Parsing epileptic events using a Markov switching process model for correlated time series

Supplementary Materials

Drausin F. Wulsin, Emily B. Fox, Brian Litt

Contents

A Conditional likelihoods

A.1 Channel i conditional likelihood at time t

Let $\mathbf{i'} \subseteq \{1, \ldots, N\}$ index the neighboring channels upon which channel i is conditioned. The conditional likelihood of observation $y_t^{(i)}$ under AR model k given the other observations $\mathbf{y}_t^{(\mathbf{i'})}$ at time t is

$$p\left(y_t^{(i)} \,|\, \widetilde{\mathbf{y}}_t^{(i)}, \mathbf{y}_t^{(\mathbf{i'})}, z_t^{(i)} = k, \mathbf{z}_t^{(\mathbf{i'})}, Z_t, \{\mathbf{a}_k\}, \{\Delta_l\}\right) = \mathcal{N}\left(y_t^{(i)}; \widetilde{\mu}_t, \widetilde{\sigma}_t^2\right) \tag{1}$$

for

$$\begin{aligned}\widetilde{\mu}_t &= \mathbf{a}_k^T \widetilde{\mathbf{y}}_t^{(i)} + \Delta_{Z_t}^{(i,\mathbf{i'})} \Delta_{Z_t}^{-1(\mathbf{i'},\mathbf{i'})} \left(\mathbf{y}_t^{(\mathbf{i'})} - \mathbf{A}_{\mathbf{z}^{(\mathbf{i'})}} \widetilde{\mathbf{Y}}_t^{(\mathbf{i'})}\right) \\ \widetilde{\sigma}_t^2 &= \Delta_{Z_t}^{(i,i)} - \Delta_{Z_t}^{(i,\mathbf{i'})} \Delta_{Z_t}^{-1(\mathbf{i'},\mathbf{i'})} \Delta_{Z_t}^{(\mathbf{i'},i)}.\end{aligned} \tag{2}$$

A.2 Channel i conditional marginal likelihood over $t = 1, \ldots, T$

The sum-product algorithm can be used to produce the conditional likelihood of channel i's observations over all $t = 1, \ldots, T$, marginalizing over the exponentially many state sequences $z_{1:T}^{(i)}$. Let $\boldsymbol{\xi}_t \in \mathbb{R}^{K^{(i)}}$ be a vector defining the forward messages for this channel at time t. Element k in this vector gives the joint probability of the observations from the first t time points and the state $z_t^{(i)} = k$,

$$\begin{aligned}\xi_{kt} &= p\left(y_{1:t}^{(i)}, z_t^{(i)} = k \,|\, \mathbf{y}_{1:t}^{(\mathbf{i'})}, \mathbf{z}_{1:t}^{(\mathbf{i'})}, Z_{1:t}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) \\ \xi_{kt} &= p\left(y_t^{(i)} \,|\, \mathbf{y}_t^{(\mathbf{i'})}, z_t^{(i)} = k, \mathbf{z}_t^{(\mathbf{i'})}, Z_{1:t}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) \cdot \\ &\quad \sum_{k'} p\left(y_{1:t-1}^{(i)} \,|\, \mathbf{y}_{1:t-1}^{(\mathbf{i'})}, z_{t-1}^{(i)} = k', \mathbf{z}_{1:t-1}^{(\mathbf{i'})}, Z_{1:t-1}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) p\left(z_t^{(i)} = k \,|\, z_{t-1}^{(i)} = k'\right) p\left(z_{t-1}^{(i)} = k'\right) \\ \xi_{kt} &= p\left(y_t^{(i)} \,|\, \mathbf{y}_t^{(\mathbf{i'})}, z_t^{(i)} = k, Z_t, \mathbf{z}_t^{(\mathbf{i'})}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) \sum_{k'} \xi_{k',t-1} \cdot p\left(z_t^{(i)} = k \,|\, z_{t-1}^{(i)} = k'\right).\end{aligned}$$

In the above, we omit the dependence on $\widetilde{\mathbf{y}}_t^{(i)}$ for notational simplicity.

If $\mathbf{u}_t \in \mathbb{R}^{K^{(i)}}$ defines the (conditional) likelihood vector of $y_t^{(i)}$ under each of the $K^{(i)}$ possible states (following from Eq. (1)), these forward messages can be written compactly in vector notation as

$$\boldsymbol{\xi}_t = \mathbf{u}_t \circ (\widetilde{\boldsymbol{\pi}}^{(i)\mathrm{T}} \boldsymbol{\xi}_{t-1})$$

with

$$\boldsymbol{\xi}_1 = \mathbf{u}_1 \circ \widetilde{\boldsymbol{\pi}}_0^{(i)}$$

where we let $\widetilde{\boldsymbol{\pi}}^{(i)} \in \mathbb{R}^{K^{(i)} \times K^{(i)}}$ be a matrix of the positive channel state transition probabilities in $\boldsymbol{\pi}^{(i)}$, which is a function of $\mathbf{f}^{(i)}$ and $\boldsymbol{\eta}^{(i)}$. The total conditional likelihood of the sequence of channel i observations given the states of the other channels $\mathbf{i'}$ and the event states is thus

$$\begin{aligned}p\left(y_{1:T}^{(i)} \,|\, \mathbf{y}_{1:T}^{(\mathbf{i'})}, \mathbf{z}_{1:T}^{(\mathbf{i'})}, Z_{1:T}, \mathbf{f}^{(i)}, \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) &= \sum_k \xi_{kT} \\ \ell(y_{1:T}^{(i)}) &= \mathbf{1}^T \boldsymbol{\xi}_T.\end{aligned} \tag{3}$$

A.3 Conditional event likelihood

Let $\mathbf{z}_t$ denote the vector of N states at time t. Since the space of $\mathbf{z}_t$ is exponentially large, we cannot integrate it out to compute the marginal conditional likelihood of the data given the event state sequence

$Z_{1:T}$ (and model parameters). Instead, we consider the conditional likelihood of an observation at time t given channel states $\mathbf{z}_t$ and event state Z_t:

$$p(\mathbf{y}_t \mid \widetilde{\mathbf{Y}}_t, \mathbf{z}_t, Z_t, \{\mathbf{a}_k\}, \{\Delta_l\}) = \mathcal{N}(\mathbf{y}_t; \mathbf{A}_{\mathbf{z}_t} \widetilde{\mathbf{Y}}_t, \Delta_{Z_t}). \tag{4}$$

We integrate over the event states $Z_{1:T}$ via the sum-product algorithm to yield the conditional event likelihood, given only the channel states. Let $\boldsymbol{\zeta}_t = [\zeta_{1t}, \ldots, \zeta_{Lt}]^T$ describe the vector of forward messages at time t for L possible event states with elements

$$\begin{aligned}
\zeta_{lt} &= p\left(\mathbf{y}_{1:t}, Z_t = l \mid \mathbf{z}_{1:t}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) \\
&= p\left(\mathbf{y}_t \mid \mathbf{z}_t, Z_t = l, \{\mathbf{a}_k\}, \{\Delta_l\}\right) \cdot \\
&\qquad \sum_{l'} p\left(\mathbf{y}_{1:t-1} \mid \mathbf{z}_{1:t-1}, Z_{t-1} = l', \{\mathbf{a}_k\}, \{\Delta_l\}\right) p\left(Z_t = l \mid Z_{t-1} = l'\right) p\left(Z_{t-1} = l'\right) \\
\zeta_{lt} &= p\left(\mathbf{y}_t \mid \mathbf{z}_t, Z_t = l, \{\mathbf{a}_k\}, \{\Delta_l\}\right) \sum_{l'} \zeta_{lt-1} \cdot p\left(Z_t = l \mid Z_{t-1} = l'\right).
\end{aligned} \tag{5}$$

Again, we omit $\widetilde{\mathbf{Y}}_t$ above for notational simplicity.

If $\mathbf{v}_t \in \mathbb{R}^L$ denotes the conditional likelihood vector of $\mathbf{y}_t$ under each of the L possible event states (following from Eq. (4)), these forward messages can be written compactly in vector notation as

$$\boldsymbol{\zeta}_t = \mathbf{v}_t \circ (\boldsymbol{\phi}^{\mathrm{T}} \boldsymbol{\zeta}_{t-1}) \tag{6}$$

with $\boldsymbol{\zeta}_1 = \mathbf{v}_1 \circ \phi_0$. The matrix $\boldsymbol{\phi} \in \mathbb{R}^{L \times L}$ gives the event state transition probabilities. The conditional likelihood of the entire event given the channel states is thus

$$\begin{aligned}
p\left(\mathbf{y}_{1:T} \mid \mathbf{z}_{1:T}, \boldsymbol{\phi}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) &= \sum_l \zeta_{lT} \\
\ell(\mathbf{y}_{1:T}) &= \mathbf{1}^T \boldsymbol{\zeta}_T.
\end{aligned} \tag{7}$$

B Details of posterior computation

B.1 Sampling individual channel variables

We sample the active features, state sequences, and transition parameters for each channel i.

B.1.1 Channel active features, $\mathbf{f}^{(i)}$

We briefly describe the active feature sampling scheme given in detail by Fox et al. (2009). Recall that for our HIW-spatial BP-AR-HMM, we need to condition on neighboring channel state sequences $\mathbf{z}_{1:T}^{(\mathbf{i}')}$ and event state sequences $Z_{1:T}$. Sampling the feature indicators $\mathbf{f}^{(i)}$ for channel i via the Indian buffet process (IBP) involves considering those features shared by other channels and those unique to channel i. We denote the set of shared features across channels not including those specific to channel i as $\mathcal{S}^{-i} \subseteq \{1, \ldots, K\}$ and the set of unique features for channel i as $\mathcal{U}^i \subseteq \{1, \ldots, K\}/\mathcal{S}^{-i}$.

Shared features The posterior for each shared feature $k \in \mathcal{S}^{-i}$ for channel i is given by

$$\begin{aligned}
&p\left(f_k^{(i)} \mid y_{1:T}^{(i)}, \mathbf{y}_{1:T}^{(\mathbf{i}')}, \mathbf{z}_{1:T}^{(\mathbf{i}')}, Z_{1:T}, \mathbf{F}^{\mathbf{i}'k}, \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}, \alpha\right) \propto \\
&\qquad p\left(f_k^{(i)} \mid \mathbf{F}^{\mathbf{i}'k}, \alpha\right) p\left(y_{1:T}^{(i)} \mid \mathbf{y}_{1:T}^{(\mathbf{i}')}, \mathbf{z}_{1:T}^{(\mathbf{i}')}, Z_{1:T}, \mathbf{f}^{(i)}, \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}\right).
\end{aligned} \tag{8}$$

The IBP implies that that $p\left(f_k^{(i)} \mid \mathbf{F}^{-ik}, \alpha\right) = m_k^{-i}/N$, where m_k^{-i} denotes the number of other channels that use feature k. The likelihood term in Eq. (8) follows from Eq. (3). We use this posterior to formulate a

Metropolis-Hastings proposal that flips the current indicator value $f_k^{(i)}$ to its complement $\bar{f}_k^{(i)}$ with probability $\rho(\bar{f}_k^{(i)} \mid f_k^{(i)})$,

$$f_k^{(i)} = \begin{cases} \bar{f}_k^{(i)}, & \text{w.p.} \quad \rho(\bar{f}_k^{(i)} \mid f_k^{(i)}) \\ f_k^{(i)}, & \text{w.p.} \quad 1 - \rho(\bar{f}_k^{(i)} \mid f_k^{(i)}) \end{cases} \tag{9}$$

where

$$\rho(\bar{f}_k^{(i)} \mid f_k^{(i)}) = \min\left(\frac{p\left(\bar{f}_k^{(i)} \mid y_{1:T}^{(i)}, \mathbf{y}_{1:T}^{(\mathbf{i}')}, \mathbf{z}_{1:T}^{(\mathbf{i}')}, Z_{1:T}, \mathbf{F}^{\mathbf{i}'k}, \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}, \alpha\right)}{p\left(f_k^{(i)} \mid y_{1:T}^{(i)}, \mathbf{y}_{1:T}^{(\mathbf{i}')}, \mathbf{z}_{1:T}^{(\mathbf{i}')}, Z_{1:T}, \mathbf{F}^{\mathbf{i}'k}, \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}, \alpha\right)}, 1 \right).$$

Unique features We either propose a new feature or remove a unique feature for channel i using a birth and death reversible jump MCMC sampler (see Fox et al. (2009) for details). We denote the number of unique features for channel i as $n_i = |\mathcal{U}^i|$. We define the vector of shared feature indicators as $\mathbf{f}_-^{(i)} = \mathbf{f}_{k' \mid k' \in \mathcal{S}^{\mathbf{i}'}}^{(i)}$ and that for unique feature indicators as $\mathbf{f}_+^{(i)} = \mathbf{f}_{k' \mid k' \in \mathcal{U}^i}^{(i)}$, which together $[\mathbf{f}_-^{(i)} \ \mathbf{f}_+^{(i)}]$ define the full feature indicator vector $\mathbf{f}^{(i)}$ for channel i. Similarly, $\mathbf{a}_+^{(i)}$ and $\boldsymbol{\eta}_+^{(i)}$ describe the model dynamics and transition parameters associated with these unique features. We propose a new unique feature vector $\mathbf{f}'_+$ and corresponding model dynamics $\mathbf{a}'_+$ and transition parameters $\boldsymbol{\eta}'_+$ (sampled from their priors in the case of feature birth) with a proposal distribution of

$$p\left(\mathbf{f}'_+, \mathbf{a}'_+, \boldsymbol{\eta}'_+ \mid \mathbf{f}_+^{(i)}, \mathbf{a}_+^{(i)}, \boldsymbol{\eta}_+^{(i)}\right) = p\left(\mathbf{f}'_+ \mid \mathbf{f}_+^{(i)}\right) p\left(\mathbf{a}'_+ \mid \mathbf{f}'_+, \mathbf{f}_+^{(i)}, \mathbf{a}_+^{(i)}\right) p\left(\boldsymbol{\eta}'_+ \mid \mathbf{f}'_+, \mathbf{f}_+^{(i)}, \boldsymbol{\eta}_+^{(i)}\right). \tag{10}$$

A new unique feature is proposed with probability 0.5 and each existing unique feature is removed with probability $0.5/n_i$. This proposal is accepted with probability

$$\rho\left(\mathbf{f}'_+, \mathbf{a}'_+, \boldsymbol{\eta}'_+ \mid \mathbf{f}_+^{(i)}, \mathbf{a}_+^{(i)}, \boldsymbol{\eta}_+^{(i)}\right) = \min\left(\frac{p\left(y_{1:T}^{(i)} \mid \mathbf{y}_{1:T}^{(\mathbf{i}')}, \mathbf{z}_{1:T}^{(\mathbf{i}')}, [\mathbf{f}_-^{(i)} \ \mathbf{f}'_+], \boldsymbol{\eta}^{(i)}, \boldsymbol{\eta}'_+, \{\mathbf{a}_k\}, \{\Delta_l\}\right) \operatorname{Poisson}\left(n'_i \mid \alpha/N\right) p\left(\mathbf{f}_+^{(i)} \mid \mathbf{f}'_+\right)}{p\left(y_{1:T}^{(i)} \mid \mathbf{y}_{1:T}^{(\mathbf{i}')}, \mathbf{z}_{1:T}^{(\mathbf{i}')}, [\mathbf{f}_-^{(i)} \ \mathbf{f}_+^{(i)}], \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) \operatorname{Poisson}\left(n_i \mid \alpha/N\right) p\left(\mathbf{f}'_+ \mid \mathbf{f}_+^{(i)}\right)}, 1 \right). \tag{11}$$

The likelihood terms again follow from Eq. (3).

B.1.2 Channel state sequence, $z_{1:T}^{(i)}$

We sample the state sequence $z_{1:T}^{(i)}$ for all the time points of channel i, given that channel's feature-constrained transition distributions $\boldsymbol{\pi}^{(i)}$, the state parameters $\{\mathbf{a}_k\}$, the observations $y_{1:T}^{(i)}$, and the event's other observations $\mathbf{y}_{1:T}^{(\mathbf{i}')}$ and current states $\mathbf{z}_{1:T}^{(\mathbf{i}')}$. The joint probability of the state sequence $z_{1:T}^{(i)}$ is given by

$$p\left(z_{1:T}^{(i)} \mid y_{1:T}^{(i)}, \mathbf{y}_{1:T}^{(\mathbf{i}')}, \mathbf{z}_{1:T}^{(\mathbf{i}')}, \mathbf{f}^{(i)}, \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) = p\left(z_1^{(i)} \mid y_1^{(i)}, \mathbf{y}_1^{(\mathbf{i}')}, \mathbf{z}_1^{(\mathbf{i}')}, \mathbf{f}^{(i)}, \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) \prod_{t=2}^{T} p\left(z_t^{(i)} \mid y_{t:T}^{(i)}, \mathbf{y}_{t:T}^{(\mathbf{i}')}, z_{t-1}^{(i)}, \mathbf{z}_{t:T}^{(\mathbf{i}')}, \mathbf{f}^{(i)}, \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}\right). \tag{12}$$

Again following the sum-product algorithm, we compute a vector $\boldsymbol{\psi}_t \in \mathbb{R}^{K^{(i)}}$ of backward messages from time point $t+1$ to t, where each element $\psi_{k,t}$ is proportional to the likelihood of future observations $y_{t+1:T}^{(i)}$ given $z_t^{(i)} = k$ at time t,

$$\psi_{k,t} \propto p\left(y_{t+1:T}^{(i)} \mid \mathbf{y}_{t+1:T}^{(\mathbf{i}')}, z_t^{(i)} = k, \mathbf{z}_{t+1:T}^{(\mathbf{i}')}, Z_{1:T}, \mathbf{f}^{(i)}, \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}\right).$$

As before, $\mathbf{u}_t \in \mathbb{R}^{K^{(i)}}$ defines the likelihood vector of $y_t^{(i)}$ under each of the $K^{(i)}$ possible states (following from Eq. (1)), so the backward message recursion can be written efficiently as

$$\boldsymbol{\psi}_t \propto \widetilde{\boldsymbol{\pi}}^{(i)}(\mathbf{u}_{t+1} \circ \boldsymbol{\psi}_{t+1}).$$

The conditional probability of $z_t^{(i)}$ is given by

$$p\left(z_t^{(i)} \,|\, y_{t:T}^{(i)}, \mathbf{y}_{t:T}^{(\mathbf{i}')}, z_{t-1}^{(i)}, Z_{1:T}, \mathbf{z}_{t:T}^{(\mathbf{i}')}, \mathbf{f}^{(i)}, \boldsymbol{\eta}^{(i)}, \{\mathbf{a}_k\}, \{\Delta_l\}\right) = \text{Multi}\left(\left(\widetilde{\boldsymbol{\pi}}_{z_{t-1}^{(i)}}^{(i)}\right)^{\mathrm{T}} \circ \mathbf{u}_t \circ \boldsymbol{\psi}_t\right). \tag{13}$$

B.1.3 Channel transition parameters, $\boldsymbol{\eta}^{(i)}$

Following the correction described by Hughes et al. (2012), the posterior for the transition variable $\eta_{jk}^{(i)}$ is given by

$$p(\eta_{jk}^{(i)} \,|\, z_{1:T}^{(i)}, f_k^{(i)}) \propto \frac{(\eta_{jk}^{(i)})^{n_{jk}^{(i)}+\gamma_c+\delta_{j,k}\kappa_c-1} e^{\eta_{jk}^{(i)}}}{\sum_{k' \,|\, f_k^{(i)}=1} \eta_{jk'}^{(i)}}, \tag{14}$$

where $n_{jk}^{(i)}$ denotes the number of times channel i transitions from state j to state k. We can sample from this posterior via two auxiliary variables,

$$\begin{aligned}
\bar{\boldsymbol{\eta}}_j^{(i)} &\sim \text{Dir}(\gamma_c + \kappa_c \mathbf{e}_j + \mathbf{n}_j) \\
C_j^{(i)} &\sim \text{Gamma}(K\gamma_c + \kappa_c, 1) \\
\boldsymbol{\eta}_j^{(i)} &= C_j^{(i)} \bar{\boldsymbol{\eta}}_j^{(i)}.
\end{aligned} \tag{15}$$

B.2 Channel state dynamic parameters

Recall that our prior on the autoregressive coefficients $\mathbf{a}_k$ is a multivariate normal with zero mean and covariance Σ_0,

$$\begin{aligned}
p(\mathbf{a}_k|\Sigma_0) &= \mathcal{N}(\mathbf{a}_k; \mathbf{0}, \Sigma_0) \\
\log p(\mathbf{a}_k|\Sigma_0) &\propto -\frac{1}{2}\mathbf{a}_k^T \Sigma_0^{-1} \mathbf{a}_k.
\end{aligned} \tag{16}$$

From Eq. (4) the conditional event likelihood given the channel states $\mathbf{z}_{1:T}$ and the event states $Z_{1:T}$ is

$$\begin{aligned}
p(\mathbf{y}_{1:T} \,|\, \mathbf{z}_{1:T}, Z_{1:T}, \{\mathbf{a}_k\}, \{\Delta_l\}) &= \prod_{t=1}^{T} \mathcal{N}(\mathbf{y}_t; \mathbf{A}_{\mathbf{z}_t} \widetilde{\mathbf{Y}}, \Delta_{Z_t}) \\
\log p(\mathbf{y}_{1:T} \,|\, \mathbf{z}_{1:T}, Z_{1:T}, \{\mathbf{a}_k\}, \{\Delta_l\}) &\propto -\frac{1}{2}\sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{A}_{\mathbf{z}_t}\widetilde{\mathbf{Y}}_t)^T \Delta_{Z_t}^{-1} (\mathbf{y}_t - \mathbf{A}_{\mathbf{z}_t}\widetilde{\mathbf{Y}}_t).
\end{aligned} \tag{17}$$

The product of these prior and likelihood terms is the joint distribution over $\mathbf{a}_k$ and $\mathbf{y}_{1:T}$,

$$\log p(\mathbf{a}_k, \mathbf{y}_{1:T} \,|\, \mathbf{z}_{1:T}, Z_{1:T}, \{\mathbf{a}_{k'}\}_{k' \neq k}, \{\Delta_l\}) \propto -\frac{1}{2}\mathbf{a}_k^T \Sigma_0^{-1} \mathbf{a}_k - \frac{1}{2}\sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{A}_{\mathbf{z}_t}\widetilde{\mathbf{Y}}_t)^T \Delta_{Z_t}^{-1} (\mathbf{y}_t - \mathbf{A}_{\mathbf{z}_t}\widetilde{\mathbf{Y}}_t). \tag{18}$$

We take a brief tangent to prove a useful identity,

Lemma B.1. *Let the column vector* $\mathbf{x} \in \mathbb{R}^m$ *and the symmetric matrix* $A \in \mathbb{S}^{m \times m}$ *be defined as*

$$\mathbf{x} = \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} \quad \textit{and} \quad A = \begin{bmatrix} B & C \\ C^T & D \end{bmatrix},$$

where $\mathbf{y} \in \mathbb{R}^p$, $\mathbf{z} \in \mathbb{R}^q$, $B \in \mathbb{S}^{p \times p}$, $D \in \mathbb{S}^{q \times q}$, $C \in \mathbb{R}^{p \times q}$, *and* $m = p + q$. *Then*

$$\mathbf{x}^T A \mathbf{x} = \mathbf{y}^T B \mathbf{y} + \mathbf{z}^T D \mathbf{z} + 2\mathbf{y}^T C \mathbf{z}. \tag{19}$$

Proof.

$$\begin{aligned} \mathbf{x}^T A \mathbf{x} &= \begin{bmatrix} \mathbf{y}^T & \mathbf{z}^T \end{bmatrix} \begin{bmatrix} B & C \\ C^T & D \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{y}^T & \mathbf{z}^T \end{bmatrix} \begin{bmatrix} B\mathbf{y} + C\mathbf{z} \\ C^T\mathbf{y} + D\mathbf{z} \end{bmatrix} \\ &= \mathbf{y}^T B \mathbf{y} + \mathbf{y}^T C \mathbf{z} + \mathbf{z}^T C^T \mathbf{y} + \mathbf{z}^T D \mathbf{z} \\ &= \mathbf{y}^T B \mathbf{y} + \mathbf{z}^T D \mathbf{z} + 2\mathbf{y}^T C \mathbf{z} \end{aligned}$$

□

Note that this identity also holds for any permutation p applied to the rows of $\mathbf{x}$ and the rows and columns of A. We now can manipulate the likelihood term of Eq. (18) into a form that separates $\mathbf{a}_k$ from $\mathbf{a}_{k' \neq k}$. Suppose that $\mathbf{k}^+$ denotes the indices of the N channels where $z_t^{(i)} = k$ and $\mathbf{k}^- = \{1, \ldots, N\}/\mathbf{k}^+$ denotes those for whom $z_t^{(i)} \neq k$. Furthermore, we use the superscript indexing on these two sets of indices to select the corresponding portions of the $\mathbf{y}_t$ vector and the $\mathbf{A}_{\mathbf{z}_t}$, $\widetilde{\mathbf{Y}}_T$, and $\Delta_{Z_t}^{-1}$ matrices. We start by decomposing the likelihood term into three parts,

$$\begin{aligned} &(\mathbf{y}_t - \mathbf{A}_{\mathbf{z}_t}\widetilde{\mathbf{Y}}_t)^T \Delta_{Z_t}^{-1} (\mathbf{y}_t - \mathbf{A}_{\mathbf{z}_t}\widetilde{\mathbf{Y}}_t) = \\ &\quad \left(\mathbf{y}_t^{(\mathbf{k}^+)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^+,\mathbf{k}^+)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^+,\mathbf{k}^+)}\right)^T \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)} \left(\mathbf{y}_t^{(\mathbf{k}^+)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^+,\mathbf{k}^+)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^+,\mathbf{k}^+)}\right) + \\ &\quad 2\left(\mathbf{y}_t^{(\mathbf{k}^+)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^+,\mathbf{k}^+)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^+,\mathbf{k}^+)}\right)^T \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)} \left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right) + \\ &\quad \left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right)^T \Delta_{Z_t}^{-1(\mathbf{k}^-,\mathbf{k}^-)} \left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right), \end{aligned} \tag{20}$$

which we then insert into our previous expression (Eq. (18)) for the joint distribution of $\mathbf{a}_k$ and $\mathbf{y}_{1:T}$,

$$\begin{aligned} &\log p(\mathbf{a}_k, \mathbf{y}_{1:T} \,|\, \mathbf{z}_{1:T}, Z_{1:T}, \{\mathbf{a}_{k'}\}_{k' \neq k}, \{\Delta_l\}) \propto -\frac{1}{2}\mathbf{a}_k^T \Sigma_0^{-1} \mathbf{a}_k - \\ &\quad \frac{1}{2}\sum_{t=1}^{T} \left\{ \left(\mathbf{y}_t^{(\mathbf{k}^+)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^+,\mathbf{k}^+)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^+,\mathbf{k}^+)}\right)^T \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)} \left(\mathbf{y}_t^{(\mathbf{k}^+)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^+,\mathbf{k}^+)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^+,\mathbf{k}^+)}\right) + \right. \\ &\quad 2\left(\mathbf{y}_t^{(\mathbf{k}^+)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^+,\mathbf{k}^+)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^+,\mathbf{k}^+)}\right)^T \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)} \left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right) + \\ &\quad \left. \left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right)^T \Delta_{Z_t}^{-1(\mathbf{k}^-,\mathbf{k}^-)} \left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right) \right\}. \end{aligned} \tag{21}$$

Conditioning on $\mathbf{y}_{1:T}$ allows us to absorb the third term of the sum into the proportionality, and after replacing $\mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^+,\mathbf{k}^+)}\widetilde{\mathbf{Y}}^{(\mathbf{k}^+,\mathbf{k}^+)}$ with a more explicit expression, we have

$$\log p(\mathbf{a}_k \,|\, \mathbf{y}_{1:T}, \mathbf{z}_{1:T}, Z_{1:T}, \{\mathbf{a}_{k'}\}_{k'\neq k}, \{\Delta_l\}) \propto -\frac{1}{2}\mathbf{a}_k^T \Sigma_0^{-1}\mathbf{a}_k - \frac{1}{2}\sum_{t=1}^{T}\Bigg\{\left(\mathbf{y}_t^{(\mathbf{k}^+)} - \left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]^T \mathbf{a}_k\right)^T \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)} \left(\mathbf{y}_t^{(\mathbf{k}^+)} - \left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]^T \mathbf{a}_k\right) + 2\left(\mathbf{y}_t^{(\mathbf{k}^+)} - \left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]^T \mathbf{a}_k\right)^T \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)} \left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right)\Bigg\}, \quad (22)$$

which we can further expand to yield

$$\log p(\mathbf{a}_k \,|\, \mathbf{y}_{1:T,}\mathbf{z}_{1:T}, Z_{1:T}, \{\mathbf{a}_{k'}\}_{k'\neq k}, \{\Delta_l\}) \propto -\frac{1}{2}\mathbf{a}_k^T \Sigma_0^{-1}\mathbf{a}_k - \frac{1}{2}\sum_{t=1}^{T}\Bigg\{\left(\mathbf{y}_t^{(\mathbf{k}^+)}\right)^T \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)}\left(\mathbf{y}_t^{(\mathbf{k}^+)}\right) + \left(\mathbf{a}_k^T\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]\right)\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)}\left(\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]^T\mathbf{a}_k\right) - 2\left(\mathbf{y}_t^{(\mathbf{k}^+)}\right)^T\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)}\left(\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]^T\mathbf{a}_k\right)\Bigg\} - \sum_{t=1}^{T}\Bigg\{\mathbf{y}_t^{(\mathbf{k}^+)}\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)}\left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right) - \left(\mathbf{a}_k^T\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]\right)\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)}\left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right)\Bigg\}. \quad (23)$$

Absorbing more terms unrelated to $\mathbf{a}_k$ into the proportionality, we have

$$\log p(\mathbf{a}_k \,|\, \mathbf{y}_{1:T,}\mathbf{z}_{1:T}, Z_{1:T}, \{\mathbf{a}_{k'}\}_{k'\neq k}, \{\Delta_l\}) \propto -\frac{1}{2}\mathbf{a}_k^T \Sigma_0^{-1}\mathbf{a}_k - \frac{1}{2}\sum_{t=1}^{T}\Bigg\{\left(\mathbf{a}_k^T\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]\right)\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)}\left(\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]^T\mathbf{a}_k\right) - 2\left(\mathbf{y}_t^{(\mathbf{k}^+)}\right)^T\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)}\left(\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]^T\mathbf{a}_k\right)\Bigg\} - \sum_{t=1}^{T}\Bigg\{-\left(\mathbf{a}_k^T\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]\right)\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)}\left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right)\Bigg\}, \quad (24)$$

which after some rearranging gives

$$\log p(\mathbf{a}_k \,|\, \mathbf{y}_{1:T,}\mathbf{z}_{1:T}, Z_{1:T}, \{\mathbf{a}_{k'}\}_{k'\neq k}, \{\Delta_l\}) \propto -\frac{1}{2}\mathbf{a}_k^T\left\{\Sigma_0^{-1} + \sum_{t=1}^{T}\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)}\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]^T\right\}\mathbf{a}_k + \mathbf{a}_k^T\left\{\sum_{t=1}^{T}\left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)}\left(\mathbf{y}_t^{(\mathbf{k}^+)}\right) + \left[\widetilde{\mathbf{y}}_t^{(k_1^+)}|\cdots|\widetilde{\mathbf{y}}_t^{(k_{|\mathbf{k}^+|}^+)}\right]\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)}\left(\mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)}\widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}\right)\right\}. \quad (25)$$

Before completing the square, we will find it useful to introduce a bit more notation to simplify the expression,

$$\bar{\mathbf{Y}}_t^{(\mathbf{k}^+)} = \left[\widetilde{\mathbf{y}}_t^{(k_1^+)} | \cdots | \widetilde{\mathbf{y}}_t^{(k^+_{|\mathbf{k}^+|})}\right] \quad \text{and} \quad \boldsymbol{\epsilon}_t^{(\mathbf{k}^-)} = \mathbf{y}_t^{(\mathbf{k}^-)} - \mathbf{A}_{\mathbf{z}_t}^{(\mathbf{k}^-,\mathbf{k}^-)} \widetilde{\mathbf{Y}}_t^{(\mathbf{k}^-,\mathbf{k}^-)}, \tag{26}$$

yielding

$$\begin{aligned}\log p(\mathbf{a}_k \,|\, \mathbf{y}_{1:T,}\mathbf{z}_{1:T}, Z_{1:T}, \{\mathbf{a}_{k'}\}_{k'\neq k}, \{\Delta_l\}) \propto \\ -\frac{1}{2}\mathbf{a}_k^T \left\{\Sigma_0^{-1} + \sum_{t=1}^T \bar{\mathbf{Y}}_t^{(\mathbf{k}^+)} \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)} \bar{\mathbf{Y}}_t^{T(\mathbf{k}^+)}\right\} \mathbf{a}_k + \\ \mathbf{a}_k^T \left\{\sum_{t=1}^T \bar{\mathbf{Y}}_t^{(\mathbf{k}^+)} \left(\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)} \mathbf{y}_t^{(\mathbf{k}^+)} + \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)} \boldsymbol{\epsilon}_t^{(\mathbf{k}^-)}\right)\right\}.\end{aligned} \tag{27}$$

We desire an expression in the form $-\frac{1}{2}(\mathbf{a}_k - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{a}_k - \boldsymbol{\mu}_k)$ for unknown $\boldsymbol{\mu}_k$ and Σ_k^{-1} so that it conforms to the multivariate normal density with mean $\boldsymbol{\mu}_k$ and precision Σ_k^{-1}. We already have our Σ_k^{-1} value from the quadratic term above,

$$\Sigma_k^{-1} = \Sigma_0^{-1} + \sum_{t=1}^T \bar{\mathbf{Y}}_t^{(\mathbf{k}^+)} \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)} \bar{\mathbf{Y}}_t^{T(\mathbf{k}^+)}, \tag{28}$$

which allows us to solve the cross-term for $\boldsymbol{\mu}_k$:

$$-\frac{1}{2}(-2\boldsymbol{\mu}_k^T \Sigma_k^{-1} \mathbf{a}_k) = \mathbf{a}_k^T \left(\sum_{t=1}^T \bar{\mathbf{Y}}_t^{(\mathbf{k}^+)} \left(\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)} \mathbf{y}_t^{(\mathbf{k}^+)} + \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)} \boldsymbol{\epsilon}_t^{(\mathbf{k}^-)} +\right)\right)$$

$$\Sigma_k^{-1} \boldsymbol{\mu}_k = \sum_{t=1}^T \bar{\mathbf{Y}}_t^{(\mathbf{k}^+)} \left(\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)} \mathbf{y}_t^{(\mathbf{k}^+)} + \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)} \boldsymbol{\epsilon}_t^{(\mathbf{k}^-)} +\right) \tag{29}$$

$$\tag{30}$$

We can pull the final required $-\frac{1}{2}\boldsymbol{\mu}_k^T \Sigma_k^{-1} \boldsymbol{\mu}_k$ term from the proportionality and thus complete the square. Thus, we have the form of the posterior for $\mathbf{a}_k$,

$$\begin{aligned}p(\mathbf{a}_k \,|\, \mathbf{y}_{1:T,}\mathbf{z}_{1:T}, Z_{1:T}, \{\mathbf{a}_{k'}\}_{k'\neq k}, \{\Delta_l\}) &\propto \exp\left(-\frac{1}{2}(\mathbf{a}_k - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{a}_k - \boldsymbol{\mu}_k)\right) \\ &= \mathcal{N}(\mathbf{a}_k; \boldsymbol{\mu}_k, \Sigma_k),\end{aligned} \tag{31}$$

where

$$\begin{aligned}\Sigma_k^{-1} &= \Sigma_0^{-1} + \sum_{t=1}^T \bar{\mathbf{Y}}_t^{(\mathbf{k}^+)} \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)} \bar{\mathbf{Y}}_t^{T(\mathbf{k}^+)} \\ \Sigma_k^{-1} \boldsymbol{\mu}_k &= \sum_{t=1}^T \bar{\mathbf{Y}}_t^{(\mathbf{k}^+)} \left(\Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^+)} \mathbf{y}_t^{(\mathbf{k}^+)} + \Delta_{Z_t}^{-1(\mathbf{k}^+,\mathbf{k}^-)} \boldsymbol{\epsilon}_t^{(\mathbf{k}^-)}\right).\end{aligned} \tag{32}$$

B.3 Event variables

We first sample the event state sequence $Z_{1:T}$ and then its state transition parameters $\boldsymbol{\phi}$.

B.3.1 Event state sequence, $Z_{1:T}$

The mechanics of sampling the event state sequence $Z_{1:T}$ directly parallel those of sampling the individual channel state sequences $z_{1:T}^{(i)}$. The joint probability of the event state sequence is given by

$$p(Z_{1:T} \,|\, \mathbf{y}_{1:T}, \mathbf{z}_{1:T}, \boldsymbol{\phi}, \{\mathbf{a}_k\}, \{\Delta_l\}) = \\ p(Z_1 \,|\, \mathbf{y}_1, \mathbf{z}_1, \boldsymbol{\phi}, \{\mathbf{a}_k\}, \{\Delta_l\}) \sum_{t=2}^{T} p(Z_t \,|\, \mathbf{y}_{t:T}, \mathbf{z}_{t:T}, Z_{t-1}, \boldsymbol{\phi}, \{\mathbf{a}_k\}, \{\Delta_l\}). \quad (33)$$

We again follow the sum-product algorithm, using a vector $\boldsymbol{\psi}_t \in \mathbb{R}^L$ of backward messages from time point $t+1$ to t with each element proportional to the likelihood of future observations $\mathbf{y}_{t+1:T}$ given $Z_t = l$,

$$\psi_{l,t} \propto p(\mathbf{y}_{t+1:T} \,|\, \mathbf{z}_{t+1:T}, Z_t = l, \boldsymbol{\phi}, \{\mathbf{a}_k\}, \{\Delta_l\}), \quad (34)$$

which we compactly represent in vector notation,

$$\boldsymbol{\psi}_t \propto \boldsymbol{\phi}(\mathbf{v}_{t+1} \circ \boldsymbol{\psi}_{t+1}), \quad (35)$$

with $\mathbf{v}_t \in \mathbb{R}^L$ now representing the conditional event likelihoods under the L possible event states (following from Eq. (4)).

B.3.2 Event transition parameters, $\boldsymbol{\phi}$

The event state transition parameters $\boldsymbol{\phi} = [\boldsymbol{\phi}_1 | \cdots | \boldsymbol{\phi}_L]^T$ parallel $\boldsymbol{\pi}$ for individual channels. The main difference, however, is that we assume all possible event states are available to each event, whereas individual channels are constrained by the beta process to only use particular states. The Dirichlet posterior for $\boldsymbol{\phi}_l$ simply involves transition counts $\mathbf{n}_l \in \mathbb{R}^L$ from event state l to all L states,

$$\begin{aligned} \boldsymbol{\phi}_l &\sim \mathrm{Dir}(\alpha_e \boldsymbol{\beta} + \mathbf{e}_l \kappa_e + \mathbf{n}_l) \\ \boldsymbol{\beta} &\sim \mathrm{Dir}(\gamma_e / L + \bar{m}_{\cdot 1}, \ldots, \gamma_e / L + \bar{m}_{\cdot L}). \end{aligned} \quad (36)$$

The auxiliary variable $\bar{m}_{ll'}$ is defined by

$$\begin{aligned} \bar{m}_{ll'} &= \begin{cases} m_{ll'}, & l \neq l' \\ m_{ll} - w_l, & l = l' \end{cases} \\ m_{ll'} &= \sum_r \theta_r \\ \theta_r &\sim \mathrm{Ber}\left(\frac{\alpha_e \beta_l + \kappa_e \mathbf{1}(l = l')}{\alpha_e \beta_l + \mathbf{1}(l = l') + r} \right), \qquad r = 1, \ldots, n_{ll'} \\ w_l &\sim \mathrm{Binomial}\left(m_{ll'}, \frac{\rho_e}{\rho_e + \beta_l (1 - \rho_e)} \right). \end{aligned} \quad (37)$$

B.4 Event state covariance parameters

B.4.1 Event state covariances, Δ_l

In the most straightforward formulation of the Gaussian graphical model (Dawid & Lauritzen, 1993), a set of variables and their conditional independencies—described by the vertices V and edges E, respectively—is decomposed into an ordered series of prime components $(P_1, P_2, \ldots, P_Q)$ of the graph $G = (V, E)$. Each prime component P_i is connected by a set of separating variables S_i, where $S_i = P_i \cup P_j$ for some $j < i$. While interesting techniques for inferring the vertices to produce both decomposable and nondecomposible graph structures exist (Jones et al., 2005; Wang et al., 2011), in this work we assume that the vertex structure is known. Specifically, we define the conditional independencies based on the spatial adjacencies of the iEEG channels, with a few exceptions to make the graphical model fully decomposable.

The sufficient statistics associated with the event states stem from the event innovations at time t,

$$\boldsymbol{\epsilon}_t = \mathbf{y}_t - \mathbf{A}_{\mathbf{z}_t} \widetilde{\mathbf{Y}}_t. \tag{38}$$

For each event state l, we have

$$b_l = b_0 + |\{t | Z_t = l\}| \quad \text{and} \quad D_l = D_0 + \sum_{t \,|\, Z_t = l} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T, \tag{39}$$

which we then use to sample from the hyper-inverse Wishart posterior for that state,

$$\Delta_l \sim \text{HIW}_G(b_l, D_l). \tag{40}$$

Details on how to efficiently sample from a HIW distribution are provided in (Carvalho et al., 2007).

B.5 Hyperparameters

We sample the various hyperparameters of the model as well. For completeness, we include the posteriors for each below, which follow those described in more detail in Fox et al. (2011, Supplementary Materials) and Fox et al. (2009).

B.5.1 Sticky HDP-HMM hyperparameters, γ_e, α_e, κ_e, ρ_e

Instead of sampling α_e and κ_e independently, we instead introduce an additional parameter $\rho_e = \kappa_e/(\alpha_e + \kappa_e)$ and sample $(\alpha_e + \kappa_e)$ and ρ_e instead, which is simpler than sampling α_e and κ_e independently. Recall that we are working with a truncated approximation to the DP that involves L discrete atoms.

$(\boldsymbol{\alpha_e + \kappa_e})$ We place a Gamma(a, b) prior on $(\alpha_e + \kappa_e)$ and use the auxiliary variables $\{r_l\}_{l=1}^L$ and $\{s_l\}_{l=1}^L$ to sample from the posterior,

$$p(\alpha_e + \kappa_e \,|\, \{r_l\}_{l=1}^L, \{s_l\}_{l=1}^L, m_{1\cdot}, \ldots, m_{1L}) \propto \text{Gamma}\left(a + m_{\cdot\cdot} - \sum_{l=1}^L s_l, b - \sum_{l=1}^L \log(r_l)\right), \tag{41}$$

where $m_{\cdot\cdot} = \sum_{l,l'=1}^L m_{ll'}$ is the sum over auxiliary variables $m_{ll'}$ defined in Eq. (37), and the auxiliary variables $\{r_l\}_{l=1}^L$ and $\{s_l\}_{l=1}^L$ are sampled as

$$r_l \sim \text{Beta}(\alpha + \kappa + 1, n_{l\cdot})$$
$$s_l \sim \text{Ber}(n_{l\cdot}/(n_{l\cdot} + \alpha + \kappa)).$$

$\boldsymbol{\rho_e}$ We place a Beta(c, d) prior on ρ_e and use the auxiliary variables $\{w_{l\cdot}\}_{l=1}^L$ to sample from the posterior,

$$p(\rho_e \,|\, \{w_{l\cdot}\}) \propto \text{Beta}\left(\sum_l w_{l\cdot} + c, m_{\cdot\cdot} - \sum_l w_{l\cdot} + d\right), \tag{42}$$

where for $w_{ls} \sim \text{Ber}(\rho)$ over $s = 1, \ldots, m_{ll}$, the posterior for $w_{l\cdot}$ is

$$p(w_{l\cdot} \,|\, m_{ll}, \beta_l, \rho_e) \propto \text{Bin}(m_{ll}, \rho_e + \beta_l(1 - \rho_e)) \tag{43}$$

$\boldsymbol{\gamma_e}$ We place a Gamma(a,b) prior γ_e and again use auxiliary variables v and q to sample from the posterior,

$$\gamma_e \sim \text{Gamma}\left(a + \bar{L} - q, b - \log v\right), \tag{44}$$

where, recalling again Eq. (37), the auxiliary variables are sampled as

$$\begin{aligned} v &\sim \text{Beta}(\gamma + 1, \bar{m}_{\cdot\cdot}) \\ q &\sim \text{Ber}(\bar{m}_{\cdot\cdot}/(\gamma + \bar{m}_{\cdot\cdot})) \\ \bar{L} &= \sum_{l=1}^{L} \mathbf{1}(\bar{m}_{\cdot l} > 0) \end{aligned}$$

and $\bar{m}_{\cdot\cdot} = \sum_{l,l'=1}^{L} \bar{m}_{ll'}$.

B.5.2 BP-AR-HMM hyperparameters, γ_c, κ_c

We use Metropolis-Hastings steps to propose a new value γ_c' from gamma distributions with fixed variance $\sigma^2_{\gamma_c}$ and accept with probability $\min(r(\gamma_c' \mid \gamma_c), 1)$,

$$\begin{aligned} r(\gamma_c' \mid \gamma_c) &= \frac{p(\{\boldsymbol{\pi}^{(i)}\} \mid \gamma_c', \kappa, \mathbf{F}) p(\gamma_c' \mid \gamma_c^2/\sigma^2_{\gamma_c}, \gamma_c/\sigma^2_{\gamma_c}) p(\gamma_c \mid \gamma_c', \sigma^2_{\gamma_c})}{p(\{\boldsymbol{\pi}^{(i)}\} \mid \gamma_c, \kappa, \mathbf{F}) p(\gamma_c \mid \gamma_c^2/\sigma^2_{\gamma_c}, \gamma_c/\sigma^2_{\gamma_c}) p(\gamma_c' \mid \gamma_c, \sigma^2_{\gamma_c})} \\ &= \frac{p(\{\boldsymbol{\pi}^{(i)}\} \mid \gamma_c', \kappa, \mathbf{F})}{p(\{\boldsymbol{\pi}^{(i)}\} \mid \gamma_c, \kappa, \mathbf{F})} \frac{\Gamma(\nu) \gamma_c^{\nu' - \nu - a}}{\Gamma(\nu') \gamma_c^{\nu - \nu' - a}} \exp\left(-b(\gamma_c' - \gamma_c) \sigma_{\gamma_c}^{2(\nu - \nu')}\right), \end{aligned} \tag{45}$$

where $\nu = \gamma_c^2/\sigma^2_{\gamma_c}$, $\nu' = {\gamma_c'}^2/\sigma^2_{\gamma_c}$, and we have a Gamma$(a,b)$ prior on γ_c. Recall that the transition parameters $\boldsymbol{\pi}^{(i)}$ are independent over i, and thus their Dirichlet likelihoods multiply. The proposal and acceptance ratio for κ_c is similar.

B.5.3 BP hyperparameter, α_c

We place a Gamma(a,b) prior on α_c, which implies a gamma posterior of the form

$$p(\alpha_c \mid \mathbf{F}, a, b) \propto \text{Gamma}(a + K_+, b + \sum_{i=1}^{N} (1/i)), \tag{46}$$

where K_+ denotes the number of unique channel states that are activated in at least one of the channels.

C Simulation experiment

Data We simulated data from six time series in a 2x3 arrangement, with vertices connecting all adjacent nodes (i.e., two cliques of 4 nodes each). We generated 2000 scalar observations using an first-order AR process with five channel states—AR coefficients linearly spaced between -0.9 and 0.9—and three event states with covariances shown in the bottom left of Fig. 1. Channel and event state transition matrices were set to 0.99 and 0.9, respectively, for a self-transition and uniform between the other states. We generated channel feature indicators using a $\alpha_c = 10$.

Results We ran the MCMC sampler for 6000 iterations, taking 500 samples after 1000-iteration burn-in and 10-sample thinning. Fig. 1 shows the generated data and its true states along with the inferred states and event state covariances for one of the posterior samples. The event state matching is almost perfect, and the channel state matching is quite good, though we see that the sampler added an additional (yellow) state in the middle of the first time series when it should have assigned that section to the cyan state. The scale and structure of the estimated event state covariances match the true covariances quite well. Furthermore, Table 1 shows how the posterior estimates of the channel state AR coefficients also center well around the true values.

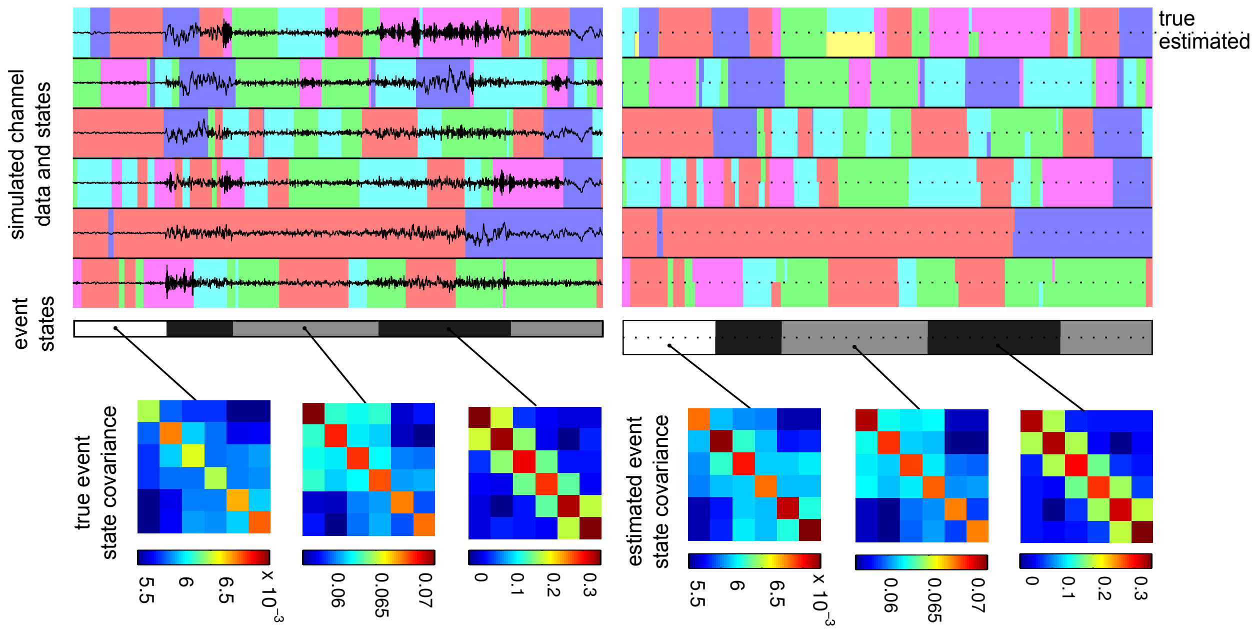

Figure 1: (**top left**) The six simulated channel time series overlaid on the five true channel states, denoted by different colors; the three true event states are shown in grayscale in the bar below. (**top right**) The true and estimated channel (color) and event (grayscale) states shown below for comparison after 6000 MCMC iterations. The true (**bottom left**) and estimated (**bottom right**) event state innovation covariances.

channel state	true $\mathbf{a}_k$	post. $\mathbf{a}_k$ mean	post. $\mathbf{a}_k$ 95% interval
1	-0.900	-0.906	[-0.917, -0.896]
2	-0.450	-0.456	[-0.474, -0.436]
3	0	-0.009	[-0.038, 0.020]
4	0.450	0.445	[0.425, 0.466]
5	0.900	0.902	[0.890, 0.913]

Table 1: The true and estimated values for the channel state coefficients in the simulated dataset.

D Model parameters used

The model parameters used in the simulation experiment and the EEG experiments are given in Tables 2 and 3.

E Seizure offset parsing

Fig. 2 shows the event state parsing at the offset of the same seizure whose onset is shown in Fig. 3 of the main paper. Though the channel states are intuitive for this offset, we have shown only the event states to illustrate how well the model is capable of distinguishing subtle transitions in the event dynamics like that from the first half to the second half of the offset. In this parsing of the transition, we see how the seizure moves from strong correlations in the spikings of a few channels to a more widespread correlation structure and synchronized discharge pattern. The automatic identification of brief intervals of synchronized spiking makes it easy for a clinician to calculate changes in the inter-spike interval, a quantity of clinical importance.

parameter	description	value
N	number of time series per event	6
r	AR model order	1
$\mathbf{m}_0$	$\mathbf{a}_k$ $\mathcal{N}$ prior mean	0
Σ_0	$\mathbf{a}_k$ $\mathcal{N}$ prior covariance	$0.1 \cdot I_{1\times 1}$
L	truncated number of event states	20
b_0	Δ_l HIW prior degrees of freedom	$N+3$
D_0	Δ_l HIW prior scale	$(b_0 - N - 1) \cdot (0.05 \cdot I_{N\times N} + 0.05)$
$(\alpha_e + \kappa_e)_0$	$\alpha_e + \kappa_e$ Gamma prior	$(1, 1)$
γ_{e0}	γ_e Gamma prior	$(1, 1)$
ρ_{e0}	ρ_e Gamma prior	$(1, 1)$
γ_{c0}	γ_c Gamma prior	$(1, 1)$
κ_{c0}	κ_c Gamma prior	$(1000, 1)$
γ_p	γ_c Metropolis-Hastings proposal variance	1
κ_p	κ_c Metropolis-Hastings proposal variance	100
α_{c0}	α_c Gamma prior	$(1, 1)$

Table 2: Parameters used in simulation experiment

F Assessing the utility of the beta process

We explored the benefit of the BP-AR-HMM (both spatial and non-spatial) relative to variants of these models that are finite Bayesian AR-HMMs without the feature-based modeling provided by the beta process. (Note: the finite AR-HMM examined here can be equated with a truncated sticky hierarchical Dirichlet process HMM, as in Fox et al. (2011). In Fig. 3, we see the improved heldout predictive log-likelihood of the BP-based models, though the incorporation of the associated feature sampling comes at a significant computational cost.

References

Carvalho, C. M., Massam, H., and West, M. Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika*, 94(3):647–659, August 2007.

Dawid, A. P. and Lauritzen, S. L. Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 21(3):1272–1317, 1993.

Fox, E. B., Sudderth, E. B, Jordan, M. I., and Willsky, A. S. Sharing features among dynamical systems with beta processes. *Advances in Neural Information Processing Systems*, 22, 2009.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics*, 5(2A):1020–1056, 2011.

Hughes, M., Fox, E., and Sudderth, E. Effective Split-Merge Monte Carlo Methods for Nonparametric Models of Sequential Data. In *Advances in Neural Information Processing Systems*, 2012.

Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., and West, M. Experiments in Stochastic Computation for High-Dimensional Graphical Models. *Statistical Science*, 20(4):388–400, November 2005.

Wang, H., Reeson, C., and Carvalho, C.M. Dynamic financial index models: Modeling conditional dependencies via graphs. *Bayesian Analysis*, 2011.

parameter	description	value
N	number of time series per event	16 and 6
r	AR model order	5
$\mathbf{m}_0$	$\mathbf{a}_k$ $\mathcal{N}$ prior mean	$\mathbf{0}$
Σ_0	$\mathbf{a}_k$ $\mathcal{N}$ prior covariance	$\mathrm{Cov}(\{y_t^{(i)}\}_{\forall t,i})$
L	truncated number of event states	30
b_0	Δ_l (H)IW prior degrees of freedom	$N+3$
D_0	Δ_l (H)IW prior scale	$(b_0-N-1)\cdot\mathrm{Cov}(\{\mathbf{y}_{t+1}-\mathbf{y}_t\}_{\forall t})$
$(\alpha_e+\kappa_e)_0$	$\alpha_e+\kappa_e$ Gamma prior	$(1,1)$
γ_{e0}	γ_e Gamma prior	$(1,1)$
ρ_{e0}	ρ_e Gamma prior	$(1,1)$
γ_{c0}	γ_c Gamma prior	$(1,1)$
κ_{c0}	κ_c Gamma prior	$(1000,1)$
γ_p	γ_c Metropolis-Hastings proposal variance	1
κ_p	κ_c Metropolis-Hastings proposal variance	100
α_{c0}	α_c Gamma prior	$(1,1)$

Table 3: Parameters used in epileptic seizures and bursts experiments. When applicable, the same parameters were used for the standard BP-AR-HMM as in the correlated BP-AR-HMMs. The analysis of two two seizures involved 16 iEEG channels, and the analysis of the 15 bursts and single seizure involved 6 iEEG channels.

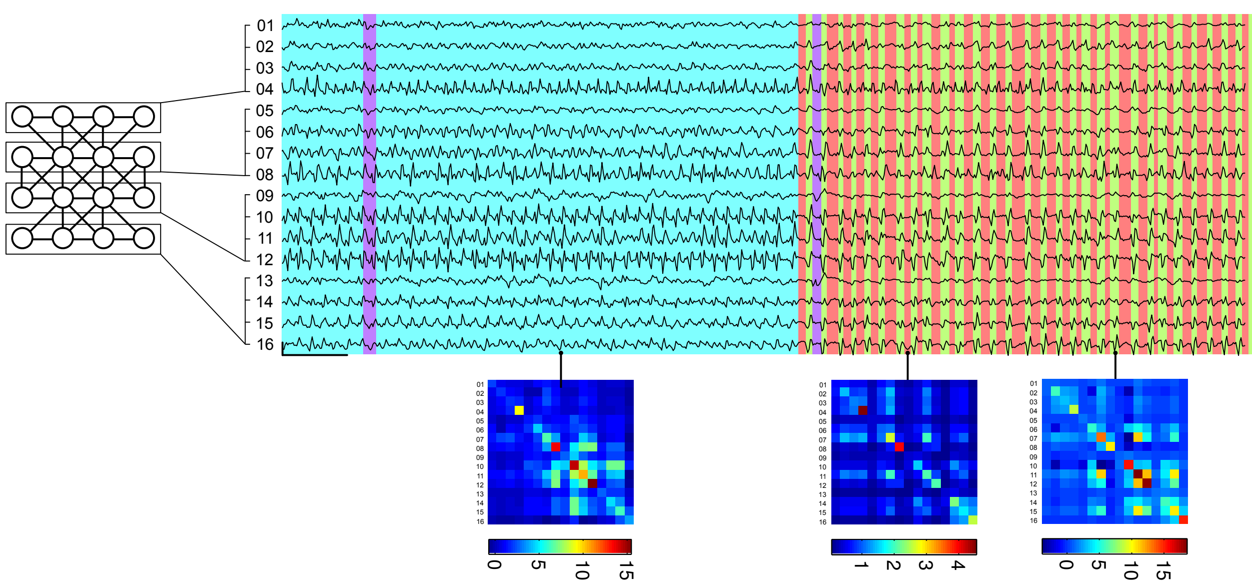

Figure 2: A representative sample showing the event state parsing (copied across all EEG channels) of a seizure by the HIW-spatial BP-AR-HMM model.

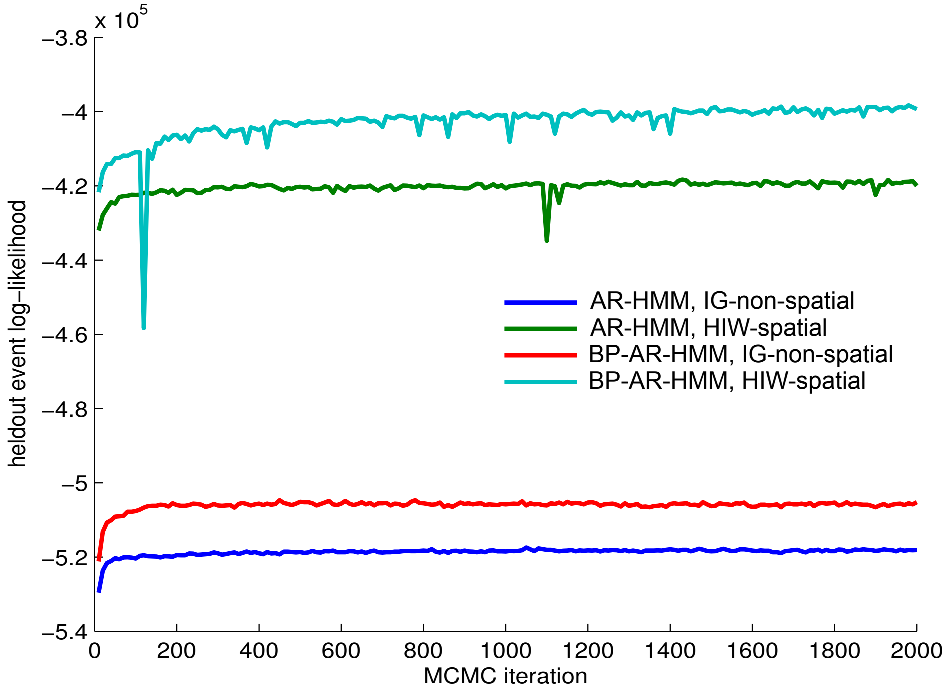

Figure 3: The heldout event log-likelihood of a single MCMC chain over the first 2000 iterations for four models with two HIW-spatial models and two invovling beta process feature sampling (BP-).