STOCHASTIC DYNAMICAL SYSTEMS BASED LATENT STRUCTURE DISCOVERY IN HIGH-DIMENSIONAL TIME SERIES

Qi She, Member, IEEE and Rosa H.M. Chan, Senior Member, IEEE

Laboratory for Computational Neuroscience
City University of Hong Kong

ABSTRACT

The brain encodes information by neural spiking activities, which can be described by time series data as spike counts. Latent Variable Models (LVMs) are widely used to study the unknown factors (i.e. the latent states) that are dependent in a network structure to modulate neural spiking activities. Yet, challenges in performing experiments to record on neuronal level commonly results in relatively short and noisy spike count data, which is insufficient to derive latent network structure by existing LVMs. Specifically, it is difficult to set the number of latent states. A small number of latent states may not be able to model the complexities of underlying systems, while a large number of latent states can lead to overfitting. Therefore, based on a specific LVMs called Linear Dynamical System (LDS), we propose a Reduced-Rank Linear Dynamical System (RRLDS) to estimate latent states and retrieve an optimal latent network structure from short, noisy spike count data. This framework estimates the model using Laplace approximation. To further handle count-valued data, we introduce the *dispersion-adaptive* distribution to accommodate over-/ equal-/ and under-dispersion nature of such data. Results on both simulated and experimental data demonstrate our model can robustly learn latent space from *short-length*, noisy, *count-valued* data and significantly improve the prediction performance over the state-of-the-art methods.

Index Terms— Laplace Approximation, Reduced-Rank Structure, Dispersion-Adaptive Distribution, Linear Dynamical Systems, Count-valued data

1. INTRODUCTION

Deciphering the latent structure from high-dimensional time series data is one of the fundamental problems of Signal Processing, which has been extensively applied in various fields from social, economics, to biology science [1, 2, 3, 4, 5]. In such settings, many studies and theories posit that high-dimensional time series are noisy observations of some underlying, low-dimensional, and time-varying signal of interest [6, 7, 8, 9]. Linear Dynamical Systems (LDS) have been employed to extract a low-dimensional *implicit* network structure from multivariate time series data [7, 8, 10, 11], which captures the variability of the observed data.

However, two main challenges exist when using LDS to retrieve an optimal *implicit* network structure. First, the existing models need a predefined latent dimensionality. In order to ensure the models' capability, it is typically set to be a large value, which leads to the difficulties in modeling the *short-length* high-dimensional time series data due to overfitting. This modeling problem is troublesome since the *short-length* time series data exist in many real-world scenarios. For example, in neuroscience, we can only obtain short sequences of high-quality neural data in experiments because of the short lifetime of some neurons [12, 13, 14]. In the clinical domain, the length of patient clinical data is usually less than 50 because the hospitalization period for most patients is less than two weeks [15]. In economics, the econometric multivariate time series, such as gross domestic product and consumer price index, are measured quarterly or yearly which results in short-length data [16]. Second, real-world time series data are often count-valued (rather than real-valued). Therefore, application of standard LDS, which assumes the observation follows Gaussian distribution, is infeasible. Examples include multiple spike trains recorded from neural populations [17], the data of trades on the S&P 100 index [18], to name just a few. Further extensions on observation model to handle count nature of the data are necessary.

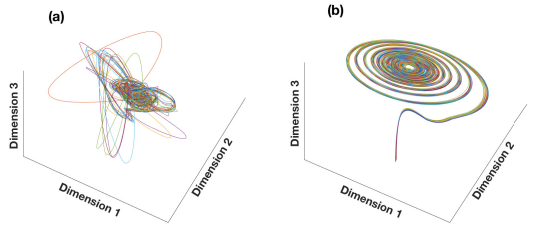


Fig. 1: Latent trajectories reconstructed from **(a)** unconstrained dynamics matrix and **(b)** reduced-rank dynamics matrix (different colors indicate different simulated trials). The low-dimensional manifold in **(b)** is smoother and better structured.

Hidden Markov Models (HMM) and Linear Dynamical System (LDS) are two latent variable models. They assume that the latent variables which make noisy measurements evolves over time. The only difference is that HMM uses a discrete state variable with arbitrary dynamics and measurements, while LDS uses a continues state variable with linear Gaussian dynamics and measurements. However, network structures with latent variables are not well studied. The dynamics matrix in LDS, which represents the dynamics of latent states, also contains structure of latent networks. We focus on the hypothesis of latent network structure, performing Expectation-Maximization algorithm to infer latent states trajectories.

Figure 1 demonstrates the advantage of reduced-rank dynamics matrix with $\mathcal{DA}$ distribution in recovering low dimensional manifolds from *short-length*, noisy *count-valued* time series data. The observation is 40-dimension (*i.e.* 40D) time series data, which is modeled with a 10D dynamics matrix (same initial states). It shows that our method successfully retrieves three intrinsic dimensionalities from the dynamics matrix, leading to a smoother and better structured manifold indicated by the three dimensional curves, while the method with unconstrained dynamics matrix fails. In summary, our contributions are three-folds:

- We can reduce redundant dimensions of latent states in the reconstruction of a more precise latent network structure.
- We introduce a count-valued exponential family distribution (called $\mathcal{DA}$ distribution) to capture the dispersion nature of

count data.

- We develop a Laplace approximation inference method to have tractable solutions for proposed model.

The rest of this paper is structured as follows. In Section 2, we generally review linear dynamical system and several works to the problems of inference and learning in LDS. In Section 3, we present our algorithm. In Section 4 and 5, we present an empirical evaluation. Finally, in Section 6, we conclude.

2. BACKGROUND AND RELATED WORK

We present a framework for both top-down assumption on latent network structure and bottom-up data-driven approach for latent states inference and parameters learning. The framework works by requesting only short-span data to efficiently train models, shutting down unnecessary latent dimensions. The hypothesis of latent network structure efficiently avoids overfitting problem with high-dimensional latent states in learning step; Inference of Non-Gaussian state space model is also a challenge. Instead of using simulation-based methods, such as particle filtering, to do the inference, Laplace approximation becomes more computationally efficient. We experimentally evaluate our framework against several baselines on synthetic data, and show that it is able to inference latent variables and learn parameters with high accuracy; also the overfitting problem is avoided with good prediction accuracy.

Linear Dynamical System (LDS) models **real-valued** multivariate time series (**MTS**) $\{\mathbf{y}_t \in \mathbb{R}^q\}_{t=1}^T$ using latent states $\{\mathbf{x}_t \in \mathbb{R}^n\}_{t=1}^T$:

$$\mathbf{x}_t|\mathbf{x}_{t-1} \sim \mathcal{N}(\mathbf{x}_t|A\mathbf{x}_{t-1}, Q), \quad (1)$$

$$\mathbf{y}_t|\mathbf{x}_t \sim \mathcal{N}(\mathbf{y}_t|C\mathbf{x}_t, R). \quad (2)$$

Eq. 1 represents state dynamics, and Eq. 2 is the observation model. Briefly, $\{\mathbf{x}_t\}$ is evolved via a dynamics matrix $A \in \mathbb{R}^{n\times n}$. Observations $\{\mathbf{y}_t\}$ are generated from $\{\mathbf{x}_t\}$ via a emission matrix $C \in \mathbb{R}^{q\times n}$. These two processes have Gaussian noise with mean $\mathbf{0}$ and covariance matrices Q and R respectively. The complete set of the LDS parameters is $\Omega = \{A, C, Q, R, \mathbf{x}_0, Q_0\}$, and learned from MTS data using Expectation-Maximization [19] or spectral learning [20].

Various regularization methods have been proposed for both time series modeling and prediction tasks with LDS [21]. These can be divided into three categories: (1) state regularization; (2) innovation regularization; and (3) parameter regularization.

Our method is different from category (1) and (2) because both of them learn a sparse representation for latent states $\{\mathbf{x}_t\}_{t=1}^T$ while they assume all parameters of LDS as a *priori*. While our method belongs to the same category (3) as the stable LDS proposed by Boots *et al.* [22], we focus on finding an appropriate state space and prevent overfitting given a small amount of MTS count data.

For learning LDS model from count observations, Busing *et al.* proposed Poisson Linear Dynamical Systems (PLDS) and to learn it using spectral learning method [20] or variational inference. The Poisson assumption of count data, while offering algorithmic conveniences, implies the conditional mean and variance of count data are equal. This property is improper in some analysis of count data, which are observed to be either over- or under-dispersed (**variance greater or less than mean**). Thus, it is crucial to develop a general observation distribution to capture over/equal/under-dispersion of count data.

To address these needs, we employ a count-valued exponential family distribution (*weighted Poisson distribution*), which is superior to current methods in two aspects: (i) adaptive dispersion, where a log-convex/linear/concave weight function can produce the expected over/equal/under-dispersion; (ii) flexible setting, with a variety of previous work are derived as special cases under this distribution.

3. REDUCED-RANK LINEAR DYNAMICAL SYSTEM

3.1. Laplace Approximation

We aim to characterize the full posterior distribution of latent states given the observations $p(\mathbf{x}_{1:T}|\mathbf{y}_{1:T})$, which is not a well-analyzed distribution such as Gaussian distribution. Hence, we implement a Gaussian approximation for it. Denote $\mathbf{x} = \mathrm{vec}(\mathbf{x}_{1:T})$ and $\mathbf{y} = \mathrm{vec}(\mathbf{y}_{1:T})$, then $p(\mathbf{x}|\mathbf{y}) \approx q(\mathbf{x}|\mu, \Sigma) = \mathcal{N}(\mathbf{x}|\mu, \Sigma)$. The mean μ and the inverse covariance matrix Σ^{-1} can be found because the log-density of approximated Gaussian distribution is unimodal,

$$\begin{aligned} \mu &= \mathrm{argmax}_{\mathbf{x}} \log q(\mathbf{x}|\mu, \Sigma) \\ \Sigma^{-1} &= -\nabla_x^2 \log q(\mathbf{x}|\mu, \Sigma). \end{aligned} \quad (3)$$

The approximated Gaussian distribution $q(\mathbf{x}|\mu, \Sigma)$ can be obtained via $p(\mathbf{x}|\mathbf{y})$.

3.2. Prior on Dynamics Matrix

In order to recover the intrinsic dimensionality from MTS datasets through the rank of dynamics matrix A, we shall choose specific priors which can induce the desired low-rank property. We have two choices of inducing a low-rank dynamics matrix: (1) a multivariate Laplacian prior and (2) a nuclear norm prior as shown in Table 1:

Prior Name	Prior Form	Regularization
Multivariate Laplacian	$\propto \exp(-\beta_1 \|\|A_i\|\|_2)$	$\beta_1 \|\|A_i\|\|_2$
Nuclear norm	$\propto \exp(-\beta_2 \|\|A\|\|_*)$	$\beta_2 \|\|A\|\|_*$

Table 1: Prior choices for dynamics matrix

[Multivariate Laplacian prior] It assumes every row in dynamics matrix A is independent of each other and has the multivariate Laplacian density. Also in order to avoid overfitting, we introduce a multivariate Gaussian prior to each element in A , which leads to the ridge regularization. Then, we combine the multivariate Laplacian prior and Gaussian prior to get a new prior $p_{\mathcal{ML}}(A)$, as

$$\log p_{\mathcal{ML}}(A) = -\beta_1 \sum_{i=1}^{n} \|A_i\|_2 - \frac{\beta_2}{2}\|A\|_F^2 + \mathrm{const}, \quad (4)$$

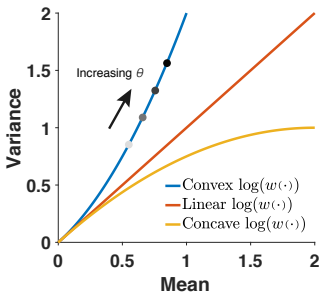
where β_1,β_2 are regularization parameters. $\{A_i\}_{i=1}^n$ indicates rows of A. $\|\cdot\|_2$ and $\|\cdot\|_F$ are ℓ_2 and Frobenius norm.

[Nuclear norm prior] It can be regarded as a convex relaxation of the number of non-zeros eigenvalues (*i.e.*,the rank) of the dynamics matrix A. We get an alternative prior $p_{\mathcal{NN}}(A)$ by applying nuclear norm density and multivariate Gaussian to dynamics matrix, as

$$\log p_{\mathcal{NN}}(A) = -\beta_3\|A\|_* - \frac{\beta_4}{2}\|A\|_F^2 + \mathrm{const}, \quad (5)$$

where $\|\cdot\|_*$ is nuclear norm. β_3, β_4 are regularization parameters. $\{\beta_i\}_{i=1}^4$ are selected (in all experiments) by the internal cross validation while optimizing model's predictive performance. We impose $p_{\mathcal{ML}}(A)$ and $p_{\mathcal{NN}}(A)$ separately to the learning process, and derive two methods to optimize a low-rank dynamics matrix.

(a)



(b)

Model	Typical parametrization	Reparametrization in $\mathcal{DA}$
Bernoulli	$p(y=k)=p^k(1-p)^{1-k}$	$w(k)=1; k=0,1$ $\theta=\text{logit}(p)$
Poisson	$p(y=k)=\frac{\lambda^k\exp(-\lambda)}{k!}$	$w(k)=1$ $\theta=\log(\lambda)$
Negative binomial	$p(y=k)=\frac{(k+r-1)!}{k!(r-1)!}(1-p)^r p^k$	$w(k)=(k+r-1)!$ $\theta=\log(p)$
COM-Poisson	$p(y=k)=\frac{\lambda^k}{(k!)^v}/\sum_{j=1}^{+\infty}\frac{\lambda^j}{(j!)^v}$	$w(k)=\exp(1-v)+k!$ $\theta=\log(\lambda)$

Fig. 2: **(a)** The mean and variance of the $\mathcal{DA}$ distribution with different choices of the function $w(\cdot)$. With a fixed $\log w(\cdot)$, increasing θ can drive mean and variance to be larger (darker dots); **(b)** Common count distributions are special cases of $\mathcal{DA}$ distribution by parameterizing θ and $w(\cdot)$.

3.3. Dispersion-Adaptive Model

For count-valued observations, we define the $\mathcal{DA}$ distribution as the family of count-valued probability distribution:

$$p_{\mathcal{DA}}(Y=k;\theta,w(\cdot))=\frac{w(k)\exp(\theta k)}{k!\mathbb{E}[w(Y)]},k\in\mathbb{N} \tag{6}$$

where $\theta\in\mathbb{R}$ and the function $w(\cdot):\mathbb{N}\to\mathbb{R}$ parameterizes the distribution, and $\mathbb{E}[w(Y)]=\sum_{k\in\mathbb{N}}\frac{w(k)\exp(\theta k)}{k!}$ is the normalizing constant. It can be viewed as an extension of Poisson distribution with a weight function $w(\cdot)$. Figure 2 (a) demonstrates different $w(\cdot)$ functions model the dispersion of count data, and controlling θ can adjust the mean value of $\mathcal{DA}$ distribution. As shown in Figure 2 (b), we derive many of the commonly used count-data distributions as special cases of $\mathcal{DA}$, by restricting the $w(\cdot)$ function and θ to have certain parametric form. Figure 2 shows that $\mathcal{DA}$ offers a rich, flexible exponential family for count data, and allows $w(\cdot)$ and θ to be interpretable for capturing statistics of count-valued data.

3.4. Reduced-Rank Linear Dynamical Systems

We apply our method to model time series data (spike counts) recorded from brain neurons, and it is straightforward to extend it to describe and interpret other count-process observations. Denoting $y^i_{t,r}$ as the spike count of neuron $i\in\{1,\dots,q\}$ at time $t\in\{1,\dots,T\}$ on experimental trial $r\in\{1,\dots,R\}$, we assume the spiking activities of neurons are noisy count observations of underlying low-dimensional latent states $\mathbf{x}_{t,r}\in\mathbb{R}^n(n<q)$ (modulating mean value of $\mathcal{DA}$ distribution) and define the $\mathcal{DA}$ observation model as:

$$y^i_{t,r}|\mathbf{x}_{t,r}\sim\mathcal{DA}(c_i^\top\mathbf{x}_{t,r},w_i(\cdot)). \tag{7}$$

We parametrize $\theta=c_i^\top\mathbf{x}_{t,r}$, where $C=[c_1,\cdots,c_q]^\top\in\mathbb{R}^{q\times n}$ is emission matrix mapping latent space to observation space. $w_i(\cdot)$ is a neuron-specific function capturing the dispersion property of each time series. The evolution of latent state $\mathbf{x}_{t,r}$ is described as:

$$\begin{aligned}\mathbf{x}_{1,r}&\sim\mathcal{N}(\mathbf{x}_{1,r}|\mathbf{x}_0,Q_0),\\ \mathbf{x}_{t,r}|\mathbf{x}_{t-1,r}&\sim\mathcal{N}(\mathbf{x}_{t,r}|A\mathbf{x}_{t-1,r}+B\mathbf{u}_{t-1,r},Q).\end{aligned} \tag{8}$$

Here, $\mathbf{x}_0$ and Q_0 are the mean and covariance of the initial state and Q is the covariance of the innovations. External input $\mathbf{u}_{t,r}$ with coupling effects B are considered in the process of latent evolution. Meanwhile, reduced-rank structures $p_{\mathcal{ML}}(A)$ and $p_{\mathcal{NN}}(A)$ are imposed on dynamics matrix A.

RRLDS is illustrated in Figure 3 along with two-stage model structure: The first stage includes reduced-rank structures composed on the dynamics matrix A, which governs the evolution of latent states $\mathbf{x}_t$. The second stage maps latent states $\mathbf{x}_t$ onto responses $\mathbf{y}_t$ via $\mathcal{DA}$ observation model, which learns the dispersion property.

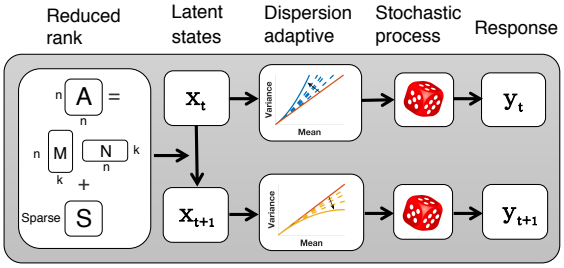


Fig. 3: Illustration of the two stages of RRLDS.

Expectation-Maximization (EM) algorithm is adopted for estimating latent states $\mathbf{x}_{1:T,r}$ (E-step) and parameters (M-step). Since the posterior distribution of $\mathbf{x}_{1:T,r}$ has no analytical solution, Laplace approximation [23] is implemented. Maximum a posteriori (MAP) estimations of dynamics matrix for two reduced-rank structures are solved using Second Order Cone Program (SOCP) [24] and generalized gradient descent algorithm [25].

3.5. Analysis of w/o Regularization

Here, we analyze the effect of regularization applied to dynamics matrix. The more data available, the influence of unnecessary dimensions will be insignificant to parameter estimation; instead, when the data sample is small, the regularization will help improve the performance remarkably. We define $(v_1,v_2,\dots,v_j)$ and $(\lambda_1,\lambda_2,\dots,\lambda_j)$ are eigenvalues/vectors of dynamics matrix $A\in\mathbb{R}^{n\times n}$. Latent variables $\mathbf{x}\in\mathbb{R}^{n\times n}$, each column of $\mathbf{x}$ can be a linear combination of $(v_1,\dots,v_j)$, as

$$\mathbf{x}=[\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n]=\left[\sum_{j=1}^n c_j^1v_j,\sum_j^n c_j^2v_j,\dots,\sum_{j=1}^n c_j^nv_j\right].$$

$\{c_j^i\}_{i=1}^n$ are the coefficients. Thus the action of A on $\mathbf{x}$ is determined by the eigenvalues/vectors as

$$\begin{aligned}A\mathbf{x}&=\left[A\sum_{j=1}^n c_j^1v_j,A\sum_{j=1}^n c_j^2v_j,\dots,A\sum_{j=1}^n c_j^nv_j\right],\\&=\left[\sum_{j=1}^n c_j^1\lambda_jv_j,\sum_{j=1}^n c_j^2\lambda_jv_j,\dots,\sum_{j=1}^n c_j^n\lambda_jv_j\right].\end{aligned}$$

Hence, the action of A on $\mathbf{x}$ is determined by the estimation of eigenvalues/vectors. The more data we have, the eigenvalue/vector pair from unnecessary dimensions of A will be estimated close to the true eigenvalue/vector, thus, the decomposition space constructed by the eigenvectors will be similar with or without regularization.

4. SIMULATED RESULTS

To demonstrate the generality of $\mathcal{DA}$ and verify our algorithmic implementation, we first test inference and learning method on extensive simulated data.

4.1. *Parameter Estimation*

Figure 4 (a) compares the eigenvalue spectrum of the estimated (by RRLDS-ML/-NN (w/o DA), PLDS) and the true dynamics matrices. The true rank of dynamics matrix is 10 (# blue circles), while the number of latent states is initialized to be 20, larger than the rank. The experiment is performed on the simulated data $\mathbf{y}$ with 40 sequences, each of which has 100 bins. It verifies that RRLDS-ML/NN (w/o DA) indeed result in a low-rank estimation of the dynamics matrix, with higher accuracy than PLDS. Given the estimated parameters, Figure 4 (b) plots the elements of stationary covariance matrix [26] for the predicted (using RRLDS-NN) and true count data, also demonstrating the accuracy of our estimation.

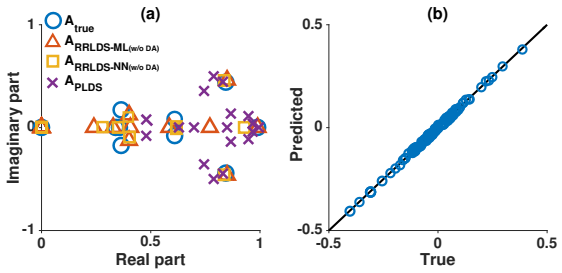


Fig. 4: **(a)** The spectrum of estimated dynamics matrices using RRLDS-ML (red triangle), RRLDS-NN (yellow square) and LDS with Poisson observation model (PLDS, purple cross). **(b)** Scatter plot of the elements in stationary covariance matrix of predicted and true count data.

5. EXPERIMENTAL RESULTS

We have taken the real dataset HC-3 [27]. The spiking activity of multiple neurons are recorded while the rat is running in a simple maze in which different visual markers are set up. The position of the rat is also recorded. Here, we have tempted to predict the neuron activity (spike count data) based on position.

5.1. Prediction performance of neural activities

Table 2 shows prediction performance (mean square error) of six LDSs (RRLDS-NN/-ML, PLDS, SubspaceID, Stable LDS, and LDS) with different predefined number of latent states for Task #1. As shown by the mean square error, while a single latent state cannot model the system well, RRLDS-NN/-ML significantly ($p < 0.001$, paired t-test) outperform the alternatives.

5.2. Retrieval of intrinsic dimensionality

We test the retrieval of intrinsic dimensionality for the complex neural system based on the *estimated* rank of dynamics matrix. In Figure 5, each subfigure plots the normalized eigenvalues of the dynamics matrices learned from different experimental trials. It is observed that given the same task, the rank of the optimized dynamics matrix consistently converges to 5 or 6 for Task #1 and 10 or 11 for Task #2, regardless of varying the number of latent states (10, 20 or 30).

This result provides a valuable insight into the internal factors of the neural system: the spiking activities in hippocampus are intrinsically characterized by an underlying low-rank dynamical system.

Table 2: Mean Square Error with Different Number of Latent States

# latent states	1	5	10	15	20	30
RRLDS-NN	**6.72**	**3.92**	**3.41**	**3.37**	**3.21**	**3.52**
RRLDS-ML	**6.74**	**3.93**	**3.43**	**3.39**	**3.22**	**3.55**
PLDS	7.35	4.50	3.96	3.93	3.45	4.32
SubspaceID	7.39	4.92	4.41	4.37	4.21	4.53
Stable LDS	7.67	5.02	5.91	5.40	5.11	5.32
LDS	8.21	7.22	8.41	8.17	7.21	7.64

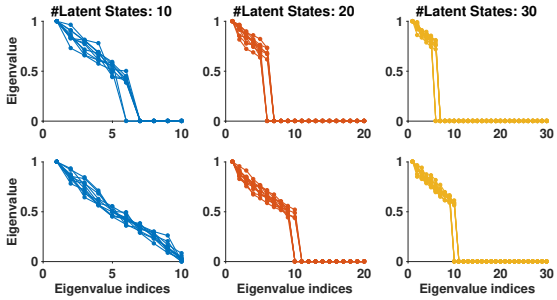


Fig. 5: Latent state space recovery from neuroscience data using RRLDS-NN. Top row: Task #1; bottom row: Task #2. Different lines in each subfigure represent different trials. 10, 20, and 30 latent states are selected for testing robustness of RRLDS.

6. CONCLUSION

Network representations derived from neural spiking activities can provide a simple measure of complex dynamics. With observed spike count data to be noisy and stochastic, a latent network composed by latent states (i.e. unknown factors) governing observed spiking activities could be utilized in modeling. Linear Dynamical System (LDS) is able to extract a low-dimensional state space which is unobserved for observed neurons. This is crucial in effective visualization and finding a latent network structure governing observations. In this paper, we analyzed latent network structure by applying prior distributions on dynamics matrix, and developed a framework for estimating states and parameters. We expect our method can benefit the learning of more concise, structured, and interpretable patterns from social science and financial data, which are often observed to be *short-length*, noisy and *count-valued*.

7. ACKNOWLEDGMENT

The work described in this paper was fully supported by grants from City University of Hong Kong (Project No. 6000555, 7004437, 7004614, 7004838, 9678060).

8. REFERENCES

[1] Qi She, Winnie KY So, and Rosa HM Chan, "Reconstruction of neural network topology using spike train data: Small-world

features of hippocampal network," in *EMBC, 2015*. IEEE, 2015, pp. 2506–2509.

[2] Winnie KY So, Lingling Yang, Beth Jelfs, Qi She, Savio WH Wong, Joseph N Mak, and Rosa HM Chan, "Cross-frequency information transfer from eeg to emg in grasping," in *EMBC, 2016*. IEEE, 2016, pp. 4531–4534.

[3] Qi She, Guanrong Chen, and Rosa HM Chan, "Evaluating the small-world-ness of a sampled network: Functional connectivity of entorhinal-hippocampal circuitry," *Scientific reports*, vol. 6, 2016.

[4] Grit Hein, Yosuke Morishima, Susanne Leiberg, Sunhae Sul, and Ernst Fehr, "The brains functional network architecture reveals human motives," *Science*, vol. 351, no. 6277, pp. 1074–1078, 2016.

[5] Qi She, Yuan Gao, Kai Xu, and Rosa HM Chan, "Reduced-rank linear dynamical systems," *AAAI*, 2018.

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

[7] Evan W Archer, Urs Koster, Jonathan W Pillow, and Jakob H Macke, "Low-dimensional models of neural population activity in sensory cortical circuits," in *NIPS*, 2014, pp. 343–351.

[8] Karthik C Lakshmanan, Patrick T Sadtler, Elizabeth C Tyler-Kabara, Aaron P Batista, and M Yu Byron, "Extracting low-dimensional latent structure from time series in the presence of delays," *Neural computation*, 2015.

[9] David Sussillo, Rafal Jozefowicz, LF Abbott, and Chethan Pandarinath, "Latent factor analysis via dynamical systems," *arXiv:1608.06315*, 2016.

[10] Yuan Gao, Jiayi Ma, and Alan L Yuille, "Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples," *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2545–2560, 2017.

[11] Mehdi Aghagolzadeh and Wilson Truccolo, "Inference and decoding of motor cortex low-dimensional dynamics via latent state-space models," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 2, pp. 272–282, 2016.

[12] Micha E Spira and Aviad Hai, "Multi-electrode array technologies for neuroscience and cardiology," *Nature nanotechnology*, vol. 8, no. 2, pp. 83–94, 2013.

[13] Qi She, Winnie KY So, and Rosa HM Chan, "Effective connectivity matrix for neural ensembles," in *EMBC, 2016*. IEEE, 2016, pp. 1612–1615.

[14] Qi She, Beth Jelfs, and Rosa HM Chan, "Modeling short over-dispersed spike count data: A hierarchical parametric empirical bayes framework," *arXiv preprint arXiv:1605.02869*, 2016.

[15] Hadi Banaee, Mobyen Uddin Ahmed, and Amy Loutfi, "Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges," *Sensors*, vol. 13, no. 12, pp. 17472–17500, 2013.

[16] Rainer Winkelmann, *Econometric analysis of count data*, Springer Science & Business Media, 2013.

[17] 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*, vol. 29, no. 1-2, pp. 107–126, 2010.

[18] Scott Linderman and Ryan Adams, "Discovering latent network structure in point process data," in *ICML*, 2014, pp. 1413–1421.

[19] Zoubin Ghahramani and Geoffrey E Hinton, "Parameter estimation for linear dynamical systems," Tech. Rep., Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science, 1996.

[20] 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*, 2012, pp. 1682–1690.

[21] Adam Charles, M Salman Asif, Justin Romberg, and Christopher Rozell, "Sparsity penalties in dynamical system estimation," in *CISS, 2011 45th Annual Conference on*. IEEE, 2011, pp. 1–6.

[22] Byron Boots, Geoffrey J Gordon, and Sajid M Siddiqi, "A constraint generation approach to learning stable linear dynamical systems," in *NIPS*, 2007, pp. 1329–1336.

[23] William D Penny, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols, *Statistical parametric mapping: the analysis of functional brain images*, Academic press, 2011.

[24] Francis Bach, Rodolphe Jenatton, Julien Mairal, Guillaume Obozinski, et al., "Convex optimization with sparsity-inducing norms," *Optimization for Machine Learning*, vol. 5, 2011.

[25] Danilo P Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 115–118, 2004.

[26] Jakob H Macke, Lars Buesing, and Maneesh Sahani, "Estimating state and parameters in state space models of spike trains," in *NIPS*, pp. 137–159. 2015.

[27] Kenji Mizuseki, Anton Sirota, Eva Pastalkova, and György Buzsáki, "Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop," *Neuron*, vol. 64, no. 2, pp. 267–280, 2009.