A Differentiable POGLM with Forward-Backward Message Passing

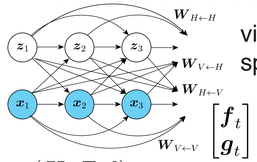
Chengrui Li, Weihan Li, Yule Wang, Anqi Wu @ GaTech CSE

Georgia Tech

ICML 2024 International Conference On Machine Learning

Partially observable generalized linear model (POGLM)

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



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

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

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: $\theta = \left\{\boldsymbol{b} \in \mathbb{R}^{V+H}, \boldsymbol{W} \in \mathbb{R}^{(V+H)\times(V+H)}\right\}$

Variational inference (VI)

Variational model $q(\boldsymbol{Z}|\boldsymbol{X};\phi)$ parameterized by ϕ :

$\boldsymbol{z}_t \sim \text{Pois}(\boldsymbol{g}_t)$ related to **inference methods**

$\boldsymbol{g}_t = \text{function}(\boldsymbol{X};\phi)$ related to **sampling schemes**

Evidence lower bound: $\text{ELBO}(\boldsymbol{X};\theta,\phi) = \mathbb{E}_q\left[\ln p(\boldsymbol{X},\boldsymbol{Z};\theta) - \ln q(\boldsymbol{Z}|\boldsymbol{X};\phi)\right]$

VI: 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	$\{\boldsymbol{Z}^{(k)}\}_{k=1}^K \sim q(\boldsymbol{Z}\|\boldsymbol{X};\phi)$	$\{\boldsymbol{\epsilon}^{(k)}\}_{k=1}^K \sim R(\boldsymbol{\epsilon})$
$\frac{\partial\,\text{ELBO}(\boldsymbol{X};\theta,\phi)}{\partial\phi} \approx$	$\frac{\partial}{\partial\phi}\frac{-1}{2K}\sum_{k=1}^K\left[\ln p\left(\boldsymbol{X},\boldsymbol{Z}^{(k)};\theta\right) - \ln q\left(\boldsymbol{Z}^{(k)}\middle\|\boldsymbol{X};\phi\right)\right]^2$	$\frac{\partial}{\partial\phi}\frac{1}{K}\sum_{k=1}^K\left[\ln p\left(\boldsymbol{X}, r\left(\boldsymbol{\epsilon}^{(k)}\|\boldsymbol{X};\phi\right);\theta\right) - \ln q\left(r\left(\boldsymbol{\epsilon}^{(k)}\|\boldsymbol{X};\phi\right)\|\boldsymbol{X};\phi\right)\right]$

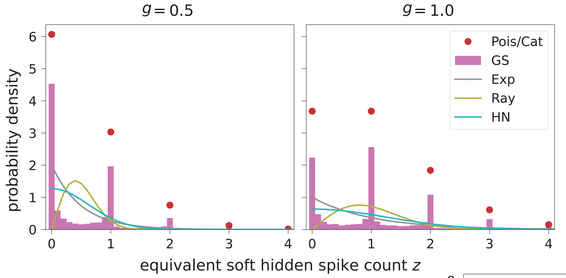
Relaxation

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

References: [1] Pillow & Latham, NeurIPS, 2007. [2] Jimenez Rezende & Gerstner, Front. Comp. Neur., 2014. [3] Linderman et al., NeurIPS, 2017. [4] Jang et al., arXiv, 2016. [5] Maddison et al., arXiv, 2016. [6] Kajino, ICML, 2021. [7] Li et al., ICLR, 2024.

Candidate distributions

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



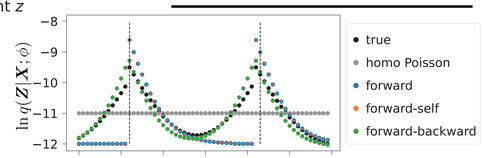
distribution	can pathwise
$\text{Pois}(g)$	✗
$\text{Cat}(\boldsymbol{\pi}(g))$	✗
$\text{GS}(\boldsymbol{\pi}(g);\tau)$	✓
$\text{Exp}\left(\frac{1}{g}\right)$	✓
$\text{Ray}\left(\sqrt{\frac{2}{\pi}}g\right)$	✓
$\text{HN}\left(\sqrt{\frac{\pi}{2}}g\right)$	✓

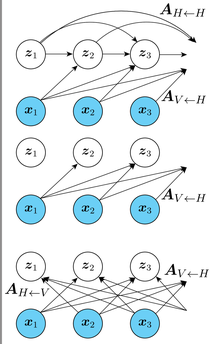
VI: Sampling schemes

Three message passing designs:

$\boldsymbol{c}$ mimicking the bias $\boldsymbol{b}$

$\boldsymbol{A}$ mimicking the weight $\boldsymbol{W}$.





Forward-self (FS)

$$\boldsymbol{g}_t = \sigma\left(\boldsymbol{c}_H + \boldsymbol{A}_{H\leftarrow V}\left(\sum_{l=1}^L \psi_l\, \boldsymbol{x}_{t-l}\right) + \boldsymbol{A}_{H\leftarrow H}\left(\sum_{l=1}^L \psi_l\, \boldsymbol{z}_{t-l}\right)\right)$$

$$\phi = \left\{\boldsymbol{c}_H \in \mathbb{R}^H, \boldsymbol{A}_{H\leftarrow V} \in \mathbb{R}^{H\times V}, \boldsymbol{A}_{H\leftarrow H} \in \mathbb{R}^{H\times H}\right\}$$

Forward (F)

$$\boldsymbol{g}_t = \sigma\left(\boldsymbol{c}_H + \boldsymbol{A}_{H\leftarrow V}\left(\sum_{l=1}^L \psi_l\, \boldsymbol{x}_{t-l}\right)\right)$$

$$\phi = \left\{\boldsymbol{c}_H \in \mathbb{R}^H, \boldsymbol{A}_{H\leftarrow V} \in \mathbb{R}^{H\times V}\right\}$$

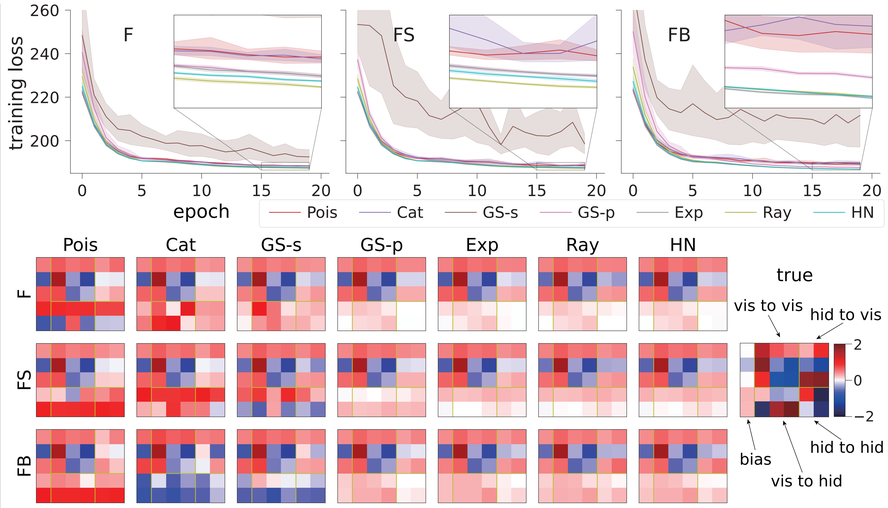
Forward-backward (FB)

$$\boldsymbol{g}_t = \sigma\left(\boldsymbol{c}_H + \boldsymbol{A}_{H\leftarrow V}\left(\sum_{l=1}^L \psi_l\, \boldsymbol{x}_{t-l}\right) + \boldsymbol{A}_{V\leftarrow H}\left(\sum_{l=1}^L \psi_l\, \boldsymbol{x}_{t+l}\right)\right)$$

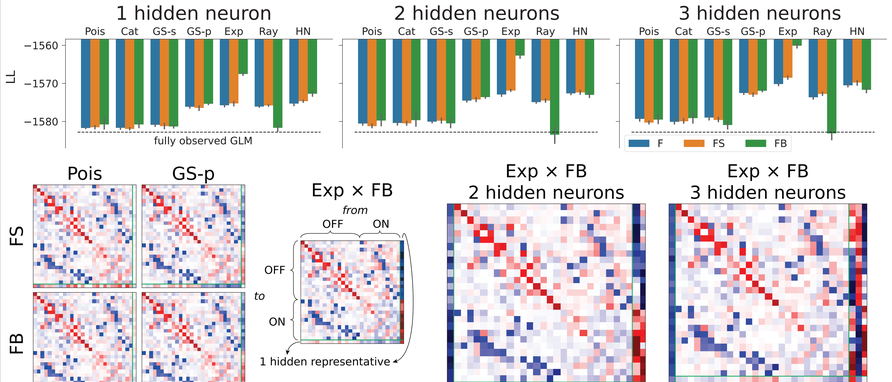
$$\phi = \left\{\boldsymbol{c}_H \in \mathbb{R}^H, \boldsymbol{A}_{H\leftarrow V} \in \mathbb{R}^{H\times V}, \boldsymbol{A}_{V\leftarrow H} \in \mathbb{R}^{V\times H}\right\}$$

Experiments

Synthetic



Retinal ganglion neurons (Pillow & Scott, 2012)



Primary visual cortex (cncrs: PVC-5)