We derive a lower bound for a switching state (non)linear dynamical system. Let $\vec{y}$ represent observations, $\vec{x}$ represent latent variables, and $\vec{z}$ represent categorical state variables.

$$\log p(\vec{y}) \geq \int_{\vec{z}} \int_{\vec{x}} q(\vec{z}, \vec{x}|\vec{y}) \log \frac{p(\vec{y}, \vec{x}, \vec{z})}{q(\vec{z}, \vec{x}|\vec{y})} d\vec{x} d\vec{z} \tag{1}$$

where $\vec{z}, \vec{x} \sim q(\vec{z}, \vec{x}|\vec{y})$. In particular, $q(\vec{z}, \vec{x}|\vec{y}) = q(\vec{z}|\vec{x})q(\vec{x}|\vec{y})$ and $p(\vec{y}, \vec{x}, \vec{z}) = p(\vec{y}|\vec{x})p(\vec{x}|\vec{z})p(\vec{z})$. We define the following factorizations:

$$p(\vec{y}|\vec{x}) = \prod_{t=1}^{T} p(y_t|x_t) \tag{2}$$

$$p(\vec{x}|\vec{z}) = \prod_{t=1}^{T} p(x_t|x_{t-1}, z_t) \tag{3}$$

$$p(\vec{z}) = \prod_{t=1}^{T} p(z_t|z_{t-1}) \tag{4}$$

$$q(\vec{z}|\vec{x}) = \prod_{t=1}^{T} q(z_t|z_{t-1}, x_{1:T}^{(1)}, ..., x_{1:T}^{(K)}) \tag{5}$$

$$q(\vec{x}|\vec{y}) = \prod_{t=1}^{T} q(x_t|x_{t-1}, y_{1:T}) \tag{6}$$

Note, $x_{1:T}^{(k)}$ for $k = 1...K$ indicates the latent variables generated from state k. The objective is then:

$$\mathcal{L} = \mathbb{E}_{q(\vec{z},\vec{x}|\vec{y})}[\log \frac{p(\vec{y}|\vec{x})p(\vec{x}|\vec{z})p(\vec{z})}{q(\vec{z}|\vec{x})q(\vec{x}|\vec{y})}] \tag{7}$$

$$= \mathbb{E}_{q(\vec{z},\vec{x}|\vec{y})}[\log p(\vec{y}|\vec{x}) + \log p(\vec{x}|\vec{z}) + \log p(\vec{z}) - \log q(\vec{z}|\vec{x}) - \log q(\vec{x}|\vec{y})] \tag{8}$$

$$= \mathbb{E}[\sum_{t=1}^{T} \{w_t^{(k)} \cdot \sum_{k=1}^{K} \{\log p(y_t|x_t^{(k)}) + \log p(x_t^{(k)}|x_{t-1}^{(k)}, z_t) + \log p(z_t|z_{t-1}) \tag{9}$$

$$- \log q(z_t|z_{t-1}, x_{1:T}^{(1)}, ..., x_{1:T}^{(K)}) - \log q(x_t^{(k)}|x_{t-1}^{(k)}, y_{1:T})\}\}] \tag{10}$$

where $w_t^{(k)}$ is the k-th entry of the soft categorical parameters for z_t. Basically, we are averaging the ELBO over K systems and weighting the loss by the contribution of that system at timestep T.