A Differentiable POGLM with Forward-Backward Message Passing

Chengrui Li, Weihan Li, Yule Wang, Anqi Wu @ GaTech CSE

Contents

1. Partially observable generalized linear model (POGLM)
2. Variational inference (VI)
    - Inference methods
    - Sampling schemes
3. Experiments

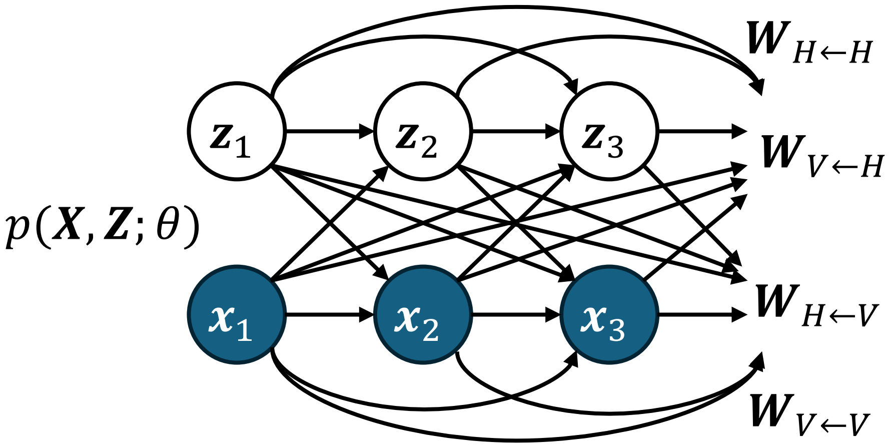
1 Partially observable generalized linear model (POGLM)

Introduction

- The partially observable generalized linear model (POGLM) is a powerful tool for understanding neural connectivity under the assumption of existing hidden neurons.
- POGLM itself is a difficult problem.
- Two main issues and our contributions:
  - The gradient estimator used in variational inference (VI).
  - The sampling scheme of the variational model.
- Comprehensive experiments on one synthetic and two real-world datasets.

POGLM

- V visible neurons and $H = N - V$ hidden neurons.



visible/hidden spike counts $\begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{z}_t \end{bmatrix} \sim \mathrm{Pois}\left(\begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{g}_t \end{bmatrix}\right)$ visible/hidden firing rates

$$\begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{g}_t \end{bmatrix} = \sigma\left(\begin{bmatrix} \boldsymbol{b}_V \\ \boldsymbol{b}_H \end{bmatrix} + \begin{bmatrix} \boldsymbol{W}_{V \leftarrow V} & \boldsymbol{W}_{V \leftarrow H} \\ \boldsymbol{W}_{H \leftarrow V} & \boldsymbol{W}_{H \leftarrow H} \end{bmatrix}\left(\sum_{l=1}^{L} \psi_l \begin{bmatrix} \boldsymbol{x}_{t-l} \\ \boldsymbol{z}_{t-l} \end{bmatrix}\right)\right)$$

bias weight convolved history

- Observed variable: visible spike train $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]^{\mathrm{T}} \in \mathbb{N}^{T \times V}$.
- Latent variable: hidden spike train $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T]^{\mathrm{T}} \in \mathbb{N}^{T \times H}$.
- Generative parameter set: $\theta = \left\{\boldsymbol{b} \in \mathbb{R}^{V+H}, \boldsymbol{W} \in \mathbb{R}^{(V+H) \times (V+H)}\right\}$.

2 Variational inference (VI)

Variational inference (VI)

- Variational model $q(\boldsymbol{Z}|\boldsymbol{X};\phi)$ parameterized by ϕ.
- **Inference methods**: $\boldsymbol{z}_t \sim \mathrm{Pois}(\boldsymbol{g}_t)$.
- **Sampling schemes**: $\boldsymbol{g}_t = \mathrm{function}(\boldsymbol{X};\phi)$.
- Evidence lower bound:

$$\mathrm{ELBO}(\boldsymbol{X};\theta,\phi) = \mathbb{E}_q[\ln p(\boldsymbol{X},\boldsymbol{Z};\theta) - \ln q(\boldsymbol{Z}|\boldsymbol{X};\phi)]$$

Inference methods: Gradient estimators

	score function	pathwise
distribution	any distribution	continuous distribution with reparameterization trick $\boldsymbol{Z}\|\boldsymbol{X};\phi = r(\boldsymbol{\epsilon}\|\boldsymbol{X};\phi)$
samples	$\{\boldsymbol{Z}^{(k)}\}_{k=1}^{K} \sim q(\boldsymbol{Z}\|\boldsymbol{X};\phi)$	$\{\boldsymbol{\epsilon}^{(k)}\}_{k=1}^{K} \sim R(\boldsymbol{\epsilon})$
$\frac{\partial \mathrm{ELBO}(\boldsymbol{X};\theta,\phi)}{\partial\phi} \approx$	$\frac{\partial}{\partial\phi}\frac{-1}{2K}\sum_{k=1}^{K}\left[\ln p\left(\boldsymbol{X},\boldsymbol{Z}^{(k)};\theta\right) - \ln q\left(\boldsymbol{Z}^{(k)}\|\boldsymbol{X};\phi\right)\right]$	$\frac{\partial}{\partial\phi}\frac{1}{K}\sum_{k=1}^{K}\left[\ln p\left(\boldsymbol{X},r\left(\boldsymbol{\epsilon}^{(k)}\|\boldsymbol{X};\phi\right);\theta\right) - \ln q\left(r\left(\boldsymbol{\epsilon}^{(k)}\|\boldsymbol{X};\phi\right)\|\boldsymbol{X};\phi\right)\right]$

Inference methods: Relaxation

Poisson $\mathrm{Pois}(g)$ —truncate at an upperbound M→ Categorical $\mathrm{Cat}(\boldsymbol{\pi}(g))$ —reparameterization approximation→ Gumbel-Softmax $\mathrm{GS}(\boldsymbol{\pi}(g);\tau)$

- Truncation:

$$\boldsymbol{\pi}(g) = \left(1 - \sum_{m=1}^{M-1} \frac{g^m \mathrm{e}^g}{m!}, \frac{g^1 \mathrm{e}^g}{1!}, \dots, \frac{g^{M-1} \mathrm{e}^g}{(M-1)!}\right)$$

- Gumbel-Softmax soft spike:

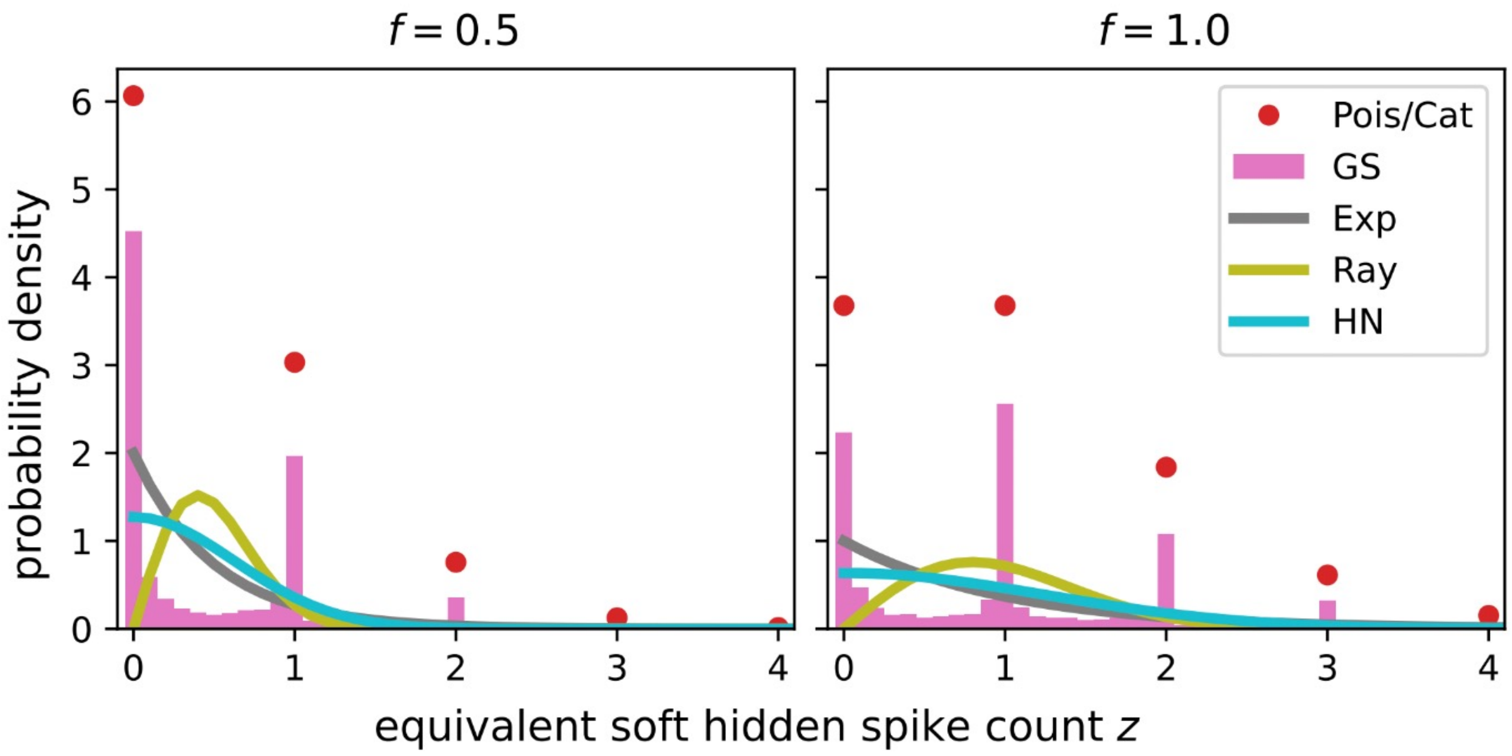
$$\tilde{\boldsymbol{z}}_{t,h} = \left(\tilde{z}_{t,h,0}, \dots, \tilde{z}_{t,h,M-1}\right) \sim \mathrm{GS}\left(\boldsymbol{\pi}\left(g_{t,h}\right);\tau\right)$$

- Equivalent soft spike count:

$$z_{t,h} = \sum_{m=0}^{M-1} m \cdot \tilde{z}_{t,h,m}$$

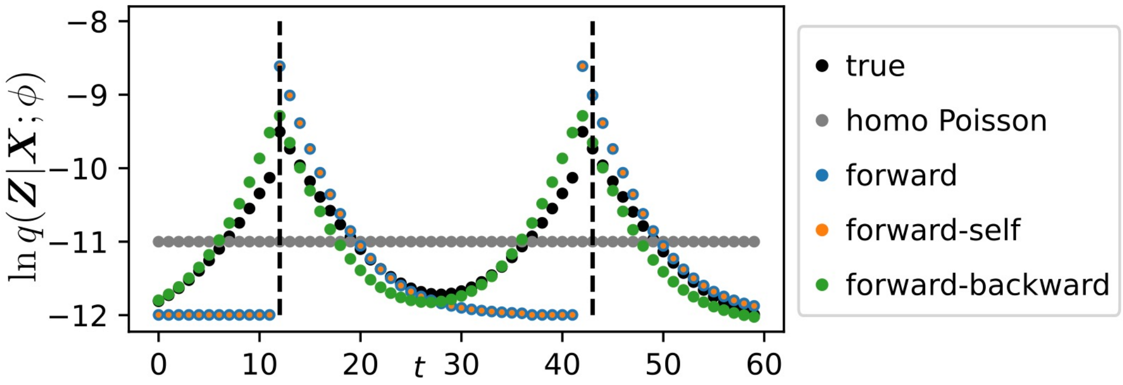
Inference methods: Candidate distributions

- Replace the distribution $\mathbb{P}[z; g]$ in the generative and variational model with the following distributions governed by $\mathbb{E}[z] = g$.

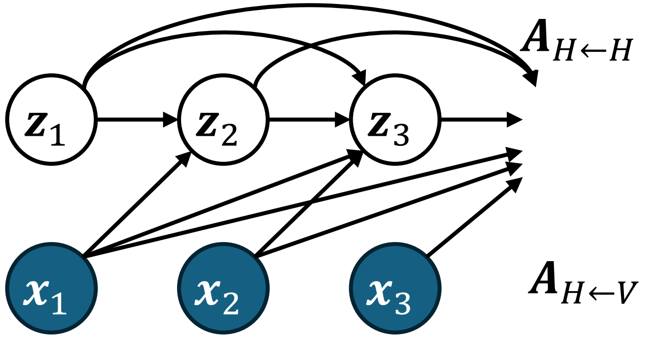


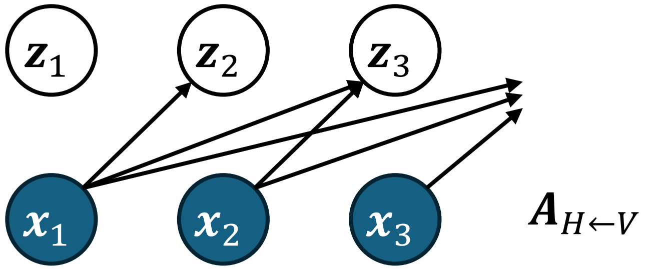
distribution	can pathwise
$\mathrm{Pois}(g)$	✗
$\mathrm{Cat}(\boldsymbol{\pi}(\mathrm{g}))$	✗
$\mathrm{GS}(\boldsymbol{\pi}(\mathrm{g}); \tau)$	✓
$\mathrm{Exp}\left(\frac{1}{g}\right)$	✓
$\mathrm{Ray}\left(\sqrt{\frac{2}{\pi}}g\right)$	✓
$\mathrm{HN}(\sqrt{\frac{\pi}{2}}g)$	✓

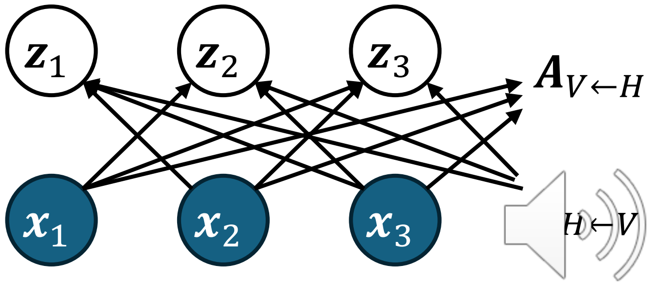
Sampling schemes



- Three message passing designs where $\boldsymbol{c}$ mimicking the bias $\boldsymbol{b}$, and $\boldsymbol{A}$ mimicking the weight $\boldsymbol{W}$.
- Forward-self (FS)
  - $\boldsymbol{g}_t = \sigma\left(\boldsymbol{c}_H + \boldsymbol{A}_{H\leftarrow V}\left(\sum_{l=1}^{L}\psi_l \boldsymbol{x}_{t-l}\right) + \boldsymbol{A}_{H\leftarrow H}\left(\sum_{l=1}^{L}\psi_l \boldsymbol{z}_{t-l}\right)\right)$
  - $\phi = \{\boldsymbol{c}_H \in \mathbb{R}^H, \boldsymbol{A}_{H\leftarrow V} \in \mathbb{R}^{H\times V}, \boldsymbol{A}_{H\leftarrow H} \in \mathbb{R}^{H\times H}\}$
- Forward (F)
  - $\boldsymbol{g}_t = \sigma\left(\boldsymbol{c}_H + \boldsymbol{A}_{H\leftarrow V}\left(\sum_{l=1}^{L}\psi_l \boldsymbol{x}_{t-l}\right)\right)$
  - $\phi = \{\boldsymbol{c}_H \in \mathbb{R}^H, \boldsymbol{A}_{H\leftarrow V} \in \mathbb{R}^{H\times V}\}$
- Forward-backward (FB)
  - $\boldsymbol{g}_t = \sigma\left(\boldsymbol{c}_H + \boldsymbol{A}_{H\leftarrow V}\left(\sum_{l=1}^{L}\psi_l \boldsymbol{x}_{t-l}\right) + \boldsymbol{A}_{V\leftarrow H}\left(\sum_{l=1}^{L}\psi_l \boldsymbol{x}_{t+l}\right)\right)$
  - $\phi = \{\boldsymbol{c}_H \in \mathbb{R}^H, \boldsymbol{A}_{H\leftarrow V} \in \mathbb{R}^{H\times V}, \boldsymbol{A}_{V\leftarrow H} \in \mathbb{R}^{V\times H}\}$







3 Experiments

Method combinations

- 7 inference methods
  - Poisson (Pois)
  - Categorical (Cat)
  - Gumbel-Softmax-score (GS-s)
  - Gumbel-Softmax-pathwise (GS-p)
  - Exponential (Exp)
  - Rayleigh (Ray)
  - Half-normal (HN)

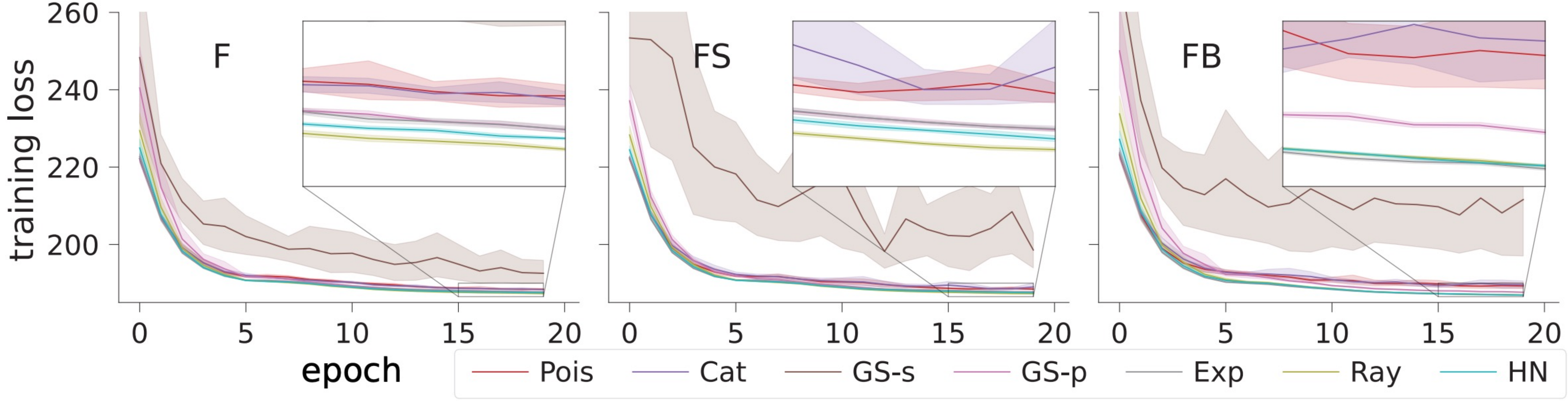
- 3 variational sampling schemes
  - Forward (F)
  - Forward-self (FS)
  - Forward-backward (FB)

Synthetic

- 3-visible-2-hidden.
- High likelihood and low weight error for differentiable inference methods (GS-pathwise, Exp, Ray, HN) with FB sampling schemes.
- Fast runtime for forward (F) and forward-backward (FB).

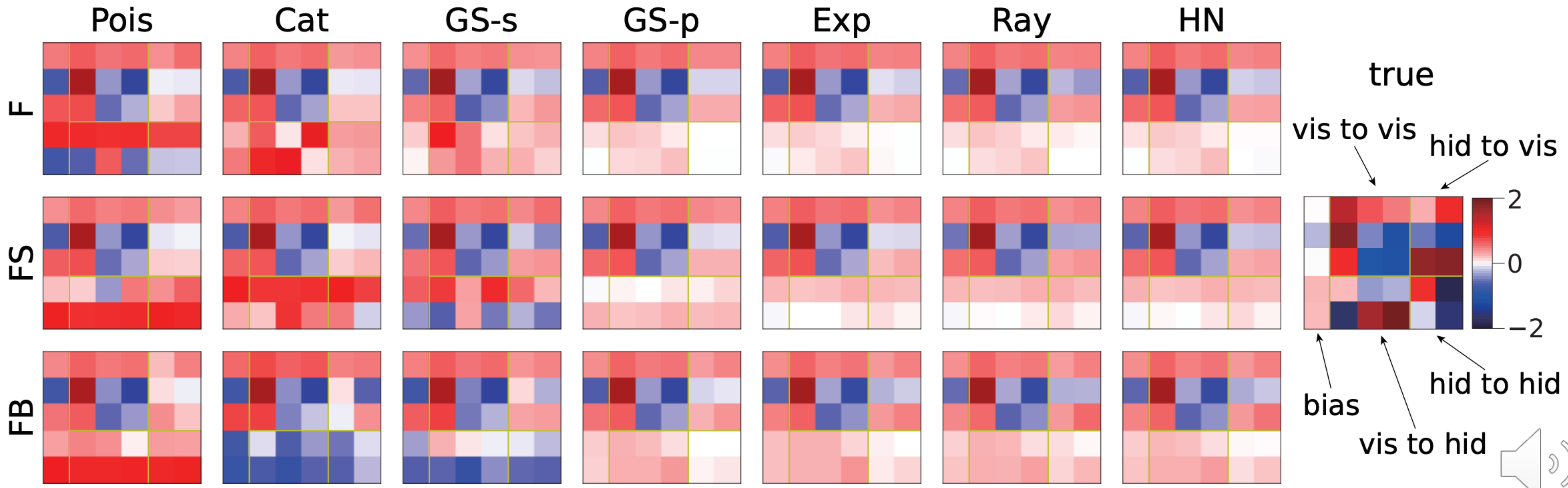
Synthetic

- Better convergence for differentiable inference methods.



Synthetic

- Differentiable inference methods × FB have better parameter estimation, especially in the hidden to visible block.



Retinal ganglion neurons (Pillow & Scott, 2012)

- Neuron 1-16: OFF cell. Neuron 17-27: ON cell.
- 20 minutes of a visual task on a mouse.
- Assuming $H \in \{1,2,3\}$ hidden neurons.

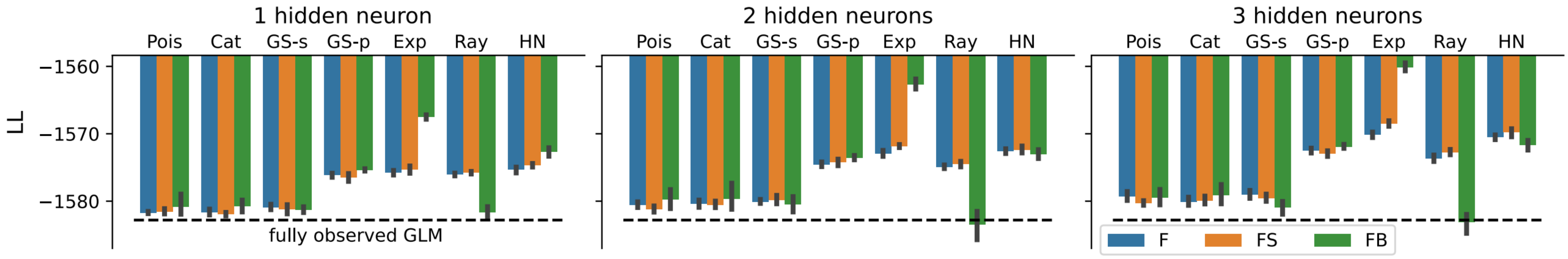
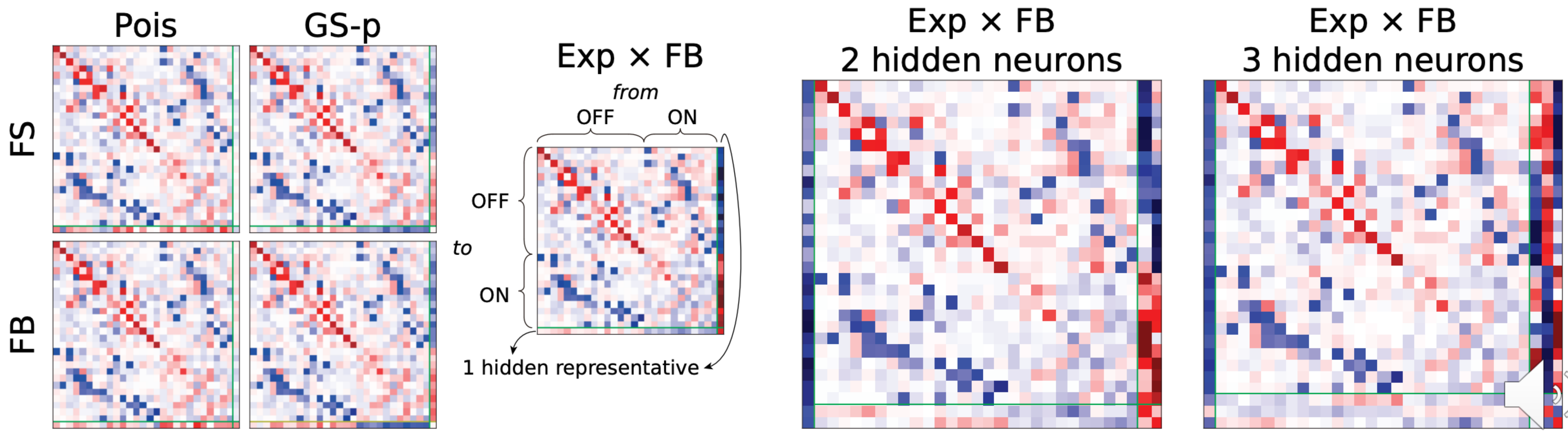


Figure 5. The test log-likelihood (LL) of different method combinations under $H \in \{1, 2, 3\}$ hidden neurons. The dashed black line represents the test LL of the fully observed GLM as the baseline.

Retinal ganglion neurons (Pillow & Scott, 2012)

- The learned weight matrix.
- The learned one hidden representative from Exp × FB serves as a negative feed back regulating unit.



Primary visual cortex (cncrs: PVC-5)

- Primary visual cortex (V1) recordings from a macaque monkey over a 15-minute duration without presenting any stimuli.
- Only 3 visible neurons
- Assuming $H \in \{1, \ldots, 9\}$ hidden neurons. Containing cases where $H \gg V$.

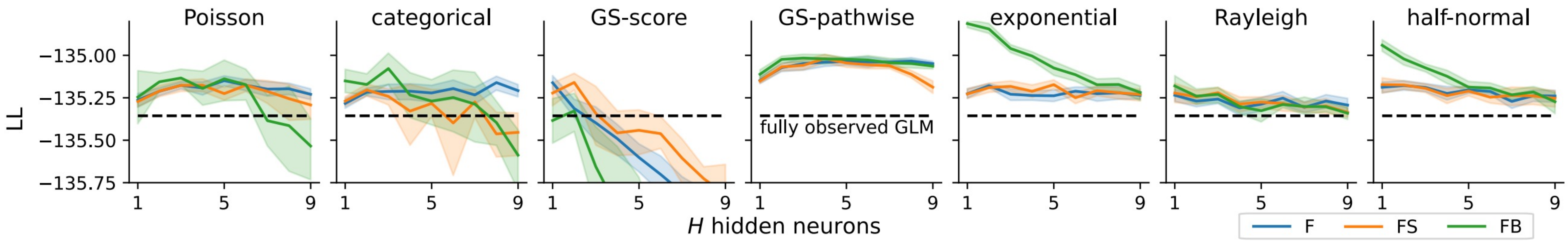


Figure 7. The curves of the test log-likelihood (LL) v.s. the number of hidden neurons H, for different method combinations.

Summary

- We proposes a differentiable version of the partially observable generalized linear model (POGLM), in which the pathwise gradient estimator becomes applicable when doing variational inference (VI).
- The new forward-backward message passing sampling scheme is faster and more expressive.
- Note that the relaxation from Gumbel-Softmax distribution to general continuous distributions loses the meaning of $\boldsymbol{Z}$ as representing spike counts, but can produce better performance. This is worth to be investigated in the future.

Thanks for listening!