Diag}(\boldsymbol{D}). \tag{A2}$$

For the KL term,

$$\mathrm{KL}\big(\mathcal{N}(\boldsymbol{m}_c,\boldsymbol{S})\,\|\,\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})\big) = \frac{1}{2}\left[\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}) + \boldsymbol{m}_c^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{m}_c - d + \log|\boldsymbol{\Sigma}| - \log|\boldsymbol{S}|\right].$$

Therefore

$$\nabla_{\boldsymbol{S}}\sum_{c=1}^{C}\mathrm{KL}\big(\mathcal{N}(\boldsymbol{m}_c,\boldsymbol{S})\,\|\,\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})\big) = \frac{C}{2}\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{S}^{-1}\right). \tag{A3}$$

At a critical point $\boldsymbol{S} \succ 0$, $\nabla_{\boldsymbol{S}}\mathcal{L}(\boldsymbol{S}) = \mathbf{0}$. Combining equation A2 and equation A3,

$$-\frac{C}{2}\operatorname{Diag}(\boldsymbol{D}) - \frac{C}{2}\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{S}^{-1}\right) = \mathbf{0}.$$

Thus

$$\boldsymbol{S}^{-1} = \boldsymbol{\Sigma}^{-1} + \operatorname{Diag}(\boldsymbol{D}),$$

and hence

$$\boldsymbol{\Sigma}_q^\star = \left(\boldsymbol{\Sigma}^{-1} + \operatorname{Diag}(\boldsymbol{D})\right)^{-1}.$$

□

Remark. The expression above is a fixed-point condition: $\boldsymbol{D}$ is evaluated under the variational distribution, and therefore depends on $\boldsymbol{\Sigma}_q^\star$.

D.2 Proof of Lemma 2

Proof. For Poisson observations with rate $g(\eta) > 0$,

$$\ell(y,\eta) = \log p(y \mid \eta) = y\log g(\eta) - g(\eta) - \log(y!). \tag{A4}$$

Thus

$$\partial_\eta^2\ell(y,\eta) = y(\log g)''(\eta) - g''(\eta), \tag{A5}$$

and hence

$$\mathbf{J}(y,\eta) \equiv -\partial_\eta^2\ell(y,\eta) = -y(\log g)''(\eta) + g''(\eta). \tag{A6}$$

Since g is log-concave, $(\log g)''(\eta) \le 0$. Therefore, for $y \ge 0$,

$$-y(\log g)''(\eta) \ge 0. \tag{A7}$$

Together with $g''(\eta) > 0$, this gives

$$\mathbf{J}(y,\eta) > 0, \qquad y \ge 0,\ \eta \in \mathbb{R}. \tag{A8}$$

□

For the softplus link used in the paper, let

$$\text{softplus}(\eta) = \log(1 + e^{\eta}), \qquad \sigma(\eta) = \frac{1}{1 + e^{-\eta}}. \tag{A9}$$

Then $\text{softplus}' = \sigma$ and

$$\text{softplus}''(\eta) = \sigma(\eta)(1 - \sigma(\eta)) > 0. \tag{A10}$$

Finally, to check the log-concavity, we note that the derivative

$$h(\eta) = \sigma'(\eta) = \sigma(\eta)(1 - \sigma(\eta)) \tag{A11}$$

is the logistic density, and

$$(\log h)''(\eta) = -2\sigma(\eta)(1 - \sigma(\eta)) < 0, \tag{A12}$$

so h is log-concave. Using Bergstrom & Bagnoli (2005, Theorem 1), the CDF σ is log-concave, and its left integral is log-concave. Since

$$\text{softplus}(\eta) = \int_{-\infty}^{\eta} \sigma(u)\, du, \tag{A13}$$

the softplus function is log-concave. Therefore the softplus link satisfies the hypotheses of Lemma 2, and so $\mathbf{J}(y, \eta) > 0$ pointwise.

D.3 Proof of Theorem 1

Proof. By Lemma 2, $\mathbf{J}(y, \eta) > 0$ pointwise, hence $D_{nt} > 0$ for every (n, t) and therefore $\gamma > 0$. By Lemma 1, replacing $\boldsymbol{D}$ by $\gamma \mathbf{1}\mathbf{1}^\top$ gives

$$\boldsymbol{\Sigma}_q^\star = \left(\boldsymbol{\Sigma}^{-1} + \gamma \boldsymbol{I}_{NT}\right)^{-1} = \left(\boldsymbol{V}^{-1} \otimes \boldsymbol{U}^{-1} + \gamma \boldsymbol{I}_T \otimes \boldsymbol{I}_N\right)^{-1}.$$

Suppose, for contradiction, that $\boldsymbol{\Sigma}_q^\star = \boldsymbol{\Psi}_V \otimes \boldsymbol{\Psi}_U$ with $\boldsymbol{\Psi}_U, \boldsymbol{\Psi}_V \succ 0$. Inverting,

$$\boldsymbol{V}^{-1} \otimes \boldsymbol{U}^{-1} + \gamma \boldsymbol{I}_T \otimes \boldsymbol{I}_N = \boldsymbol{\Psi}_V^{-1} \otimes \boldsymbol{\Psi}_U^{-1}. \tag{A14}$$

Apply the Kronecker rearrangement map Loan & Pitsianis (1992, Theorem 2.1) $\mathcal{R}(\boldsymbol{A} \otimes \boldsymbol{B}) = \text{vec}(\boldsymbol{A})\,\text{vec}(\boldsymbol{B})^\top$. The right-hand side has rearrangement rank one, while the left-hand side is

$$\text{vec}(\boldsymbol{V}^{-1})\,\text{vec}(\boldsymbol{U}^{-1})^\top + \gamma\ \text{vec}(\boldsymbol{I}_T)\,\text{vec}(\boldsymbol{I}_N)^\top.$$

Because neither $\boldsymbol{U}$ nor $\boldsymbol{V}$ is a scalar multiple of the identity, the two left factors are linearly independent and the two right factors are linearly independent. Since $\gamma > 0$, this sum has rank two. This contradicts the rank-one rearrangement of a single Kronecker product. Hence

$$\left(\boldsymbol{\Sigma}^{-1} + \gamma \boldsymbol{I}_{NT}\right)^{-1} \neq \boldsymbol{\Psi}_V \otimes \boldsymbol{\Psi}_U$$

for all $\boldsymbol{\Psi}_U, \boldsymbol{\Psi}_V \succ 0$. □

E Synthetic

E.1 Extended Simulated Results

In this section we expand upon the benchmark results shown in section 4 to demonstrate performance across all combinations of the Kronecker factors for the proposed methods as well as across comparable methods.

E.1.1 PROPOSED METHOD BENCHMARKS

Scenario $(\mathbf{U}, \mathbf{V})$		**relative error** ε		**correlation** ρ	
		VEM	PC-VEM	VEM	PC-VEM
(LR,LR)	$\widehat{\mathbf{U}}$	0.238 ±0.004	**0.183 ±0.004**	0.985 ±0.001	**0.987 ±0.001**
	$\widehat{\mathbf{V}}$	0.287 ±0.005	**0.196 ±0.011**	0.967 ±0.003	0.967 ±0.004
	$\hat{f}$	0.064 ±0.001	**0.063 ±0.001**	0.989 ±0.000	0.989 ±0.000
(LR,M)	$\widehat{\mathbf{U}}$	0.434 ±0.005	**0.320 ±0.004**	0.951 ±0.002	**0.960 ±0.002**
	$\widehat{\mathbf{V}}$	0.468 ±0.006	**0.247 ±0.016**	0.949 ±0.005	**0.960 ±0.004**
	$\hat{f}$	**0.065 ±0.001**	0.067 ±0.002	0.989 ±0.000	0.989 ±0.000
(M,LR)	$\widehat{\mathbf{U}}$	0.228 ±0.004	**0.198 ±0.004**	0.966 ±0.002	**0.968 ±0.002**
	$\widehat{\mathbf{V}}$	0.229 ±0.004	**0.128 ±0.005**	0.983 ±0.001	**0.986 ±0.001**
	$\hat{f}$	0.067 ±0.001	**0.066 ±0.001**	0.989 ±0.000	0.989 ±0.000
(M,M)	$\widehat{\mathbf{U}}$	0.409 ±0.006	**0.342 ±0.006**	0.896 ±0.004	**0.906 ±0.004**
	$\widehat{\mathbf{V}}$	0.357 ±0.006	**0.133 ±0.008**	0.968 ±0.001	**0.970 ±0.003**
	$\hat{f}$	**0.068 ±0.001**	0.069 ±0.001	0.989 ±0.000	0.989 ±0.000

Table A1: Recovery of covariance factors and tuning curves by VEM and PC-VEM with N=45, C=12, T=75; mean ± SEM over 11 independent data draws from fixed $\mathbf{U}, \mathbf{V}$. $\varepsilon(\cdot)$: relative Frobenius error after unit-trace normalization; $\rho(\cdot)$: Pearson correlation of off-diagonal entries $(\mathbf{U}, \mathbf{V})$ or per-neuron correlation averaged over neurons (f). Best per row in bold; ties at the reported precision unmarked.

with N=45, C=12, T=75 (single seed). $\varepsilon(\cdot)$: relative Frobenius error to ground truth matrix; $\rho(\cdot)$: off-diagonal correlation with ground truth. Best per row in bold; tied values are not bolded.

E.1.2 METHOD COMPARISON BENCHMARKS

scenario	**predictive performance**					
	oracle	PC-VEM	GPFA	P-GPFA	SNP-GPFA	GP-LVM
(LR, LR)	+0.139	**+0.134 (0.96)**	+0.120 (0.87)	+0.107 (0.77)	+0.120 (0.86)	+0.126 (0.91)
(LR, M)	+0.149	**+0.144 (0.96)**	+0.125 (0.83)	+0.109 (0.73)	+0.140 (0.94)	+0.132 (0.88)
(M, M)	+0.154	**+0.147 (0.95)**	+0.118 (0.77)	+0.101 (0.66)	+0.138 (0.90)	+0.123 (0.80)
(SNP)	+0.061	**+0.055 (0.91)**	+0.038 (0.63)	+0.027 (0.45)	+0.054 (0.89)	+0.043 (0.70)
(PLDS)	+0.183	**+0.178 (0.97)**	+0.138 (0.75)	+0.119 (0.65)	+0.167 (0.91)	+0.154 (0.84)

Table A2: Complete co-smoothing results: bits/spike of held-out neurons relative to a flat per-neuron rate baseline (fraction of oracle in parentheses; higher is better)