Notations (common)

Each row is the recording for neuron i, $\boldsymbol{y}_i = (y_{i1}, \dots, y_{iT})'$, $i = 1, \dots N$. Denote the cluster index for neuron i as $z_i \in \{1, \dots\}$. The number of neurons in cluster j is $n_j = \sum_{i=1}^{N} I(z_i = j)$, and $\sum_{j=1,2,\dots} n_j = N$.

Model

Denote the latent vector in cluster j as $\boldsymbol{x}_t^{(j)} \in R^{p_j}$. For simplicity, assume all $p_j = p$. Each observation follows a Poisson distribution as follows:

$$\log \lambda_{it} = d_i + \boldsymbol{c}_i' \boldsymbol{x}_t^{(z_i)}$$

$$y_{it} \sim Poisson(\lambda_{it})$$

Where $\boldsymbol{c}_i \in R^p$ and $\boldsymbol{x}_t^{(z_i)} \in R^p$.

In this version, the loading d_i and $\boldsymbol{c}_i$ are also cluster dependent. That is,

$$(d_i, \boldsymbol{c}_i')' \sim N(\boldsymbol{\mu}_{dc}^{(z_i)}, \boldsymbol{\Sigma}_{dc}^{(z_i)})$$

Denote all latent states as $\boldsymbol{x}_t = \left(\boldsymbol{x}_t^{(1)'}, \boldsymbol{x}_t^{(2)'}, \dots\right)'$ and they evolve linearly with Gaussian noise:

$$\boldsymbol{x}_1 \sim N(\boldsymbol{x}_0, \boldsymbol{Q}_0)$$

$$\boldsymbol{x}_{t+1} | \boldsymbol{x}_t \sim N(\boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{b}, \boldsymbol{Q})$$

To simplify, assume $\boldsymbol{Q}_0$ is known (e.g. $\boldsymbol{Q}_0 = \boldsymbol{I}$).

If we assume block diagonal (as in Joshua et al., 2020) for process noise covariance, we can write things as:

$$\boldsymbol{x}_{t+1}^{(j)} | \boldsymbol{x}_t^{(1)}, \boldsymbol{x}_t^{(2)}, \dots \sim N(\sum_{l=1,\dots} \boldsymbol{A}_{j\leftarrow l}\, \boldsymbol{x}_t^{(l)} + \boldsymbol{b}_j, \boldsymbol{Q}^{(j)})$$

Notice $\{\boldsymbol{A}_{j\leftarrow l}\}$ forms the full transition matrix as:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{1\leftarrow 1} & \boldsymbol{A}_{1\leftarrow 2} & \dots \\ \boldsymbol{A}_{2\leftarrow 1} & \boldsymbol{A}_{2\leftarrow 2} & \dots \\ \dots & \dots & \dots \end{pmatrix}$$

If the j^{th} row block of $\boldsymbol{A}$ is $\boldsymbol{A}_j = (\boldsymbol{A}_{j\leftarrow 1} \quad \boldsymbol{A}_{j\leftarrow 2} \quad \dots)$. Then, $\sum_{l=1,\dots} \boldsymbol{A}_{j\leftarrow l}\, \boldsymbol{x}_t^{(l)} + \boldsymbol{b}_j = \boldsymbol{A}_j \boldsymbol{x}_t + \boldsymbol{b}_j$.

If we further let $\boldsymbol{Q}$ be diagonal: denote the k^{th} row of $\boldsymbol{x}_t$, $\boldsymbol{A}$, $\boldsymbol{b}$ as x_{kt}, $\boldsymbol{a}_k$, b_k. The corresponding process noise variance as q_k. Then:

$$x_{k,t+1} | x_{kt} \sim N(\boldsymbol{a}_k' \boldsymbol{x}_t + b_k, q_k)$$

The parameters need to estimate:

(1) Latent vectors: $\boldsymbol{x}_t$
(2) Initials: $\boldsymbol{x}_0$

(3) Linear mapping for latent vectors: $\{d_i\}_{i=1}^N$ and $\{\boldsymbol{c}_i\}_{i=1}^N$
(4) Mean and covariance for linear mapping in each cluster: $\left\{\boldsymbol{\mu}_{dc}^{(j)}\right\}_j$ and $\left\{\boldsymbol{\Sigma}_{dc}^{(j)}\right\}_j$
(5) Linear dynamics for latent vectors: $\boldsymbol{A}$ and $\boldsymbol{b}$
(6) Process noise: $\boldsymbol{Q}$

Since the progress noise is independent in the model, $f\left(\boldsymbol{y}_i|\boldsymbol{\Theta}_j\right) = \prod_{t=1}^T P(y_{it}|\boldsymbol{\Theta}_j)$, where $P(\cdot)$ is the Poisson density and $\boldsymbol{\Theta}_j$ is the parameters in cluster j.

Conditional Priors

Others are the same as v3, but modify loading related ones, i.e. mean and covariance for linear mapping in each cluster $\left\{\boldsymbol{\mu}_{dc}^{(j)}\right\}_j$ and $\left\{\boldsymbol{\Sigma}_{dc}^{(j)}\right\}_j$:

$$\boldsymbol{\mu}_{dc}^{(j)} \sim N(\boldsymbol{\delta}_{dc0}, \mathbf{T}_{dc0})$$

Where $\boldsymbol{\delta}_{dc0} = \mathbf{0}_{p+1}$ and $\mathbf{T}_{dc0} = 0.1\boldsymbol{I}_{p+1}$

$$\boldsymbol{\Sigma}_{dc}^{(j)} \sim W^{-1}(\Psi_{dc0}, \nu_{dc0})$$

Where $\nu_{dc0} = p + 1 + 2$ and $\Psi_{dc} \ = \boldsymbol{I}_{p+1} \times 10^{-2}$

MCMC iteration

Others are the same as v3, but modify loading related ones.

(1) Update $\{d_i\}_{i=1}^N$ and $\{\boldsymbol{c}_i\}_{i=1}^N$:
Denote $(d_i, \boldsymbol{c}_i')' = \boldsymbol{\zeta}_i \in R^{p+1}$ and $\left(1, \boldsymbol{x}_t^{(z_i)'}\right) = \widetilde{\boldsymbol{x}}_t^{(z_i)'}$.

$$P\left(\boldsymbol{\zeta}_i \middle| \boldsymbol{y}_i, \left\{\boldsymbol{x}_t^{(z_i)}\right\}_{t=1}^T, \dots\right) = \exp f(\boldsymbol{\zeta}_i) \approx N\left(\boldsymbol{\zeta}_i|\boldsymbol{\mu}_{\zeta_i}, \boldsymbol{\Sigma}_{\zeta_i}\right)$$

$$\frac{\partial f}{\partial \boldsymbol{\zeta}_i} = \frac{\partial l}{\partial \boldsymbol{\zeta}_i} - \boldsymbol{\Sigma}_{dc}^{(z_i)^{-1}}\left(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_{dc}^{(z_i)}\right) = \left[\sum_{t=1}^T \widetilde{\boldsymbol{x}}_t^{(z_i)}(y_{it} - \lambda_{it})\right] - \boldsymbol{\Sigma}_{dc}^{(z_i)^{-1}}\left(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_{dc}^{(z_i)}\right)$$

$$\frac{\partial^2 f}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} = \frac{\partial^2 l}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'} - \boldsymbol{\Sigma}_{dc}^{(z_i)^{-1}} = -\left[\sum_{t=1}^T \lambda_{it}\widetilde{\boldsymbol{x}}_t^{(z_i)}\widetilde{\boldsymbol{x}}_t^{(z_i)'}\right] - \boldsymbol{\Sigma}_{dc}^{(z_i)^{-1}}$$

Where l is Poisson log-likelihood.

Use Newton-Raphson to find $\boldsymbol{\mu}_{\zeta_i} = argmax_{\zeta_i}\left(f(\boldsymbol{\zeta}_i)\right)$ and $\boldsymbol{\Sigma}_{\zeta_i} = -\left[\frac{\partial^2 f}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_i'}|_{\zeta_i = \boldsymbol{\mu}_{\zeta_i}}\right]^{-\mathbf{1}}$

(2) Update $\left\{\boldsymbol{\mu}_{dc}^{(j)}\right\}_j$ and $\left\{\boldsymbol{\Sigma}_{dc}^{(j)}\right\}_j$:
Again, denote $(d_i, \boldsymbol{c}_i')' = \boldsymbol{\zeta}_i \in R^{p+1}$.
Mean $\left\{\boldsymbol{\mu}_{dc}^{(j)}\right\}_j$: by conjugacy, $\boldsymbol{\mu}_{dc}^{(j)} \sim N(\boldsymbol{\delta}_{dc}, \mathbf{T}_{dc})$

$$\mathbf{T}_{dc}^{-1} = \left(\mathbf{T}_{dc0}^{-1} + n_j \boldsymbol{\Sigma}_{dc}^{(j)^{-1}}\right)^{-1}$$

$$\boldsymbol{\delta}_{dc} = \mathbf{T}_{dc}\left(\mathbf{T}_{dc0}^{-1}\boldsymbol{\delta}_{dc0} + \boldsymbol{\Sigma}_{dc}^{(j)^{-1}}\sum_{i:z_i=j}\boldsymbol{\zeta}_i\right)$$

Covariance: by conjugacy, $\boldsymbol{\Sigma}_{dc}^{(j)} \sim W^{-1}(\Psi_{dc}, \nu_{dc})$

$$\nu_{dc} = n_j + \nu_{dc0}$$

$$\Psi_{dc} = \Psi_{dc0} + \sum_{i:z_i=j}\left(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_{dc}^{(j)}\right)\left(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_{dc}^{(j)}\right)'$$

If assume $\boldsymbol{\Sigma}_{dc}^{(j)} = \boldsymbol{\Sigma}_{dc}$, then $\boldsymbol{\Sigma}_{dc} \sim W^{-1}(\Psi_{dc}, \nu_{dc})$

$$\nu_{dc} = N + \nu_{dc}$$

$$\Psi_{dc} = \Psi_{dc} \;\; + \sum_{i=1}^{N}\left(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_{dc}^{(z_i)}\right)\left(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_{dc}^{(z_i)}\right)'$$