times between sampling $\boldsymbol{\alpha}^{(k)}$ given $\mathbf{A}^{(k)}$ and $\mathbf{A}^{(k)}$ given $\boldsymbol{\alpha}^{(k)}$ before moving to the next sampling stage.

2) Sampling Measurement Noise R (HDP-SLDS Only): For the HDP-SLDS, we additionally sample the measurement noise covariance R conditioned on the sampled state sequence $\boldsymbol{x}_{1:T}$.

3) Block Sampling $z_{1:T}$: As shown in [21], the mixing rate of the Gibbs sampler for the HDP-HMM can be dramatically improved by using a *truncated* approximation to the HDP and jointly sampling the mode sequence using a variant of the forward-backward algorithm. In the case of our switching dynamical systems, we must account for the direct correlations in the observations in our likelihood computation. The variant of the forward-backward algorithm we use here then involves computing backward messages $m_{t+1,t}(z_t) \propto p(\boldsymbol{\psi}_{t+1:T}|z_t, \bar{\boldsymbol{\psi}}_t, \boldsymbol{\pi}, \boldsymbol{\theta})$ for each $z_t \in \{1, \ldots, L\}$ with L the chosen truncation level, followed by recursively sampling each z_t conditioned on z_{t-1} from

$$p\left(z_t \mid z_{t-1}, \boldsymbol{\psi}_{1:T}, \boldsymbol{\pi}, \boldsymbol{\theta}\right) \propto \pi_{z_{t-1},z_t} p\left(\boldsymbol{\psi}_t \mid \bar{\boldsymbol{\psi}}_{t-1}, \mathbf{A}^{(z_t)}, \Sigma^{(z_t)}\right) m_{t+1,t}(z_t). \quad (30)$$

Joint sampling of the mode sequence is especially important when the observations are directly correlated via a dynamical process since this correlation further slows the mixing rate of the sequential sampler of Teh *et al.* [16]. Note that using an order L weak limit approximation to the HDP still encourages the use of a sparse subset of the L possible dynamical modes.

4) Block Sampling $\boldsymbol{x}_{1:T}$ (HDP-SLDS Only): Conditioned on the mode sequence $z_{1:T}$ and the set of SLDS parameters $\boldsymbol{\theta} = \{\mathbf{A}^{(k)}, \Sigma^{(k)}, R\}$, our dynamical process simplifies to a time-varying linear dynamical system. We can then block sample $\boldsymbol{x}_{1:T}$ by first running a backward Kalman filter to compute $m_{t+1,t}(\boldsymbol{x}_t) \propto p(\boldsymbol{y}_{t+1:T}|\boldsymbol{x}_t, z_{t+1:T}, \boldsymbol{\theta})$ and then recursively sampling each $\boldsymbol{x}_t$ conditioned on $\boldsymbol{x}_{t-1}$ from

$$p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{y}_{1:T}, z_{1:T}, \boldsymbol{\theta}) \propto p\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, A^{(z_t)}, \Sigma^{(z_t)}\right) p(\boldsymbol{y}_t \mid \boldsymbol{x}_t, R) m_{t+1,t}(\boldsymbol{x}_t). \quad (31)$$

The messages are given in information form by $m_{t,t-1}(\boldsymbol{x}_{t-1}) \propto \mathcal{N}^{-1}(\vartheta_{t,t-1}, \Lambda_{t,t-1})$, where the information parameters are recursively defined as

$$\begin{aligned}\vartheta_{t,t-1} &= A^{(z_t)^T}\Sigma^{-(z_t)}\left(\Sigma^{-(z_t)} + \Lambda^b_{t|t}\right)^{-1}\vartheta^b_{t|t}\\ \Lambda_{t,t-1} &= A^{(z_t)^T}\Sigma^{-(z_t)}A^{(z_t)} - A^{(z_t)^T}\Sigma^{-(z_t)}\\ &\quad\times\left(\Sigma^{-(z_t)} + \Lambda^b_{t|t}\right)^{-1}\Sigma^{-(z_t)}A^{(z_t)}. \end{aligned} \quad (32)$$

The standard $\vartheta^b_{t|t}$ and $\Lambda^b_{t|t}$ updated information parameters for a backward running Kalman filter are given by

$$\begin{aligned}\Lambda^b_{t|t} &= C^T R^{-1} C + \Lambda_{t+1,t}\\ \vartheta^b_{t|t} &= C^T R^{-1} y_t + \vartheta_{t+1,t}. \end{aligned} \quad (33)$$

See [42] for a derivation and for a more numerically stable version of this recursion.

5) Sequentially Sampling $z_{1:T}$ (HDP-SLDS Only): For the HDP-SLDS, iterating between the previous sampling stages can lead to slow mixing rates since the mode sequence is sampled conditioned on a sample of the state sequence. For high-dimensional state spaces $\mathbb{R}^n$, this problem is exacerbated. Instead, one can analytically marginalize the state sequence and sequentially sample the mode sequence from $p(z_t \mid z_{\backslash t}, \boldsymbol{y}_{1:T}, \boldsymbol{\pi}, \boldsymbol{\theta})$. This marginalization is accomplished by once again harnessing the fact that conditioned on the mode sequence, our model reduces to a time-varying linear dynamical system. When sampling z_t and conditioning on the mode sequence at all *other* time steps, we can run a forward Kalman filter to marginalize the state sequence $\boldsymbol{x}_{1:t-2}$ producing $p(\boldsymbol{x}_{t-1} \mid \boldsymbol{y}_{1:t-1}, z_{1:t-1}, \boldsymbol{\theta})$ and a backward filter to marginalize $\boldsymbol{x}_{t+1:T}$ producing $p(\boldsymbol{y}_{t+1:T} \mid \boldsymbol{x}_t, z_{t+1:T}, \boldsymbol{\theta})$. Then, for each possible value of z_t, we combine these forward and backward messages with the local likelihood $p(\boldsymbol{y}_t \mid \boldsymbol{x}_t)$ and local dynamic $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{\theta}, z_t = k)$ and marginalize over $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t-1}$ resulting in the likelihood of the observation sequence $\boldsymbol{y}_{1:T}$ as a function of $\boldsymbol{z}_t$. This likelihood is combined with the prior probability of transitioning from z_{t-1} to $z_t = k$ and from $z_t = k$ to z_{t+1}. The resulting distribution is given by

$$\begin{aligned}p(z_t = k \mid z_{\backslash t}, \boldsymbol{y}_{1:T}, \boldsymbol{\pi}, \boldsymbol{\theta}) &\propto \pi_{z_{t-1},k}\pi_{k,z_{t+1}}\\ &\frac{|\Lambda_t^{(k)}|^{1/2}}{|\Lambda_t^{(k)} + \Lambda^b_{t|t}|^{1/2}}\exp\left(-\frac{1}{2}\vartheta_t^{(k)^T}\Lambda_t^{-(k)}\vartheta_t^{(k)}\right.\\ &\left.+\frac{1}{2}\left(\vartheta_t^{(k)} + \vartheta^b_{t|t}\right)^T\left(\Lambda_t^{(k)} + \Lambda^b_{t|t}\right)^{-1}\left(\vartheta_t^{(k)} + \vartheta^b_{t|t}\right)\right)\end{aligned} \quad (34)$$

with

$$\begin{aligned}\Lambda_t^{(k)} &= \left(\Sigma^{(k)} + \mathbf{A}^{(z_t)}\Lambda^{-f}_{t-1|t-1}\mathbf{A}^{(z_t)^T}\right)^{-1}\\ \vartheta_t^{(k)} &= \left(\Sigma^{(k)} + \mathbf{A}^{(z_t)}\Lambda^{-f}_{t-1|t-1}\mathbf{A}^{(z_t)^T}\right)^{-1}\\ &\quad\times\mathbf{A}^{(z_t)}\Lambda^{-f}_{t-1|t-1}\vartheta^f_{t-1|t-1}. \end{aligned} \quad (35)$$

See [42] for full derivations. Here, $\vartheta^f_{t|t}$ and $\Lambda^f_{t|t}$ are the updated information parameters for a forward running Kalman filter, defined recursively as

$$\begin{aligned}\Lambda^f_{t|t} &= C^T R^{-1} C + \Sigma^{-(z_t)} - \Sigma^{-(z_t)}\mathbf{A}^{(z_t)}\\ &\quad\times\left(\mathbf{A}^{(z_t)^T}\Sigma^{-(z_t)}\mathbf{A}^{(z_t)} + \Lambda^f_{t-1|t-1}\right)^{-1}\\ &\quad\times\mathbf{A}^{(z_t)^T}\Sigma^{-(z_t)}\\ \vartheta^f_{t|t} &= C^T R^{-1} y_t + \Sigma^{-(z_t)}\mathbf{A}^{(z_t)}\\ &\quad\times\left(\mathbf{A}^{(z_t)^T}\Sigma^{-(z_t)}\mathbf{A}^{(z_t)} + \Lambda^f_{t-1|t-1}\right)^{-1}\\ &\quad\times\vartheta^f_{t-1|t-1}. \end{aligned} \quad (36)$$

Note that a sequential node ordering for this sampling step allows for efficient updates to the recursively defined filter parameters. However, this sequential sampling is still computationally intensive, so our Gibbs sampler iterates between blocked sampling of the state and mode sequences many times before interleaving a sequential mode sequence sampling step.

The resulting Gibbs sampler is outlined in Algorithm 1.