Idea of Latent Variable Model

In this model, the latent variables $z_i^{(t)}$ influence the explanatory variables $x_i^{(t)}$, which, in turn, affect the observed values $y_i^{(t)}$. Moreover, we introduce a discrete latent variable $c_i^{(t)}$ representing the cluster assignment of time series i at time t. The complete data consists of observed variables $(x_i^{(t)}, y_i^{(t)})$ and latent variables $(z_i^{(t)}, c_i^{(t)})$.

Model Definition

- **Observed Data**:
  - $y_i^{(t)}$: Observed value for time series i at time t.
  - $x_i^{(t)}$: Observed explanatory variables for time series i at time t.
- **Latent Variables**:
  - $z_i^{(t)}$: Unobserved latent variables influencing $x_i^{(t)}$.
  - $c_i^{(t)}$: Unobserved cluster assignment variable of time series i at time t.

Cluster Assignment: Each time series i at time t belongs to a cluster $c_i^{(t)}$:

$$p(c_i^{(t)}) = \pi_{c_i^{(t)}} \tag{1}$$

where π_c is the prior probability of cluster c.

The model equations are:

$$x_i^{(t)} = \mathbf{B}_{\mathbf{c_i}} z_i^{(t)} + \eta_i^{(t)}, \quad \eta_i^{(t)} \sim \mathcal{N}(0, \mathbf{\Sigma}_{\mathbf{c_i}}), \tag{2}$$

Therefore, we have :

$$x_i^{(t)} | z_i^{(t)}, c_i^{(t)} \sim \mathcal{N}(\mathbf{B}_{c_i} z_i^{(t)}, \Sigma_{c_i}) \tag{3}$$

where $\mathbf{B}_{c_i}$ maps latent variables $z_i^{(t)}$ to the explanatory variables $x_i^{(t)}$, and $\eta_i^{(t)}$ is a Gaussian noise. The observed values $y_i^{(t)}$ are modeled as:

$$y_i^{(t)} = \theta_{c_i}^\top x_i^{(t)} + \sum_{j \neq i, c^*(j) = c^*(i)} \gamma_{ij}^\top x_j^{(t)} + \beta_{c_i} u_{c_i}^{(t)} + \epsilon_i^{(t)}, \tag{4}$$

where:

- $\epsilon_i^{(t)} \sim \mathcal{N}(0, \sigma^2)$: Gaussian noise.
- $u_{c^*(i)}^{(t)}$: Cluster-level control.

Likelihood Function

To estimate the parameters $\{\theta_i, \gamma_{ij}, \beta_{c^*(i)}, \mathbf{B}, \mathbf{\Sigma}\}$, we maximize the likelihood of the observed variables $y_i^{(t)}$ and $x_i^{(t)}$ given the latent variables $z_i^{(t)}$.

Joint Distribution

The joint density is given by:

$$p(y_i^{(t)}, x_i^{(t)}, z_i^{(t)}) = p(y_i^{(t)}|x_i^{(t)}) \cdot p(x_i^{(t)}|z_i^{(t)}) \cdot p(z_i^{(t)}). \tag{5}$$

where

$$p(y_i^{(t)}|x_i^{(t)}, c_i^{(t)}) = \mathcal{N}\left(y_i^{(t)} \middle| \theta_{c_i}^\top x_i^{(t)} + \sum_{j \neq i} \gamma_{ij}^\top x_j^{(t)} + \beta_{c_i} u_{c_i}^{(t)}, \sigma_{c_i}^2\right), \tag{6}$$

$$p(x_i^{(t)}|z_i^{(t)}, c_i^{(t)}) = \mathcal{N}(x_i^{(t)}|\mathbf{B}_{c_i} z_i^{(t)}, \mathbf{\Sigma}_{c_i}), \tag{7}$$

$$p(z_i^{(t)}|c_i^{(t)}) = \mathcal{N}(z_i^{(t)}|0, \mathbf{I}). \tag{8}$$

Likelihood

The observed data likelihood is given by:

$$p(y_i^{(t)}, x_i^{(t)}) = \int p(y_i^{(t)}|x_i^{(t)}) \cdot p(x_i^{(t)}|z_i^{(t)}) \cdot p(z_i^{(t)})\, dz_i^{(t)} = \int p(y_i^{(t)}, x_i^{(t)}, z_i^{(t)})\, dz_i^{(t)}. \tag{9}$$

Therefore, the log-likelihood becomes:

$$logL(\Theta, \Gamma, \beta, \sigma^2, \mathbf{B}, \mathbf{\Sigma}) = \sum_{i=1}^{n} \sum_{t=1}^{T} log\left(\int p(y_i^{(t)}, x_i^{(t)}, z_i^{(t)})\, dz_i^{(t)}\right). \tag{10}$$

Hence, the maximum likelihood estimation is given by :

$$\hat{\Theta}, \hat{\Gamma}, \hat{\beta}, \hat{\sigma}^2, \hat{\mathbf{B}}, \hat{\mathbf{\Sigma}} = \arg \max_{\Theta, \Gamma, \beta, \sigma^2, \mathbf{B}, \mathbf{\Sigma}} logL(\Theta, \Gamma, \beta, \sigma^2, \mathbf{B}, \mathbf{\Sigma}). \tag{11}$$

Cluster-Level Optimal Control

To find the optimal control $u_c^{(t)}$ for a cluster c at time t, we minimize the prediction error across all time series i in the cluster:

$$u_c^{(t)} = \arg\min_u \sum_{i,\, c^*(i)=c} p(c_i^{(t)} = c|x_i^{(t)}, y_i^{(t)}) \left\| y_i^{(t)} - \left(\theta_i^\top x_i^{(t)} + \sum_{j \neq i,\, c^*(j)=c} \gamma_{ij}^\top x_j^{(t)} + \beta_c u\right) \right\|_2^2. \tag{12}$$

On line Expectation-Maximization (EM) with Optimal Control Learning Algorithm

We introduce a discrete latent variable $c_i^{(t)}$ representing the cluster assignment of data point i at time t. The complete data consists of observed variables $(x_i^{(t)}, y_i^{(t)})$ and latent variables $(z_i^{(t)}, c_i^{(t)})$.

0.1 Model overview

1. **Cluster Assignment:** Each data point i at time t belongs to a cluster $c_i^{(t)}$:

$$p(c_i^{(t)}) = \pi_{c_i^{(t)}} \tag{13}$$

where π_c is the prior probability of cluster c.
2. **Latent Variable Prior:**

$$p(z_i^{(t)}|c_i^{(t)}) = \mathcal{N}(0, I) \tag{14}$$

3. **Observation Model:**

$$x_i^{(t)}|z_i^{(t)}, c_i^{(t)} \sim \mathcal{N}(B_{c_i} z_i^{(t)}, \Sigma_{c_i}) \tag{15}$$

4. **Output Model:**

$$y_i^{(t)}|x_i^{(t)}, c_i^{(t)} \sim \mathcal{N}\left(\theta_{c_i}^\top x_i^{(t)} + \sum_{j\neq i, c_j^{(t)}=c_i^{(t)}} \gamma_{c_i}^\top x_j^{(t)} + \beta_{c_i} u_{c_i}^{(t)}, \sigma_{c_i}^2\right) \tag{16}$$

0.2 Online EM algorithm with Control Learning

The algorithm goes as follows:

I. Initialize Parameters

Start with an initial values for the model parameters:

$$\Theta^{(0)} = \{\theta_c, \gamma_c, \beta_c, \mathbf{B}_c, \mathbf{\Sigma}_c, \sigma_c^2\}.$$

II. Iterative Steps

For each iteration $t \geq 0$:

E-step : Posterior Distribution

First of all, we compute the **cluster assignment probabilities:**

$$p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) = \frac{\pi_{c_i^{(t)}} p(y_i^{(t)}, x_i^{(t)}|c_i^{(t)})}{\sum_c \pi_c p(y_i^{(t)}, x_i^{(t)}|c)} \tag{17}$$

where

$$p(y_i^{(t)}, x_i^{(t)}|c_i^{(t)}) = \mathcal{N}\left(\theta_{c_i}^\top x_i^{(t)} + \sum_{j\neq i, c^*(j)=c_i} \gamma_{c_i}^\top x_j^{(t)} + \beta_{c_i} u_{c_i}^{(t)}, B_{c_i} B_{c_i}^\top + \Sigma_{c_i} + \sigma_{c_i}^2\right). \tag{18}$$

We compute the expected value of the complete-data log-likelihood given the observed data $y_i^{(t)}, x_i^{(t)}$ and the current parameter estimates $\Theta^{(t)}$:

$$Q(\Theta|\Theta^{(t)}) = \mathbb{E}_{p(z_i^{(t)}|x_i^{(t)},y_i^{(t)})}[\log p(y_i^{(t)}, x_i^{(t)}, z_i^{(t)})].$$

Using Gaussian conditioning formulas, we find that the posterior distribution $p(z_i^{(t)}|x_i^{(t)}, c_i^{(t)})$ remains Gaussian, that is $z_i^{(t)}|x_i^{(t)}, c_i^{(t)} \sim \mathcal{N}(\mu_{z|x}, \Sigma_{z|x})$. Therefore, the E-step consists in computing the posterior mean and posterior covariance, which are given by :

- $\mu_{z|x}^{(i,t)} = B_{c_i}^\top (B_{c_i} B_{c_i}^\top + \Sigma_{c_i})^{-1} x_i^{(t)}$
- $\Sigma_{z|x} = I - B_{c_i}^\top (B_{c_i} B_{c_i}^\top + \Sigma_{c_i})^{-1} B_{c_i}$

M-Step: Parameter Updates

First, we **update the Cluster Priors:**

$$\pi_c^{new} = \frac{1}{nT} \sum_{i,t} p(c_i^{(t)} = c|x_i^{(t)}, y_i^{(t)}). \tag{19}$$

Then, we maximize $Q(\Theta|\Theta^{(t)})$ with respect to the parameters :

1. **Update $B_c^{(t)}$:**

$$B_c^{new} = \left(\sum_{i,t} p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) x_i^{(t)} \mu_{z|x}^{(i,t)\top} \right) \left(\sum_{i,t} p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) (\Sigma_{z|x} + \mu_{z|x}^{(i,t)} \mu_{z|x}^{(i,t)\top}) \right)^{-1} \tag{20}$$

2. **Update $\Sigma_c^{(t)}$:**

$$\Sigma_c^{new} = \frac{1}{\sum_{i,t} p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)})} \sum_{i,t} p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) \Big(x_i^{(t)} x_i^{(t)\top} - B_c^{new} \mu_{z|x}^{(i,t)} x_i^{(t)\top} - x_i^{(t)} \mu_{z|x}^{(i,t)\top} B_c^{new\top} + B_c^{new} (\Sigma_{z|x} + \mu_{z|x}^{(i,t)} \tag{21}$$

3. **Update $\theta_i^{(t)}$:**

$$\theta_{c_i}^{new} = \left(\sum_t p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) x_i^{(t)} x_i^{(t)\top} \right)^{-1} \left(\sum_t p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) x_i^{(t)} y_i^{(t)} \right) \tag{22}$$

4. **Update $\gamma_{ij}^{(t)}$:**

$$\gamma_{ij}^{new} = \left(\sum_t p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) x_j^{(t)} x_j^{(t)\top} \right)^{-1} \left(\sum_t p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) x_j^{(t)} (y_i^{(t)} - \theta_i^\top x_i^{(t)}) \right) \tag{23}$$

5. **Update $\beta_{c^*(i)}^{(t)}$:**

$$\beta_c^{new} = \frac{\sum_{i,c^*(i)=c} \sum_t p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) u_c^{(t)} (y_i^{(t)} - \theta_i^{new} {}^\top x_i^{(t)} - \sum_{j \neq i, c^*(j)=c^*(i)} \gamma_{ij}^{new} {}^\top x_j^{(t)})}{\sum_{i,c^*(i)=c} \sum_t p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) u_c^{(t)} u_c^{(t)}} \tag{24}$$

6. **Update $\sigma^{2,(t)}$:**

$$\sigma^{2,new} = \frac{1}{nT} \sum_{i,t} p(c_i^{(t)}|x_i^{(t)}, y_i^{(t)}) \left(y_i^{(t)} - \theta_i new^\top x_i^{(t)} - \sum_{j \neq i, c^*(j)=c^*(i)} \gamma_{ij}^{new} {}^\top x_j^{(t)} - \beta_{c^*(i)}^{new} u_{c^*(i)}^{(t)} \right)^2 \tag{25}$$

Final-Step: Optimal Control Learning

Once the parameters are updated in the M-step, we use the updated model to learn the optimal control u for each cluster. To find the optimal control $u_c^{(t)}$ for cluster c, we solve the following minimization problem:

$$u_c^{(t)} = \arg\min_u \sum_{i,\, c^*(i)=c} p(c_i^{(t)} = c | x_i^{(t)}, y_i^{(t)}) \left\| y_i^{(t)} - \left(\theta_i^\top x_i^{(t)} + \sum_{j\neq i,\, c^*(j)=c^*(i)} \gamma_{ij}^\top x_j^{(t)} + \beta_c u \right) \right\|_2^2 . \quad (26)$$

This is a least square problem, and solving for $u_c^{(t)}$ explicitly we obtain the optimal $u_c^{(t)}$ given by:

$$u_c^{(t)} = \frac{\sum_{i,c^*(i)=c} p(c_i^{(t)} | x_i^{(t)}, y_i^{(t)}) \left(y_i^{(t)} - \theta_i^{new\top} x_i^{(t)} - \sum_{j\neq i, c^*(j)=c} \gamma_{ij}^{new\top} x_j^{(t)} \right)}{\sum_{i,c^*(i)=c} p(c_i^{(t)} | x_i^{(t)}, y_i^{(t)}) \beta_c^{new\top} \beta_c^{new}} . \quad (27)$$

After computing $u_c^{(t)}$ for all clusters, we apply these updated control values in the next iteration of the system. The process is repeated at each time step t to continuously adapt the control based on new observations.