Clustered factor analysis of multineuronal spike data

Lars Buesing[1], **Timothy A. Machado**[1,2], **John P. Cunningham**[1] **and Liam Paninski**[1]
[1] Department of Statistics, Center for Theoretical Neuroscience
& Grossman Center for the Statistics of Mind
[2] Howard Hughes Medical Institute & Department of Neuroscience
Columbia University, New York, NY
{lars,cunningham,liam}@stat.columbia.edu

Abstract

High-dimensional, simultaneous recordings of neural spiking activity are often explored, analyzed and visualized with the help of latent variable or factor models. Such models are however ill-equipped to extract structure beyond shared, distributed aspects of firing activity across multiple cells. Here, we extend unstructured factor models by proposing a model that discovers subpopulations or groups of cells from the pool of recorded neurons. The model combines aspects of mixture of factor analyzer models for capturing clustering structure, and aspects of latent dynamical system models for capturing temporal dependencies. In the resulting model, we infer the subpopulations and the latent factors from data using variational inference and model parameters are estimated by Expectation Maximization (EM). We also address the crucial problem of initializing parameters for EM by extending a sparse subspace clustering algorithm to integer-valued spike count observations. We illustrate the merits of the proposed model by applying it to calcium-imaging data from spinal cord neurons, and we show that it uncovers meaningful clustering structure in the data.

1 Introduction

Recent progress in large-scale techniques for recording neural activity has made it possible to study the joint firing statistics of 10^2 up to 10^5 cells at single-neuron resolution. Such data sets grant unprecedented insight into the temporal and spatial structure of neural activity and will hopefully lead to an improved understanding of neural coding and computation.

These recording techniques have spurred the development of statistical analysis tools which help to make accessible the information contained in simultaneously recorded activity time-series. Amongst these tools, latent variable models prove to be particularly useful for analyzing such data sets [1, 2, 3, 4]. They aim to capture shared structure in activity across different neurons and therefore provide valuable summary statistics of high-dimensional data that can be used for exploratory data analysis as well as for visualization purposes. The majority of latent variable models, however, being relatively general purpose tools, are not designed to extract additional structure from the data. This leads to latent variables that can be hard to interpret biologically. Furthermore, additional information from other sources, such as spatial structure or genetic cell type information, cannot be readily integrated into these models.

An approach to leveraging simultaneous activity recordings that is complementary to applying unstructured factor models, is to infer detailed circuit properties from the data. By modelling the detailed interactions between neurons in a local micro-circuit, multiple tools aim at inferring the existence, type, and strength of synaptic connections between neurons [5, 6]. In spite of algorithmic progress [7], the feasibility of this approach has only been demonstrated in circuits of up to three

neurons [8], as large scale data with ground truth connectivity is currently only rarely available. This lack of validation data sets also makes it difficult to asses the impact of model mismatch and unobserved, highly-correlated noise sources ("common input").

Here, we propose a statistical tool for analyzing multi-cell recordings that offers a middle ground between unstructured latent variable models and models for inferring detailed network connectivity. The basic goal of the model is to cluster neurons into groups based on their joint activity statistics. Clustering is a ubiquitous and valuable tool in statistics and machine learning as it often yields interpretable structure (a partition of the data), and is of particular relevance in neuroscience because neurons often can be categorized into distinct groups based on their morphology, physiology, genetic identity or stimulus-response properties. In many experimental setups, side-information allowing for a reliable supervised partitioning of the recorded neurons is not available. Hence, the main goal of the paper is to develop a method for clustering neurons based on their activity recordings.

We model the firing time-series of a cluster of neurons using latent factors, assuming that different clusters are described by disjoint sets of factors. The resulting model is similar to a mixture of factor analyzers [9, 10] with Poisson observations, where each mixture component describes a subpopulation of neurons. In contrast to a mixture of factor analyzers model which assumes independent factors, we put a Markovian prior over the factors, capturing temporal dependencies of neural activity as well as interactions between different clusters over time. The resulting model, which we call mixture of Poisson linear dynamical systems (mixPLDS) model, is able to capture more structure using the cluster assignments compared to latent variable models previously applied to neural recordings, while at the same time still providing low-dimensional latent trajectories for each cluster for exploratory data analysis and visualization. In contrast to the lack of connectivity ground truth for neurons from large-scale recordings, there are indeed large-scale activity recordings available that exhibit rich and biologically interpretable clustering structure, allowing for a validation of the mixPLDS model in practice.

2 Mixture of Poisson linear dynamical systems for modelling neural subpopulations

2.1 Model definition

Let y_{kt} denote the observed spike count of neuron $k = 1, \ldots, K$ in time-bin $t = 1, \ldots, T$. For the mixture of Poisson linear dynamical systems (mixPLDS) model, we assume that each neuron k belongs to exactly one of M groups (subpopulations, clusters), indicated by the discrete (categorical) variable $s_k \in \{1, \ldots, M\}$. The s_k are modelled as i.i.d.:

$$p(\mathbf{s}) \quad = \quad \prod_{k=1}^{K} p(s_k) = \prod_{k=1}^{K} \mathrm{Disc}(s_k|\phi_0), \tag{1}$$

where $\phi_0 := (\phi_0^1, \ldots, \phi_0^M)$ are the natural parameters of the categorical distribution. In the remainder of the paper we use the convention that the group-index $m = 1, \ldots, M$ is written as superscript. The activity of each subpopulation m at time t is modeled by a latent variable $\mathbf{x}_t^m \in \mathbb{R}^{d^m}$. We assume that these latent variables (we will also call them factors) are jointly normal and we model interactions between different groups by a linear dynamical system (LDS) prior:

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{x}_t^1 \\ \vdots \\ \mathbf{x}_t^M \end{pmatrix} \quad = \quad A\mathbf{x}_{t-1} + \eta_t = \begin{pmatrix} A^{11} & \cdots & A^{1M} \\ \vdots & & \vdots \\ A^{M1} & \cdots & A^{MM} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-1}^1 \\ \vdots \\ \mathbf{x}_{t-1}^M \end{pmatrix} + \eta_t, \tag{2}$$

where the block-matrices $A^{ml} \in \mathbb{R}^{d^m \times d^l}$ capture the interactions between groups m and l. The innovations η_t are i.i.d. from $\mathcal{N}(0, Q)$ and the starting distribution is given by $\mathbf{x}_1 \sim \mathcal{N}(\mu_1, Q_1)$. If neuron k belongs to group m, i.e. $s_k = m$, we model its activity y_{kt} at time t as Poisson distributed spike count with a log-rate given by an affine combination of the factors of group m:

$$z_{kt} \mid s_k = m \quad = \quad C_{k:}^m \mathbf{x}_t^m \tag{3}$$

$$y_{kt} \mid z_{kt},\ s_k \quad \sim \quad \mathrm{Poisson}(\exp(z_{kt} + b_k)), \tag{4}$$

where $\mathbf{b} \in \mathbb{R}^K$ captures the baseline of the firing rates. We denote with $C^m \in \mathbb{R}^{K \times d^m}$ the group loading matrix with rows $C^m_{k:}$ for neurons k in group m and fill in the remaining rows with 0s for all neurons not in group m. We concatenate these into the total loading matrix $C := (C^1 \ \cdots \ C^M) \in \mathbb{R}^{K \times d}$, where $d := \sum_{m=1}^M d^m$ is the total latent dimension. If the neurons are sorted with respect to their group membership, then the total loading C has block-diagonal structure. Further, we denote with $\mathbf{y}_{k:} := (y_{k,1} \ \cdots \ y_{k,T})$ the activity time series of neuron k and use an analogous notation for $\mathbf{x}^m_n := (\mathbf{x}^m_{n,1} \ \cdots \ \mathbf{x}^m_{n,T}) \in \mathbb{R}^{1 \times T}$ for $n = 1, \ldots, d^m$. The model parameters are $\theta := (A, Q, Q_1, \mu_1, C, \mathbf{b})$; we consider the hyper-parameters ϕ_0 to be given and fixed.

For known clusters $\mathbf{s}$, the mixPLDS model can be regarded as a special case of the Poisson linear dynamical system (PLDS) model [3], where the loading C is block-diagonal. For unknown group memberships $\mathbf{s}$, the mixPLDS model defined above is similar to a mixture of factor analyzers (e.g. see [9, 10]) with Poisson observations over neurons $k = 1, \ldots, K$. In the mixPLDS model however, we do not restrict the factors of the mixture components to be independent but allow for interactions over time which are modeled by a LDS.

2.2 Variational inference and parameter estimation for the mixPLDS model

When applying the mixPLDS model to data $\mathbf{y}$, we are interested in inferring the group memberships $\mathbf{s}$ and the latent trajectories $\mathbf{x}$ as well as estimating the parameters θ. For known parameters θ, the posterior $p(\mathbf{x}, \mathbf{s}|\mathbf{y}, \theta)$ (even in the special case of a single mixture component $M = 1$) is not available in closed form and needs approximating. Here we propose to approximate the posterior using variational inference with the following factorization assumption:

$$p(\mathbf{x}, \mathbf{s}|\mathbf{y}, \theta) \quad \approx \quad q(\mathbf{x})q(\mathbf{s}). \tag{5}$$

We further restrict $q(\mathbf{x})$ to be a normal distribution $q(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mathbf{m}, V)$ with mean $\mathbf{m}$ and covariance V. Under the assumption (5), $q(\mathbf{s})$ further factorizes into the product $\prod_k q(s_k)$ where $q(s_k)$ is a categorical distribution with natural parameters $\phi_k = (\phi_k^1, \ldots, \phi_k^M)$. The variational parameters $\mathbf{m}, V$ and $\phi = (\phi_1, \ldots, \phi_K)$ are obtained by maximizing the variational lower bound of the log marginal likelihood $\log p(\mathbf{y}|\theta)$:

$$\begin{aligned}
\mathcal{L}(\mathbf{m}, V, \phi, \theta) \quad = \quad & \frac{1}{2}\left(\log|V| - \mathrm{tr}[\Sigma^{-1}V] - (\mathbf{m}-\mu)^\top \Sigma^{-1}(\mathbf{m}-\mu)\right) + \sum_{k=1}^K \mathrm{D}_{\mathrm{KL}}[q(s_k)\|p(s_k)] \\
& + \sum_{m=1}^M \sum_{k=1}^K \sum_{t=1}^T \pi_k^m \left(y_{kt} h_{kt}^m - \exp(h_{kt}^m + \rho_{kt}^m/2)\right) + \text{const} \qquad (6) \\
\mathbf{h}_t^m := C^m \mathbf{m}_t + \mathbf{b}, \qquad & \rho_t^m := \mathrm{diag}(C^m V_t C^{m\top}), \qquad \pi_k^m \propto \exp(\phi_k^m),
\end{aligned}$$

where $V_t = \mathrm{Cov}_{q(\mathbf{x})}[\mathbf{x}_t]$ and $\mu \in \mathbb{R}^{dT}$, $\Sigma \in \mathbb{R}^{dT \times dT}$ are the mean and covariance of the LDS prior over $\mathbf{x}$. The first two terms in (6) are the Kullback-Leibler divergence between the prior $p(\mathbf{x}, \mathbf{s}) = p(\mathbf{x})p(\mathbf{s})$ and its approximation $q(\mathbf{x})q(\mathbf{s})$, penalizing a variational posterior that is far away from the prior. The third term in (6) is given by the expected log-likelihood of the data, promoting a posterior approximation that explains the observed data well. We optimize $\mathcal{L}$ in a coordinate ascent manner, i.e. we hold ϕ fixed and optimize jointly over $\mathbf{m}, V$ and vice versa. A naive implementation of the optimization of $\mathcal{L}$ over $\{\mathbf{m}, V\}$ is prohibitively costly for data sets with large T, as the posterior covariance V has $\mathcal{O}((dT)^2)$ elements and has to be optimized over the set of semi-definite matrices. Instead of solving this large program, we apply a method proposed in [11], where the authors show that Gaussian variational inference for latent Gaussian models with Poisson observations can be solved more efficiently using the dual problem. We generalize their approach to the mixture of Poisson observation model (3) considered here, and we also leverage the Markovian structure of the LDS prior to speed up computations (see below). In the supplementary material, we derive this approach to inference in the mixPLDS model in detail. The optimization over ϕ is available in closed form and is also given in the supplementary material. We iterate updates over $\mathbf{m}, V$ and ϕ. In practice, this method converges very quickly, often requiring only two or three iterations to reach a reasonable convergence criterion.

The most computationally intensive part of the proposed variational inference method is the update of $\mathbf{m}, V$. Using properties of the LDS prior (i.e. the prior precision Σ^{-1} is block-tri-diagonal),

we can show that evaluation of $\mathcal{L}$, its dual and the gradient of the latter all cost $\mathcal{O}(KTd + Td^3)$, which is the same complexity as Kalman smoothing in a LDS with Gaussian observations or a single iteration of Laplace inference over $\mathbf{x}$. While having the same cost as Laplace approximation, variational inference has the advantage of a non-deceasing variational lower bound $\mathcal{L}$, which can be used for monitoring convergence as well as for model comparison.

We can also get estimates for the model parameters by maximizing the lower bound $\mathcal{L}$ over θ. To this end, we interleave updates of ϕ and $\mathbf{m}, V$ with maximizations over θ. The latter corresponds to standard parameter updates in a LDS model with Poisson observations and are discussed e.g. in [3]. This procedure implements variational Expectation Maximization (VEM) in the mixPLDS model.

2.3 Initialization by Poisson subspace clustering

In principle, for a given number of groups M with given dimensions $d^1, \ldots, d^M$ one can estimate the parameters of the mixPLDS using VEM as described above. In practice we find however that this yields poor results without having reasonable initial membership assignments $\mathbf{s}$, i.e. reasonable initial values for the variational parameters ϕ. Furthermore, VEM requires the a priori specification of the latent dimensions $d^1, \ldots, d^M$. Here we show that a simple extension to an existing subspace clustering algorithm provides, given the number of groups M, a sufficiently accurate initializer for ϕ and allows for an informed choice for the dimensions $d^1, \ldots, d^M$.

We first illustrate the connection of the mixPLDS model to the subspace clustering problem (for a review of the latter see e.g. [12]). Assume that we observe the log-rates z_{kt} defined in equation (3) directly; we denote the corresponding data matrix as $Z \in \mathbb{R}^{K \times T}$. For unknown loading C, the row $Z_{k:}$ lies on a d^m-dimensional subspace spanned by the "basis-trajectories" $\mathbf{x}^m_{1,:}, \ldots, \mathbf{x}^m_{d^m,:}$, if neuron k is in group m. If $\mathbf{s}$ and $\mathbf{x}$ are unobserved, we only know that the rows of Z lie on a union of M subspaces of dimensions $d^1, \ldots, d^m$ in an ambient space of dimension T. Reconstructing the subspaces and the subspace assignments is known as a subspace clustering problem and connections to mixtures of factor analyzers have been pointed out in [13]. The authors of [13] propose to solve the subspace clustering problem by the means of the following sparse regression problem:

$$\begin{aligned} \min_{W \in \mathbb{R}^{K \times K}} \quad & \frac{1}{2}\|Z - WZ\|_F^2 + \lambda \|W\|_1 \\ \text{s.t.} \quad & \operatorname{diag}(W) = 0. \end{aligned} \tag{7}$$

This optimization can be interpreted as trying to reconstruct each row $Z_{k:}$ by the remaining rows $Z_{\setminus k:}$ using sparse reconstruction weights W. Intuitively, a point on a subspace can be reconstructed using the fewest reconstruction weights by points on the same subspace, i.e. $W_{kl} = 0$ if k and l lie on different subspaces. The symmetrized, sign-less weights $|W| + |W|^\top$ are then interpreted as the adjacency matrix of a graph and spectral clustering, with a user defined number of clusters M, is applied to obtain a subspace clustering solution. In the noise-free case (and taking $\lambda \to 0$ in eqn 7), under linear independence assumptions on the subspaces, [13] shows that this procedure recovers the correct subspace assignments.

If the matrix Z is not observed directly but only through the observation model (3), the subspace clustering approach does not directly apply. The observed data Y generated from the model (3) is corrupted by Poisson noise and furthermore the non-linear link function transforms the union of subspaces into a union of manifolds. We can circumvent these problems using the simple observation that not only Z but also the rows $C_{k:}$ of the loading matrix C lie on a union of subspaces of dimensions $d^1, \ldots, d^m$ (where the ambient space has dimension d). This can be easily seen from the block-diagonal structure of C (if the neurons are sorted by their true cluster assignments) mentioned in section 2.1. Hence we can use an estimate $\tilde{C}$ of the loading C as input to the subspace clustering optimization (7). In order to get an initial estimate $\tilde{C}$ we can use a variety of dimensionality reduction methods with exp-Poisson observations, e.g. exponential family PCA [14], a nuclear norm based method [15], subspace identification methods [16] and EM-based PLDS learning [16]; here we use the nuclear norm based method [15] for reasons that will become obvious below. Because of the non-identifiability of latent factor models, these methods only yield an estimate of $C \cdot D$ with an unknown, invertible transformation $D \in \mathbb{R}^{d \times d}$. Nevertheless, the rows of $C \cdot D$ still lie on a union of subspaces (which are however not axis-aligned anymore as is the case for C), and therefore the cluster assignments can still be recovered. Given these cluster assignments, we can get initial estimates of the non-zero rows of C^m by applying nuclear norm minimization to the individual clusters. This

method also returns a singular value spectrum associated with each subspace, which can be used to determine the dimension d^m. One can specify e.g. a threshold $\sigma_{\min}$, and determine the dimension d^m as the number of singular values $> \sigma_{\min}$.

2.4 The full parameter estimation algorithm

We briefly summarize the proposed parameter estimation algorithm for the mixPLDS model. The procedure requires the user to define the number of groups M. This choice can either be informed by biological prior knowledge or one can use standard model selection methods, such as cross-validation on the variational approximation of the marginal likelihood. We first get an initial estimate $\tilde{C}$ of the total loading matrix by nuclear-norm-penalized Poisson dimensionality reduction. Then, subspace clustering on $\tilde{C}$ yields initial group assignments. Based on these assignments, for each cluster we estimate the group dimension d^m and the group loading $\tilde{C}^m$. Keeping the cluster assignments fixed, we do a few VEM steps in the mixPLDS model with an initial estimation for the loading matrix given by $(\tilde{C}^1, \ldots, \tilde{C}^M)$. This last step provides reasonable initial parameters for the parameters A, Q, Q_1, μ_1 of the dynamical system prior. Finally, we do full VEM iterations in the mixPLDS model to refine the initial parameters. We monitor the increase of the variational lower bound $\mathcal{L}$ and use its increments in a termination criterion for the VEM iterations.

2.5 Non-negativity constraints on the loading C

Each component m of the mixPLDS model, representing a subpopulation of neurons, can be a very flexible model by itself (depending on the latent dimension d^m). This flexibility can in some situations lead to counter-intuitive clustering results. Consider the following example. Let half of the recorded neurons oscillate in phase and the remaining neurons oscillate with a phase shift of π relative to the first half. Depending on the context, we might be interested in clustering the first and second half of the neurons into separate groups reflecting oscillation phase. The mixPLDS model could however end up putting all neurons into a single cluster, by modelling them with one oscillating latent factor that has positive loadings on the first half of neurons and negative on the second half (or vice versa). We can prevent this behavior, by imposing element-wise non-negativity constraints on the loading matrix C, denoted as $C \geq 0$ (and by simultaneously constraining the latent dimensions of each group). The constraints guarantee that the influence of each factor on its group has the same sign across all neurons. The suitability of these constraints strongly depends on the biological context. In the application of the mixPLDS model in section 3.2, we found them to be essential for obtaining meaningful results.

We modify the subspace clustering initialization to respect the constraints $C \geq 0$ in the following way. Instead of solving the unconstrained reconstruction problem (7) with respect to W, we add non-negativity constraints $W \geq 0$. These sign constraints restrict the points that can be reconstructed from a given set of points to the convex cone of these points (instead of the subspace containing these points). Hence, under these assumptions, all data points in a cluster can be approximately reconstructed by a (non-negative) convex combination of some "time-series basis". We empirically observed that this yields initial loading matrix estimates with only very few negative elements (after possible row-wise sign inversions). For the full mixPLDS model we enforce $C \geq 0$ by the reparametrization $C = \exp(\chi)$ and doing VEM updates on χ.

3 Experiments

3.1 Artificial data

Here we validate the parameter estimation procedure for the mixPLDS model on artificial data. We generate 35 random ground truth mixPLDS models with $M = 3$, $d^1 = d^2 = d^3 = 2$ and 20 observed neurons per cluster. We sampled from each ground truth model a data set consisting of 4 i.i.d. trials with $T = 250$ times steps each. Ground truth parameters were generated such that the resulting data was sparse (12% of the bins non-empty). We compared the ability of different clustering methods to recover the 3 clusters from each data set. We report the results in fig. 1**A** in terms of the fraction of misclassified neurons (class labels were determined by majority vote in each cluster). We applied K-Means with careful initialization of the cluster centers [17] to the data. For K-Means, we pre-

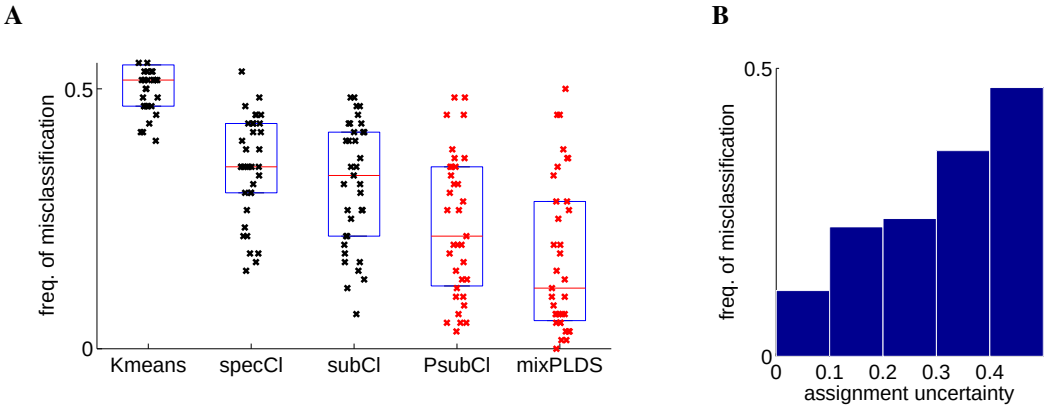

Figure 1: **Finding clusters of neurons in artificial data. A:** Performance of different clustering algorithms, reported in terms of frequency of misclassified neurons, on artificial data sampled from ground truth mixPLDs models. Red bars indicate medians and blue boxes the 25% and 75% percentiles. Standard clustering methods (data plotted in black) such as K-Means, spectral clustering ("specCl"), and subspace clustering ("subCl") are substantially outperformed by the two methods proposed here (data plotted in red). Poisson subspace clustering ("PsubCl") yielded accurate initial cluster estimates that were significantly improved by application of the full mixPLDs model. **B:** Misclassification rate as a function of the cluster assignment uncertainty for the mixPLDS model. This shows that the posterior over cluster assignments returned by the mixPLDS model is well calibrated, as neurons with low assignment uncertainty as rarely misclassified.

processed the data in a standard way by smoothing (Gaussian kernel, standard deviation 10 time-steps), mean-centering and scaling (such that each dimension $k = 1, \ldots, K$ has variance 1). We found K-Means yielded reasonable clusters when all populations are one-dimensional (i.e. $\forall m\, d^m = 1$, data not shown) but it fails when clustering multi-dimensional groups of neurons. An alternative approach is to cluster the cross-correlation matrix of neurons (computed from pre-processed data as above) with standard spectral clustering [18]. We found that this approach works well when all the factors have small variances, as in this case the link function of the observation model is only mildly non-linear. However, with growing variances of the factors (larger dynamic ranges of neurons) spectral clustering performance quickly degrades. Standard sparse subspace clustering [13] on the spike trains (pre-processed as above) yielded very similar results to spectral clustering. We found our novel Poisson subspace clustering algorithm proposed in section 2.3 to robustly outperform the other approaches, as long as reasonable amounts of data were available (roughly $T > 100$ for the above system). The mixPLDS model initialized with the Poisson subspace clustering consistently yielded the best results, as it is able to integrate information over time and denoise the observations. One advantage of the mixPLDS model is that it not only returns cluster assignments for neurons but also provides a measure of uncertainty over these assignments. However, variational inference tends to return over-confident posteriors in general and the factorization approximation (5) might yield posterior uncertainty that is uninformative. To show that the variational posterior uncertainty is well-calibrated we computed the entropy of the posterior cluster assignment $q(s_k)$ for all neurons as a measure for assignment uncertainty. We binned the neurons according to their assignment uncertainty and report the misclassification rate for each bin in fig. 1**B**. 89% of the neurons have low posterior uncertainty and reside in the first bin having a low misclassification rate of ≈ 0.1, whereas few neurons (5%) have an assignment uncertainty larger than 0.3 nats and they are misclassified with a rate of ≈ 0.4.

3.2 Calcium imaging of spinal cord neurons

We tested the mixPLDS model on calcium imaging data obtained from an in vitro, neonatal mouse spinal cord that expressed the calcium indicator GCaMP3 in all motor neurons. When an isolated spinal cord is tonically excited by a cocktail of rhythmogenic drugs (5 μM NMDA, 10 μM 5-HT, 50 μM DA), motor neurons begin to fire rhythmically. In this network state, spatially clustered ensembles of motor neurons fire in phase with each other [19]. Since multiple ensembles that have

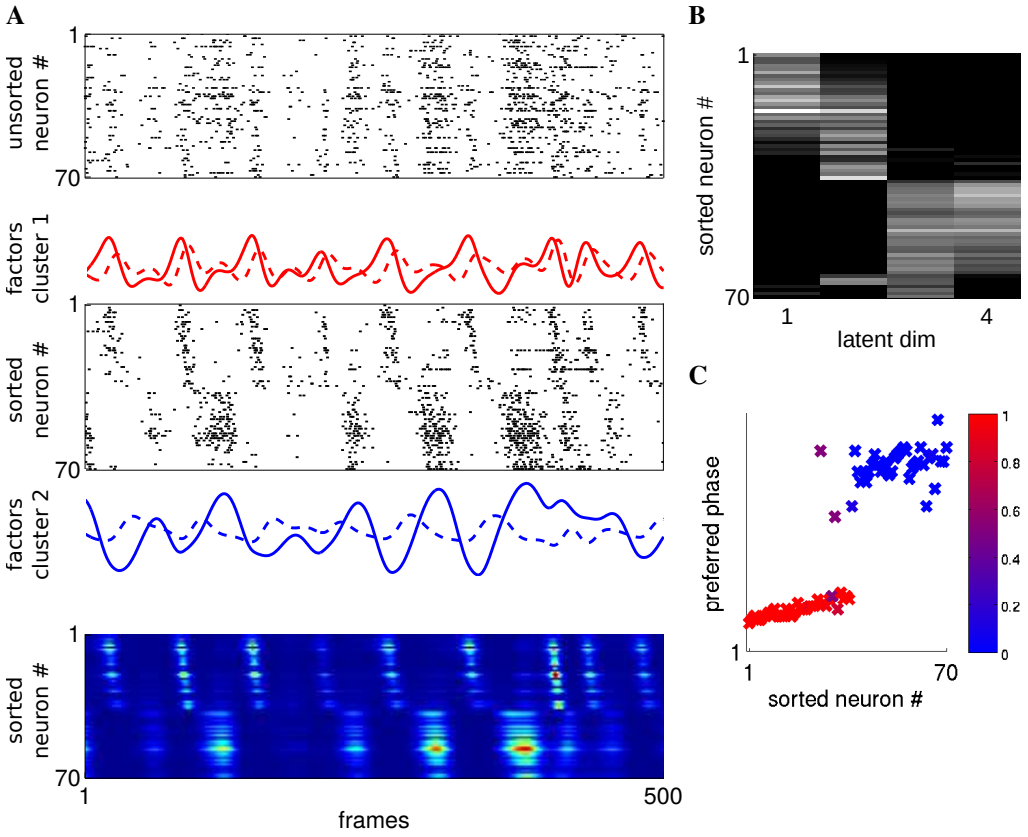

Figure 2: **Application of the mixPLDS model to recordings from spinal cord neurons. A,** top panel: 500 frames of input data to the mixPLDS model. Middle panel: Same data as in upper panel, but rows are sorted by mixPLDS clusters and factor loading. Inferred latent factors (red: cluster 1, blue: cluster 2, solid: factor 1, dashed: factor 2) are also shown. Bottom panel: Inferred (smoothed) firing rates. **B:** Loading matrix C of the mixPLDS model showing how factors 1,2 of cluster 1 and factors 3,4 of cluster 2 influence the neurons. **C:** Preferred phases shown as a function of (sorted) neuron index and colored by posterior probability of belonging to cluster 1. Clearly visible are two clusters as well as an (approximately) increasing ordering within a cluster.

distinct phase tunings can be visualized in a single imaging field, this data represents a convenient setting for testing our algorithm. The data (90 second long movies) were acquired at 15 Hz from a custom two-photon microscope equipped with a resonant scanner (downsampled from 60 Hz to boost SNR). The frequency of the rhythmic activity was typically 0.2 Hz. In addition, aggregate motor neuron activity was simultaneously acquired with each movie using a suction electrode attached to a ventral root. This electrophysiology recording (referred to here as ephys-trace) was used as an external phase reference point to compute phase tuning curves for imaged neurons, which we used to validate our mixPLDS results.

A deconvolution algorithm [20] was applied to the recorded calcium time-series to estimate the spiking activity of 70 motor neurons. The output of the deconvolution, a 70×1140 (neurons $\times$ frames) matrix of posterior expected number of spikes, was used as input to the mixPLDS model. The non-empty bins of the the first 500 out of the 1140 frames of input data (thresholded at 0.1) are shown in fig. 2A (upper panel). We used a mixPLDS model with $M = 2$ groups with two latent dimensions each, i.e. $d^1 = d^2 = 2$. We imposed the non-negativity constraints $C \geq 0$ on the loading matrix; these were found to be crucial for finding a meaningful clustering of the neurons, as discussed above. The mixPLDS clustering reveals two groups with strongly periodic but phase-shifted population activities, as can be seen from the inferred latent factors shown in fig. 2A (middle panel, factors of cluster 1 shown in red, factors of cluster 2 in blue). For each cluster, the model learned a stronger (higher variance) latent factor (solid line) and a weaker one (dashed line); we

interpret the former as capturing the main activity structure in a cluster and the latter as describing deviations. Based on the estimated mixPLDS model, we sorted the neurons for visualization into two clusters according to their most likely cluster assignment $\mathrm{argmax}_{s_k=1,2}\, q(s_k)$. Within each cluster, we sorted the neurons according to the ratio of the loading coefficient onto the stronger factor over the loading onto the weaker factor. Re-plotting the spike-raster with this sorting in fig. 2A (middle panel) reveals interesting structure. First, it shows that the initial choice of two clusters was well justified for this data set. Second, the sorting reveals that the majority of neurons tend to fire at a preferred phase relative to the oscillation cycle, and the mixPLDS-based sorting corresponds to an increasing ordering of preferred phases. Fig. 2B shows the loading matrix C of the mixPLDS, which is found to be approximately block-diagonal.

On this data set we also have the opportunity to validate the unsupervised clustering by taking into account the simultaneously recorded ephys-trace. We computed for each neuron a phase tuning curve based on the ephys-trace history of the last 80 times steps (estimated via L_2 regularized generalized linear model estimation, with an exp-Poisson observation model). For each neuron, we extracted the peak location of this phase tuning curve, which we call the preferred phase. Fig. 2C shows these preferred phases as a function of (sorted) neuron index, revealing that the two clusters found by the mixPLDS model coincide well with the two modes of the bi-model distribution of preferred phases. Furthermore, within each cluster, the preferred phases are (approximately) increasing, showing that the mixPLDS-sorting of neurons reflects the phase-relation of the neurons to the global, oscillatory ephys-trace. We emphasize that the latter was not used for fitting the mixPLDS; i.e., this constitutes an independent validation of our results.

We conclude that the mixPLDS model successfully uncovered clustering structure from the recordings that can be validated using the side information from electrophysiological tuning, and furthermore allowed for a meaningful sorting within each cluster capturing neural response properties. In addition, the mixPLDS model leverages the temporal structure in recordings, automatically optimizing for the temporal smoothness level and revealing the main time-constants in the data (in the above data set 1.8 and 6.5 sec) as well as main oscillation frequencies (0.2 and 0.45Hz). Furthermore, either the latent trajectories or the inferred firing rates shown in fig. 2A can be used as smoothed proxies for their corresponding population activities for subsequent analyses.

4 Discussion

One can generalize the mixPLDS model in several ways. Here we assumed that, given the latent factors, all neurons fire independently. This is presumably a good assumption if the recorded neurons are spatially distant, but it might break down if neurons are densely sampled from a local population and have strong, monosynaptic connections. This more general case can be accounted for by incorporating direct interaction terms between neurons into the observation model in the spirit of coupled GLMs (see [21]); inference and parameter learning are still tractable in this model using VEM. Furthermore, in addition to the activity recordings, one might have access to other covariates that are informative about the clustering structure of the population, such as cell location, genetic markers, or cell morphology. We can add such data as additional observations into the mixPLDS model to facilitate clustering of the cells. An especially relevant example are stimulus-response properties of cells. We can add a mixture model over receptive-field parameters using the cluster assignments $\mathbf{s}$. This extension would provide a clustering of neurons based on their joint activity statistics (such as shared trial-to-trial variability) as well as on their receptive field properties.

We presented three technical contributions, that we expect to be useful outside the context of the mixPLDS model. First, we proposed a simple extension of the sparse subspace clustering algorithm to Poisson observations. We showed that if the dimension of the union of subspaces is much smaller than the ambient dimension, our method substantially outperforms other approaches. Second, we introduced a version of subspace clustering with non-negativity constraints on the reconstruction weights, which therefore clusters points into convex cones. We expect this variant to be particularly useful when clustering activity traces of cells, allowing for separating anti-phasic oscillations. Third, we applied the dual variational inference approach of [11] to a model with a Markovian prior and with mixtures of Poisson observations. The resulting inference method proved itself numerically robust, and we expect it to be a valuable tool for analyzing time-series of sparse count variables.

References

[1] Anne C Smith and Emery N Brown. Estimating a state-space model from point process observations. *Neural Computation*, 15(5):965–991, 2003.

[2] 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.

[3] Jakob H Macke, Lars Buesing, John P Cunningham, M Yu Byron, Krishna V Shenoy, and Maneesh Sahani. Empirical models of spiking in neural populations. In *NIPS*, pages 1350–1358, 2011.

[4] 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. In *NIPS*, pages 1881–1888, 2008.

[5] Murat Okatan, Matthew A Wilson, and Emery N Brown. Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Computation*, 17(9):1927–1961, 2005.

[6] Yuriy Mishchenko, Joshua T Vogelstein, Liam Paninski, et al. A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data. *The Annals of Applied Statistics*, 5(2B):1229–1261, 2011.

[7] Suraj Keshri, Eftychios Pnevmatikakis, Ari Pakman, Ben Shababo, and Liam Paninski. A shotgun sampling solution for the common input problem in neural connectivity inference. *arXiv preprint arXiv:1309.3724*, 2013.

[8] 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.

[9] Michael E Tipping and Christopher M Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999.

[10] Zoubin Ghahramani, Geoffrey E Hinton, et al. The EM algorithm for mixtures of factor analyzers. Technical report, Technical Report CRG-TR-96-1, University of Toronto, 1996.

[11] Mohammad Emtiyaz Khan, Aleksandr Aravkin, Michael Friedlander, and Matthias Seeger. Fast dual variational inference for non-conjugate latent gaussian models. In *Proceedings of The 30th International Conference on Machine Learning*, pages 951–959, 2013.

[12] René Vidal. A tutorial on subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2010.

[13] Ehsan Elhamifar and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(11):2765–2781, Nov 2013.

[14] Michael Collins, Sanjoy Dasgupta, and Robert E Schapire. A generalization of principal component analysis to the exponential family. In *NIPS*, volume 13, page 23, 2001.

[15] David Pfau, Eftychios A Pnevmatikakis, and Liam Paninski. Robust learning of low-dimensional dynamics from large neural ensembles. In *NIPS*, pages 2391–2399, 2013.

[16] Lars Buesing, Jakob H Macke, and Maneesh Sahani. Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. In *NIPS*, pages 1691–1699, 2012.

[17] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[18] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering1 analysis and an algorithm. *Proceedings of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press*, 14:849–856, 2001.

[19] Timothy A. Machado, Eftychios Pnevmatikakis, Liam Paninski, Thomas M. Jessell, and Andrew Miri. Functional organization of spinal motor neurons revealed by ensemble imaging. In *79th Cold Spring Harbor Symposium on Quantitative Biology Cognition*, 2014.

[20] E. A. Pnevmatikakis, Y. Gao, D. Soudry, D. Pfau, C. Lacefield, K. Poskanzer, R. Bruno, R. Yuste, and L. Paninski. A structured matrix factorization framework for large scale calcium imaging data analysis. *ArXiv e-prints*, September 2014.

[21] 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.