

FULLY BAYESIAN INFERENCE OF INTERPRETABLE STRUCTURE UNDERLYING NEURAL SPIKE TRAINS

BY SCOTT W. LINDERMAN,^{*} RYAN P. ADAMS,[†] AND JONATHAN W. PILLOW[‡]

Harvard University, Harvard University, Princeton University

With the steady expansion of neural recording capability comes the promise that unexpected patterns will be found and novel insights will be gleaned about neural computation. Toward realizing this goal, we present a negative binomial spike train model parameterized by a network of functional interactions between neurons. Under this network we introduce prior distributions characterized by interpretable latent variables associated with each neuron, for example a latent class or location. We leverage recent innovations in negative binomial regression to perform efficient fully-Bayesian inference, and show how the underlying network variables can be inferred alongside the low-dimensional latent variables of each neuron.

1. Introduction.

2. Model. We begin with a matrix of observed spike counts, $\mathbf{S} \in \mathbb{N}^{T \times N}$, where T is the number of time bins and N is the number of recorded neurons. Typically, the bin size is taken to be small enough that a only small number of spikes may be observed, for example 10ms. We will model these spike counts as random draws from a negative binomial distribution,

$$(1) \quad s_{t,n} \sim \text{NB}(\xi, p_{t,n}) = \binom{s_{t,n} + \xi - 1}{s_{t,n}} (1 - p_{t,n})^\xi p_{t,n}^{s_{t,n}},$$

for $\xi > 0$ and $p_{t,n} \in (0, 1)$. Letting $\lambda_{t,n} = \frac{p_{t,n}}{1-p_{t,n}}$, we have that $\mathbb{E}[s_{t,n}] = \xi \lambda_{t,n}$ and $\text{Var}[s_{t,n}] = \xi \lambda_{t,n} (1 + \lambda_{t,n})$. Unlike the Poisson distribution, here the variance can exceed the mean and model overdispersed spike counts. We may arrive at the negative binomial distribution by marginalizing out the firing rate, f , of the Poisson distribution in the following generative model,

$$(2) \quad f_{t,n} \sim \text{Gamma}(\xi, \lambda_{t,n})$$

$$(3) \quad s_{t,n} \sim \text{Poisson}(f_{t,n}).$$

We proceed by modeling $\lambda_{t,n}$ with a log linear model. As shown by [Zhou et al. \(2012\)](#) and previously leveraged by [Scott and Pillow \(2012\)](#), we may then capitalize on a number of nice properties of the negative binomial distribution to perform efficient inference. Let,

$$(4) \quad \psi_{t,n} \equiv \log \lambda_{t,n} = x_{t,n} + \epsilon_{t,n},$$

$$(5) \quad x_{t,n} = b_n + \sum_{n'=1}^n \sum_{\Delta t=1}^{\Delta t_{\max}} h_{n' \rightarrow n}[\Delta t] \cdot s_{t-\Delta t, n'},$$

where $b_n \in \mathbb{R}$ is a constant bias; $h_{n' \rightarrow n}(\Delta t) : (0, \Delta t_{\max}) \rightarrow \mathbb{R}$ is a causal, directed impulse response capturing the effect that spikes on neuron n' have on the activation of neuron n ; and $\epsilon_{t,n} \in \mathbb{R}$ is an additive noise term.

Following [Zhou et al. \(2012\)](#), we introduce Gaussian priors,

$$(6) \quad b_n \sim \mathcal{N}(\mu_b, \sigma_b^2),$$

$$(7) \quad \epsilon_{t,n} \sim \mathcal{N}(0, \sigma_\epsilon^2).$$

As they have shown, by augmenting our data with a set of Polya-Gamma mixing weights, the likelihood will be rendered conjugate with these Gaussian priors.

^{*}Supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216.

[†]

[‡]

(a) Erdős-Renyi

(b) SBM

(c) Latent Distance

Fig 1: Examples from a variety of network models. (a) An Erdős-Renyi random graph with independent Bernoulli entries in \mathbf{A} and Gaussian-distributed weights. (b) A stochastic block model (SBM) with 3 blocks. The block affiliations govern both the weight and the connection probability. (c) A latent distance model in which nearby neurons have dense, excitatory interactions whereas distant neurons have sparse, inhibitory interactions.

2.1. *Modeling the Network.* Traditionally, an impulse response $h_{n' \rightarrow n}(\Delta t)$ is represented by a weighted set of basis functions with Gaussian or group sparsity inducing priors on the coefficients (Pillow et al., 2008). In order to lend greater interpretability to the model we introduce a directed adjacency matrix \mathbf{A} that models the sparsity of the functional network. This *spike-and-slab* model (Mitchell and Beauchamp, 1988) allows us to reason about the sparsity and strength of the functional interactions. We begin by decomposing the directed impulse responses as,

$$(8) \quad h_{n' \rightarrow n}[\Delta t] \equiv A_{n' \rightarrow n} \cdot \mathbf{w}_{n' \rightarrow n}^T \mathbf{r}[\Delta t].$$

Here, $A_{n' \rightarrow n} \in \{0, 1\}$ is a binary random variable indicating the presence of a direct interaction from neuron n' to neuron n , $\mathbf{w}_{n' \rightarrow n} \in \mathbb{R}^B$ is a vector of weights corresponding to the B basis functions $r_b[\Delta t]$. The basis functions are specified a priori. Typical choices of basis functions include exponentially decaying functions and rectified cosine functions (Pillow et al., 2008).

2.2. *Interpretable Latent Network Structure.* Unfortunately, our reparameterized model still has $O(N^2)$ parameters. For large N , this not only exceeds the limits of what we can reasonably interpret, but fitting this model may also require more data than we can handle computationally.

To address this issue, we propose to model the network with latent variables that reflect low dimensional underlying structure. During inference, we will marginalize out the weights of the individual interactions and deal directly with the sparsity pattern and the low dimensional latent variables. We will consider three types of well known latent network models (Hoff, 2008; Lloyd et al., 2012) in detail: Erdős-Renyi graphs, stochastic block models (SBMs), and latent distance models. Figure 1 illustrates examples from these three classes.

Erdős-Renyi Networks. Erdős-Renyi networks intuitively capture sparse networks. Each weighted connection is an independent, identically distributed random variable. For example, in our model we would have,

$$(9) \quad A_{n' \rightarrow n} \sim \text{Bernoulli}(\rho),$$

$$(10) \quad \mathbf{w}_{n' \rightarrow n} \sim \mathcal{N}(\boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W).$$

In this case, the entire network can be represented by only three parameters. Though this dramatically reduces the complexity of our model, it provides little insight into the patterns of functional interaction. In a fully-Bayesian model we place beta and normal-inverse Wishart priors on the parameters, $\rho \sim \text{Beta}(a_0, b_0)$ and $\boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W \sim \text{NIW}(\boldsymbol{\mu}_0, \kappa_0, \boldsymbol{\Sigma}_0, \nu_0)$.

Stochastic Block Models (SBM's). One type of structure we might hope to find in functional networks is latent classes of neurons. For example, consider the following generative model:

$$(11) \quad c_n \sim \text{Categorical}(\mathbf{p}),$$

$$(12) \quad A_{n' \rightarrow n} \sim \text{Bernoulli}(\rho_{c_{n'} \rightarrow c_n}),$$

$$(13) \quad \mathbf{w}_{n' \rightarrow n} \sim \mathcal{N}(\boldsymbol{\mu}_{c_{n'} \rightarrow c_n}, \boldsymbol{\Sigma}_{c_{n'} \rightarrow c_n}),$$

where $\{c_n\} \in [1, \dots, C]^N$ is a set of latent “class” assignments that determine the distribution over interactions with other neurons. These latent classes might correspond to excitatory versus inhibitory neurons, or to neurons that perform similar computational roles and hence have similar patterns of interaction. For example, in the retina these classes could distinguish “ON” and “OFF” cells. Here we place a Dirichlet prior over the class probabilities, $\mathbf{p} \sim \text{Dirichlet}(\boldsymbol{\alpha})$, and conjugate priors on the remaining parameters as above.

Latent Distance Models. Finally, consider a model in which each neuron has a latent location. This location may correspond to a measured location, such as the location of a neuron on the retina, or an abstract location, such as the center of a neuron’s tuning curve in some feature space. In this generative model, the probability and strength of an interaction then depends on the distance between a pair of neurons. Let,

$$(14) \quad \ell_n \sim \mathcal{N}(\mu_L, \Sigma_L),$$

$$(15) \quad A_{n' \rightarrow n} \sim \text{Bernoulli}(f_\rho(\|\ell_n - \ell_{n'}\|)),$$

$$(16) \quad \mathbf{w}_{n' \rightarrow n} \sim \mathcal{N}(f_\mu(\|\ell_n - \ell_{n'}\|), f_\Sigma(\|\ell_n - \ell_{n'}\|)),$$

where $\ell_n \in \mathbb{R}^D$ is the latent location of the neuron, and f_\bullet are functions that map distances into probabilities, means, or variances, as appropriate. For example, we may specify that the probability of an interaction decays exponentially with distance, $f_\rho(\Delta\ell) = \exp\{-\Delta\ell/\tau\}$. More generally, we could learn these functions under a generic Gaussian process prior.

Composing Network Models. The network models described above can easily be composed to capture more complicated structure. Erdős-Renyi networks are already clearly a special case of SBM’s., but there may be cases where we would like to set an global sparsity level with an Erdős-Renyi model and then learn block structure on top of that by letting $\rho_{n' \rightarrow n} = \rho_{ER} \cdot \rho_{c_{n'} \rightarrow c_n}$. This multiplicative combination implements a logical “AND;” a logical “OR” could be implemented with a softmax function.

We could also combine SBM’s and latent distance models by, for example, we could place a latent distance model on \mathbf{A} and a SBM on \mathbf{W} by endowing neurons with both a latent class and a latent location. This is a natural model for retinal ganglion cells in which neurons have latent types that determine the sign of their interaction weights, but the probability of an interaction is largely determined by how far apart the two cells are.

3. Bayesian Inference. Zhou et al. (2012) have shown how the negative binomial distribution can be seen as a Polya-Gamma mixture of normals, thereby enable efficient Bayesian inference by data augmentation. The fundamental property they exploit is that the likelihood as a function of $\psi_{t,n} = \log \lambda_{t,n}$ can be written as,

$$(17) \quad p(s_{t,n} | \psi_{t,n}, \xi) \propto \frac{(e^{\psi_{t,n}})^{s_{t,n}}}{(1 + e^{\psi_{t,n}})^{s_{t,n} + \xi}}$$

$$(18) \quad \propto \exp\left(\frac{s_{t,n} - \xi}{2} \psi_{t,n}\right) \mathbb{E}_{\omega_{t,n}} [\exp(-\omega_{t,n} \psi_{t,n}^2 / 2)],$$

where $\omega_{t,n} \sim \text{PG}(s_{t,n} + \xi, 0)$ and PG denotes the Polya-Gamma distribution. Hence, if we augment our data with Polya-Gamma random variables $\omega_{t,n}$, then conditioned on these variables the log likelihood is a quadratic function of $\psi_{t,n}$ and conjugate with a Gaussian prior. Specifically, we have,

$$(19) \quad \begin{aligned} \psi_n | \mathbf{s}_n, \boldsymbol{\omega}_n, \mathbf{x}_n, \sigma_\epsilon^2 &\sim \mathcal{N}(\mathbf{m}_\psi, \mathbf{C}_\psi), \\ \mathbf{m}_\psi &= \mathbf{C}_\psi \left[\frac{\mathbf{s}_n - \xi}{2} + \frac{1}{\sigma_\epsilon^2} \mathbf{x}_n \right], \\ \mathbf{C}_\psi &= \left(\frac{1}{\sigma_\epsilon^2} \mathbf{I} + \text{diag}(\boldsymbol{\omega}_n) \right)^{-1}. \end{aligned}$$

Notice that the conditional covariance matrix, \mathbf{C}_ψ , is diagonal, and hence the individual time points are uncorrelated. Thus, we do not need to invert the entire matrix and incur an $O(T^3)$ cost. Rather, each entry can be sampled in parallel given the appropriate computational hardware.

Similarly, the posterior distribution over $\omega_{t,n}$ can be shown to be Polya-Gamma distributed as well. Gibbs sampling from the posterior distribution then begins as follows:

$$(20) \quad \omega_{t,n} \sim \text{PG}(s_{t,n} + \xi, \psi_{t,n}).$$

Again, notice that these updates are independent of one another, and hence could be sampled in parallel. The negative binomial parameter ξ may also be sampled as in (Zhou et al., 2012).

3.1. *Gibbs Sampling the Latent Network Parameters.* Recall that,

$$(21) \quad x_{t,n} = b_n + \sum_{n'=1}^n A_{n' \rightarrow n} \left(\sum_{\Delta t=1}^{\Delta t_{\max}} s_{t-\Delta t, n'} \cdot \mathbf{r}[\Delta t] \mathbf{w}_{n' \rightarrow n} \right)$$

$$(22) \quad = b_n + \sum_{n'=1}^n \mathbf{S}_{t, n', :} \cdot \hat{\mathbf{w}}_{n' \rightarrow n},$$

where $\mathbf{S} \in \mathbb{R}^{T \times N \times B}$ is a tensor of covariates that can be precomputed given the spike train and fixed basis functions with $\mathbf{S}_{t, n', :} = \sum_{\Delta t=1}^{\Delta t_{\max}} \mathbf{r}[\Delta t] \cdot s_{t-\Delta t, n'}$. To simplify notation, we have rewritten our model in terms of the *effective weights*,

$$(23) \quad \hat{\mathbf{w}}_{n' \rightarrow n} \equiv A_{n' \rightarrow n} \cdot \mathbf{w}_{n' \rightarrow n} \sim \rho \cdot \delta(\hat{\mathbf{w}}_{n' \rightarrow n}) + (1 - \rho) \cdot \mathcal{N}(\boldsymbol{\mu}_{n' \rightarrow n}, \boldsymbol{\Sigma}_{n' \rightarrow n}).$$

Joint Gibbs update of $A_{n' \rightarrow n}$ and $\mathbf{w}_{n' \rightarrow n}$. We can improve sampling performance in this spike-and-slab model by jointly updating the binary variables $A_{n' \rightarrow n}$ and $\mathbf{w}_{n' \rightarrow n}$. To do so, we first sample,

$$(24) \quad A_{n' \rightarrow n} \sim \int p(A_{n' \rightarrow n}, \mathbf{w}_{n' \rightarrow n} \mid \boldsymbol{\psi}, \{\hat{\mathbf{w}}_{n'' \rightarrow n}\}_{n'' \neq n'}, \boldsymbol{\mu}_{n' \rightarrow n}, \boldsymbol{\Sigma}_{n' \rightarrow n}) d\mathbf{w}_{n' \rightarrow n}.$$

This integral is trivial when $A_{n' \rightarrow n} = 0$, and, since we are working with a conjugate Gaussian model, we can evaluate this integral in closed form for $A_{n' \rightarrow n} = 1$. Given on $A_{n' \rightarrow n}$, the conditional distribution of $\mathbf{w}_{n' \rightarrow n}$ is Gaussian.

The effective weights for each incoming connection to neuron n must be sampled serially. We can, however, parallelize the updates for each receiving neuron. That is, the effective weights $\{\hat{\mathbf{w}}_{n' \rightarrow n}\}_{n'=1}^N$ are conditionally independent of the effective weights $\{\hat{\mathbf{w}}_{n' \rightarrow n''}\}_{n''=1}^N$.

Gibbs updates of $\boldsymbol{\mu}_{n' \rightarrow n}$ and $\boldsymbol{\Sigma}_{n' \rightarrow n}$. Given a sample of the effective weights, the means and covariances in Erdős-Renyi and SBM networks are conjugate with a normal-inverse Wishart (NIW) prior. For latent distance models the conjugacy will depend upon the relation between means, covariances, and latent distances.

4. Collapsing Gibbs Sampling of the Latent Network Parameters. This is a work in progress. The gist is that we can integrate out the weights and sample the binary indicators given $\boldsymbol{\psi}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$, but the conditional distributions are complicated because all of the time points become correlated and the posterior distribution of $\boldsymbol{\Sigma}$ doesn't have a nice form. In short, instantiating \mathbf{w} makes conditional updates easier. Conditioning on the presence or absence of an interaction, we may replace the spike-and-slab prior with a degenerate Gaussian distribution,

$$(25) \quad \hat{\mathbf{w}}_{n' \rightarrow n} \mid A_{n' \rightarrow n}, \boldsymbol{\mu}_{n' \rightarrow n}, \boldsymbol{\Sigma}_{n' \rightarrow n} \sim \mathcal{N}(A_{n' \rightarrow n} \cdot \boldsymbol{\mu}_{n' \rightarrow n}, A_{n' \rightarrow n} \cdot \boldsymbol{\Sigma}_{n' \rightarrow n}).$$

Effectively, when $A_{n' \rightarrow n}$ equals zero the conditional distribution of the effective weights reduces to a zero-mean, zero-covariance Gaussian. Concatenating the weight vectors into $\hat{\mathbf{w}} = [\hat{\mathbf{w}}_{1 \rightarrow n}^\top, \dots, \hat{\mathbf{w}}_{N \rightarrow n}^\top]^\top \in \mathbb{R}^{NB}$, we have,

$$(26) \quad p(\hat{\mathbf{w}} \mid \{A_{n' \rightarrow n}, \boldsymbol{\mu}_{n' \rightarrow n}, \boldsymbol{\Sigma}_{n' \rightarrow n}\}_{n'=1}^N) = \mathcal{N}(\boldsymbol{\Gamma} \boldsymbol{\mu}, \boldsymbol{\Gamma} \boldsymbol{\Sigma} \boldsymbol{\Gamma}^\top)$$

where

$$(27) \quad \boldsymbol{\mu} = \begin{bmatrix} \hat{\boldsymbol{\mu}}_{1 \rightarrow n} \\ \vdots \\ \hat{\boldsymbol{\mu}}_{N \rightarrow n} \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{1 \rightarrow n} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\Sigma}_{N \rightarrow n} \end{bmatrix}, \quad \boldsymbol{\Gamma} = \begin{bmatrix} A_{1 \rightarrow n} \mathbf{I}_B & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & A_{N \rightarrow n} \mathbf{I}_B \end{bmatrix}$$

Since Equation 22 now reduces to a linear function of Gaussians, we can rewrite the prior distribution of $\boldsymbol{\psi}_n \equiv [\psi_{1,n}, \dots, \psi_{T,n}]^\top \in \mathbb{R}^T$ as a multivariate Gaussian,

$$(28) \quad p(\boldsymbol{\psi}_n | \hat{\mathbf{s}}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mu_b, \sigma_b^2) = \mathcal{N}(\boldsymbol{\mu}_\psi, \boldsymbol{\Sigma}_\psi)$$

$$(29) \quad \boldsymbol{\mu}_\psi = \mu_b \mathbf{1}_T + \hat{\mathbf{S}} \boldsymbol{\Gamma} \boldsymbol{\mu}$$

$$(30) \quad \boldsymbol{\Sigma}_\psi = \sigma_b^2 \mathbf{1} \mathbf{1}^\top + \hat{\mathbf{S}} \boldsymbol{\Gamma} \boldsymbol{\Sigma} \boldsymbol{\Gamma}^\top \hat{\mathbf{S}}^\top + \sigma_\epsilon^2 \mathbf{I}_T$$

$$(31) \quad \hat{\mathbf{S}} = [\hat{\mathbf{s}}_1, \dots, \hat{\mathbf{s}}_N] \in \mathbb{R}^{T \times NB},$$

Now, having collapsed out the weights $\mathbf{w}_{n' \rightarrow n}$, we are left with the following conditional distributions:

$$(32) \quad p(\boldsymbol{\psi}_n | \mathbf{s}_n, \boldsymbol{\omega}_n, \xi, \hat{\mathbf{S}}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mu_b, \sigma_b^2) = \mathcal{N}(\boldsymbol{\psi} | \mathbf{m}_\psi, b\mathbf{C}_\psi),$$

where $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\omega}_n)$ and

$$(33) \quad \mathbf{m}_\psi = \mathbf{C}_\psi \left(\boldsymbol{\Sigma}_\psi^{-1} \boldsymbol{\mu}_\psi + \left(\frac{\mathbf{s}_n - \xi}{2} \right) \right)$$

$$(34) \quad \mathbf{C}_\psi = \left(\boldsymbol{\Sigma}_\psi^{-1} + \boldsymbol{\Omega} \right)^{-1}$$

Unfortunately this is difficult to sample since \mathbf{C}_ψ is not diagonal. At first glance, it seems that this will require an $O(T^3)$ cost to factorize the covariance matrix. However, it also seems promising that we could leverage the fact that the covariance matrix is diagonal plus low rank to sample more efficiently. [TODO: Work out the details.]

Sampling $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, σ_b^2 , and σ_ϵ^2 . Given $\boldsymbol{\psi}$ and $\boldsymbol{\Gamma}$, the likelihood as a function of $\boldsymbol{\mu}$ is conjugate with a Gaussian prior. Sampling $\boldsymbol{\mu}$ is effectively like sampling the coefficients in a Bayesian linear regression. [TODO: Work out the details.]

Gibbs sampling $\boldsymbol{\Sigma}$ seems more challenging since $\boldsymbol{\Sigma}_\psi$ is a quadratic function of $\boldsymbol{\Sigma}$. We'll consider the case where $\boldsymbol{\Sigma} = \sigma_w^2 \mathbf{I}$. Then,

$$(35) \quad p(\boldsymbol{\psi}_n | \hat{\mathbf{s}}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mu_b, \sigma_b^2) = \mathcal{N}(\boldsymbol{\mu}_\psi, \sigma_b^2 \mathbf{1} \mathbf{1}^\top + \sigma_w^2 \hat{\mathbf{S}} \boldsymbol{\Gamma} \boldsymbol{\Gamma}^\top \hat{\mathbf{S}}^\top + \sigma_\epsilon^2 \mathbf{I}_T).$$

This does not have a simple conjugate prior for σ_w , but we can use an inverse gamma prior and slice sample this parameter. We use the same approach to sample σ_b^2 and σ_ϵ^2 .

5. Results.

6. Conclusion.

References.

- ENGELHARDT, B. E. and ADAMS, R. P. (2014). Bayesian Structured Sparsity from Gaussian Fields. *arXiv preprint arXiv:1407.2235*.
- HOFF, P. D. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in Neural Information Processing Systems 20* **20** 1–8.
- LLOYD, J. R., ORBANZ, P., GHAHRAMANI, Z. and ROY, D. M. (2012). Random function priors for exchangeable arrays with applications to graphs and relational data. *Advances in Neural Information Processing Systems* 1–9.
- MITCHELL, T. J. and BEAUCHAMP, J. J. (1988). Bayesian Variable Selection in Linear Regression. *Journal of the American Statistical Association* **83** 1023–1032.
- PILLOW, J. W., SHLENS, J., PANINSKI, L., SHER, A., LITKE, A. M., CHICHILNISKY, E. and SIMONCELLI, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454** 995–999.
- SCOTT, J. and PILLOW, J. W. (2012). Fully Bayesian inference for neural models with negative-binomial spiking. In *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds.) 1898–1906.
- ZHOU, M., LI, L., CARIN, L. and DUNSON, D. B. (2012). Lognormal and Gamma Mixed Negative Binomial Regression. In *Proceedings of the 29th International Conference on Machine Learning* 1343–1350.

SCOTT W. LINDERMAN
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
HARVARD UNIVERSITY
CAMBRIDGE, MA 02138, USA
E-MAIL: swl@seas.harvard.edu

RYAN P. ADAMS
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
HARVARD UNIVERSITY
CAMBRIDGE, MA 02138, USA
E-MAIL: rpa@seas.harvard.edu

JONATHAN PILLOW
PRINCETON NEUROSCIENCE INSTITUTE
PRINCETON UNIVERSITY
PRINCETON, NJ 08544, USA
E-MAIL: pillow@princeton.edu