5

Networks with Nonlinear Autoregressive Dynamics

The last two chapters combined network models and Hawkes processes to construct probabilistic models for dynamic neural spike trains. The key assumption of Hawkes processes, the assumption we leveraged to derive efficient inference algorithms, was that the firing rate was the sum of nonnegative impulse responses from preceding spikes. In other words, the interactions in Hawkes processes were additive and purely excitatory. While this led to interpretable network models for some systems, in many cases it is more natural to expect a mix of excitatory and inhibitory interactions. Unfortunately, we cannot have inhibitory interactions within a simple additive model because they could yield negative firing rates. Instead, we need to revisit the assumption of linear dynamics.

As we discussed in Chapter 2, there is a simple way to generalize the linear autoregressive dynamics. We retain the linear combination of impulse responses from preceding events, but then we apply a rectifying nonlinear function to obtain a firing rate. Formally, we assume the following model:

$$\psi_{t,n} = \psi_n^{(0)} + \sum_{n'=1}^{N} \sum_{d=1}^{D} s_{t-d,n'} \cdot h_{n' \to n}[d],$$

$$\lambda_{t,n} = g(\psi_{t,n}),$$

where $\psi_{t,n}$ is a real-valued signal called the *activation*, and $g : \mathbb{R} \to \mathbb{R}_+$ is a rectifying nonlinear function that converts the activation into a firing rate. Once this nonlinearity has been introduced, the impulse response functions, $h_{n' \to n}[d]$, are free to be both positive and negative.

As before, we assume the spike count is randomly drawn from a discrete distribution parameterized by $\lambda_{t,n}$. In this chapter, however, we consider a more general class of observation distributions with neuron-specific global parameters ν_n,

$$s_{t,n} \sim p(s_{t,n} \mid \lambda_{t,n}, \nu_n).$$

This nonlinear autoregressive model is also known as a generalized linear model (GLM), and it is widely used in neuroscience (Paninski, 2004; Truccolo et al., 2005). For example, these models have proven useful in elucidating correlated patterns of activity in simultaneously recorded populations of retinal ganglion cells (RGCs) (Pillow et al., 2008). By elaborating upon the basic generalized linear model, more sophisticated structure of the underlying bipolar cells has been revealed from RGC activity (Freeman et al., 2015). While the weights of the GLM cannot typically be interpreted as synaptic connections (Vidne et al., 2012), these models have nevertheless had some success extracting underlying synaptic connectivity from spike trains (Gerhard et al., 2013; Fletcher et al., 2011; Soudry et al., 2015). As we move toward larger and more complete recordings of neural circuits, generalized linear models are likely to play a major role in guiding our understanding of neural systems.

If the spike counts are Poisson distributed, then under general conditions (Paninski, 2004) the likelihood of the spike counts will be log concave and amenable to efficient, exact *maximum a posteriori* (MAP) estimation. Moreover, fully Bayesian estimation procedures based on expectation-propagation (Gerwinn et al., 2008) and Hamiltonian Monte Carlo (Ahmadian et al., 2011) have given some insight into the posterior distribution of network weights. Unfortunately, once we introduce nontrivial prior distributions, like network models, the posterior becomes multimodal and Bayesian inference becomes more challenging. However, see Soudry et al. (2015) for some recent advances in approximate MAP estimation in models that combine network priors and generalized linear models of spike trains.

Name	Parameters	$p(s \mid \psi, \nu)$	$\mathbb{E}[s]$	$\mathrm{Var}(s)$
Gaussian	$\mathcal{N}(\psi, \nu)$	$\frac{1}{\sqrt{2\pi\nu}} e^{-\frac{1}{2\nu}(s-\psi)^2}$	ψ	ν
Bernoulli	$\mathrm{Bern}(\sigma(\psi))$	$\frac{(e^{\psi})^s}{1+e^{\psi}}$	$\sigma(\psi)$	$\sigma(\psi)\,\sigma(-\psi)$
Binomial	$\mathrm{Bin}(\nu, \sigma(\psi))$	$\binom{\nu}{s} \frac{(e^{\psi})^s}{(1+e^{\psi})^{\nu}}$	$\nu\sigma(\psi)$	$\nu\sigma(\psi)\,\sigma(-\psi)$
Neg. Binomial	$\mathrm{NB}(\nu, \sigma(\psi))$	$\binom{\nu+s-1}{s} \frac{(e^{\psi})^s}{(1+e^{\psi})^{\nu+s}}$	νe^{ψ}	$\nu e^{\psi}/\sigma(-\psi)$

Table 5.1: List of observation distributions.

This chapter develops efficient MCMC algorithms for approximating the posterior distribution of weights in a GLM with prior distributions on the impulse responses that are derived from probabilistic network models. Unfortunately, the augmentation scheme developed for linear Hawkes processes is no longer viable due to the nonlinear interactions. Instead, we develop an inference algorithm based on a recently developed scheme known as Pólya-gamma augmentation (Polson et al., 2013). A recent application of these methods to factor analysis models for neural spike trains provided the motivation for the work in this chapter (Pillow and Scott, 2012). The basic idea behind the Pólya-gamma augmentation is to introduce a set of auxiliary variables conditioned upon which the discrete spike counts actually "look" like Gaussian observations. Once we have reduced the problem to inference in a linear Gaussian model, a host of efficient inference algorithms are at our disposal.

Probabilistic Model

The underlying model is very similar to the linear autoregressive models of the preceding chapter. Rather than considering Poisson distributed spike counts, however, here we work with other discrete observation models that are amenable to Pólya-gamma augmentation. As we generalize to allow both excitatory and inhibitory interactions, we revisit the form of our impulse response and introduce a simpler form. Finally, we outline a variety of network models that provide multivariate Gaussian prior distributions on the interaction weights.

Spike count models

The spike counts are stochastically drawn from a distribution parameterized by an underlying, real-valued *activation*, $\psi_{t,n}$, and a static parameter, ν_n. The Pólya-gamma augmentation can be applied to observation models that depend on a logistic transformation of the activation,

$$\sigma(\psi) = \frac{e^{\psi}}{1 + e^{\psi}},$$

which has the property $\sigma(-\psi) = 1 - \sigma(\psi)$. The Bernoulli, binomial, and negative binomial distributions are all natural choices. While the Gaussian distribution is not a proper model for discrete spike counts, we include it in our list since our inference procedure will reduce the other models to the Gaussian case. Table 5.1 lists the basic properties of these observation distributions.

The Bernoulli distribution is appropriate for binary spike counts, whereas the binomial and negative binomial have support for integer $s \in \{0, \ldots, \nu\}$ and $s \in \{0, 1, \ldots\}$, respectively. Notably missing from this list is the Poisson distribution, which is not amenable to the Pólya-gamma augmentation. Nevertheless, both the binomial and negative binomial distributions converge to the Poisson under certain limits. For example, $\lim_{r\to\infty} \mathrm{NB}(r, \sigma(\psi - \log r)) = \mathrm{Poisson}(e^{\psi})$. Moreover, the binomial and negative binomial distributions afford the added flexibility of modeling under- and over-dispersed spike counts. Specifically, while the Poisson has unit dispersion (its mean is equal to its variance), the binomial distribution is always under-dispersed, with variance less than its mean, and the negative binomial is always over-dispersed, with variance greater than its mean.

Linear Gaussian Activation Model

As in the Chapters 3 and 4, we model the impulse response as a weighted sum of basis functions,

$$h_{n'\to n}[d] = a_{n'\to n} \sum_{b=1}^{B} w_{n'\to n}^{(b)} \cdot \phi_b[d]. \tag{5.1}$$

We dispense with the normalized basis function, $\hbar[d]$, because, in the nonlinear model, the normalized basis functions can no longer be seen as distributions over child spike times. Indeed the notion of "parent" and "child" spikes is no longer warranted since the Poisson superposition principle does not apply to nonlinear models. Moreover, in some cases, we may wish to model interactions that are both excitatory *and* inhibitory — for example, the effect of a spike may be inhibitory at short time scales and excitatory after a delay. While these types of interactions may not correspond to actual biological synapses, they can still capture salient correlations in neural spike trains.

Plugging Eq. 5.1 into the activation model, we have,

$$\begin{aligned}\psi_{t,n} &\triangleq \psi_n^{(0)} + \sum_{n'=0}^{N}\sum_{b=1}^{B} a_{n'\to n}\, w_{n'\to n}^{(b)} \left(\sum_{d=1}^{D} \phi_b[d] \cdot s_{t-d,n'}\right) \\ &= \psi_n^{(0)} + \sum_{n'=0}^{N}\sum_{b=1}^{B} a_{n'\to n}\, w_{n'\to n}^{(b)}\, \widehat{s}_{t,n'}^{(b)} \\ &= (\boldsymbol{a}_n \odot \boldsymbol{w}_n)^{\mathsf{T}}\, \widehat{\boldsymbol{s}}_t, \end{aligned} \tag{5.2}$$

where $\psi_n^{(0)}$ is the baseline activation of neuron n, $\boldsymbol{\phi}_b$ is a basis function that weights the previous spike counts for offsets $d \in \{1, \ldots, D\}$, the binary variable $a_{n'\to n} \in \{0, 1\}$ indicates whether or not there exists a directed connection from neuron n' to neuron n, and $w_{n'\to n}^{(b)}$ captures the influence that spikes on neuron n' exert on neuron n at offsets weighted by the b-th basis function. Since the basis function and the signal are assumed to be fixed, we precompute the inner sum, which is simply the convolution of the signal with the basis function, to get $\widehat{s}_{t,n}^{(b)}$. Since this is a linear function, we combine the connections, weights, and filtered spike trains into vectors to get the linear form in (5.2). Here, we have,

$$\begin{aligned}
\boldsymbol{a}_n &= \Big[1, \quad a_{1\to n}, \quad \ldots, \quad a_{1\to n}, \quad \ldots, \quad a_{N\to n}, \quad \ldots, \quad a_{N\to n}\Big]^{\mathsf{T}}, \\
\boldsymbol{w}_n &= \Big[\psi_n^{(0)}, \quad w_{1\to n}^{(1)}, \quad \ldots, \quad w_{1\to n}^{(B)}, \quad \ldots, \quad w_{N\to n}^{(1)}, \quad \ldots, \quad w_{N\to n}^{(B)}\Big]^{\mathsf{T}}, \\
\widehat{\boldsymbol{s}}_t &= \Big[1, \quad \widehat{s}_{t,1}^{(1)}, \quad \ldots, \quad \widehat{s}_{t,1}^{(B)}, \quad \ldots, \quad \widehat{s}_{t,N}^{(1)}, \quad \ldots, \quad \widehat{s}_{t,N}^{(B)}\Big]^{\mathsf{T}},
\end{aligned}$$

and $\odot$ denotes the elementwise product. Note that each $a_{n' \to n}$ is repeated B times in the vector $\boldsymbol{a}_n$. For convenience, we let $\boldsymbol{A}$ and $\boldsymbol{W}$ refer to the $N \times NB + 1$ matrices obtained by stacking the vectors $\boldsymbol{a}_n^\mathsf{T}$ and $\boldsymbol{w}_n^\mathsf{T}$, and we let $\widehat{\boldsymbol{S}}$ denote the $T \times NB + 1$ matrix with rows given by $\widehat{\boldsymbol{s}}_t^\mathsf{T}$. The major difference between this formulation and that of the standard GLM is that here we have explicitly modeled the sparsity of the weights via the "adjacency matrix" $\boldsymbol{A}$. Under the standard formulation, all weights are present, that is, $a_{n' \to n} \equiv 1$.

Consider a model with one basis function ($B = 1$) defined by, $\phi_{1,d} = e^{-d/\tau}$. Then $\widehat{s}_{t,n}^{(1)}$ is a weighted sum of spikes in the window $[t - D, t - 1]$, where the weights decay according to an exponential function with time constant τ. If $a_{n' \to n} = 1$, indicating a connection from neuron n' to neuron n, and the weight, $w_{n' \to n}^{(1)}$, is positive, the influence will be excitatory. If it is negative, the effect will be inhibitory. Together, the weights $\boldsymbol{W}$ define a functional *network* of interactions.

Network Models

With this new impulse response model, each edge of the network is now associated with a B-dimensional weight vector. We can use the same adjacency matrix models as in the previous chapters, but we need to consider new models for the weight matrix. A multivariate Gaussian prior is a natural choice. We consider weight models of the form,

$$\begin{aligned} p(\boldsymbol{w}_n \mid \{\boldsymbol{z}_n\}, \boldsymbol{\vartheta}) &= \mathcal{N}(\psi_n^{(0)} \mid \mu_0, \sigma_0^2) \prod_{m=1}^{N} \mathcal{N}(\boldsymbol{w}_{n' \to n} \mid \boldsymbol{\mu}_{n' \to n}, \boldsymbol{\Sigma}_{n' \to n}), \\ &= \mathcal{N}(\boldsymbol{w}_n \mid \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n), \end{aligned}$$

where $\boldsymbol{\mu}_{n' \to n}$ and $\boldsymbol{\Sigma}_{n' \to n}$ are the mean and covariance, and they implicitly depend on the latent variables, $\boldsymbol{z}_{n'}$ and $\boldsymbol{z}_n$, and the parameters of the network model, $\boldsymbol{\vartheta}$. The last line

Name	$\text{dom}(\boldsymbol{z}_n)$	$\boldsymbol{\mu}_{n' \to n}$	$\boldsymbol{\Sigma}_{n' \to n}$
Gaussian Model	—	$\boldsymbol{\mu}$	$\boldsymbol{\Sigma}$
Stochastic Block Model	$\{1, \ldots, K\}$	$\boldsymbol{\mu}_{z_{n'} \to z_n}$	$\boldsymbol{\Sigma}_{z_{n'} \to z_n}$
Latent Distance Model ($B = 1$)	$\mathbb{R}^K$	$-\|\|\boldsymbol{z}_n - \boldsymbol{z}_{n'}\|\|_2^2 + \gamma_0$	σ^2

Table 5.2: Gaussian Weight Models

combines the Gaussian factors into single multivariate Gaussian prior with parameters,

$$\boldsymbol{\mu}_n = \begin{bmatrix} \mu_0 \\ \boldsymbol{\mu}_{1 \to n} \\ \vdots \\ \boldsymbol{\mu}_{N \to n} \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{\Sigma}_n = \begin{bmatrix} \sigma_0^2 & & & \\ & \boldsymbol{\Sigma}_{1 \to n} & & \\ & & \ddots & \\ & & & \boldsymbol{\Sigma}_{N \to n} \end{bmatrix}. \tag{5.3}$$

Table 5.2 defines the three weight models considered in this chapter. Each model defines the mean and variance of a multivariate normal distribution, $\boldsymbol{w}_{n' \to n} \sim \mathcal{N}(\boldsymbol{\mu}_{n' \to n}, \boldsymbol{\Sigma}_{n' \to n})$. In the Gaussian model, all weights are independent and identically distributed. The stochastic block model, has parameters for each pair of classes, each drawn from a normal inverse-Wishart prior. Finally, we consider a latent distance model, but only for the case where the weights are scalar, i.e. $B = 1$. In this case, the distance between points is inversely proportional to the mean weight. For higher order weights, additional assumptions would be required in order to relate distance to vector weights. In this model, we assume standard normal priors on the parameters $\boldsymbol{z}_n$ and γ_0. The variance is given an inverse gamma prior, $\sigma^2 \sim \text{IGa}(\alpha, \beta)$.

Inference via Gibbs Sampling

Inference is the process of evaluating the posterior distribution over latent variables given the observed signal, which is related to the joint distribution by Bayes' rule:

$$p(\{\nu_n, \boldsymbol{a}_n, \boldsymbol{w}_n, \boldsymbol{z}_n\}_{n=1}^N, \boldsymbol{\vartheta} \mid \boldsymbol{S}) = \frac{p(\boldsymbol{S}, \{\nu_n, \boldsymbol{a}_n, \boldsymbol{w}_n, \boldsymbol{z}_n\}_{n=1}^N, \boldsymbol{\vartheta})}{p(\boldsymbol{S})}.$$

It is computationally intractable to compute this posterior exactly and it has no simple closed form solution, so we must instead resort to approximate methods. We use Markov chain Monte Carlo (MCMC) methods to collect samples from this posterior distribution.

Collapsed Gibbs Network Updates

The most challenging aspect of inference is sampling the posterior distribution over connections, $\boldsymbol{A}$. In the dense model, where $a_{n' \to n} \equiv 1$, the posterior distribution over weights is often log concave, which makes it easy to find the *maximum a posteriori* (MAP) estimate and characterize the local uncertainty around the most likely weights. When the connectivity matrix is sparse, there are instead many modes corresponding to different patterns of connectivity. While this makes inference more challenging, sparse connectivity is an important feature that contributes to the interpretability of the model.

Fortunately, we can make posterior inference of the network considerably more efficient by integrating over possible weights and sampling the binary adjacency matrix from its marginal distribution. First, consider the Gaussian observation model. Since $\psi_{t,n}$ is linear in $\boldsymbol{w}_n$, the likelihood is conjugate with the Gaussian prior, and hence the posterior is Gaussian as well. We compute the posterior distribution in closed form:

$$\begin{aligned} p(\boldsymbol{w}_n \mid \boldsymbol{S}, \boldsymbol{a}_n, \{\boldsymbol{z}_n\}, \boldsymbol{\vartheta}) &\propto \mathcal{N}(\boldsymbol{w}_n \mid \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) \prod_{t=1}^{T} \mathcal{N}(s_{t,n} \mid (\boldsymbol{a}_n \odot \boldsymbol{w}_n)^\mathsf{T} \widehat{\boldsymbol{s}}_t,\, \nu_n) \\ &= \mathcal{N}(\boldsymbol{w}_n \mid \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)\, \mathcal{N}(\boldsymbol{s}_n \mid (\boldsymbol{a}_n \odot \boldsymbol{w}_n)^\mathsf{T} \widehat{\boldsymbol{S}},\, \nu_n \boldsymbol{I}) \\ &\propto \mathcal{N}(\boldsymbol{w}_n \mid \widetilde{\boldsymbol{\mu}}_n, \widetilde{\boldsymbol{\Sigma}}_n), \end{aligned} \tag{5.4}$$

where

$$\begin{aligned} \widetilde{\boldsymbol{\Sigma}}_n &= \left[\boldsymbol{\Sigma}_n^{-1} + \left(\widehat{\boldsymbol{S}}^\mathsf{T} (\nu_n^{-1} \boldsymbol{I}) \widehat{\boldsymbol{S}}\right) \odot (\boldsymbol{a}_n \boldsymbol{a}_n^\mathsf{T})\right]^{-1}, \\ \widetilde{\boldsymbol{\mu}}_n &= \widetilde{\boldsymbol{\Sigma}}_n \left[\boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n + \left(\widehat{\boldsymbol{S}}^\mathsf{T} (\nu_n^{-1} \boldsymbol{I}) \boldsymbol{s}_n\right) \odot \boldsymbol{a}_n\right]. \end{aligned}$$

Given this closed-form Gaussian conditional, we can also compute the conditional distri-

bution over just $\boldsymbol{a}_n$, integrating out the corresponding weights, $\boldsymbol{w}_n$:

$$\begin{aligned} p(\boldsymbol{a}_n \mid \widehat{\boldsymbol{S}}, \boldsymbol{\rho}_n, \{\boldsymbol{z}_n\}, \boldsymbol{\vartheta}) &= \int p(\boldsymbol{a}_n, \boldsymbol{w}_n \mid \boldsymbol{S}, \{\boldsymbol{z}_n\}, \boldsymbol{\vartheta}) \, \mathrm{d}\boldsymbol{w}_n \\ &\propto p(\boldsymbol{a}_n \mid \{\boldsymbol{z}_n\}, \boldsymbol{\vartheta}) \int p(\boldsymbol{w}_n \mid \boldsymbol{S}, \boldsymbol{a}_n, \{\boldsymbol{z}_n\}, \boldsymbol{\vartheta}) \, \mathrm{d}\boldsymbol{w}_n \\ &= p(\boldsymbol{a}_n \mid \{\boldsymbol{z}_n\}, \boldsymbol{\vartheta}) \, \frac{\left|\boldsymbol{\Sigma}_n\right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \boldsymbol{\mu}_n^\mathsf{T} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\mu}_n \right\}}{\left|\widetilde{\boldsymbol{\Sigma}}_n\right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \widetilde{\boldsymbol{\mu}}_n^\mathsf{T} \widetilde{\boldsymbol{\Sigma}}_n^{-1} \widetilde{\boldsymbol{\mu}}_n \right\}}. \end{aligned} \tag{5.5}$$

Thus, we can efficiently sample from the conditional distribution of $\boldsymbol{a}_n$ and $\boldsymbol{w}_n$ by first iterating over each neuron $n' \in \{1, \ldots, N\}$ and sampling a new value of $a_{n' \to n}$, fixing the values of $a_{n'' \to n}$ for $n'' \neq n'$ and integrating out the value of $\boldsymbol{w}_n$. To do so, we simply evaluate the marginal probability in Eq. 5.5 for both values of $a_{n' \to n}$ and sample accordingly. Moreover, note that $\widetilde{\boldsymbol{\Sigma}}_n$ reduces to the prior where $a_{n' \to n} = 0$. This will lead to $B \times B$ diagonal blocks that are equal in both the numerator and the denominator. Thus, the terms for which $a_{n' \to n} = 0$ will cancel in the log determinant and quadratic form of 5.5. As a result, if $\boldsymbol{A}$ is p-sparse (i.e. each neuron has at most p incoming edges) evaluating the marginal probability of $\boldsymbol{a}_n$ has complexity an $O(p^3)$. Once $\boldsymbol{a}_n$ has been completely resampled, we can sample a new value of $\boldsymbol{w}_n$ from its multivariate Gaussian conditional distribution, given by Eq. 5.4.

Pólya-gamma augmentation for discrete observations

When the observations are not Gaussian, the conditional distribution of $\boldsymbol{w}_n$ cannot be computed in closed form and the collapsed updates are intractable. To circumvent this problem, we leverage recently developed augmentation schemes for Gaussian models with discrete observations (Polson et al., 2013; Pillow and Scott, 2012). The idea is to augment the observations, $s_{t,n}$, with auxiliary variables, $\omega_{t,n}$, such that conditioned upon the auxiliary variables, the discrete likelihood appears Gaussian.

First, notice that the discrete likelihoods in Table 5.1 can all be put into a "standard" form

in which the probability mass function can be written,

$$p(s \mid \psi, \nu) = c(s, \nu) \, \frac{(e^{\psi})^{a(s,\nu)}}{(1 + e^{\psi})^{b(s,\nu)}},$$

for some functions, a, b, and c that do not depend on ψ. The integral identity at the heart of the Pólya-gamma augmentation scheme is

$$\frac{(e^{\psi})^{a}}{(1 + e^{\psi})^{b}} = 2^{-b} e^{\kappa \psi} \int_0^{\infty} e^{-\omega \psi^2 / 2} \, p_{\mathrm{PG}}(\omega \mid b, 0) \, \mathrm{d}\omega, \tag{5.6}$$

where $\kappa = a - b/2$ and $p(\omega \mid b, 0)$ is the density of the Pólya-gamma distribution $\mathrm{PG}(b, 0)$, which does not depend on ψ.

Using Eq. 5.6 along with priors $p(\psi)$ and $p(\nu)$, we can write the joint density of (ψ, s, ν) as

$$\begin{aligned} p(s, \nu, \psi) &= p(\nu) \, p(\psi) \, c(s, \nu) \frac{(e^{\psi})^{a(s,\nu)}}{(1 + e^{\psi})^{b(s,\nu)}} \\ &= \int_0^{\infty} p(\nu) \, p(\psi) \, c(s, \nu) \, 2^{-b(s,\nu)} e^{\kappa(s,\nu)\psi} e^{-\omega \psi^2 / 2} \, p_{\mathrm{PG}}(\omega \mid b(s, \nu), 0) \, \mathrm{d}\omega. \end{aligned} \tag{5.7}$$

The integrand of Eq. 5.7 defines a joint density on (s, ν, ψ, ω) which admits $p(s, \nu, \psi)$ as a marginal density. Conditioned on these auxiliary variables ω, we have that the likelihood as a function of ψ is,

$$p(s \mid \psi, \nu, \omega) \propto e^{\kappa(s,\nu)\psi} e^{-\omega \psi^2 / 2} \propto \mathcal{N}\left(\frac{\kappa(s, \nu)}{\omega} \,\middle|\, \psi, \, \frac{1}{\omega} \right).$$

Thus, we effectively have a Gaussian likelihood for ψ, after conditioning on s and ω. Now we can apply this augmentation scheme to the full model, introducing auxiliary variables, $\omega_{t,n}$ for each spike count, $s_{t,n}$. Given these variables, the conditional distribution of $\boldsymbol{w}_n$ can be computed in closed form, as before. Let,

$$\boldsymbol{\kappa}_n = \Big[\kappa(s_{1,n}, \nu_n), \quad \ldots, \quad \kappa(s_{T,n}, \nu_n) \Big]^{\mathsf{T}},$$

and

$$\boldsymbol{\Omega}_n = \mathrm{diag}\left(\begin{bmatrix}\omega_{1,n}, & \ldots, & \omega_{T,n}\end{bmatrix}\right).$$

Then we have $p(\boldsymbol{w}_n \mid \boldsymbol{S}, \boldsymbol{a}_n, \{\boldsymbol{z}_n\}, \boldsymbol{\vartheta}, \boldsymbol{\omega}_n, \nu_n) \propto \mathcal{N}(\boldsymbol{w}_n \mid \widetilde{\boldsymbol{\mu}}_n, \widetilde{\boldsymbol{\Sigma}}_n)$, where

$$\begin{aligned}\widetilde{\boldsymbol{\Sigma}}_n &= \left[\boldsymbol{\Sigma}_n^{-1} + \left(\widehat{\boldsymbol{S}}^{\mathsf{T}}\boldsymbol{\Omega}_n\widehat{\boldsymbol{S}}\right) \odot (\boldsymbol{a}_n\boldsymbol{a}_n^{\mathsf{T}})\right]^{-1}, \\ \widetilde{\boldsymbol{\mu}}_n &= \widetilde{\boldsymbol{\Sigma}}_n\left[\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n + \left(\widehat{\boldsymbol{S}}^{\mathsf{T}}\boldsymbol{\kappa}_n\right) \odot \boldsymbol{a}_n\right].\end{aligned}$$

Having introduced auxiliary variables, we must now also derive Markov transitions to update them as well. Fortunately, the Pólya-gamma distribution is designed such that the conditional density of the auxiliary variables is just a "tilted" Pólya-gamma density,

$$p(\omega_{t,n} \mid s_{t,n}, \nu_n, \psi_{t,n}) = p_{\mathrm{PG}}(\omega_{t,n} \mid b(s_{t,n}, \nu_n),\ \psi_{t,n}).$$

These auxiliary variables are conditionally independent and hence can be updated in parallel. Moreover, efficient algorithms are available to generate Pólya-gamma random variates (Windle et al., 2014), and we have ported these to Python.*

Observation Parameter Updates

The observation parameter updates depend on the particular distribution. For Gaussian observations, ν_n is the observation variance, and it is conjugate with an inverse gamma prior. Bernoulli observations have no parameters. In the binomial model, ν_n corresponds to the maximum number of possible spikes — this may be treated as a hyperparameter. For negative binomial spike counts, the shape parameter ν_n can be sampled as in (Zhou et al., 2012).

*https://github.com/slinderman/pypolyagamma

Sampling Network Variables and Parameters

As before, the latent variables and parameters of the network are relatively easy to resample for a given network. The stochastic block model priors are conjugate with the Gaussian distributed weights. The locations of the latent distance model are not conjugate, but we can update them with hybrid Monte Carlo (Neal, 2010).

Model Selection

We have constructed a probabilistic model that supports a variety of network models, including the four adjacency models and the three weight models described above. How can we compare these models in a principled manner? We argue that the typical approach of measuring predictive log likelihood on held-out time bins is insufficient because it relies only on having accurately estimated the network, $\boldsymbol{A}$ and $\boldsymbol{W}$. It does not matter how likely that network is under the latent variable model, predictive likelihood on held-out time bins only measures the quality of the network at making predictions. Instead, we advocate for an alternative measure based on predicting the activity of held-out neurons. To perform well on this task, we must first learn an accurate model for the structure underlying the network so that we can sample latent variables for the new neuron, which in turn allow us to sample a weighted set of functional connections for that neuron and finally compute the predictive log likelihood.

The objective we measure is the probability of a new spike train $\boldsymbol{s}_{n^*} = [s_{1,n^*}, \ldots, s_{T,n^*}]$, given the observed spike train. To compute this, we must integrate over the latent variables and parameters underlying the observed spike train, as well as those underlying the new spike train. Let $\boldsymbol{\Theta} = \{\{\boldsymbol{w}_n, \boldsymbol{a}_n, \nu_n, \boldsymbol{z}_n\}_{n=1}^N, \boldsymbol{\vartheta}\}$, and let $\boldsymbol{\theta}_{n^*} = \{\nu_{n^*}, \boldsymbol{w}_{n^*}, \boldsymbol{a}_{n^*}, \boldsymbol{z}_{n^*}\}$. This objective can be written,

$$\begin{aligned} p(\boldsymbol{s}_{n^*} \mid \boldsymbol{S}) &\approx \int p(\boldsymbol{s}_{n^*} \mid \boldsymbol{\theta}_{n^*}, \boldsymbol{S})\, p(\boldsymbol{\theta}_{n^*} \mid \boldsymbol{\Theta})\, p(\boldsymbol{\Theta} \mid \boldsymbol{S})\, \mathrm{d}\boldsymbol{\theta}_{n^*}\, \mathrm{d}\boldsymbol{Z} \\ &\approx \frac{1}{L} \sum_{\ell=1}^{L} p(\boldsymbol{s}_{n^*} \mid \boldsymbol{\theta}_{n^*}^{(\ell)}, \boldsymbol{S}), \end{aligned}$$

where

$$\boldsymbol{\theta}_{n^*}^{(\ell)} \sim p(\boldsymbol{\theta}_{n^*} \mid \boldsymbol{\Theta}^{(\ell)}), \qquad \text{and} \qquad \boldsymbol{\Theta}^{(\ell)} \sim p(\boldsymbol{\Theta} \mid \boldsymbol{S}).$$

The samples $\{\boldsymbol{\Theta}^{(\ell)}\}_{\ell=1}^{L}$ are the posterior samples generated by the MCMC algorithm presented above. For each sample, we generate a new set of latent variables and connections for neuron n^*, given the parameters included in $\boldsymbol{\Theta}^{(\ell)}$. These, combined with the spike train, enable us to compute the likelihood of $\boldsymbol{s}_{n^*}$.

This approach constitutes a minor approximation: the new spike train and the original spike train are not conditionally independent. It is possible that there are significant connections from n^* to neurons in the training population, and if we had known those connections, we would have inferred different latent variables and parameters for the training population. We assume that these effects are small, i.e. we would find similar class assignments even without observing n^*. This is reasonable if we are only considering a single neuron n^* and the training population is relatively large. In fact, this assumption is fundamental to the generalized linear model. Without it, the inferred weights and predictions would be highly sensitive to the addition of a single neuron. In practice this is rarely the case.

Results

We demonstrate the efficacy of this approach by applying our framework to two neural populations for which we have longstanding experimental evidence in favor of a particular latent variable representation. First, we consider a population of simultaneously recorded retinal ganglion cells (RGCs), which can be characterized by their type (either *on* or *off* in this dataset) and by the location of their receptive field centers. We then consider a population of hippocampal place cells, which encode positions in a two dimensional environment. In both cases, our approach recovers these latent representations given only the neural spike trains, without any knowledge of the stimulus or the true location. Finally, we assess the advantages of our Bayesian approach and the scalability of our algorithm with synthetic data.

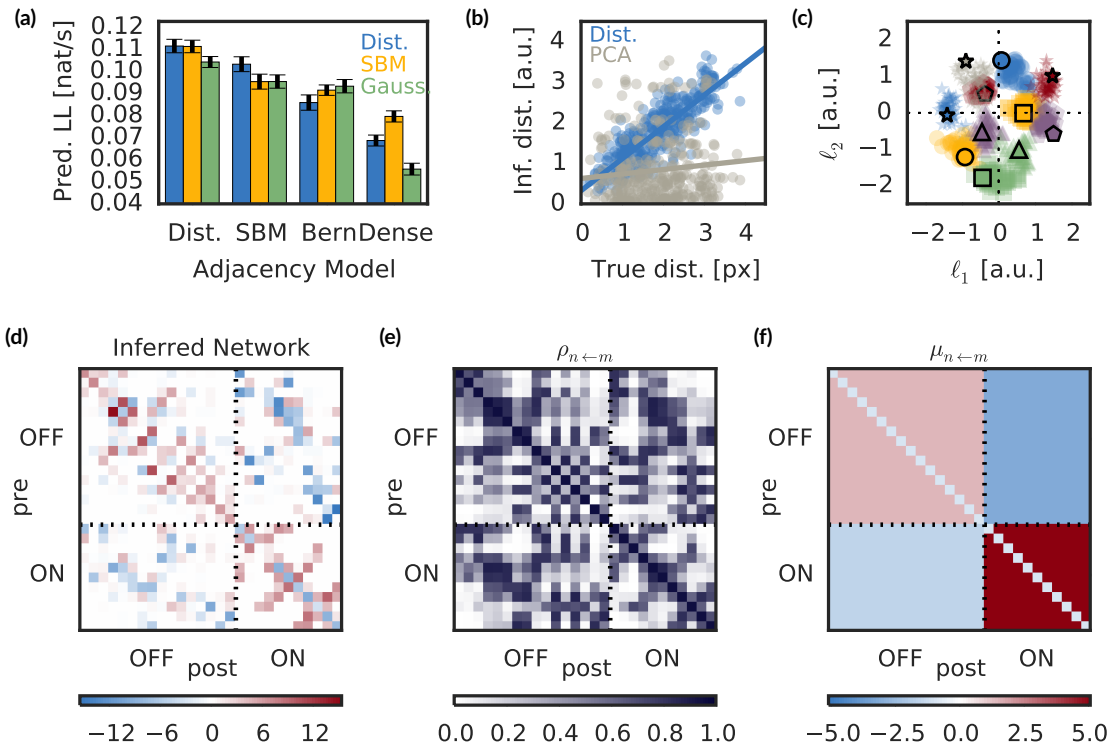

Figure 5.1: Retinal ganglion cell types and locations can be inferred from spike trains alone. **(a)** The combined distance model and block model yields the highest predictive log likelihood (units: nats/spike), compared to other combinations of adjacency models (groups of bars) and weight models (*blue*: latent distance weight model, *yellow*: stochastic block model for weights, *green*: Gaussian weights). **(b)** The inferred distances under the latent distance model are highly correlated with the true distances, as measured by stimulus sensitivity, whereas inferred embeddings under PCA are not. **(c)** Moreover, the inferred locations (semi-transparent markers) recover the true receptive field centers (solid markers with black outline). Shown here only for ON cells. **(d)** Inferred network, $\boldsymbol{A} \odot \boldsymbol{W}$, under a latent distance model of connection probability and a stochastic block model for connection weight, averaged over 500 posterior samples. **(e)** Expected probability of connection under the latent distance model. **(f)** Expected connection strength under the stochastic block model. The inferred cell classes perfectly match the true ON and OFF types, and reflect within-class excitation and between-class inhibition.

Retinal Ganglion Cells

We applied our model to simultaneously recorded multi-neuronal spike trains from a population of 27 primate retinal ganglion cells (RGCs). This dataset was previously analyzed with a standard generalized linear model by Pillow et al. (2008). In this dataset, the population is presented with a Gaussian white noise video with pixel sizes tuned to approximately

the width of a receptive field. We trained our models on one minute of spiking activity, binned at one millisecond time resolution. There were approximately 50,000 spikes in this recording. We also held out one minute of data for model evaluation.

Retinal ganglion cells respond to light (or the absence thereof) shown upon their receptive field. The cells are roughly evenly distributed across the two-dimensional retinal plane. Thus, it is natural to characterize these cells by the location of their receptive field center. Moreover, this population is comprised of two types of cells, *on* and *off* cells, characterized by their response to visual stimuli. *On* cells increase their firing when light is shone upon their receptive field; *off* cells decrease their firing rate in response to light in their receptive field. Cell types are identified by similarity in their response properties as well as morphological and physiological properties (Sanes and Masland, 2015). With knowledge of the stimulus, these cells can be clearly separated into their respective types by clustering the spike-triggered average stimulus. This characterization in terms of latent locations and cell types has been validated through decades of experiments (Kuffler, 1953), and has been made possible by the relative ease with which relevant stimuli can be identified and controlled. To what extent can these latent representations be discovered from the spike trains alone?

We fit our model to the measured spike trains for each of the twelve probabilistic network priors shown in Fig. 3.2 — three weight models and four adjacency models. Predictive likelihood comparisons on the held-out data reveal that the adjacency matrix, i.e. the pattern of connectivity, is well-characterized by a two-dimensional latent distance model (Fig. 5.1c), as expected given the localized receptive fields of RGCs. Moreover, the pairwise distances between the latent locations are highly correlated with the true distances between receptive field centers (Fig. 5.1b), even though the stimulus was never used during training. By contrast, the inferred locations given by the top two principal components of the spike trains are highly uncorrelated, indicating that PCA does not recover a meaningful spatial embedding. Finally, the inferred locations can be rotated and scaled such that they match the true locations almost perfectly (Fig. 5.1c). Rotation does not change the pairwise distances and scaling simply changes the units. This matching is highly unlikely for randomly distributed locations. In Fig. 5.1c, the true locations are shown as solid markers with black outlines, and the inferred locations sampled from the posterior distribution are shown as semi-transparent markers of the same color and shape.

For the weight matrix, a latent distance model (Dist) and a stochastic block model (SBM) both yield similar predictive likelihoods. Looking into the inferred types under the SBM, we find that the neurons are grouped according to their true *on* and *off* cells, again without any knowledge of the stimulus. These types determine the expected interaction weight for each pair of neurons.

Figure 5.1d shows the inferred functional network under the latent distance adjacency model and stochastic block model for weights. The neurons are sorted first by their type (*off* then *on*) and then by their x-location. This yields the three bands in the matrix of connection probabilities 5.1e. Nearby cells have much higher probability of connection. The mean connection strength (Fig. 5.1f) shows the characteristic pattern of positive weights between cells of the same type and negative weights between cells of opposite types. The diagonal of this matrix shows the weights of self-connections, which are typically negative due to refractory effects.

Together, these inferred types and locations provide compelling evidence for a highly structured pattern of functional connectivity. Given the extensive work on characterizing retinal ganglion cell responses, we have considerable evidence that the representation we learn from spike trains alone is indeed the optimal way to characterize this population of cells. This lends us confidence that we may trust the representations learned from spike trains recorded from deeper brain areas, where traditional stimulus-response correlations are less valuable.

Hippocampal Place Cells

We also applied our framework to simultaneously recorded multi-neuronal spike trains from the hippocampal recordings studied in Chapter 3. Here, however, we preprocess the data by binning the spikes into 250ms time bins. Furthermore, we only consider the spike counts from time bins in which the rat was moving at a velocity of at least 10cm/s, since these hippocampal place cells fire primarily during locomotion. The resulting dataset is almost ten minutes long. Finally, we only consider the 25 neurons with the most precise place fields.

As in the retina, we have strong intuitions about the latent structure underlying hippocampal place cell activity, namely, we expect cells representing nearby locations to be

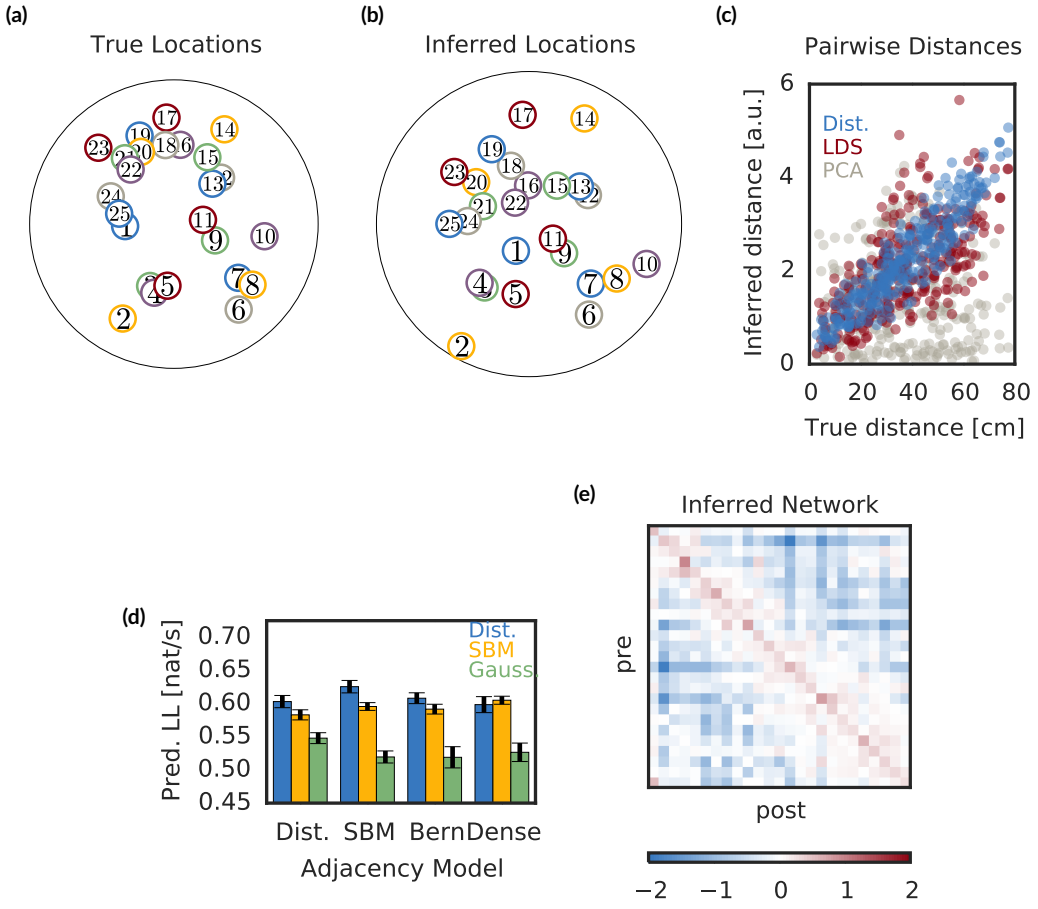

Figure 5.2: Hippocampal place fields are inferred from population spike trains.

correlated and cells with disjoint place fields to be anti-correlated. Can we extract the spatial layout of place fields from spike trains alone?

We apply our framework, fitting all twelve network models and evaluating them on the basis of predictive likelihood. We use a negative binomial observation model to allow multiple spikes per time bin. We find that the distance dependent weight models yield the highest predictive likelihood (Fig. 5.2d). The stochastic block model yields the highest predictive likelihood, but upon further investigation we find that almost all neurons are connected in this model. There is one outlier, neuron #1, which is only sparsely connected. The SBM assigns 24 neurons are assigned to one cluster, and assigns neuron #1 to its own group. Thus, the stochastic block model is quite similar to an independent Bernoulli model, as is evident from the similar predictive likelihood of these two models.

Looking into the inferred locations under the latent distance model for the weight matrix, we find that the inferred locations (Fig 5.2b) are (up to rotation and scale) highly similar to the true place field centers (Fig 5.2a) measured using the rat's true location. This is quantified by plotting the pairwise distances between inferred locations against the pairwise distances between place field centers. The latent distance model's pairwise distances are highly correlated with the ground truth, whereas distances between PCA embeddings are very uncorrelated (Fig. 5.2c).

For further comparison, we fit a Poisson linear dynamical system (PLDS) (Macke et al., 2011) with a two dimensional latent state space, and used the rows of the $N \times 2$ emission matrix as the locations of the neurons. Chapter 8 will discuss this class of models in more detail. The PLDS yields pairwise distances that are quite correlated with the true distances, but we find that the variance of this fit grows with the true distance. This reflects the fact that the PLDS does not explicitly parameterize interaction as a function of distance, but rather induces correlation due to similarity in the emission matrix as well as shared dynamics.

Finally, we investigated the expected network under the posterior and found that upon sorting the neurons by their inferred locations, an intuitive banded structure emerges (Fig. 5.2e). The positive diagonal indicates strong autocorrelation between a neurons spiking in consecutive 250ms bins. Over these time scales, self-refractory effects are not evident. The primary connections are inhibitory in nature, such that active cells suppress the activity of cells with distal place fields.

These inferred representations again confirm our intuitive beliefs about the structure of hippocampal place cell responses. The inferred latent variables recover meaningful structure in the neural activity, without any access to the location. Instead, these representations arise from the neural activity alone.

Synthetic Data

To assess the robustness and scalability of our framework, we apply our methods to simulated data with known ground truth. First, we show that our Bayesian approach can recover structure in larger populations of neurons than those studied above. Moreover, we show that our approach outperforms alternative methods that separate the problems of

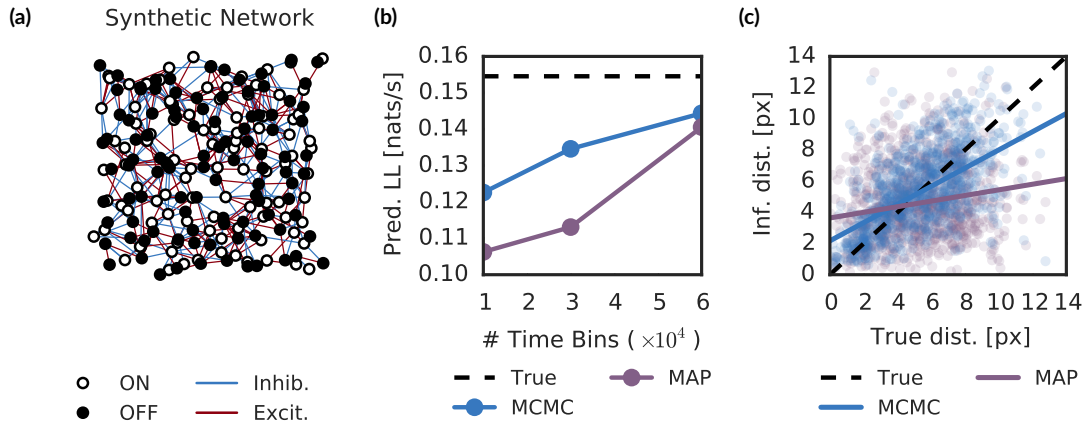

Figure 5.3: Synthetic results based on the retinal ganglion cell population studied above. **(a)** Locations of the 200 neurons (100 *on* and 100 *off* cells) along with a subset of the connections. For clarity, 15% of the connections are randomly chosen. Though the network is not symmetric, arrowheads are dropped for clarity. **(b)** Predictive likelihood on 10,000 held-out time bins as a function of the number of time bins in the training dataset. "MCMC" denotes our Bayesian approach with a latent distance adjacency model and a SBM weight model. "MAP" denotes the L_1-regularized standard GLM. **(c)** Inferred locations found by our model (MCMC) and by fitting a latent distance model to the thresholded standard GLM network (MAP).

finding the network and discovering the underlying structure. Then, we show that for sparse networks, our approach only scales quadratically with the number of neurons — a significant improvement over naïve methods that incur an additional $O(N^3)$ cost per potential connection.

Advantages of being Bayesian

For these analyses, we simulate a population of 200 neurons with structure that mimics that of the retinal ganglion cell population analyzed in Section 5.3.1. Figure 5.3a illustrates the underlying network. As before, we have *on* cells and *off* cells, each centered at a location in the 2D retinal plane. Nearby cells are more likely to be functionally connected, and cells of the same type will excite each other whereas cells of different types will inhibit one another. Again, these functional connections reflect correlations and anti-correlations between cells that arise from common input. We simulate 60000 time bins and tune the network parameters such that if each time bin were one millisecond, the firing rates would range from 10Hz to 70Hz with a mean of about 30Hz. Rather than simulating an external

white noise stimulus, here we simulate spontaneous activity of the network.

Since we know the generative model that gave rise to the data, we can quantify the improvement from using our fully Bayesian approach rather than a standard GLM, and compare these improvements to the upper bound provided by the true model. Figure 5.3b shows the predictive log likelihood for 10000 bins of held-out data as a function of the number of time bins of training data. For shorter training datasets, our Bayesian approach yields considerably higher predictive likelihood than the standard, L_1-regularized GLM. As the length of the training data increases, both our method and the standard GLM improve, and ultimately approach the likelihood of the true model that generated the data. With more data, the network inference is driven less by the prior and more by the observations. Thus, in the limit of infinite data,the GLM will converge to the same network, and hence the same predictive likelihood, as our Bayesian method. However, with hundreds of neurons and tens of thousands of time bins, these results suggest that the Bayesian approach can yield substantial benefits.

Rather than simultaneously fitting the network and the underlying latent variables, one may instead fit the network with an L_1-regularized GLM and *then* fit a probabilistic network model to the GLM connection weights. This is the approach taken by Stevenson et al. (2009). The advantage is that the network is only fit once, and since this is usually the most expensive aspect of inference, it can save considerable time. However, when the data is limited, both the network and the latent variables are uncertain. Our Bayesian approach finds joint assignments with high posterior probability, whereas the standard GLM finds a single network that does not account for prior beliefs about structure underlying the network. In this example, subsequently fitting a latent distance model to the adjacency matrix of a thresholded GLM network finds an embedding that differs considerably from the embedding found by our Bayesian approach. This is illustrated by the decreased correlation between true and inferred pairwise distances under the standard GLM compared to that of our joint, Bayesian approach, as shown in Figure 5.3c.

Figure 5.4 further illustrates this point. Here, the true network is shown as a weighted adjacency matrix (Fig. 5.4a), along with its binary adjacency matrix (Fig. 5.4d). The inferred network and connection probabilities from our fully Bayesian method with a latent distance model prior are shown in Fig. 5.4b and Fig. 5.4e, respectively. Our MCMC algorithm

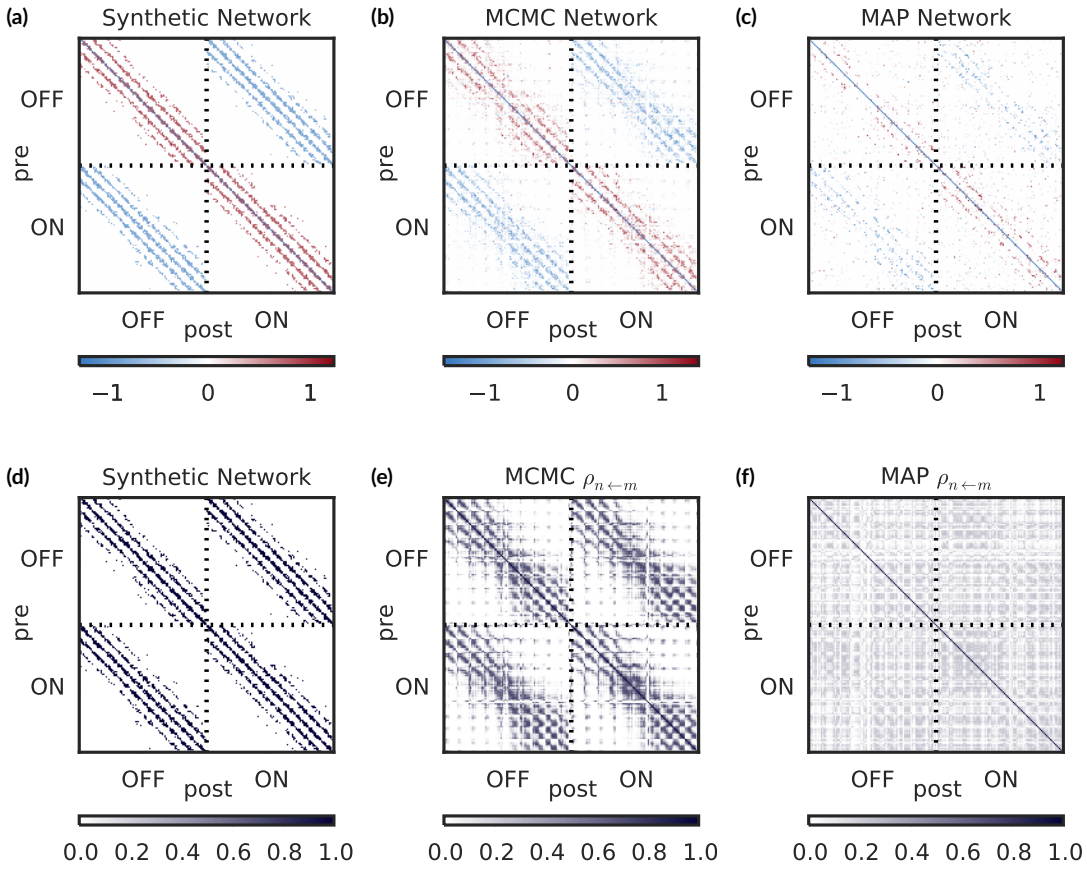

Figure 5.4: **(a)** Weighted adjacency matrix for the synthetic RGC network shown in Fig. 5.3a. **(b)** Network inferred by our fully Bayesian approach. **(c)** Thresholded network found by an L_1-regularized GLM. **(d)** Adjacency matrix corresponding to the true network in Fig. 5.4a. **(e)** Inferred connection probability found by our model with a latent distance prior. The diagonal bands reflect the increased probability for nearby neurons.. **(f)** Inferred connection probability found by fitting a latent distance model to the network in Fig. 5.4c. Noise in this network leads to poor location estimates and nearly uniform connection probabilities.

does a good job recovering the underlying locations, as shown by the diagonally banded connection probabilities reflecting the distance dependence. By contrast, the MAP network found via an L_1-regularized standard GLM is considerably noisier (Fig. 5.4c), since it knows nothing of the underlying distance dependent connectivity. A latent distance model fit to a thresholded copy of this network fails to capture the distance dependence (Fig. 5.4f), since the spurious connections force otherwise well-separated neurons to lie close in the embedding. By jointly fitting the network and the latent locations, we weed out these spurious

connections.

Scalability Analysis

Finally, we address the scalability of our Bayesian inference algorithms. With large-scale recordings becoming the norm, efficiency is paramount. There are three major parameters that govern the complexity of inference: the number of time bins, T; the number of neurons, N; and the level of sparsity, ρ. The natural unit of measure is the wall-clock time required to perform one iteration of our MCMC algorithm. The following experiments were run on a quad-core Intel i5 with 6GB of RAM. We break down the wall-clock time into time spent updating the auxiliary variables of the observation model and time spent updating the network. Resampling the latent variables incurs a lesser cost.

The wall clock time scales linearly with T, as shown in Figure 5.5a. Recall that the observation model contains NT auxiliary variables, each of which must be resampled. Additionally, each neuron must compute its sufficient statistics, which involves a rectangular matrix multiplication costing $O(TN^2)$ time.

Since there are $O(N^2)$ possible connections in a network of N neurons, there is no way to avoid at least a quadratic cost in N (unless some connections can be ruled out *a priori*). Figure 5.5b shows that this quadratic penalty is indeed evident. The cost of updating the auxiliary variables is only linear in N, but the cost of updating the network is quadratic. In Section 5.4, we discuss potential strategies for ruling out connections and limiting this cost.

The total cost could actually be worse than quadratic because the cost of updating each connection could depend on N. Fortunately, the complexity of our collapsed Gibbs sampling algorithm only depends on the number of incident connections received by each neuron, p. Specifically, we must invert a $p \times p$ matrix, which incurs an $O(p^3)$ cost. Figure 5.5c illustrates this for a Bernoulli network with $N = 200$ neurons. The expected number of incoming connections is related to the probability of connection by $\mathbb{E}[p] = \rho N$. If we increased the number of neurons but kept the average in-degree constant, the total cost would scale as $\mathcal{O}(TN + N^2 p^3)$.

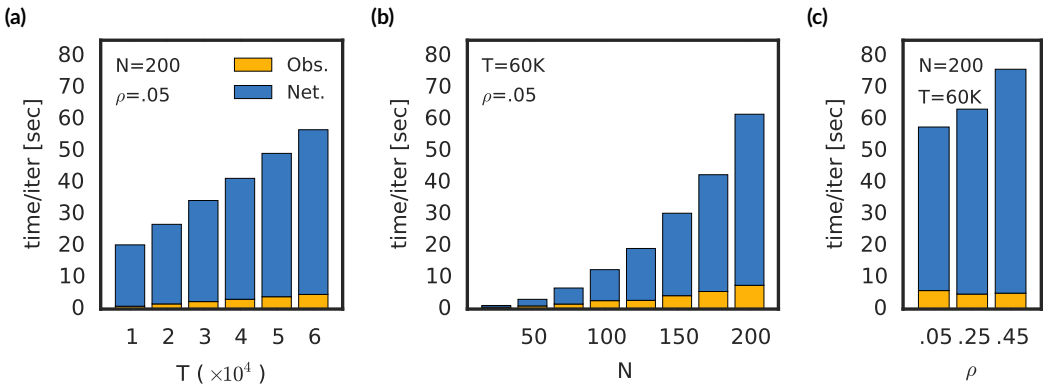

Figure 5.5: Scalability of our inference algorithm as a function of: **(a)** the number of time bins, T; **(b)** the number of neurons, N; and **(c)** the average sparsity of the network, ρ. The complexity grows linear with T and quadratically with N (for fixed ρ). The cost is broken down into the cost of resampling the Pólya-gamma auxiliary variables ("Obs.") and the cost of resampling the sparse network ("Net."). Color scheme shared across all panels.

Discussion

We have shown that the functional networks underlying spike trains may provide insight into the low-dimensional structure of neural populations. Looking forward, there are two main concerns that we would like to address. First, ideally we would like algorithms that scale better than quadratically with the number of neurons. Second, we would like to automate the process of relating the latent structure to simultaneously measured environmental cues, stimuli, and behavioral correlates.

Further scalability improvements

For large populations, the bulk of the running time is spent sampling the network variables, $\boldsymbol{A}$ and $\boldsymbol{W}$. As the number of neurons grows, this complexity scales as $O(N^2)$. To what extent can we minimize this complexity? As we have shown, naïvely applying clustering algorithms and PCA directly to spike trains does not necessarily yield meaningful features. These results can be somewhat improved by smoothing the spike trains (e.g. by taking a moving average), but they are still fundamentally limited when the low-dimensional structure lies in the pattern of functional connectivity.

Moreover, our results in Section 5.3.3 show that separating the problem of inferring the network and discovering its latent structure may yield to inferior results. This is particularly

common for large neural populations, where there is substantial posterior uncertainty. In practice, we initialize our method with a standard GLM and find that a small number of iterations of our MCMC algorithm can yield large improvements in the predictive likelihood of our model. These improvements come from pruning connections that are weakly supported by the data, but which are very unlikely under the prior. This improvement in predictive likelihood is consistent with existing literature on Bayesian spike-and-slab modeling (Mohamed et al., 2012).

An alternative approach is to consider models that explicitly limit the dimensionality of the network. For example, if we know the physical location of neurons (or their approximate location from electrode number), we may rule out long-range connections. Alternatively, we may constrain the weighted adjacency matrix to be low rank, that is $\boldsymbol{W} \odot \boldsymbol{A} \equiv \boldsymbol{U}\boldsymbol{V}^{\mathsf{T}}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are rank $K < N$. The effect is similar to that of a linear dynamical system model (Paninski et al., 2010; Macke et al., 2011). Unfortunately, it is less clear how to incorporate interpretable structure into this type of model, although see Buesing et al. (2014) for some initial steps in this direction.

Finally, we have developed a Markov chain Monte Carlo inference algorithm, and while we have shown reasonable performance on recordings of hundreds of neurons over tens of thousands of time bins, as we scale to larger recordings alternative algorithms may be preferable. When the complexity is dominated by the number of time bins, stochastic variational inference (SVI) (Hoffman et al., 2013) provides an attractive alternative. Since the time bins are conditionally independent given the network, we can get unbiased estimates of sufficient statistics from *mini-batches* of time bins and use them to inform stochastic mean field updates. While we have not derived this mean field algorithm here, it is relatively straightforward to derive these algorithms for Pólya-gamma augmented models (Pillow and Scott, 2012; Zhou et al., 2012). Still, the limiting factor will often be the number of neurons, in which case SVI is not immediately applicable. Developing approximate inference algorithms that only consider subsets of neurons is an active area of research (Soudry et al., 2015).

Relating circuit structure to stimuli and behavior

So far we have reserved the externally measured covariates for validation. We used the white noise stimulus in the retinal ganglion cell experiment to determine the neurons' receptive fields, and we used the measured location in the hippocampal experiment to determine the place cell locations. We did not, however, attempt to jointly model the neural activity *and* the external covariates, as is often done in GLM analyses. The simplest way to incorporate these external covariates, $\boldsymbol{Y} = \{\boldsymbol{y}_t\}$, into the model is via a linear term in the activation. That is, let,

$$\psi_{t,n} \triangleq \psi_n^{(0)} + \boldsymbol{u}_n^\mathsf{T} \boldsymbol{y}_t + (\boldsymbol{a}_n \odot \boldsymbol{w}_n)^\mathsf{T} \widehat{\boldsymbol{s}}_t,$$

where $\boldsymbol{u}_n$ is the "stimulus filter." Just as we modeled $\boldsymbol{a}_n$ and $\boldsymbol{w}_n$ with hierarchical network models, we may model $\boldsymbol{u}_n$ in terms of a set of neuron-specific latent variables.

As we explore deeper brain circuits that are further removed from sensory inputs or motor outputs, the relationship between neural activity and external covariates becomes less clear. In these cases, the unsupervised methods for extracting structured representations can provide useful hints as to how neural activity may be understood. That the representations found in these well-studied circuits of the retina and the hippocampus match our expectations provides reason to believe that they can be fruitfully applied to more enigmatic circuits as well.

Conclusion

This chapter extended the linear autoregressive models of previous chapters to include both excitatory and inhibitory interactions by developing a nonlinear firing rate model. We derived an efficient, fully-conjugate Gibbs sampling algorithm that integrate out the weights in order to sample sparsity patterns. The key was the Pólya-gamma augmentation, which makes discrete observations appear as Gaussian likelihoods. This led to dramatic improvements in performance and allowed us to discover interesting structure in a variety of real and synthetic datasets.

Nevertheless, these methods are still limited in that they are parameterized in terms of

stationary network. There is no aspect of the model that captures time-varying properties of the neural circuit under study. In the next chapter, we consider extensions of this model that capture dynamics in the network itself. This provides a step toward dynamical systems models that attempt to explain data in terms of a latent state that evolves over time.

References

Yashar Ahmadian, Jonathan W Pillow, and Liam Paninski. Efficient Markov chain Monte Carlo methods for decoding neural spike trains. *Neural Computation*, 23(1):46–96, 2011.

Misha B Ahrens, Michael B Orger, Drew N Robson, Jennifer M Li, and Philipp J Keller. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature Methods*, 10(5):413–420, 2013.

Laurence Aitchison and Peter E Latham. Synaptic sampling: A connection between PSP variability and uncertainty explains neurophysiological observations. *arXiv preprint arXiv:1505.04544*, 2015.

Laurence Aitchison and Máté Lengyel. The Hamiltonian brain. *arXiv preprint arXiv:1407.0973*, 2014.

David J Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598, 1981.

Charles H Anderson and David C Van Essen. Neurobiological computational systems. *Computational Intelligence Imitating Life*, pages 1–11, 1994.

Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1-2):5–43, 2003.

Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

Michael J Barber, John W Clark, and Charles H Anderson. Neural representation of probabilistic information. *Neural Computation*, 15(8):1843–64, August 2003.

Leonard E Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563, 1966.

Matthew J. Beal, Zoubin Ghahramani, and Carl E. Rasmussen. The infinite hidden Markov model. *Advances in Neural Information Processing Systems 14*, pages 577–585, 2002.

Jeffrey M Beck and Alexandre Pouget. Exact inferences in a neural implementation of a hidden Markov model. *Neural Computation*, 19(5):1344–1361, 2007.

Jeffrey M Beck, Peter E Latham, and Alexandre Pouget. Marginalization in neural circuits with divisive normalization. *The Journal of Neuroscience*, 31(43):15310–15319, 2011.

Jeffrey M Beck, Katherine A Heller, and Alexandre Pouget. Complex inference in neural circuits with probabilistic population codes and topic models. *Advances in Neural Information Processing Systems*, pages 3059–3067, 2012.

Yoshua Bengio and Paolo Frasconi. An input output HMM architecture. *Advances in Neural Information Processing Systems*, pages 427–434, 1995.

Pietro Berkes, Gergo Orbán, Máté Lengyel, and József Fiser. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331(6013): 83–7, January 2011.

Gordon J Berman, Daniel M Choi, William Bialek, and Joshua W Shaevitz. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99):20140672, 2014.

Philippe Biane, Jim Pitman, and Marc Yor. Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bulletin of the American Mathematical Society*, 38(4):435–465, 2001.

Christopher M Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

David M Blei. Build, compute, critique, repeat: Data analysis with latent variable models. *Annual Review of Statistics and Its Application*, 1:203–232, 2014.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

Carolyn R Block and Richard Block. *Street gang crime in Chicago*. US Department of Justice, Office of Justice Programs, National Institute of Justice, 1993.

Carolyn R Block, Richard Block, and Illinois Criminal Justice Information Authority. Homicides in Chicago, 1965-1995. ICPSR06399-v5. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], July 2005.

Charles Blundell, Katherine A Heller, and Jeffrey M Beck. Modelling reciprocating relationships with Hawkes processes. *Advances in Neural Information Processing Systems*, pages 2600–2608, 2012.

George EP Box. Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A (General)*, pages 383–430, 1980.

David H Brainard and William T Freeman. Bayesian color constancy. *Journal of the Optical Society of America A*, 14(7):1393–1411, 1997.

Kevin L Briggman, Henry DI Abarbanel, and William B Kristan. Optical imaging of neuronal populations during decision-making. *Science*, 307(5711):896–901, 2005.

David R. Brillinger. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, 59(3):189–200, August 1988.

David R Brillinger, Hugh L Bryant Jr, and Jose P Segundo. Identification of synaptic interactions. *Biological Cybernetics*, 22(4):213–228, 1976.

Michael Bryant and Erik B Sudderth. Truly nonparametric online variational inference for hierarchical Dirichlet processes. *Advances in Neural Information Processing Systems 25*, pages 2699–2707, 2012.

Lars Buesing, Johannes Bill, Bernhard Nessler, and Wolfgang Maass. Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS Computational Biology*, 7(11):e1002211, November 2011.

Lars Buesing, Jakob H. Macke, and Maneesh Sahani. Learning stable, regularised latent models of neural population dynamics. *Network: Computation in Neural Systems*, 23: 24–47, 2012a.

Lars Buesing, Jakob H Macke, and Maneesh Sahani. Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. *Advances in Neural Information Processing Systems*, pages 1682–1690, 2012b.

Lars Buesing, Timothy A Machado, John P Cunningham, and Liam Paninski. Clustered factor analysis of multineuronal spike data. *Advances in Neural Information Processing Systems*, pages 3500–3508, 2014.

Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

Santiago Ramón Cajal. *Textura del Sistema Nervioso del Hombre y los Vertebrados*, volume 1. Imprenta y Librería de Nicolás Moya, Madrid, Spain, 1899.

Natalia Caporale and Yang Dan. Spike timing-dependent plasticity: a Hebbian learning rule. *Annual Review of Neuroscience*, 31:25–46, 2008.

Nick Chater and Christopher D Manning. Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, 10(7):335–344, 2006.

Zhe Chen, Fabian Kloosterman, Emery N Brown, and Matthew A Wilson. Uncovering spatial topology represented by rat hippocampal population neuronal codes. *Journal of Computational Neuroscience*, 33(2):227–255, 2012.

Zhe Chen, Stephen N Gomperts, Jun Yamamoto, and Matthew A Wilson. Neural representation of spatial topology in the rodent hippocampus. *Neural Computation*, 26(1): 1–39, 2014.

Sharat Chikkerur, Thomas Serre, Cheston Tan, and Tomaso Poggio. What and where: A Bayesian inference theory of attention. *Vision Research*, 50(22):2233–2247, 2010.

Yoon Sik Cho, Aram Galstyan, Jeff Brantingham, and George Tita. Latent point process models for spatial-temporal networks. *arXiv:1302.2671*, 2013.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945, 2004.

Aaron C Courville, Nathaniel D Daw, and David S Touretzky. Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, 10(7):294–300, 2006.

Ronald L Cowan and Charles J Wilson. Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. *Journal of Neurophysiology*, 71(1):17–32, 1994.

W Maxwell Cowan, Thomas C Südhof, and Charles F Stevens. *Synapses*. Johns Hopkins University Press, 2003.

Mary Kathryn Cowles and Bradley P Carlin. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91: 883–904, 1996.

John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, 2014.

Paul Dagum and Michael Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60(1):141–153, 1993.

Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: Volume I: Elementary Theory and Methods*. Springer Science & Business Media, 2 edition, 2003.

Peter Dayan and Larry F Abbott. *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press, 2001.

Peter Dayan and Joshua A Solomon. Selective Bayes: Attentional load and crowding. *Vision Research*, 50(22):2248–2260, 2010.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

Sophie Deneve. Bayesian spiking neurons I: inference. *Neural Computation*, 20(1):91–117, January 2008.

Luc Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, USA, 1986.

Christopher DuBois, Carter Butts, and Padhraic Smyth. Stochastic block modeling of relational event dynamics. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 238–246, 2013.

Seif Eldawlatly, Yang Zhou, Rong Jin, and Karim G Oweiss. On the use of dynamic Bayesian networks in reconstructing functional neuronal networks from spike train ensembles. *Neural Computation*, 22(1):158–189, 2010.

Marc O Ernst and Martin S Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002.

Sean Escola, Alfredo Fontanini, Don Katz, and Liam Paninski. Hidden Markov models for the stimulus-response relationships of multistate neural systems. *Neural Computation*, 23(5):1071–1132, 2011.

Warren John Ewens. Population genetics theory—the past and the future. In S. Lessard, editor, *Mathematical and Statistical Developments of Evolutionary Theory*, pages 177–227. Springer, 1990.

Daniel E Feldman. The spike-timing dependence of plasticity. *Neuron*, 75(4):556–71, August 2012.

Daniel J Felleman and David C Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47, 1991.

Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230, 1973.

Christopher R Fetsch, Amanda H Turner, Gregory C DeAngelis, and Dora E Angelaki. Dynamic reweighting of visual and vestibular cues during self-motion perception. *The Journal of Neuroscience*, 29(49):15601–15612, 2009.

Christopher R Fetsch, Alexandre Pouget, Gregory C DeAngelis, and Dora E Angelaki. Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, 15(1):146–154, 2012.

József Fiser, Pietro Berkes, Gergő Orbán, and Máté Lengyel. Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, 14(3):119–130, 2010.

Alyson K Fletcher, Sundeep Rangan, Lav R Varshney, and Aniruddha Bhargava. Neural reconstruction with approximate message passing (neuramp). *Advances in Neural Information Processing Systems*, pages 2555–2563, 2011.

Emily B Fox. *Bayesian nonparametric learning of complex dynamical phenomena*. PhD thesis, Massachusetts Institute of Technology, 2009.

Emily B Fox, Erik B Sudderth, Michael I Jordan, and Alan S Willsky. An HDP-HMM for systems with state persistence. *Proceedings of the International Conference on Machine Learning*, pages 312–319, 2008.

Jeremy Freeman, Greg D Field, Peter H Li, Martin Greschner, Deborah E Gunning, Keith Mathieson, Alexander Sher, Alan M Litke, Liam Paninski, Eero P Simoncelli, et al. Mapping nonlinear receptive field structure in primate retina at single cone resolution. *eLife*, 4:e05241, 2015.

Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, 11(2):127–38, February 2010.

Karl J Friston. Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.

Deep Ganguli and Eero P Simoncelli. Implicit encoding of prior probabilities in optimal neural populations. *Advances in Neural Information Processing Systems*, pages 6–9, 2010.

Peiran Gao and Surya Ganguli. On simplicity and complexity in the brave new world of large-scale neuroscience. *Current Opinion in Neurobiology*, 32:148–155, 2015.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis*. CRC press, 3rd edition, 2013.

Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741, 1984.

Felipe Gerhard, Tilman Kispersky, Gabrielle J Gutierrez, Eve Marder, Mark Kramer, and Uri Eden. Successful reconstruction of a physiological circuit with known connectivity from spiking activity alone. *PLoS Computational Biology*, 9(7):e1003138, 2013.

Samuel J Gershman, Matthew D Hoffman, and David M Blei. Nonparametric variational inference. *Proceedings of the International Conference on Machine Learning*, pages 663–670, 2012a.

Samuel J Gershman, Edward Vul, and Joshua B Tenenbaum. Multistability and perceptual inference. *Neural Computation*, 24(1):1–24, 2012b.

Sebastian Gerwinn, Jakob Macke, Matthias Seeger, and Matthias Bethge. Bayesian inference for spiking neuron models with a sparsity prior. *Advances in Neural Information Processing Systems*, pages 529–536, 2008.

Charles J Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, pages 473–483, 1992.

Walter R Gilks. *Markov Chain Monte Carlo*. Wiley Online Library, 2005.

Anna Goldenberg, Alice X Zheng, Stephen E Fienberg, and Edoardo M Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233, 2010.

Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1019–1028, 2010.

Noah Goodman, Vikash Mansinghka, Daniel M Roy, Keith Bonawitz, and Joshua B Tenenbaum. Church: a language for generative models. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 220––229, 2008.

Noah D Goodman, Joshua B Tenenbaum, and Tobias Gerstenberg. Concepts in a probabilistic language of thought. Technical report, Center for Brains, Minds and Machines (CBMM), 2014.

Agnieszka Grabska-Barwinska, Jeff Beck, Alexandre Pouget, and Peter Latham. Demixing odors-fast inference in olfaction. *Advances in Neural Information Processing Systems*, pages 1968–1976, 2013.

SG Gregory, KF Barlow, KE McLay, R Kaul, D Swarbreck, A Dunham, CE Scott, KL Howe, K Woodfine, CCA Spencer, et al. The DNA sequence and biological annotation of human chromosome 1. *Nature*, 441(7091):315–321, 2006.

Thomas L Griffiths, Charles Kemp, and Joshua B Tenenbaum. Bayesian models of cognition. In Ron Sun, editor, *The Cambridge Handbook of Computational Psychology*. Cambridge University Press, 2008.

Roger B Grosse, Chris J Maddison, and Ruslan R Salakhutdinov. Annealing between distributions by averaging moments. *Advances in Neural Information Processing Systems*, pages 2769–2777, 2013.

Roger B Grosse, Zoubin Ghahramani, and Ryan P Adams. Sandwiching the marginal likelihood using bidirectional Monte Carlo. *arXiv preprint arXiv:1511.02543*, 2015.

Yong Gu, Dora E Angelaki, and Gregory C DeAngelis. Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, 11(10):1201–1210, 2008.

Fangjian Guo, Charles Blundell, Hanna Wallach, and Katherine A Heller. The Bayesian echo chamber: Modeling influence in conversations. *arXiv preprint arXiv:1411.2674*, 2014.

Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83, 1971.

Moritz Helmstaedter, Kevin L Briggman, Srinivas C Turaga, Viren Jain, H Sebastian Seung, and Winfried Denk. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461):168–174, 2013.

Geoffrey E Hinton. How neural networks learn from experience. *Scientific American*, 1992.

Geoffrey E Hinton and Terrence J Sejnowski. Optimal perceptual inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1983.

Daniel R Hochbaum, Yongxin Zhao, Samouil L Farhi, Nathan Klapoetke, Christopher A Werley, Vikrant Kapoor, Peng Zou, Joel M Kralj, Dougal Maclaurin, Niklas Smedemark-Margulies, et al. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nature Methods*, 2014.

Peter D Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in Neural Information Processing Systems*, 20:1–8, 2008.

Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

Douglas N. Hoover. Relations on probability spaces and arrays of random variables. Technical report, Institute for Advanced Study, Princeton, 1979.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

Patrik O Hoyer and Aapo Hyvarinen. Interpreting neural response variability as Monte Carlo sampling of the posterior. *Advances in neural information processing systems*, pages 293–300, 2003.

Yanping Huang and Rajesh P. N. Rao. Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5):580–593, September 2011.

David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962.

Hemant Ishwaran and Mahmoud Zarepour. Exact and approximate sum representations for the Dirichlet process. *Canadian Journal of Statistics*, 30(2):269–283, 2002.

Tomoharu Iwata, Amar Shah, and Zoubin Ghahramani. Discovering latent influence in online social activities via shared cascade Poisson processes. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 266–274, 2013.

Mehrdad Jazayeri and Michael N Shadlen. Temporal context calibrates interval timing. *Nature Neuroscience*, 13(8):1020–1026, 2010.

Mehrdad Jazayeri and Michael N Shadlen. A neural mechanism for sensing and reproducing a time interval. *Current Biology*, 25(20):2599–2609, 2015.

Matthew J Johnson. *Bayesian time series models and scalable inference*. PhD thesis, Massachusetts Institute of Technology, June 2014.

Matthew J Johnson and Alan S Willsky. Bayesian nonparametric hidden semi-Markov models. *Journal of Machine Learning Research*, 14(1):673–701, 2013.

Matthew J Johnson and Alan S Willsky. Stochastic variational inference for Bayesian time series models. *Proceedings of the International Conference on Machine Learning*, 32:1854–1862, 2014.

Matthew J Johnson, Scott W Linderman, Sandeep R Datta, and Ryan P Adams. Discovering switching autoregressive dynamics in neural spike train recordings. *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2015.

Lauren M Jones, Alfredo Fontanini, Brian F Sadacca, Paul Miller, and Donald B Katz. Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proceedings of the National Academy of Sciences*, 104(47):18772–18777, 2007.

Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, 1999.

Eric R Kandel, James H Schwartz, Thomas M Jessell, et al. *Principles of neural science*, volume 4. McGraw-Hill New York, 2000.

David Kappel, Stefan Habenschuss, Robert Legenstein, and Wolfgang Maass. Network plasticity as Bayesian inference. *PLoS Computational Biology*, 11(11):e1004485, 2015a.

David Kappel, Stefan Habenschuss, Robert Legenstein, and Wolfgang Maass. Synaptic sampling: A Bayesian approach to neural network plasticity and rewiring. *Advances in Neural Information Processing Systems*, pages 370–378, 2015b.

Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

Jason ND Kerr and Winfried Denk. Imaging in vivo: watching the brain in action. *Nature Reviews Neuroscience*, 9(3):195–205, 2008.

Roozbeh Kiani and Michael N Shadlen. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, 324(5928):759–64, May 2009.

John F. C. Kingman. *Poisson Processes (Oxford Studies in Probability)*. Oxford University Press, January 1993. ISBN 0198536933.

David C Knill and Whitman Richards. *Perception as Bayesian inference*. Cambridge University Press, 1996.

Konrad P Körding and Daniel M Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–7, January 2004.

Alp Kucukelbir, Rajesh Ranganath, Andrew Gelman, and David Blei. Automatic variational inference in Stan. *Advances in Neural Information Processing Systems*, pages 568–576, 2015.

Stephen W Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1):37–68, 1953.

Harold W Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.

Kenneth W Latimer, Jacob L Yates, Miriam LR Meister, Alexander C Huk, and Jonathan W Pillow. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*, 349(6244):184–187, 2015.

Tai Sing Lee and David Mumford. Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 20(7):1434–1448, 2003.

Robert Legenstein and Wolfgang Maass. Ensembles of spiking neurons with noise support optimal probabilistic inference in a dynamically changing environment. *PLoS Computational Biology*, 10(10):e1003859, 2014.

William C Lemon, Stefan R Pulver, Burkhard Höckendorf, Katie McDole, Kristin Branson, Jeremy Freeman, and Philipp J Keller. Whole-central nervous system functional imaging in larval Drosophila. *Nature Communications*, 6, 2015.

Michael S Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):R53–R78, 1998.

Percy Liang, Slav Petrov, Michael I Jordan, and Dan Klein. The infinite PCFG using hierarchical Dirichlet processes. *Proceedings of Empirical Methods in Natural Language Processing*, pages 688–697, 2007.

David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, 2007.

Jeff W Lichtman, Jean Livet, and Joshua R Sanes. A technicolour approach to the connectome. *Nature Reviews Neuroscience*, 9(6):417–422, 2008.

Scott W Linderman and Ryan P. Adams. Discovering latent network structure in point process data. *Proceedings of the International Conference on Machine Learning*, pages 1413–1421, 2014.

Scott W Linderman and Ryan P Adams. Scalable Bayesian inference for excitatory point process networks. *arXiv preprint arXiv:1507.03228*, 2015.

Scott W Linderman and Ryan P Johnson, Matthew Jand Adams. Dependent multinomial models made easy: Stick-breaking with the Pólya-gamma augmentation. *Advances in Neural Information Processing Systems*, pages 3438–3446, 2015.

Scott W Linderman, Christopher H Stock, and Ryan P Adams. A framework for studying synaptic plasticity with neural spike train data. *Advances in Neural Information Processing Systems*, pages 2330–2338, 2014.

Scott W Linderman, Ryan P Adams, and Jonathan W Pillow. Inferring structured connectivity from spike trains under negative-binomial generalized linear models. *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2015.

Scott W Linderman, Matthew J Johnson, Matthew W Wilson, and Zhe Chen. A nonparametric Bayesian approach to uncovering rat hippocampal population codes during spatial navigation. *Journal of Neuroscience Methods*, 263:36–47, 2016a.

Scott W Linderman, Aaron Tucker, and Matthew J Johnson. Bayesian latent state space models of neural activity. *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2016b.

Fredrik Lindsten, Michael I Jordan, and Thomas B Schön. Ancestor sampling for particle Gibbs. *Advances in Neural Information Processing Systems*, pages 2600–2608, 2012.

Shai Litvak and Shimon Ullman. Cortical circuitry implementing graphical models. *Neural Computation*, 21(11):3010–3056, 2009.

James Robert Lloyd, Peter Orbanz, Zoubin Ghahramani, and Daniel M Roy. Random function priors for exchangeable arrays with applications to graphs and relational data. *Advances in Neural Information Processing Systems*, 2012.

Wei Ji Ma and Mehrdad Jazayeri. Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, 37:205–220, 2014.

Wei Ji Ma, Jeffrey M Beck, Peter E Latham, and Alexandre Pouget. Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11):1432–8, November 2006.

David JC MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992.

Jakob H Macke, Lars Buesing, John P Cunningham, M Yu Byron, Krishna V Shenoy, and Maneesh Sahani. Empirical models of spiking in neural populations. *Advances in neural information processing systems*, pages 1350–1358, 2011.

Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.

Vikash Mansinghka, Daniel Selsam, and Yura Perov. Venture: a higher-order probabilistic programming platform with programmable inference. *arXiv preprint arXiv:1404.0099*, 2014.

David Marr. *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press, 1982.

Paul Miller and Donald B Katz. Stochastic transitions between neural states in taste processing and decision-making. *The Journal of Neuroscience*, 30(7):2559–2570, 2010.

T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023—-1032, 1988.

Shakir Mohamed, Zoubin Ghahramani, and Katherine A Heller. Bayesian and L1 approaches for sparse unsupervised learning. *Proceedings of the International Conference on Machine Learning*, pages 751–758, 2012.

Jesper Møller, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482, 1998.

Michael L Morgan, Gregory C DeAngelis, and Dora E Angelaki. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, 59(4):662–673, 2008.

Abigail Morrison, Markus Diesmann, and Wulfram Gerstner. Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics*, 98(6):459–478, 2008.

Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

Radford M Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001.

Radford M. Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, pages 113–162, 2010.

John A Nelder and R Jacob Baker. Generalized linear models. *Encyclopedia of Statistical Sciences*, 1972.

Bernhard Nessler, Michael Pfeiffer, Lars Buesing, and Wolfgang Maass. Bayesian computation emerges in generic cortical microcircuits through spike-timing-dependent plasticity. *PLoS Computational Biology*, 9(4):e1003037, 2013.

Mark EJ Newman. The structure and function of complex networks. *Society for Industrial and Applied Mathematics (SIAM) Review*, 45(2):167–256, 2003.

Krzysztof Nowicki and Tom A B Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

Seung Wook Oh, Julie A Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M Henry, et al. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214, 2014.

Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, 1982.

John O'Keefe and Lynn Nadel. *The Hippocampus as a Cognitive Map*, volume 3. Clarendon Press, 1978.

Peter Orbanz and Daniel M Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):437–461, 2015.

Peter Orbanz and Yee Whye Teh. Bayesian nonparametric models. In *Encyclopedia of Machine Learning*, pages 81–89. Springer, 2011.

Adam M Packer, Darcy S Peterka, Jan J Hirtz, Rohit Prakash, Karl Deisseroth, and Rafael Yuste. Two-photon optogenetics of dendritic spines and neural circuits. *Nature Methods*, 9(12):1202–1205, 2012.

Liam Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4):243–262, January 2004.

Liam Paninski, Yashar Ahmadian, Daniel Gil Ferreira, Shinsuke Koyama, Kamiar Rahnama Rad, Michael Vidne, Joshua Vogelstein, and Wei Wu. A new look at state-space models for neural data. *Journal of Computational Neuroscience*, 29(1-2):107–126, 2010.

Andrew V Papachristos. Murder by structure: Dominance relations and the social structure of gang homicide. *American Journal of Sociology*, 115(1):74–128, 2009.

Il Memming Park and Jonathan W Pillow. Bayesian spike-triggered covariance analysis. *Advances in Neural Information Processing Systems*, pages 1692–1700, 2011.

Patrick O Perry and Patrick J Wolfe. Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2013.

Biljana Petreska, Byron Yu, John P Cunningham, Gopal Santhanam, Stephen I Ryu, Krishna V Shenoy, and Maneesh Sahani. Dynamical segmentation of single trials from population neural data. *Advances in Neural Information Processing Systems*, pages 756–764, 2011.

David Pfau, Eftychios A Pnevmatikakis, and Liam Paninski. Robust learning of low-dimensional dynamics from large neural ensembles. *Advances in Neural Information Processing Systems*, pages 2391–2399, 2013.

Jonathan W. Pillow and James Scott. Fully Bayesian inference for neural models with negative-binomial spiking. *Advances in Neural Information Processing Systems*, pages 1898–1906, 2012.

Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

Eftychios A Pnevmatikakis, Daniel Soudry, Yuanjun Gao, Timothy A Machado, Josh Merel, David Pfau, Thomas Reardon, Yu Mu, Clay Lacefield, Weijian Yang, et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron*, 2016.

Nicholas G Polson, James G Scott, and Jesse Windle. Bayesian inference for logistic models using Pólya-gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013.

Ruben Portugues, Claudia E Feierstein, Florian Engert, and Michael B Orger. Whole-brain activity maps reveal stereotyped, distributed networks for visuomotor behavior. *Neuron*, 81(6):1328–1343, 2014.

Alexandre Pouget, Jeffrey M Beck, Wei Ji Ma, and Peter E Latham. Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, 16(9):1170–1178, 2013.

Robert Prevedel, Young-Gyu Yoon, Maximilian Hoffmann, Nikita Pak, Gordon Wetzstein, Saul Kato, Tina Schrödel, Ramesh Raskar, Manuel Zimmer, Edward S Boyden, et al. Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy. *Nature Methods*, 11(7):727–730, 2014.

Lawrence R Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

Adrian E Raftery and Steven Lewis. How many iterations in the Gibbs sampler? *Bayesian Statistics*, pages 763–773, 1992.

Rajesh Ranganath, Sean Gerrish, and David M Blei. Black box variational inference. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 33:275––283, 2014.

Rajesh P. N. Rao. Bayesian computation in recurrent neural circuits. *Neural Computation*, 16(1):1–38, January 2004.

Rajesh P. N. Rao. Neural models of Bayesian belief propagation. In *Bayesian brain: Probabilistic approaches to neural computation*, pages 236–264. MIT Press Cambridge, MA, 2007.

Rajesh P. N. Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1): 79–87, January 1999.

Danilo J Rezende, Daan Wierstra, and Wulfram Gerstner. Variational learning for recurrent spiking networks. *Advances in Neural Information Processing Systems*, pages 136–144, 2011.

Fred Rieke, David Warland, Rob de Ruyter van Steveninck, and William Bialek. *Spikes: exploring the neural code*. MIT press, 1999.

Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

Dan Roth. On the hardness of approximate reasoning. *Artificial Intelligence*, 82(1):273–302, 1996.

Maneesh Sahani. *Latent variable models for neural data analysis*. PhD thesis, California Institute of Technology, 1999.

Maneesh Sahani and Peter Dayan. Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity. *Neural Computation*, 2279:2255–2279, 2003.

Joshua R Sanes and Richard H Masland. The types of retinal ganglion cells: current status and implications for neuronal classification. *Annual Review of Neuroscience*, 38:221–246, 2015.

Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4: 639–650, 1994.

Ben Shababo, Brooks Paige, Ari Pakman, and Liam Paninski. Bayesian inference and online experimental design for mapping neural microcircuits. *Advances in Neural Information Processing Systems*, pages 1304–1312, 2013.

Vahid Shalchyan and Dario Farina. A non-parametric Bayesian approach for clustering and tracking non-stationarities of neural spikes. *Journal of Neuroscience Methods*, 223: 85–91, 2014.

Lei Shi and Thomas L Griffiths. Neural implementation of hierarchical Bayesian inference by importance sampling. *Advances in Neural Information Processing Systems*, 2009.

Yousheng Shu, Andrea Hasenstaub, and David A McCormick. Turning on and off recurrent balanced cortical activity. *Nature*, 423(6937):288–293, 2003.

Jack W Silverstein. The spectral radii and norms of large dimensional non-central random matrices. *Stochastic Models*, 10(3):525–532, 1994.

Aleksandr Simma and Michael I Jordan. Modeling events with cascades of Poisson processes. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2010.

Eero P Simoncelli. Optimal estimation in sensory systems. *The Cognitive Neurosciences, IV*, 2009.

Anne C Smith and Emery N Brown. Estimating a state-space model from point process observations. *Neural Computation*, 15(5):965–91, May 2003.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, pages 2951–2959, 2012.

Sen Song, Kenneth D Miller, and Lawerence F Abbott. Competitive Hebbian learning through spike-timing-dependent synaptic plasticitye. *Nature Neuroscience*, 3(9):919–26, September 2000. ISSN 1097-6256.

Daniel Soudry, Suraj Keshri, Patrick Stinson, Min-hwan Oh, Garud Iyengar, and Liam Paninski. Efficient "shotgun" inference of neural connectivity from highly sub-sampled

activity data. *PLoS Computational Biology*, 11(10):1–30, 10 2015. doi: 10.1371/journal.pcbi.1004464.

Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. *PLoS Computational Biology*, 1(4):e42, 2005.

Olav Stetter, Demian Battaglia, Jordi Soriano, and Theo Geisel. Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. *PLoS Computational Biology*, 8(8):e1002653, 2012.

Ian Stevenson and Konrad Koerding. Inferring spike-timing-dependent plasticity from spike train data. *Advances in Neural Information Processing Systems*, pages 2582–2590, 2011.

Ian H Stevenson, James M Rebesco, Nicholas G Hatsopoulos, Zach Haga, Lee E Miller, and Konrad P Körding. Bayesian inference of functional connectivity and network structure from spikes. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 17(3):203–213, 2009.

Alan A Stocker and Eero P Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–85, April 2006.

Yee Whye Teh and Michael I Jordan. Hierarchical Bayesian nonparametric models with applications. *Bayesian Nonparametrics*, pages 158–207, 2010.

Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.

Joshua B Tenenbaum, Thomas L Griffiths, and Charles Kemp. Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10(7):309–318, 2006.

Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022):1279–1285, 2011.

Luke Tierney and Joseph B Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, 1986.

Wilson Truccolo, Uri T. Eden, Matthew R. Fellows, John P. Donoghue, and Emery N. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, 93(2):1074–1089, 2005. doi: 10.1152/jn.00697.2004.

Philip Tully, Matthias Hennig, and Anders Lansner. Synaptic and nonsynaptic plasticity approximating probabilistic inference. *Frontiers in Synaptic Neuroscience*, 6(8), 2014.

Srini Turaga, Lars Buesing, Adam M Packer, Henry Dalgleish, Noah Pettit, Michael Hausser, and Jakob Macke. Inferring neural population dynamics from multiple partial recordings of the same neural circuit. *Advances in Neural Information Processing Systems*, pages 539–547, 2013.

Leslie G Valiant. *Circuits of the Mind*. Oxford University Press, Inc., 1994.

Leslie G Valiant. Memorization and association on a realistic neural model. *Neural Computation*, 17(3):527–555, 2005.

Leslie G Valiant. A quantitative theory of neural computation. *Biological Cybernetics*, 95 (3):205–211, 2006.

Jurgen Van Gael, Yunus Saatci, Yee Whye Teh, and Zoubin Ghahramani. Beam sampling for the infinite hidden Markov model. *Proceedings of the International Conference on Machine Learning*, pages 1088–1095, 2008.

Michael Vidne, Yashar Ahmadian, Jonathon Shlens, Jonathan W Pillow, Jayant Kulkarni, Alan M Litke, EJ Chichilnisky, Eero Simoncelli, and Liam Paninski. Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. *Journal of Computational Neuroscience*, 33(1):97–121, 2012.

Joshua T Vogelstein, Brendon O Watson, Adam M Packer, Rafael Yuste, Bruno Jedynak, and Liam Paninski. Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophysical Journal*, 97(2):636–655, 2009.

Joshua T Vogelstein, Adam M Packer, Timothy A Machado, Tanya Sippy, Baktash Babadi, Rafael Yuste, and Liam Paninski. Fast nonnegative deconvolution for spike train

inference from population calcium imaging. *Journal of Neurophysiology*, 104(6):3691–3704, 2010.

Hermann von Helmholtz and James Powell Cocke Southall. *Treatise on Physiological Optics: Translated from the 3rd German Ed.* Optical Society of America, 1925.

Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

Yair Weiss, Eero P Simoncelli, and Edward H Adelson. Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6):598–604, 2002.

Mike West, P Jeff Harrison, and Helio S Migon. Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, 80(389):73–83, 1985.

John G White, Eileen Southgate, J Nichol Thomson, and Sydney Brenner. The structure of the nervous system of the nematode Caenorhabditis elegans: the mind of a worm. *Philosophical Transactions of the Royal Society of London: Series B (Biological Sciences)*, 314:1–340, 1986.

Louise Whiteley and Maneesh Sahani. Attention in a Bayesian framework. *Frontiers in Human Neuroscience*, 6, 2012.

Alexander B Wiltschko, Matthew J Johnson, Giuliano Iurilli, Ralph E Peterson, Jesse M Katon, Stan L Pashkovski, Victoria E Abraira, Ryan P Adams, and Sandeep Robert Datta. Mapping sub-second structure in mouse behavior. *Neuron*, 88(6):1121–1135, 2015.

Jesse Windle, Nicholas G Polson, and James G Scott. Sampling Pólya-gamma random variates: alternate and approximate techniques. *arXiv preprint arXiv:1405.0506*, 2014.

Frank Wood and Michael J Black. A nonparametric Bayesian alternative to spike sorting. *Journal of Neuroscience Methods*, 173(1):1–12, 2008.

Frank Wood, Jan Willem van de Meent, and Vikash Mansinghka. A new approach to probabilistic programming inference. *arXiv preprint arXiv:1507.00996*, 2015.

Tianming Yang and Michael N Shadlen. Probabilistic reasoning by neurons. *Nature*, 447 (7148):1075–80, June 2007.

Byron M. Yu, John P. Cunningham, Gopal Santhanam, Stephen I. Ryu, Krishna V. Shenoy, and Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of Neurophysiology*, 102:614–635, 2009.

Alan Yuille and Daniel Kersten. Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10(7):301–308, 2006.

Richard S Zemel, Peter Dayan, and Alexandre Pouget. Probabilistic interpretation of population codes. *Neural Computation*, 10(2):403–30, February 1998.

Ke Zhou, Hongyuan Zha, and Le Song. Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 16, 2013.

Mingyuan Zhou, Lingbo Li, Lawrence Carin, and David B Dunson. Lognormal and gamma mixed negative binomial regression. *Proceedings of the International Conference on Machine Learning*, pages 1343–1350, 2012.