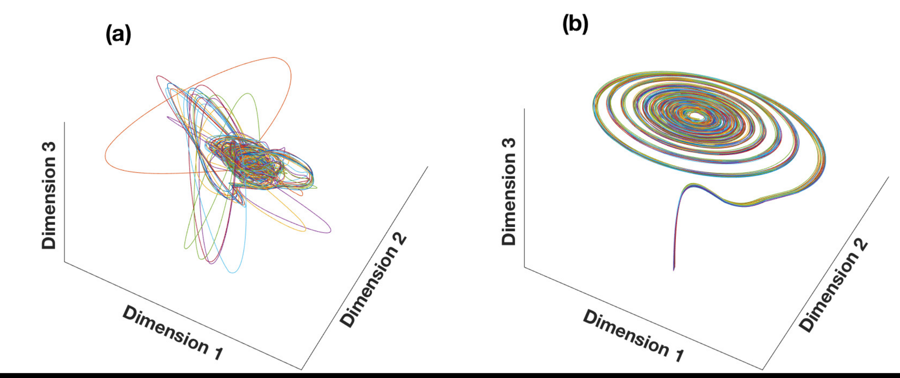
Stochastic Dynamical Systems Based Latent Structure Discovery in High-dimensional Time Series

Qi She and Rosa H.M. Chan

The Department of Electronic Engineering, City University of Hong Kong

http://sheqi.strikingly.com/



Summary

- The dynamical complexity of multivariate time series can be characterized by some underlying, low-dimensional and time-varying latent states.
- We can obtain insightful dynamical portrait of circuit computation through dimensionality reduction.
- Intractable posterior of Bayesian inference can be solved by deriving new Evidence Lower Bound.

Latent trajectories from high-dimensional time series

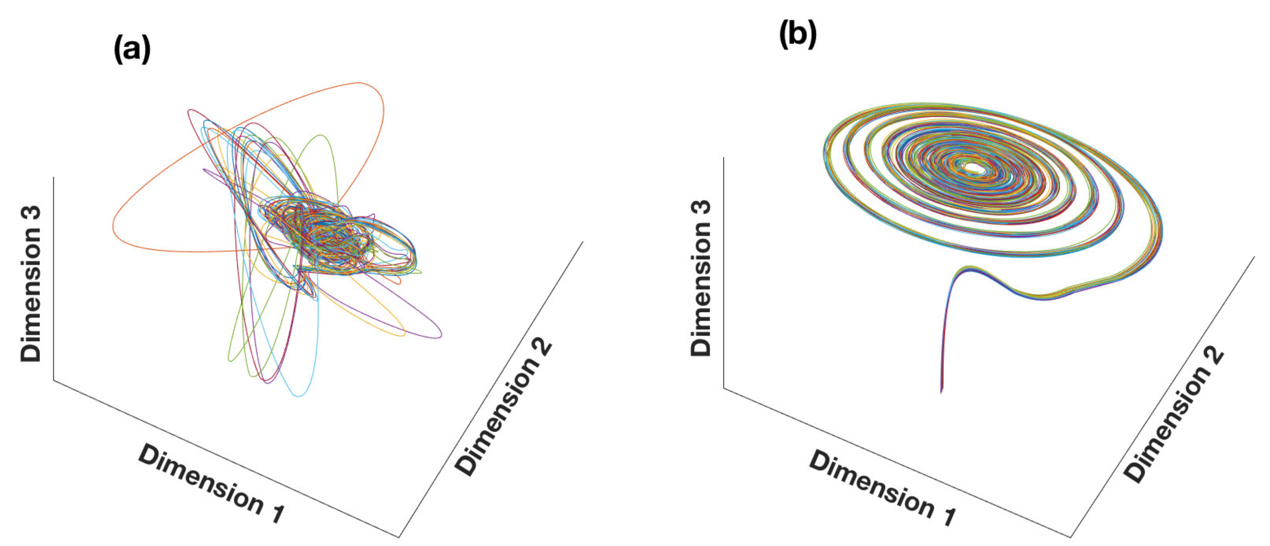
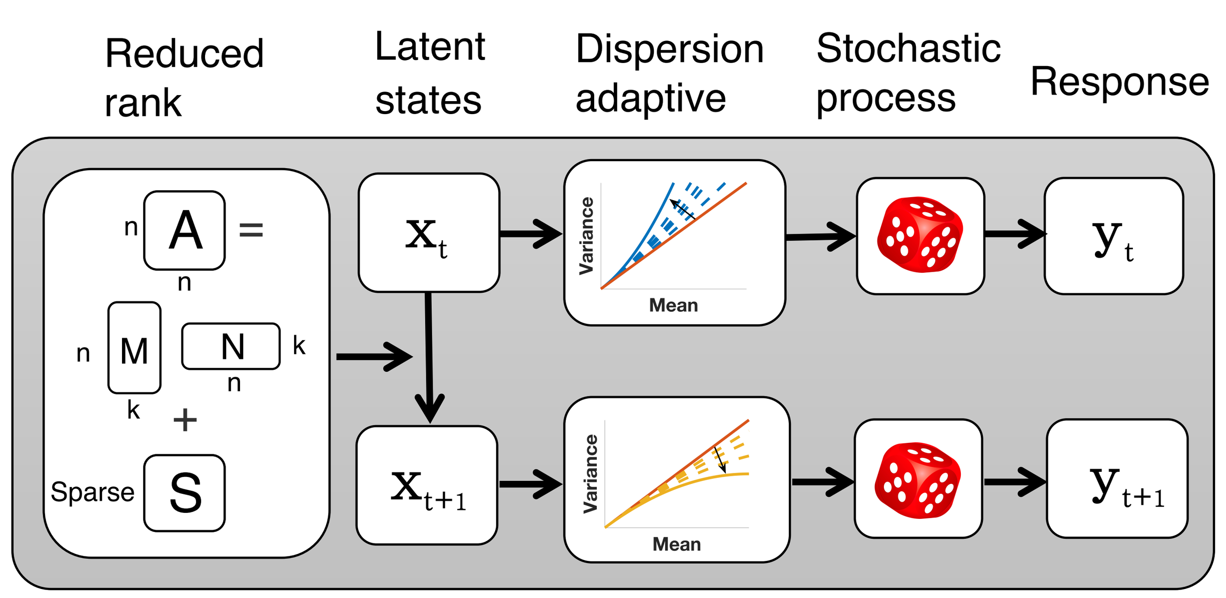


Figure 1: Latent trajectories reconstructed from **(a)** unconstrained dynamics matrix and **(b)** reduced-rank dynamics matrix (different colors indicate different simulated trials). The low-dimensional manifold in **(b)** is smoother and better structured.

Two contributions of our model



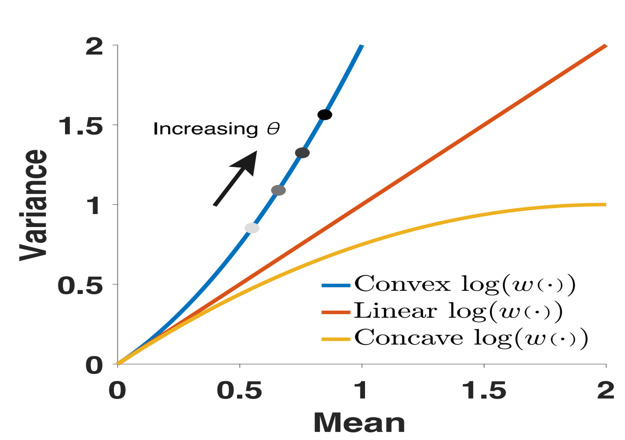
$$
\begin{aligned}
\mathbf{x}_1 &\sim \mathcal{N}(\mathbf{x}_1|\mathbf{x}_0, Q_0),\\
\mathbf{x}_{t,r}|\mathbf{x}_{t-1} &\sim \mathcal{N}(\mathbf{x}_t|A\mathbf{x}_{t-1,r} + B\mathbf{u}_{t-1,r}, Q).\\
\mathbf{y}^i_{t,r}|\mathbf{x}_{t,r} &\sim DA(c_i^\top \mathbf{x}_{t,r}, w_i(\cdot)).
\end{aligned}
$$

C1 **Reduced-rank structures** composed on the dynamics matrix A, which governs the evolution of latent states y_t.

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

C2 **Dispersion adaptive (DA)** observation model maps Latent states x_t onto responses y_t, which learns the dispersion property. Figure 2 shows the mean and variance of the 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)



$$
p(Y = k; \theta, w(\cdot)) = \frac{w(k)\exp(\theta k)}{k!\mathbb{E}[w(Y)]},\quad k \in \mathbb{N}
$$

Dispersion-adaptive distribution offers a rich, flexible exponential family for count data !

Model	Distributions setting	DA distribution setting
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	$p(y=k) = \frac{\lambda^k}{(k!)^\nu} / \sum_{j=1}^{+\infty} \frac{\lambda^j}{(j!)^\nu}$	$w(k) = \exp(1-\nu) + k!$ $\theta = \log(\lambda)$

- Common count distributions are special cases of DA distribution by parameterizing θ and $w(\cdot)$.

Model fitting: Laplace approximation and second order cone program

1 Inference: Laplace approximation
We characterize the full posterior distribution of latent states given the observations $p(\mathbf{x}_{1:T}|\mathbf{y}_{1:T})$ with a Gaussian approximation for it. Denote $\mathbf{x} = \text{vec}(\mathbf{x}_{1:T})$ and $\mathbf{y} = \text{vec}(\mathbf{y}_{1:T})$, 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,

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

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

2 Learning: maximization of $\mathscr{L}^*$ over Θ.

$$
\text{argmax}_{\Theta} \mathscr{L}(\mu, \Sigma; \bar{\mathbf{x}})
$$

3 Dynamics matrix A: Second Order Cone Program and Generalized Gradient Descent are utilized for two reduced-rank structures.

$$
\begin{aligned}
\log p_{\mathscr{ML}}(A) &= -\beta_1 \sum_{i=1}^{n} \|A_i\|_2 - \frac{\beta_2}{2}\|A\|_F^2,\\
\log p_{\mathscr{NN}}(A) &= -\beta_3 \|A\|_* - \frac{\beta_4}{2}\|A\|_F^2,
\end{aligned}
$$

4 Other dynamical system parameters $\{B, Q, Q_0, \mathbf{x}_0\}$: closed-form solution beneficial from linear Gaussian dynamics.

5 DA model parameters $\{C, \{w_i(\cdot)\}_i^p\}$: proof to have concavity property under DA setting.

Simulation results: estimation of dynamics matrix

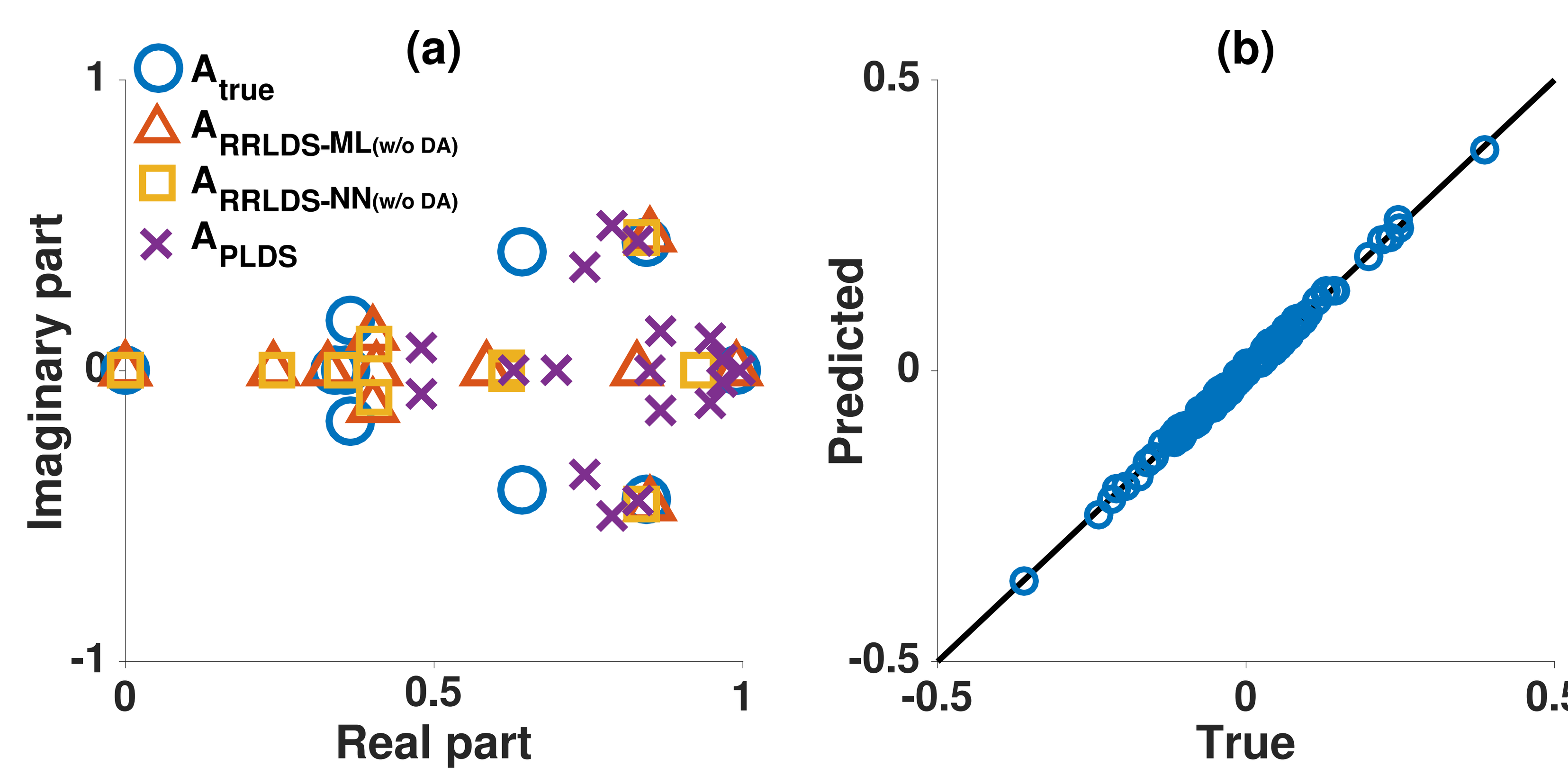


Figure 2: **(a)** The spectrum of estimated dynamics matrices using our methods (red and yellow) and LDS with Poisson observation model (PLDS, purple cross). The true complex eigenvalue spectrum is indicated with blue circles. Results of both 1&2 methods (w/o DA) are close to true eigenvalues. **(b)** Scatter plot of the elements in covariance of predicted and true count data.

Prediction performance of neural activities

(a)

	with Reduced Rank	w/o Reduced Rank
with DA	RRLDS-ML /-NN	DALDS
w/o DA	RRLDS (w/o DA)	alternative LDSs

Table 2: Abbreviations for our method and several baselines. ML stands for Multivariable Laplacian and NN for Nuclear Norm. Alternative LDS methods include vanilla LDS, Poisson-LDS, Subspace Identification and Stable LDS.

(b)

# 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

Figure 3: Mean Square Error with # Latent States

Experimental results: model comparison of neural activities

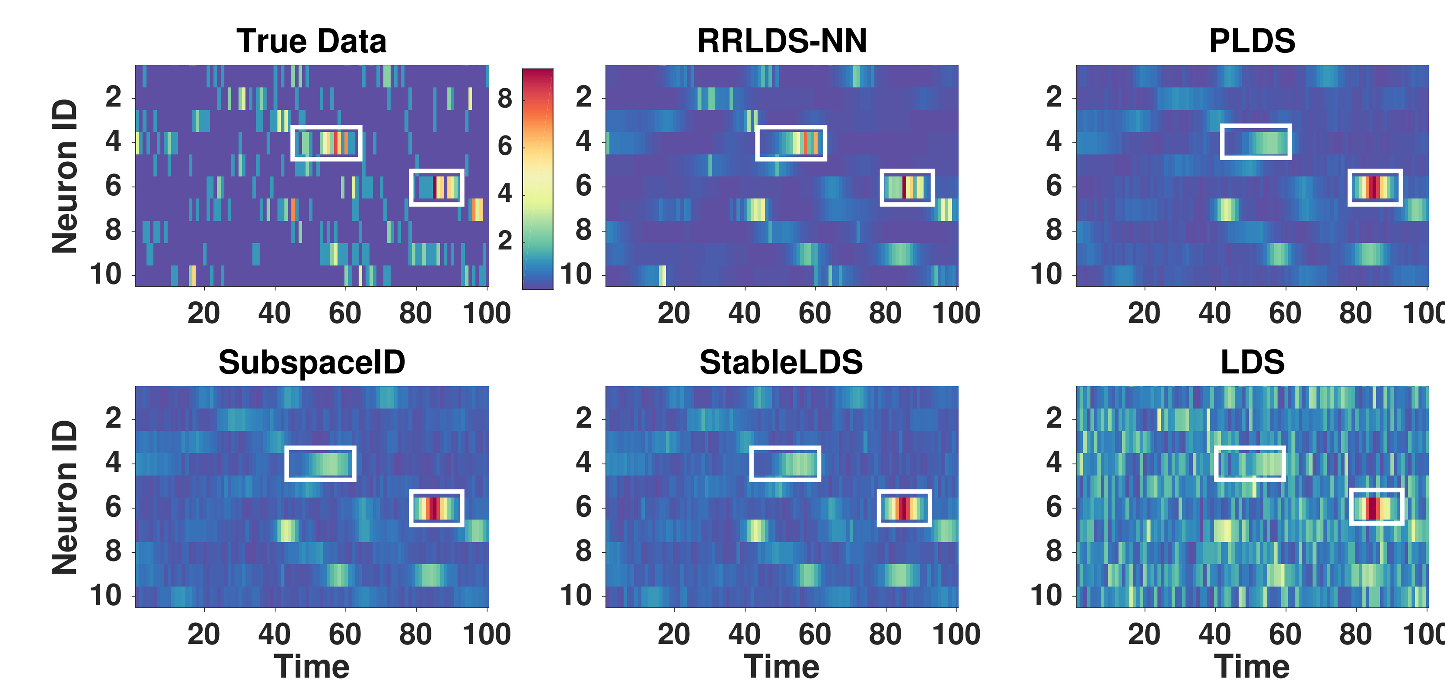


Figure 4: Prediction performance of five models for neurons' spike counts (Task #2). The rows of each subfigure indicate spiking sequence of neurons. The color highlights count values recorded/predicted at each time step.

Experimental results: retrieval of intrinsic dimensionality

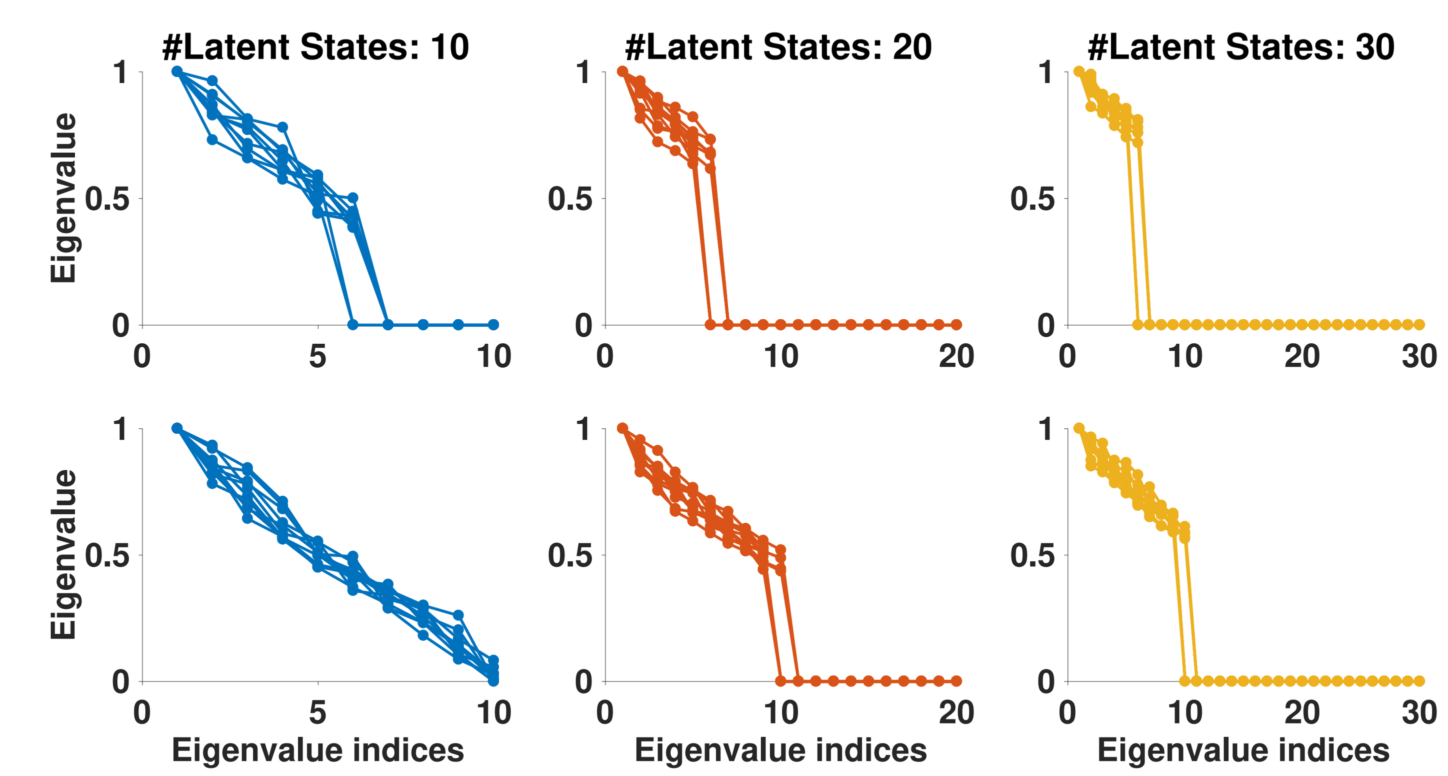


Figure 5: Latent state space recovery from neuroscience data using RRLDS-NN. Top row: Task #1; bottom row: Task #2. Different lines in each subfigure represent different trials. 10, 20, and 30 latent states are selected for testing robustness of RRLDS for retrieving intrinsic dimensionality.

Conclusion

- Linear Dynamical System (LDS) is able to extract a low-dimensional state space which is unobserved for observed neurons.
- 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*