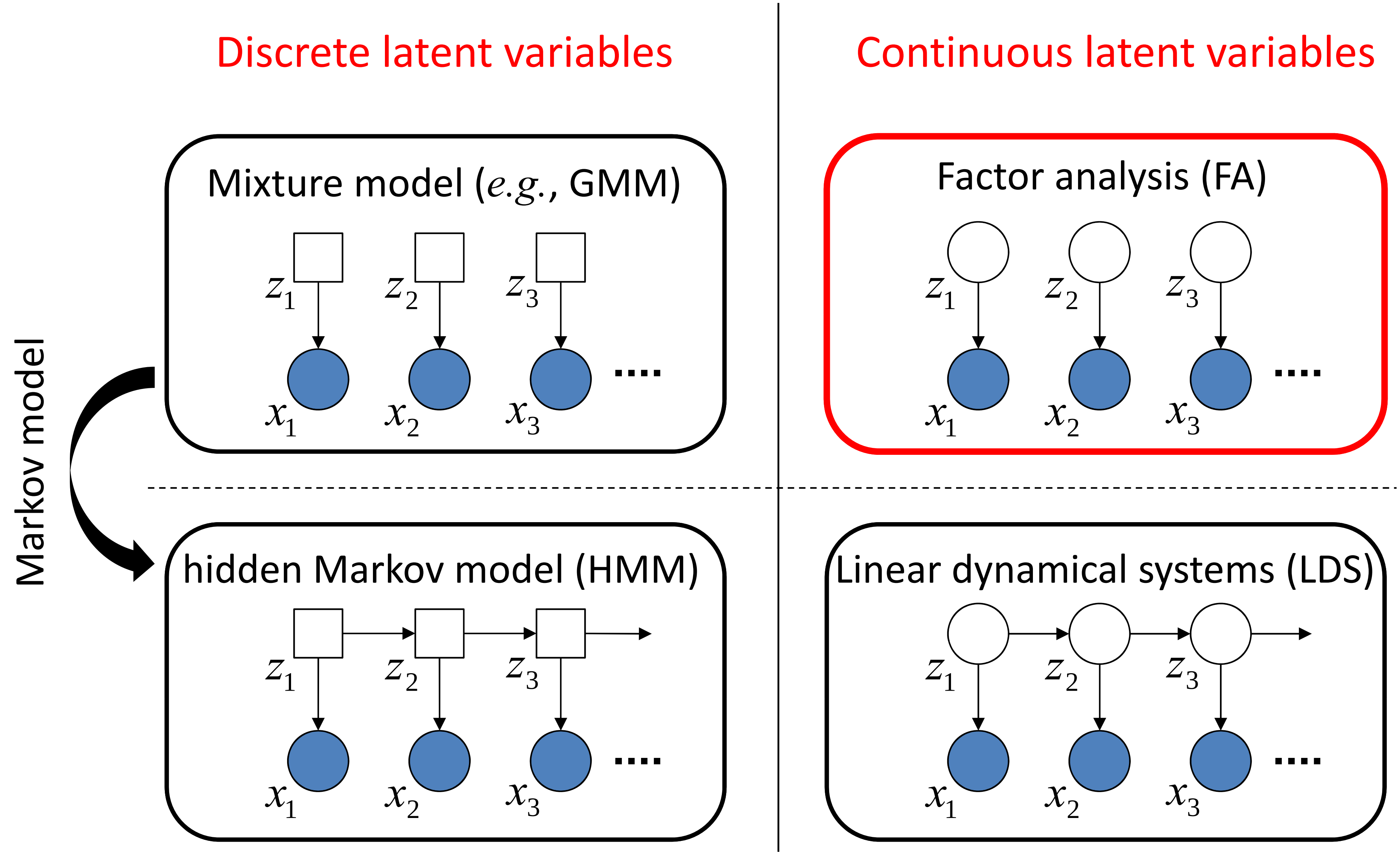
Sequential Data Modeling

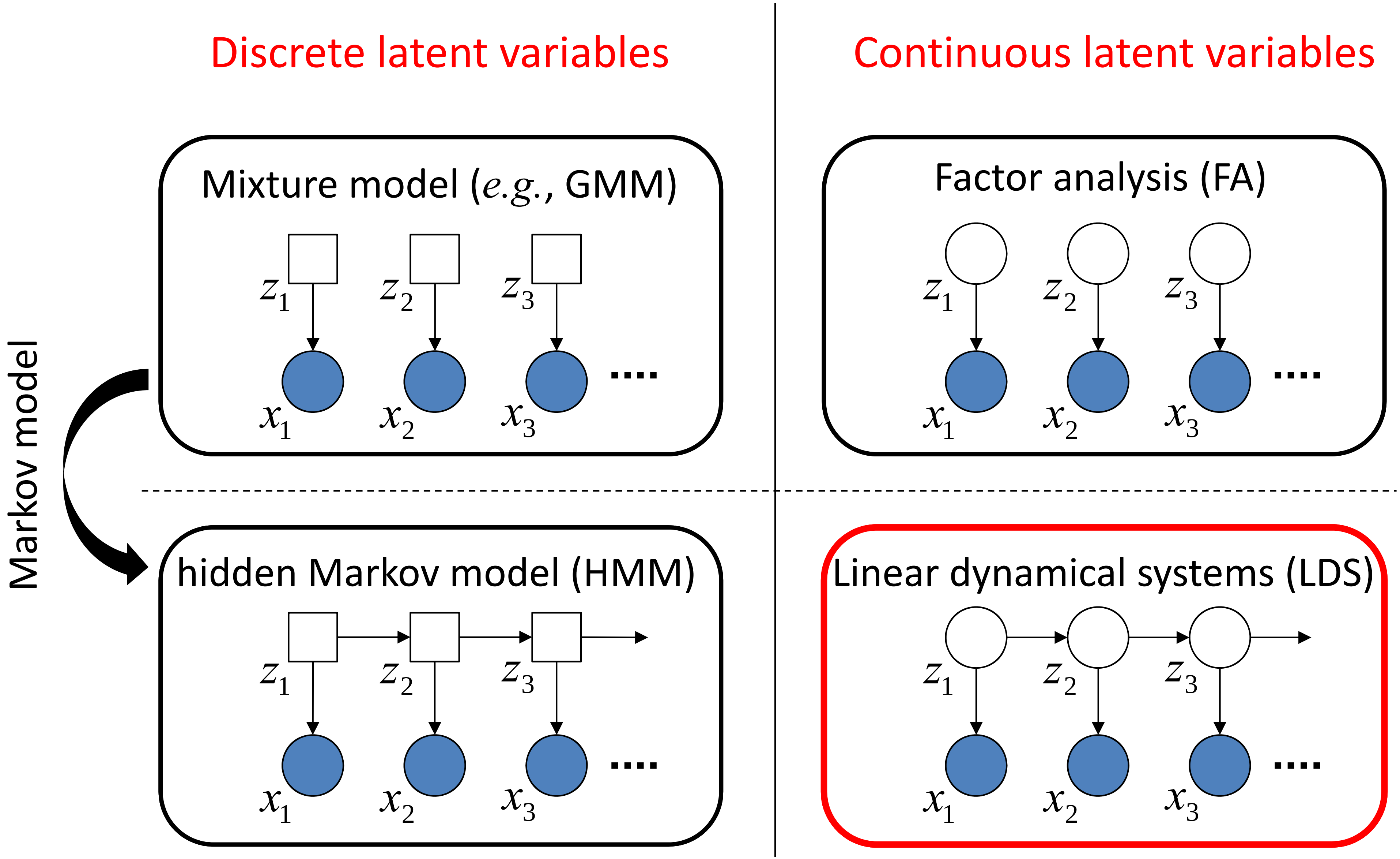
“ Linear Dynamical Systems”

Basic Techniques



Linear Dynamical Systems (LDS)

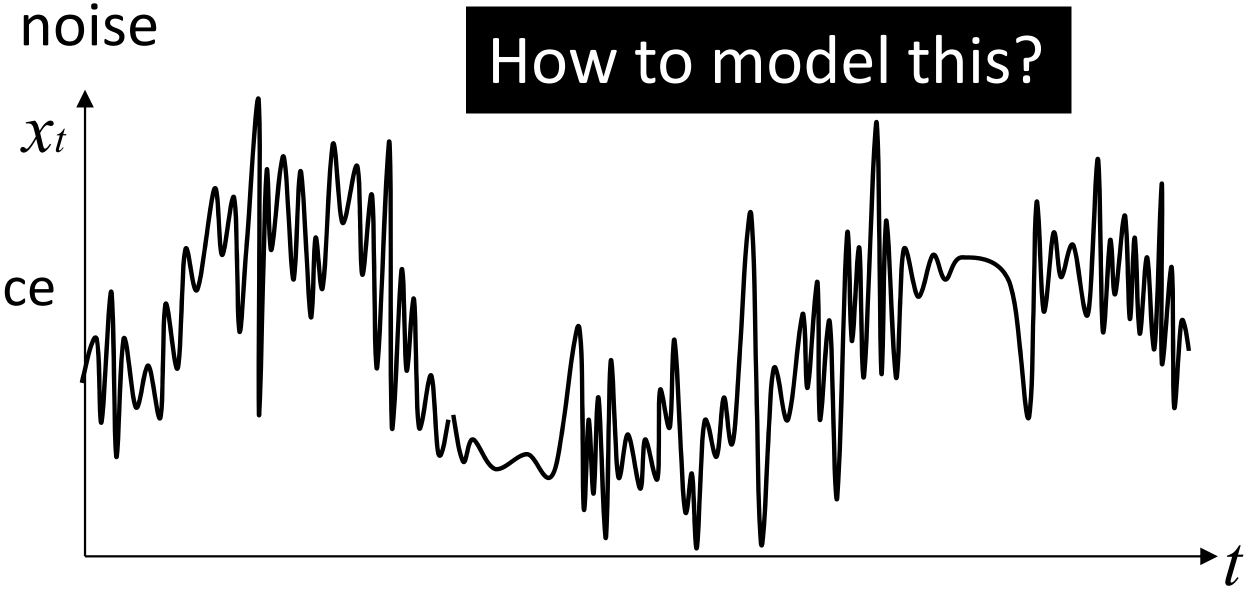
Basic Techniques



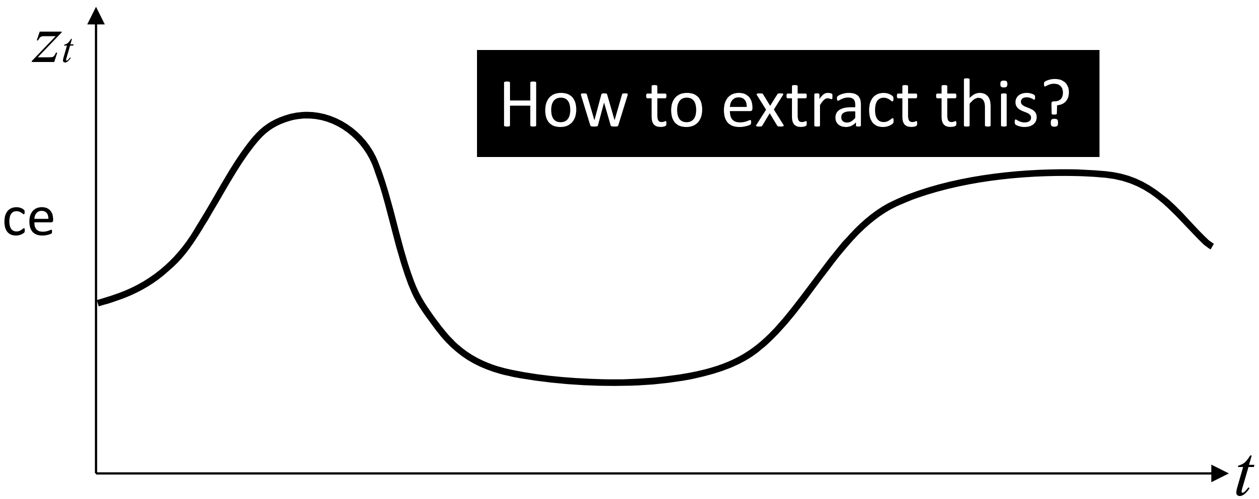
Assume Unobservable Data Sequence

- Extraction of an underlying data sequence from observable data sequence suffering from noise

Observation data sequence
w/ noise

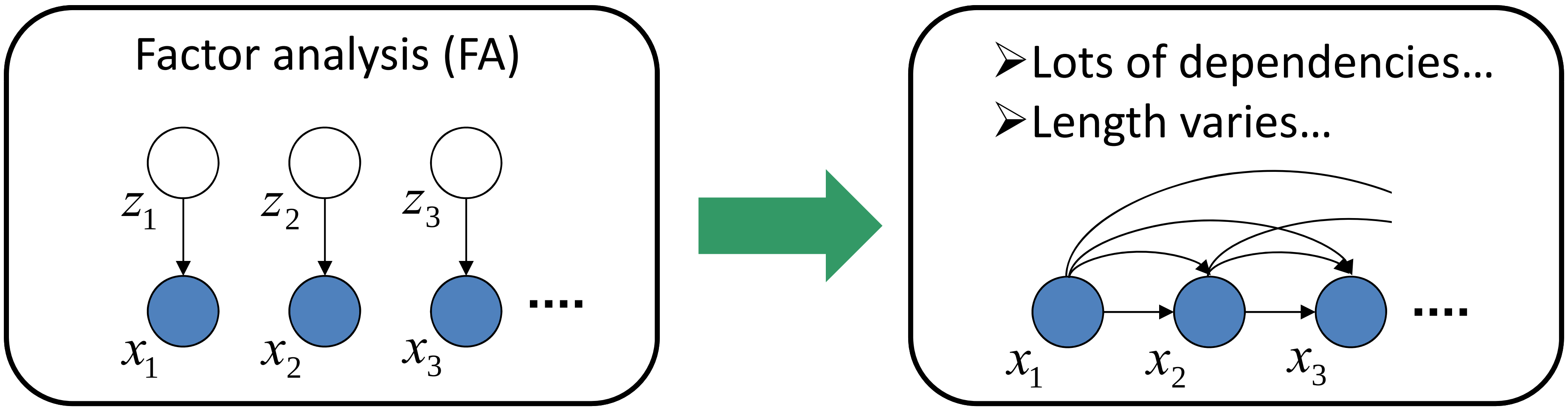


Underlying data sequence
(**unobservable**)



How to Model Sequential Data?

- To model sequential data…
  - Need to consider sample order
  - Need to model a very high-dimensional space of joint data over a sequence
  - Need to deal with various lengths of sequential data



$$p(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T \mid \boldsymbol{\lambda}) = \prod_{t=1}^{T} p(\boldsymbol{x}_t \mid \boldsymbol{\lambda})$$

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T \mid \boldsymbol{\lambda}) = \textbf{?}$$

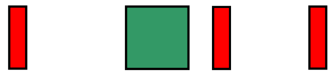
How to model this $p.d.f.$?

Linear Dynamical Systems

- **Markov process** to model a sequence of continuous latent variables
- **Linear equation** to model state transition and mapping from a state space into an observation space

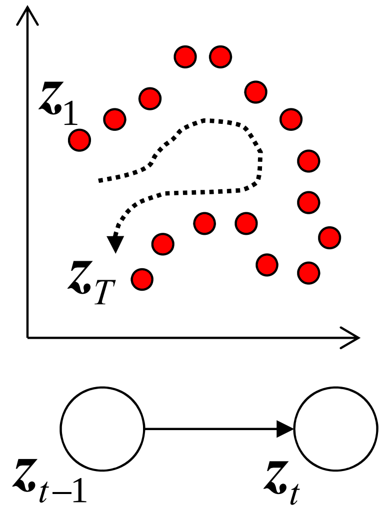
State space

$$\boldsymbol{z}_t = \boldsymbol{A}\boldsymbol{z}_{t-1} + \boldsymbol{n}_t$$



Transition noise:

$$\boldsymbol{n}_t \sim \mathcal{N}\left(\boldsymbol{n}_t; \boldsymbol{0}, \boldsymbol{\Gamma}\right)$$

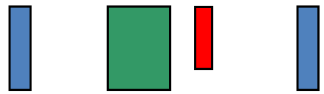


State transition *p.d.f.*

$$p\left(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}\right) = \mathcal{N}\left(\boldsymbol{z}_t; \boldsymbol{A}\boldsymbol{z}_{t-1}, \boldsymbol{\Gamma}\right)$$

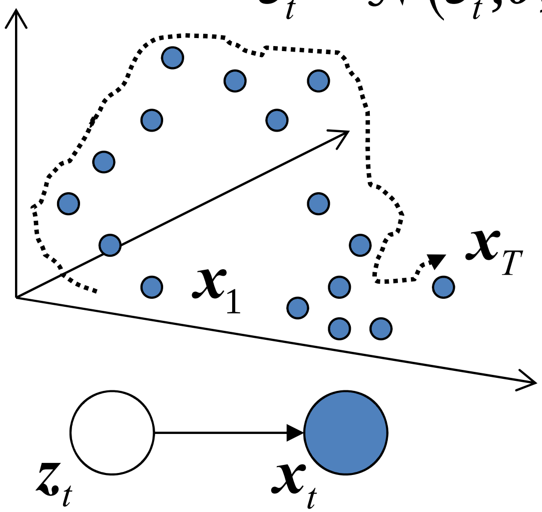
Observation space

$$\boldsymbol{x}_t = \boldsymbol{W}\boldsymbol{z}_t + \boldsymbol{e}_t$$



Observation noise:

$$\boldsymbol{e}_t \sim \mathcal{N}\left(\boldsymbol{e}_t; \boldsymbol{0}, \boldsymbol{\Sigma}\right)$$



Emission *p.d.f.*

$$p\left(\boldsymbol{x}_t \mid \boldsymbol{z}_t\right) = \mathcal{N}\left(\boldsymbol{x}_t; \boldsymbol{W}\boldsymbol{z}_t, \boldsymbol{\Sigma}\right)$$

$p.d.f.$s in Linear Dynamical Systems

- Sequence of observation data : $\boldsymbol{x}_{1:T} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T\}$
- Sequence of latent variables : $\boldsymbol{z}_{1:T} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_T\}$
- $p.d.f.$ of observation data :

$$
\begin{aligned}
p(\boldsymbol{x}_{1:T}) &= \int p(\boldsymbol{x}_{1:T} \mid \boldsymbol{z}_{1:T})\, p(\boldsymbol{z}_{1:T})\, \mathrm{d}\boldsymbol{z}_{1:T} \\
&= \int \left[\prod_{t=1}^{T} p(\boldsymbol{x}_t \mid \boldsymbol{z}_t) \right] \left[p(\boldsymbol{z}_1) \prod_{t=2}^{T} p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}) \right] \mathrm{d}\boldsymbol{z}_{1:T}
\end{aligned}
$$

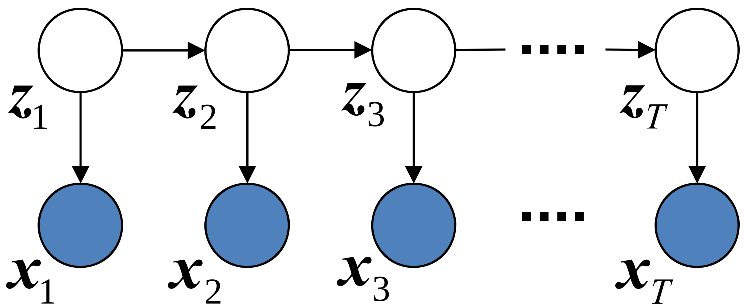
Marginalization over a sequence of latent variables

Emission $p.d.f.$s:

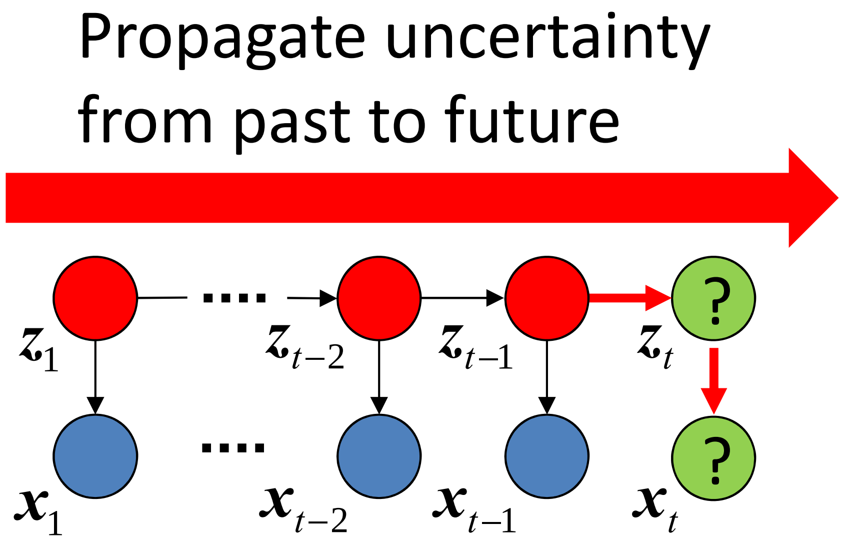
$$p(\boldsymbol{x}_t \mid \boldsymbol{z}_t) = \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{W}\boldsymbol{z}_t, \boldsymbol{\Sigma})$$

Transition $p.d.f.$s:

$$
\begin{cases}
p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}) = \mathcal{N}(\boldsymbol{z}_t; \boldsymbol{A}\boldsymbol{z}_{t-1}, \boldsymbol{\Gamma}) & t \geq 2 \\
p(\boldsymbol{z}_1) = \mathcal{N}(\boldsymbol{z}_1; \boldsymbol{\mu}_0, \boldsymbol{P}_0) & t = 1
\end{cases}
$$



Kalman Filtering



How to Recursively Calculate Likelihood?

- Likelihood function for the observation data sequence:

See appendix

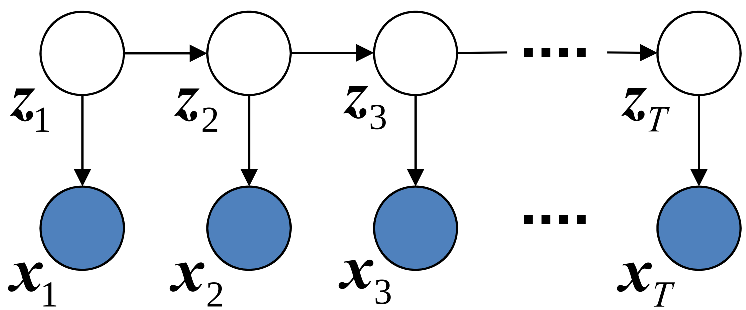
$$p(\boldsymbol{x}) = \int p(\boldsymbol{x} \mid \boldsymbol{z}) p(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} = \mathcal{N}\left(\boldsymbol{x}; \tilde{\boldsymbol{W}} \tilde{\boldsymbol{A}}^{-1} \tilde{\boldsymbol{z}}_0, \tilde{\boldsymbol{W}} \tilde{\boldsymbol{A}}^{-1} \tilde{\boldsymbol{\Gamma}} \tilde{\boldsymbol{A}}^{-\top} \tilde{\boldsymbol{W}}^{\top} + \tilde{\boldsymbol{\Sigma}}\right)$$

Batch-type calculation (w/ all data over an sequence) is assumed but **frame-by-frame calculation** will be required in some applications, such as real-time signal processing…

- Original form of the likelihood function:

$$p(\boldsymbol{x}_{1:T}) = \int p(\boldsymbol{x}_{1:T} \mid \boldsymbol{z}_{1:T}) p(\boldsymbol{z}_{1:T}) \mathrm{d}\boldsymbol{z}_{1:T}$$

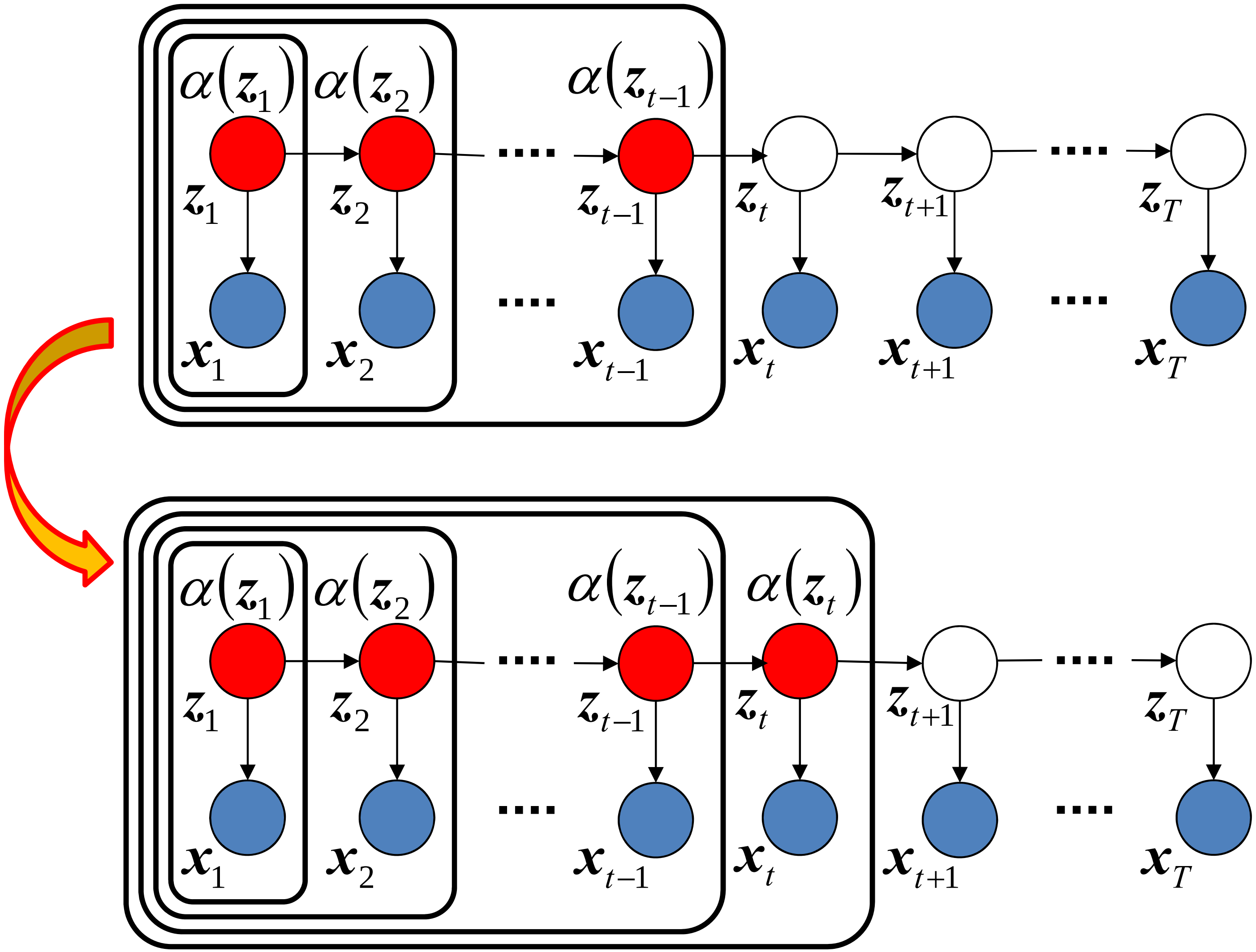
$$= \int \left[\prod_{t=1}^{T} p(\boldsymbol{x}_t \mid \boldsymbol{z}_t) \right] \left[p(\boldsymbol{z}_1) \prod_{t=2}^{T} p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}) \right] \mathrm{d}\boldsymbol{z}_{1:T}$$



Kalman Filtering (Forward Algorithm)

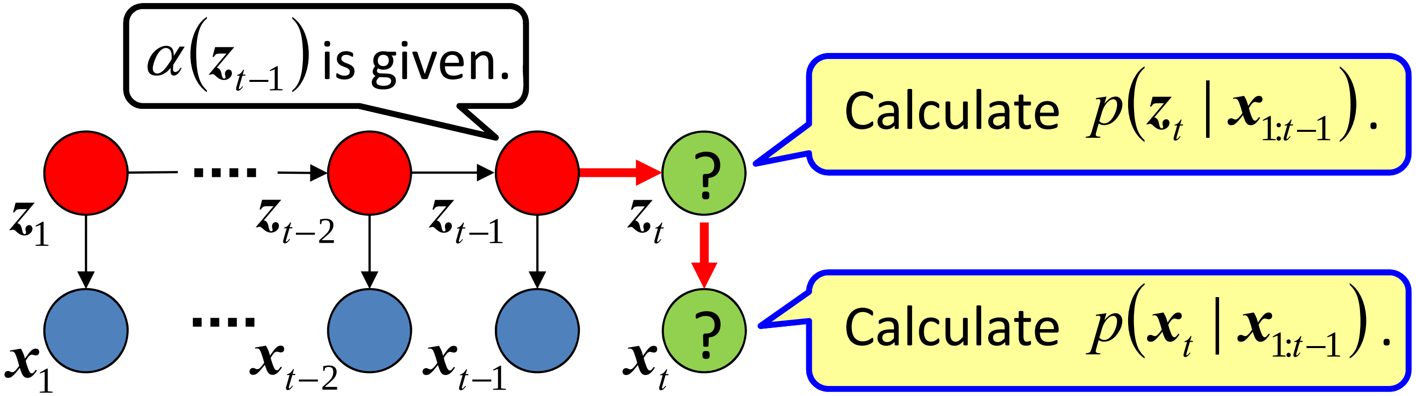
- Posterior $p.d.f.$s given all past observation data ($i.e.$, $p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:t}) = \alpha(\boldsymbol{z}_t)$) determined in Kalman filtering

Kalman filtering

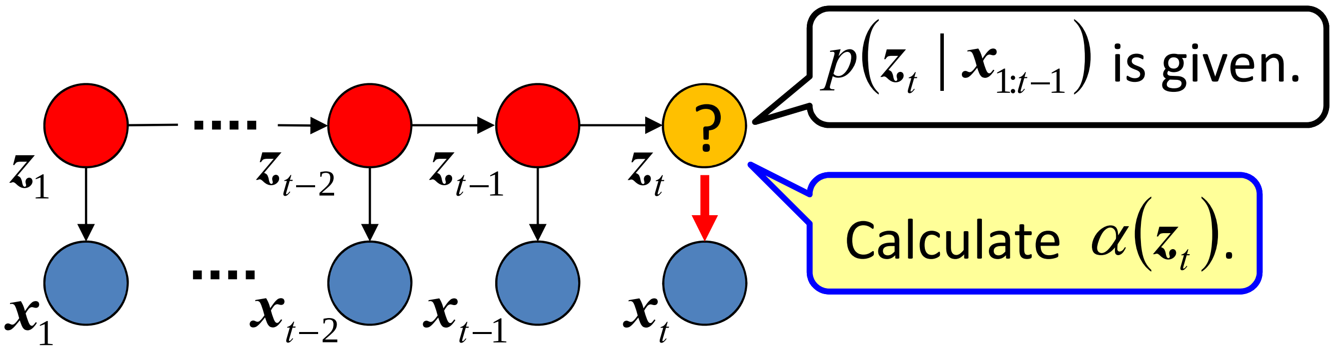


Prediction and Update

- Prediction step
  - Predict distribution of latent variables at frame t from all past observation data



- Update step
  - Update distribution of latent variables at frame t using current observation data as well as all past observation data



Predicted and Updated *p.d.f.*s

See appendix

- Predicted *p.d.f.*

$$p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:t-1}) = \int p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}) p(\boldsymbol{z}_{t-1} \mid \boldsymbol{x}_{1:t-1}) \mathrm{d}\boldsymbol{z}_t = \mathcal{N}(\boldsymbol{z}_t ; \boldsymbol{\mu}_{t|t-1}, \boldsymbol{P}_{t|t-1})$$

Predicted mean : $\boldsymbol{\mu}_{t|t-1} = \boldsymbol{A}\boldsymbol{\mu}_{t-1}$

Predicted covariance : $\boldsymbol{P}_{t|t-1} = \boldsymbol{A}\boldsymbol{P}_{t-1}\boldsymbol{A}^{\mathsf{T}} + \boldsymbol{\Gamma}$

- Updated *p.d.f.*

Posterior ∝ Likelihood x Prior

$$\alpha(\boldsymbol{z}_t) = p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:t}) = \mathcal{N}(\boldsymbol{z}_t ; \boldsymbol{\mu}_t, \boldsymbol{P}_t) \propto p(\boldsymbol{x}_t \mid \boldsymbol{z}_t) p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:t-1})$$

Kalman gain matrix : $\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\boldsymbol{W}^{\mathsf{T}}\left(\boldsymbol{W}\boldsymbol{P}_{t|t-1}\boldsymbol{W}^{\mathsf{T}} + \boldsymbol{\Sigma}\right)^{-1}$

Updated mean : $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \boldsymbol{K}_t\left(\underline{\boldsymbol{x}_t - \boldsymbol{W}\boldsymbol{\mu}_{t|t-1}}\right)$

Updated covariance : $\boldsymbol{P}_t = (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{W})\boldsymbol{P}_{t|t-1}$

Error between predicted and observed data

Likelihood Calculation

See appendix

- Conditional $p.d.f.$ of observation data

$$p(\boldsymbol{x}_t \mid \boldsymbol{x}_{1:t-1}) = \int p(\boldsymbol{x}_t \mid \boldsymbol{z}_t)\, p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:t-1})\, \mathrm{d}\boldsymbol{z}_t = \mathcal{N}\left(\boldsymbol{x}_t ; \boldsymbol{W}\boldsymbol{\mu}_{t|t-1}, \boldsymbol{W}\boldsymbol{P}_{t|t-1}\boldsymbol{W}^{\top} + \boldsymbol{\Sigma}\right)$$

Emission $p.d.f.$: $p(\boldsymbol{x}_t \mid \boldsymbol{z}_t) = \mathcal{N}(\boldsymbol{x}_t ; \boldsymbol{W}\boldsymbol{z}_t, \boldsymbol{\Sigma})$

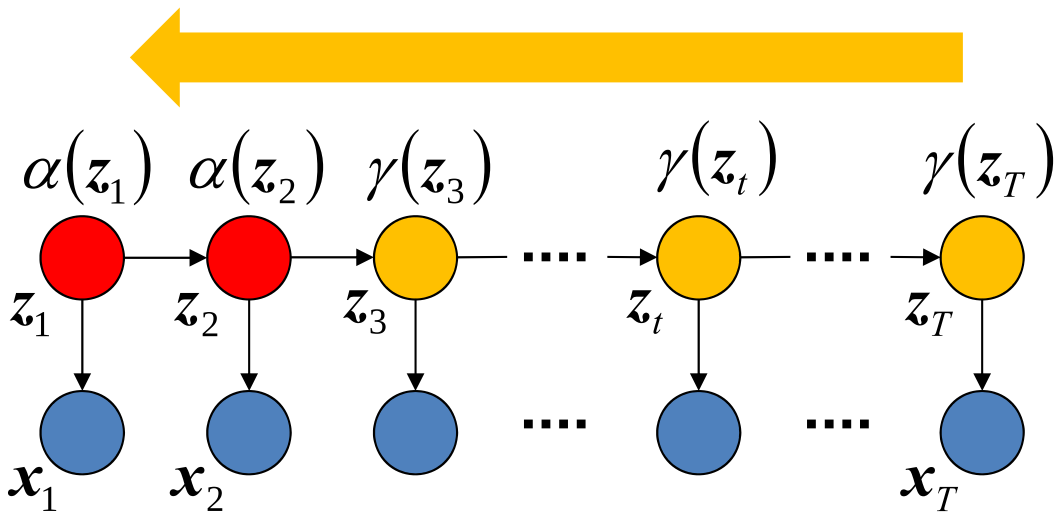
Predicted $p.d.f.$: $p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:t-1}) = \mathcal{N}(\boldsymbol{z}_t ; \boldsymbol{\mu}_{t|t-1}, \boldsymbol{P}_{t|t-1})$

- Recursive likelihood calculation

$$p(\boldsymbol{x}_{1:T}) = \underbrace{\underbrace{\underbrace{p(\boldsymbol{x}_1)\, p(\boldsymbol{x}_2 \mid \boldsymbol{x}_1)}_{p(\boldsymbol{x}_1, \boldsymbol{x}_2)}\, p(\boldsymbol{x}_3 \mid \boldsymbol{x}_1, \boldsymbol{x}_2)}_{p(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)} \cdots\, p(\boldsymbol{x}_{T-1} \mid \boldsymbol{x}_{1:T-2})}_{p(\boldsymbol{x}_{1:T-1})}\, p(\boldsymbol{x}_T \mid \boldsymbol{x}_{1:T-1})$$

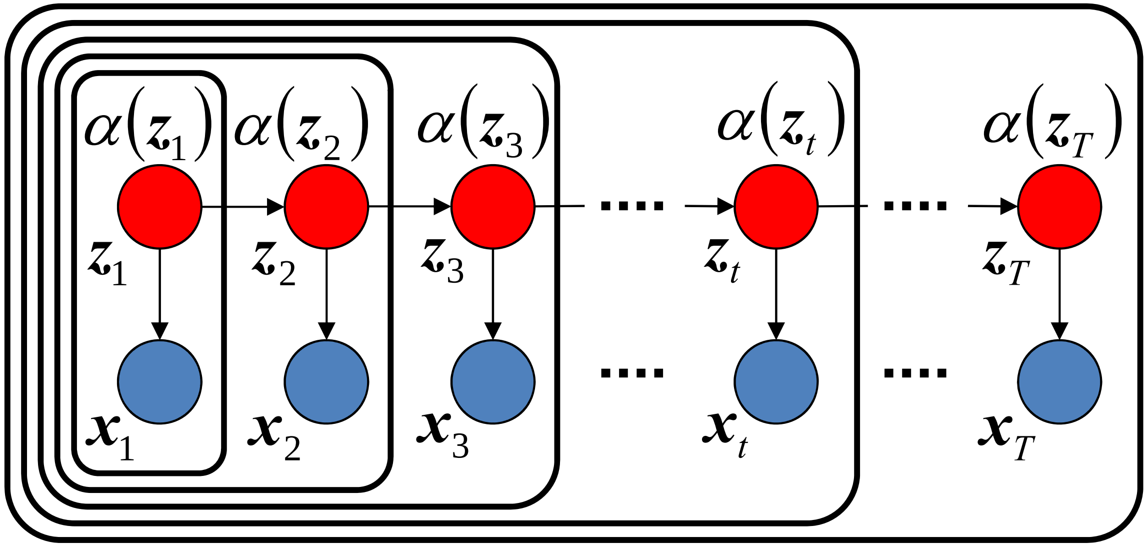
Kalman Smoothing

Improve inference considering all data

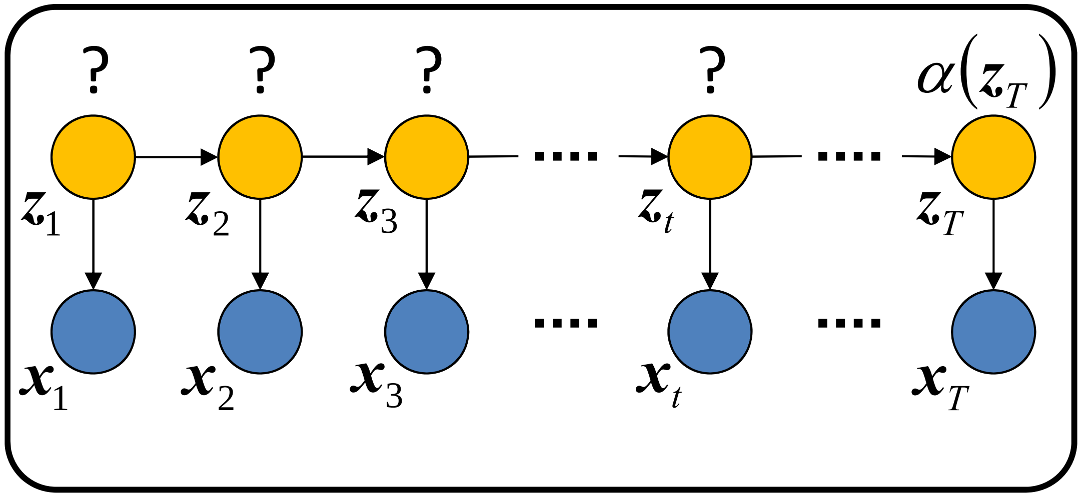


How to Calculate Posterior w/ All Data?

- Posterior *p.d.f.*s given all past observation data determined in Kalman filtering

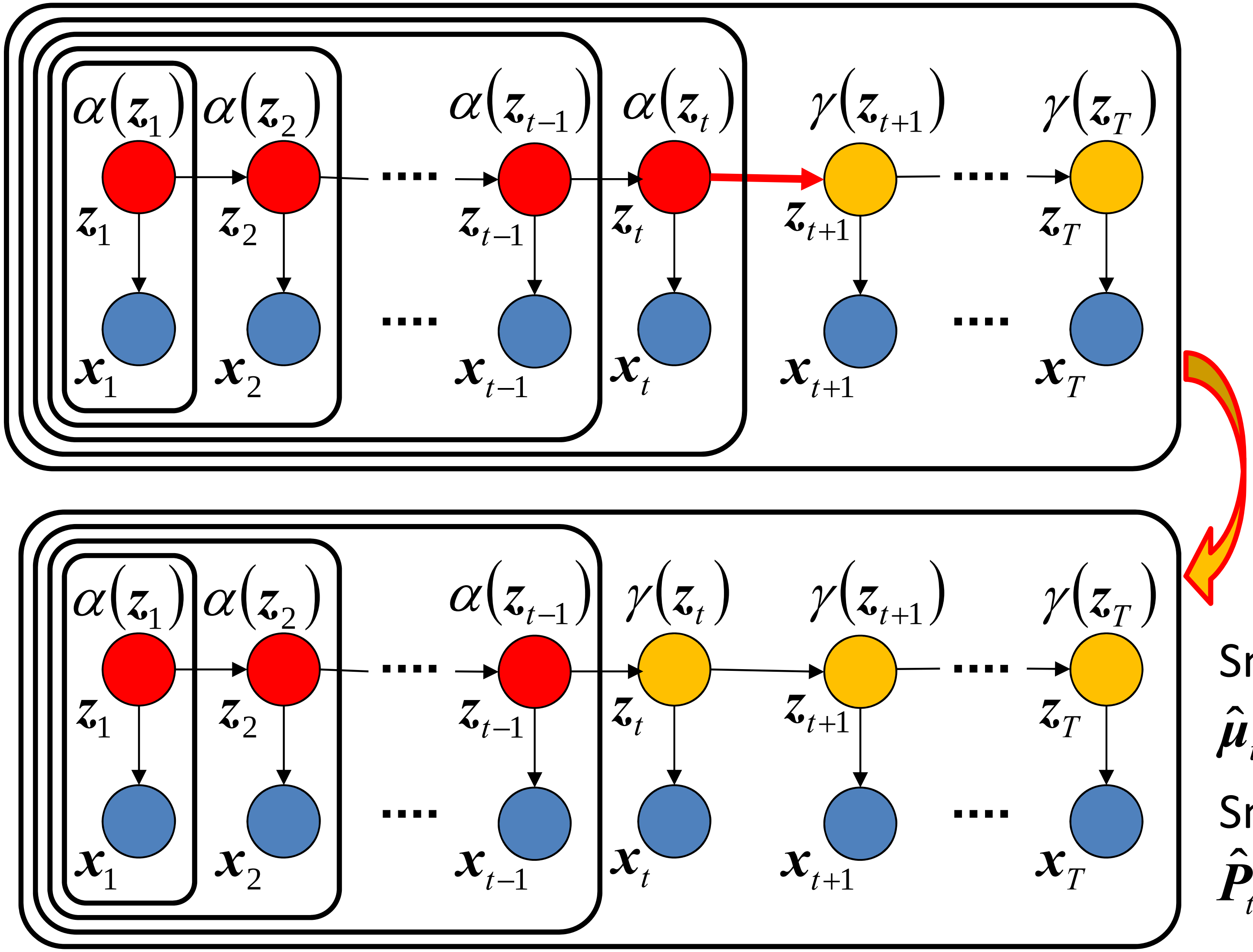


- How to calculate posterior *p.d.f.*s calculated w/ all observation data?



Kalman Smoothing (Backward Algorithm)

- Calculate smoothed *p.d.f.* at frame t using both smoothed *p.d.f.* at frame t+1 and updated *p.d.f.* at frame t determined in Kalman filtering



See appendix

Kalman smoothing

$$\gamma(\boldsymbol{z}_t) = \mathcal{N}\left(\boldsymbol{z}_t ; \hat{\boldsymbol{\mu}}_t , \hat{\boldsymbol{P}}_t\right)$$

$$\boldsymbol{J}_t = \boldsymbol{P}_t \boldsymbol{A}^{\mathsf{T}} \boldsymbol{P}_{t+1|t}^{-1}$$

Smoothed mean:

$$\hat{\boldsymbol{\mu}}_t = \boldsymbol{\mu}_t + \boldsymbol{J}_t\left(\hat{\boldsymbol{\mu}}_{t+1} - \boldsymbol{\mu}_{t+1|t}\right)$$

Smoothed covariance :

$$\hat{\boldsymbol{P}}_t = \boldsymbol{P}_t + \boldsymbol{J}_t\left(\hat{\boldsymbol{P}}_{t+1} - \boldsymbol{P}_{t+1|t}\right)\boldsymbol{J}_t^{\mathsf{T}}$$

Model Training

EM Algorithm

- Likelihood function

$$p(\boldsymbol{x}_{1:T} \mid \boldsymbol{\lambda}) = \int p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T} \mid \boldsymbol{\lambda})\, \mathrm{d}\boldsymbol{z}_{1:T}$$

$$= \int \left[\prod_{t=1}^{T} p(\boldsymbol{x}_t \mid \boldsymbol{z}_t, \{\boldsymbol{W}, \boldsymbol{\Sigma}\}) \right] \left[p(\boldsymbol{z}_1 \mid \{\boldsymbol{\mu}_0, \boldsymbol{P}_0\}) \prod_{t=2}^{T} p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \{\boldsymbol{A}, \boldsymbol{\Gamma}\}) \right] \mathrm{d}\boldsymbol{z}_{1:T}$$

- Iterative maximization of **lower bound**

$$\ln p(\boldsymbol{x}_{1:T} \mid \boldsymbol{\lambda}) = \ln \int p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T} \mid \boldsymbol{\lambda})\, \mathrm{d}\boldsymbol{z}_{1:T}$$

$$\geq \int q(\boldsymbol{z}_{1:T}) \ln \frac{p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T} \mid \boldsymbol{\lambda})}{q(\boldsymbol{z}_{1:T})}\, \mathrm{d}\boldsymbol{z}_{1:T} = \mathcal{L}(q, \boldsymbol{\lambda})$$

E-step: Set q to the posterior $p.d.f.$ calculated w/ current model parameters

$$\hat{q}(\boldsymbol{z}_{1:T}) = p(\boldsymbol{z}_{1:T} \mid \boldsymbol{x}_{1:T}, \boldsymbol{\lambda}_{\mathrm{old}})$$

M-step: Maximize auxiliary function with respect to model parameters

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}} \int \hat{q}(\boldsymbol{z}_{1:T}) \ln\{p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T} \mid \boldsymbol{\lambda})\}\, \mathrm{d}\boldsymbol{z}_{1:T}$$

E-Step: Update q

- Calculation of posterior $p.d.f.$s using a model parameter set $\boldsymbol{\lambda}_{\text{old}}$
  - Posterior $p.d.f.$ of $\boldsymbol{z}_t$ calculated in Kalman smoothing

$$\hat{q}(\boldsymbol{z}_t) = p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:T}, \boldsymbol{\lambda}_{\text{old}}) = \gamma(\boldsymbol{z}_t) = \mathcal{N}(\boldsymbol{z}_t ; \hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{P}}_t)$$

  - Joint posterior $p.d.f.$ of $\boldsymbol{z}_{t-1}$ and $\boldsymbol{z}_t$ also calculated in Kalman smoothing

$$\hat{q}(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t) = p(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t \mid \boldsymbol{x}_{1:T}, \boldsymbol{\lambda}_{\text{old}})$$
$$= \mathcal{N}\left(\begin{bmatrix} \boldsymbol{z}_{t-1} \\ \boldsymbol{z}_t \end{bmatrix} ; \begin{bmatrix} \hat{\boldsymbol{\mu}}_{t-1} \\ \hat{\boldsymbol{\mu}}_t \end{bmatrix}, \begin{bmatrix} \hat{\boldsymbol{P}}_{t-1} & \boldsymbol{J}_{t-1} \hat{\boldsymbol{P}}_t \\ \hat{\boldsymbol{P}}_t \boldsymbol{J}_{t-1}^{\mathsf{T}} & \hat{\boldsymbol{P}}_t \end{bmatrix} \right)$$

Note that $\hat{q}(\boldsymbol{z}_t) = \int \hat{q}(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t) \, \mathrm{d}\boldsymbol{z}_{t-1}$

M-Step: Update $\boldsymbol{\lambda}$

- Maximization of the following auxiliary function

$$Q(\boldsymbol{\lambda}_{\text{old}}, \boldsymbol{\lambda}) = \int \hat{q}(\boldsymbol{z}_{1:T}) \left\{ \sum_{t=1}^{T} \underbrace{\ln p(\boldsymbol{x}_t \mid \boldsymbol{z}_t, \boldsymbol{\lambda})}_{①} + \underbrace{\ln p(\boldsymbol{z}_1 \mid \boldsymbol{\lambda})}_{②} + \sum_{t=2}^{T} \underbrace{\ln p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \boldsymbol{\lambda})}_{③} \right\} \mathrm{d}\boldsymbol{z}_{1:T}$$

$$= \int \hat{q}(\boldsymbol{z}_{1:T}) \left\{ \sum_{t=1}^{T} \underbrace{\frac{1}{2} \ln \left| \boldsymbol{\Sigma}^{-1} \right| - \frac{1}{2} (\boldsymbol{x}_t - \boldsymbol{W}\boldsymbol{z}_t)^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_t - \boldsymbol{W}\boldsymbol{z}_t)}_{①} \right.$$

$$+ \underbrace{\frac{1}{2} \ln \left| \boldsymbol{P}_0^{-1} \right| - \frac{1}{2} (\boldsymbol{z}_1 - \boldsymbol{\mu}_0)^{\mathsf{T}} \boldsymbol{P}_0^{-1} (\boldsymbol{z}_1 - \boldsymbol{\mu}_0)}_{②}$$

$$\left. + \sum_{t=2}^{T} \underbrace{\frac{1}{2} \ln \left| \boldsymbol{\Gamma}^{-1} \right| - \frac{1}{2} (\boldsymbol{z}_t - \boldsymbol{A}\boldsymbol{z}_{t-1})^{\mathsf{T}} \boldsymbol{\Gamma}^{-1} (\boldsymbol{z}_t - \boldsymbol{A}\boldsymbol{z}_{t-1})}_{③} \right\} \mathrm{d}\boldsymbol{z}_{1:T}$$

Expansion of Auxiliary Function

$$
\begin{aligned}
\mathrm{Q}(\lambda_{\text{old}},\lambda) = \frac{1}{2}\Big\{ & T\ln\left|\boldsymbol{\Sigma}^{-1}\right| - \mathrm{tr}\Big[\boldsymbol{\Sigma}^{-1}\left\langle \boldsymbol{x}_t\boldsymbol{x}_t^{\mathsf{T}}\right\rangle_{1:T} + \boldsymbol{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}\left\langle \boldsymbol{z}_t\boldsymbol{z}_t^{\mathsf{T}}\right\rangle_{1:T} \\
& \qquad - \boldsymbol{\Sigma}^{-1}\boldsymbol{W}\left\langle\left\langle \boldsymbol{z}_t\right\rangle \boldsymbol{x}_t^{\mathsf{T}}\right\rangle_{1:T} - \left\langle\left\langle \boldsymbol{z}_t\right\rangle \boldsymbol{x}_t^{\mathsf{T}}\right\rangle_{1:T}^{\mathsf{T}}\boldsymbol{W}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\Big] \\
& + \ln\left|\boldsymbol{P}_0^{-1}\right| - \mathrm{tr}\left[\boldsymbol{P}_0^{-1}\left\langle \boldsymbol{z}_1\boldsymbol{z}_1^{\mathsf{T}}\right\rangle + \boldsymbol{P}_0^{-1}\boldsymbol{\mu}_0\boldsymbol{\mu}_0^{\mathsf{T}}\right] + \left\langle \boldsymbol{z}_1\right\rangle^{\mathsf{T}}\boldsymbol{P}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0^{\mathsf{T}}\boldsymbol{P}_0^{-1}\left\langle \boldsymbol{z}_1\right\rangle \\
& + (T-1)\ln\left|\boldsymbol{\Gamma}^{-1}\right| - \mathrm{tr}\Big[\boldsymbol{\Gamma}^{-1}\left\langle \boldsymbol{z}_t\boldsymbol{z}_t^{\mathsf{T}}\right\rangle_{2:T} + \boldsymbol{A}^{\mathsf{T}}\boldsymbol{\Gamma}^{-1}\boldsymbol{A}\left\langle \boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}^{\mathsf{T}}\right\rangle_{2:T} \\
& \qquad - \boldsymbol{\Gamma}^{-1}\boldsymbol{A}\left\langle \boldsymbol{z}_{t-1}\boldsymbol{z}_t^{\mathsf{T}}\right\rangle_{2:T} - \boldsymbol{A}^{\mathsf{T}}\boldsymbol{\Gamma}^{-1}\left\langle \boldsymbol{z}_{t-1}\boldsymbol{z}_t^{\mathsf{T}}\right\rangle_{2:T}^{\mathsf{T}}\Big]\Big\}
\end{aligned}
$$

Expectation:

$$\left\langle \boldsymbol{z}_t\right\rangle = \hat{\boldsymbol{\mu}}_t$$

$$\left\langle \boldsymbol{z}_t\boldsymbol{z}_t^{\mathsf{T}}\right\rangle = \hat{\boldsymbol{P}}_t + \hat{\boldsymbol{\mu}}_t\hat{\boldsymbol{\mu}}_t^{\mathsf{T}}$$

$$\left\langle \boldsymbol{z}_{t-1}\boldsymbol{z}_t^{\mathsf{T}}\right\rangle = \boldsymbol{J}_{t-1}\hat{\boldsymbol{P}}_t + \hat{\boldsymbol{\mu}}_{t-1}\hat{\boldsymbol{\mu}}_t^{\mathsf{T}}$$

Sufficient statistics:

$$\left\langle \boldsymbol{z}_t\boldsymbol{z}_t^{\mathsf{T}}\right\rangle_{1:T} = \sum_{t=1}^{T}\left\langle \boldsymbol{z}_t\boldsymbol{z}_t^{\mathsf{T}}\right\rangle \qquad \left\langle\left\langle \boldsymbol{z}_t\right\rangle \boldsymbol{x}_t^{\mathsf{T}}\right\rangle_{1:T} = \sum_{t=1}^{T}\left\langle \boldsymbol{z}_t\right\rangle \boldsymbol{x}_t^{\mathsf{T}}$$

$$\left\langle \boldsymbol{z}_{t-1}\boldsymbol{z}_t^{\mathsf{T}}\right\rangle_{2:T} = \sum_{t=2}^{T}\left\langle \boldsymbol{z}_{t-1}\boldsymbol{z}_t^{\mathsf{T}}\right\rangle \qquad \left\langle \boldsymbol{x}_t\boldsymbol{x}_t^{\mathsf{T}}\right\rangle_{1:T} = \sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t^{\mathsf{T}}$$

ML Estimates of Model Parameters

Initial parameters of transition *p.d.f.*

$$\hat{\boldsymbol{\mu}}_0 = \langle \boldsymbol{z}_1 \rangle$$

$$\hat{\boldsymbol{P}}_0 = \langle \boldsymbol{z}_1 \boldsymbol{z}_1^\top \rangle - \hat{\boldsymbol{\mu}}_0 \hat{\boldsymbol{\mu}}_0^\top$$

Parameters of transition *p.d.f.*

$$\hat{\boldsymbol{A}} = \langle \boldsymbol{z}_t \boldsymbol{z}_{t-1}^\top \rangle_{2:T} \langle \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}^\top \rangle_{2:T}^{-1}$$

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{T-1}\left(\langle \boldsymbol{z}_t \boldsymbol{z}_t^\top \rangle_{2:T} + \hat{\boldsymbol{A}} \langle \boldsymbol{z}_{t-1} \boldsymbol{z}_{t-1}^\top \rangle_{2:T} \hat{\boldsymbol{A}}^\top - \hat{\boldsymbol{A}} \langle \boldsymbol{z}_{t-1} \boldsymbol{z}_t^\top \rangle_{2:T} - \langle \boldsymbol{z}_t \boldsymbol{z}_{t-1}^\top \rangle_{2:T} \hat{\boldsymbol{A}}^\top \right)$$

Parameters of emission *p.d.f.*

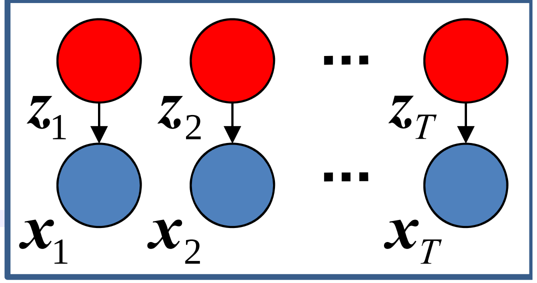
$$\hat{\boldsymbol{W}} = \left\langle \boldsymbol{x}_t \langle \boldsymbol{z}_t \rangle^\top \right\rangle_{1:T} \langle \boldsymbol{z}_t \boldsymbol{z}_t^\top \rangle_{1:T}^{-1}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T}\left(\langle \boldsymbol{x}_t \boldsymbol{x}_t^\top \rangle_{1:T} + \hat{\boldsymbol{W}} \langle \boldsymbol{z}_t \boldsymbol{z}_t^\top \rangle_{1:T} \hat{\boldsymbol{W}}^\top - \hat{\boldsymbol{W}} \left\langle \langle \boldsymbol{z}_t \rangle \boldsymbol{x}_t^\top \right\rangle_{1:T} - \left\langle \boldsymbol{x}_t \langle \boldsymbol{z}_t \rangle^\top \right\rangle_{1:T} \hat{\boldsymbol{W}}^\top \right)$$

Appendix

Derivation of *p.d.f.* of observation data

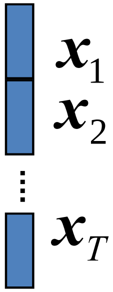
Emission $p.d.f.$ $p(\boldsymbol{x}_{1:T}|\boldsymbol{z}_{1:T})$



- Vector form to represent a data sequence

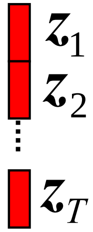
Observation data:

$$\boldsymbol{x} = \left[\boldsymbol{x}_1^{\mathsf{T}}, \boldsymbol{x}_2^{\mathsf{T}}, \cdots, \boldsymbol{x}_T^{\mathsf{T}}\right]^{\mathsf{T}}$$



Latent variables:

$$\boldsymbol{z} = \left[\boldsymbol{z}_1^{\mathsf{T}}, \boldsymbol{z}_2^{\mathsf{T}}, \cdots, \boldsymbol{z}_T^{\mathsf{T}}\right]^{\mathsf{T}}$$



- Emission $p.d.f.$ of the observation sequence vector

Modeled w/ a single Gaussian distribution

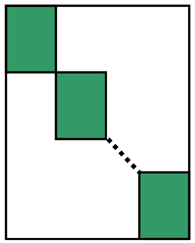
$$p(\boldsymbol{x}_{1:T} \mid \boldsymbol{z}_{1:T}) = \prod_{t=1}^{T} p(\boldsymbol{x}_t \mid \boldsymbol{z}_t)$$

$$= \mathcal{N}\left(\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_T \end{bmatrix} ; \underbrace{\begin{bmatrix} \boldsymbol{W} & & & \\ & \boldsymbol{W} & & \\ & & \ddots & \\ & & & \boldsymbol{W} \end{bmatrix}}_{\tilde{\boldsymbol{W}}} \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \\ \vdots \\ \boldsymbol{z}_T \end{bmatrix} , \underbrace{\begin{bmatrix} \boldsymbol{\Sigma} & & & \\ & \boldsymbol{\Sigma} & & \\ & & \ddots & \\ & & & \boldsymbol{\Sigma} \end{bmatrix}}_{\tilde{\boldsymbol{\Sigma}}} \right)$$

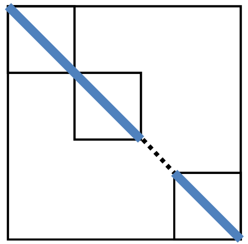
$$= \mathcal{N}\left(\boldsymbol{x}; \tilde{\boldsymbol{W}}\boldsymbol{z}, \tilde{\boldsymbol{\Sigma}}\right)$$

$$= p(\boldsymbol{x} \mid \boldsymbol{z})$$

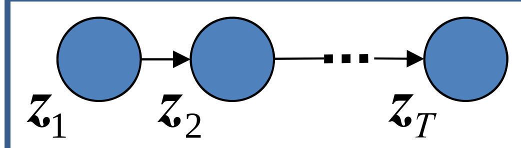
$\tilde{\boldsymbol{W}}$



$\tilde{\boldsymbol{\Sigma}}$



Transition *p.d.f.* $p(\boldsymbol{z}_{1:\mathrm{T}})$ (1)



- Transition *p.d.f.* of the latent variable sequence vector $\boldsymbol{z}$

$$p(\boldsymbol{z}_{1:T}) = p(\boldsymbol{z}_1)\prod_{t=2}^{T} p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1})$$

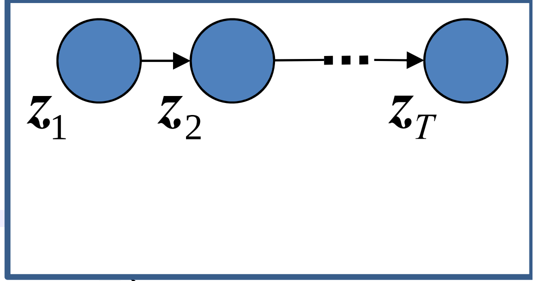
$$= \mathcal{N}\left(\begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \\ \vdots \\ \boldsymbol{z}_T \end{bmatrix} ; \overbrace{\begin{bmatrix} \boldsymbol{I} & & & \\ & \boldsymbol{A} & & \\ & & \ddots & \\ & & & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_0 \\ \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_{T-1} \end{bmatrix}}^{\text{Mean vector}} , \begin{bmatrix} \boldsymbol{P}_0 & & & \\ & \boldsymbol{\Gamma} & & \\ & & \ddots & \\ & & & \boldsymbol{\Gamma} \end{bmatrix} \right)$$

Subtraction of mean vector from $\boldsymbol{z}$ (*i.e.*, $\boldsymbol{z}$ − {mean vector}) :

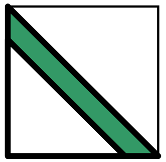
$$\begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \\ \vdots \\ \boldsymbol{z}_T \end{bmatrix} - \begin{bmatrix} \boldsymbol{I} & & & \\ & \boldsymbol{A} & & \\ & & \ddots & \\ & & & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_0 \\ \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_{T-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & & & \\ -\boldsymbol{A} & \boldsymbol{I} & & \\ & \ddots & \ddots & \\ & & -\boldsymbol{A} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \\ \vdots \\ \boldsymbol{z}_T \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu}_0 \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}$$

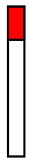
Same variables

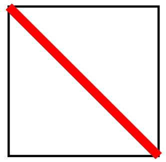
Transition $p.d.f.\ p(\boldsymbol{z}_{1:\mathrm{T}})$ (2)



$$p(\boldsymbol{z}_{1:T}) = \mathcal{N}\left(\underbrace{\begin{bmatrix} \boldsymbol{I} & & & \\ -\boldsymbol{A} & \boldsymbol{I} & & \\ & \ddots & \ddots & \\ & & -\boldsymbol{A} & \boldsymbol{I} \end{bmatrix}}_{\tilde{\boldsymbol{A}}} \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \\ \vdots \\ \boldsymbol{z}_T \end{bmatrix}; \underbrace{\begin{bmatrix} \boldsymbol{\mu}_0 \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}}_{\tilde{\boldsymbol{\mu}}_0}, \underbrace{\begin{bmatrix} \boldsymbol{P}_0 & & & \\ & \boldsymbol{\Gamma} & & \\ & & \ddots & \\ & & & \boldsymbol{\Gamma} \end{bmatrix}}_{\tilde{\boldsymbol{\Gamma}}}\right)$$







$$= \mathcal{N}\left(\tilde{\boldsymbol{A}}\boldsymbol{z}; \tilde{\boldsymbol{\mu}}_0, \tilde{\boldsymbol{\Gamma}}\right)$$

$$= (2\pi)^{-DT/2}\left|\tilde{\boldsymbol{\Gamma}}\right|^{-1/2} \exp\left(-\frac{1}{2}\left(\tilde{\boldsymbol{A}}\boldsymbol{z} - \tilde{\boldsymbol{\mu}}_0\right)^{\mathsf{T}} \tilde{\boldsymbol{\Gamma}}^{-1}\left(\tilde{\boldsymbol{A}}\boldsymbol{z} - \tilde{\boldsymbol{\mu}}_0\right)\right)$$

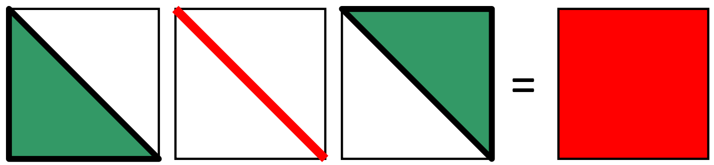
$\left|\tilde{\boldsymbol{A}}\right| = 1$

$$= (2\pi)^{-DT/2}\left|\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{A}}^{-\mathsf{T}}\right|^{-1/2} \exp\left(-\frac{1}{2}\left(\boldsymbol{z} - \tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{\mu}}_0\right)^{\mathsf{T}} \tilde{\boldsymbol{A}}^{\mathsf{T}}\tilde{\boldsymbol{\Gamma}}^{-1}\tilde{\boldsymbol{A}}\left(\boldsymbol{z} - \tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{\mu}}_0\right)\right)$$

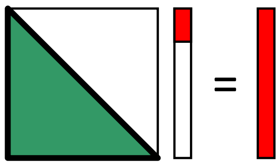
$$= \mathcal{N}\left(\boldsymbol{z}; \tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{\mu}}_0, \ \tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{A}}^{-\mathsf{T}}\right)$$

$$= p(\boldsymbol{z})$$

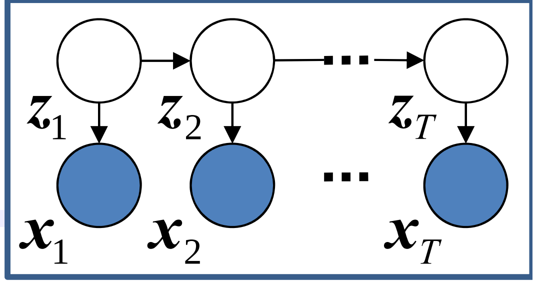
Covariance matrix



Mean vector



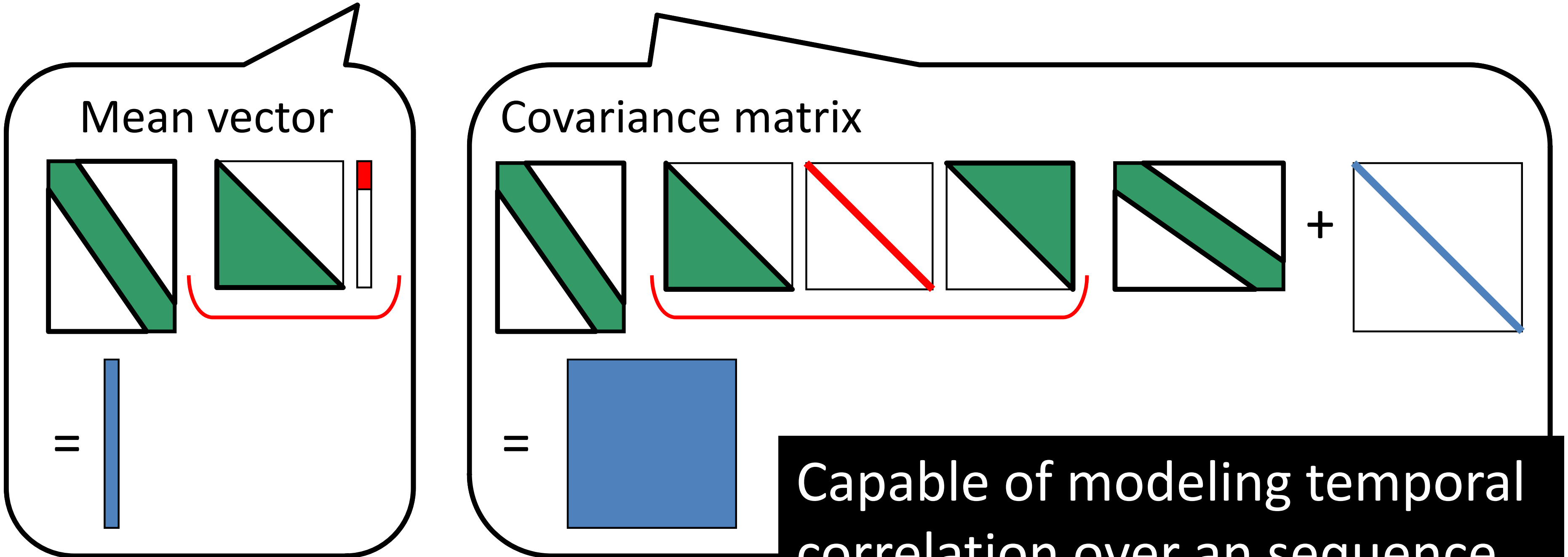
Likelihood Function: *p.d.f.* $p(\boldsymbol{x}_{1:\mathrm{T}})$



- *p.d.f.* of the observation sequence vector $\boldsymbol{x}$:

$$\begin{cases} \boldsymbol{x} = \left[\boldsymbol{x}_1^\top, \boldsymbol{x}_2^\top, \cdots, \boldsymbol{x}_T^\top\right]^\top \\ \boldsymbol{z} = \left[\boldsymbol{z}_1^\top, \boldsymbol{z}_2^\top, \cdots, \boldsymbol{z}_T^\top\right]^\top \end{cases}$$

$$\begin{aligned} p(\boldsymbol{x}) &= \int p(\boldsymbol{x} \mid \boldsymbol{z}) p(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \\ &= \int \mathcal{N}\left(\boldsymbol{x}; \tilde{\boldsymbol{W}}\boldsymbol{z}, \tilde{\boldsymbol{\Sigma}}\right) \mathcal{N}\left(\boldsymbol{z}; \underbrace{\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{\mu}}_0}, \underbrace{\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{A}}^{-\top}}\right) \mathrm{d}\boldsymbol{z} \\ &= \mathcal{N}\left(\boldsymbol{x};\ \tilde{\boldsymbol{W}}\underbrace{\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{\mu}}_0},\ \tilde{\boldsymbol{W}}\underbrace{\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{A}}^{-\top}}\tilde{\boldsymbol{W}}^\top + \tilde{\boldsymbol{\Sigma}}\right) \end{aligned}$$



Capable of modeling temporal correlation over an sequence

Appendix

Derivation of $p.d.f.$s in Kalman Filtering

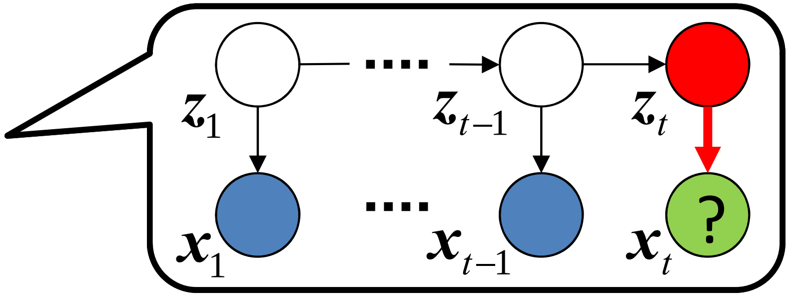
Forward Algorithm (Kalman Filtering)

- Likelihood function factorized into conditional *p.d.f.*s

$$p(\boldsymbol{x}_{1:T}) = p(\boldsymbol{x}_1)p(\boldsymbol{x}_2 \mid \boldsymbol{x}_1)p(\boldsymbol{x}_3 \mid \boldsymbol{x}_{1:2})p(\boldsymbol{x}_4 \mid \boldsymbol{x}_{1:3})\cdots \underline{p(\boldsymbol{x}_t \mid \boldsymbol{x}_{1:t-1})}\cdots p(\boldsymbol{x}_T \mid \boldsymbol{x}_{1:T-1})$$

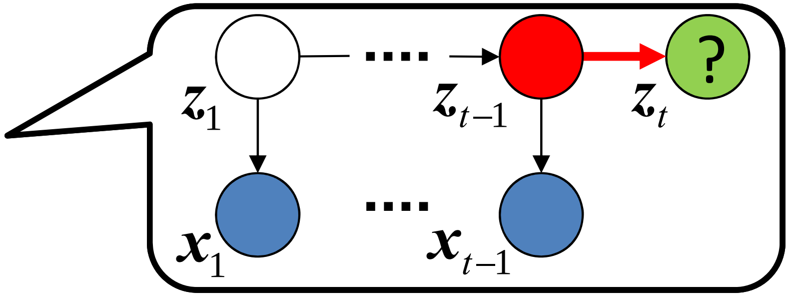
$$p(\underline{\boldsymbol{x}_t} \mid \underline{\boldsymbol{x}_{1:t-1}}) = \int p(\boldsymbol{x}_t, z_t \mid \boldsymbol{x}_{1:t-1})\,\mathrm{d}z_t$$

$$= \int p(\boldsymbol{x}_t \mid z_t)\underline{p(z_t \mid \boldsymbol{x}_{1:t-1})}\,\mathrm{d}z_t$$



$$p(\underline{z_t} \mid \underline{\boldsymbol{x}_{1:t-1}}) = \int p(z_t, z_{t-1} \mid \boldsymbol{x}_{1:t-1})\,\mathrm{d}z_{t-1}$$

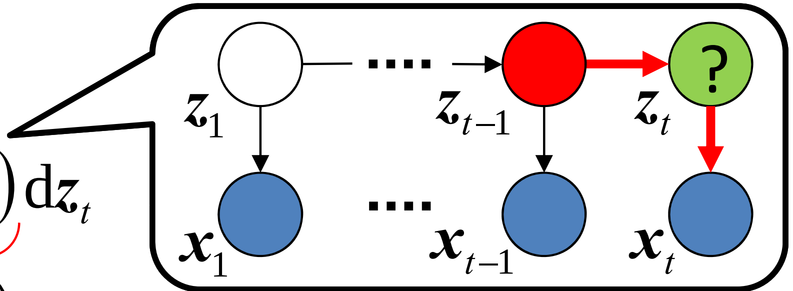
$$= \int p(z_t \mid z_{t-1})\underline{p(z_{t-1} \mid \boldsymbol{x}_{1:t-1})}\,\mathrm{d}z_{t-1}$$



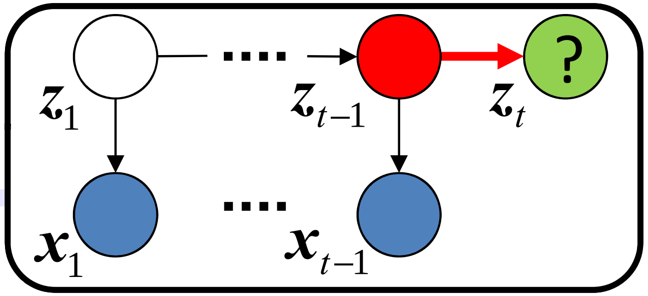
$$\underbrace{p(\underline{z_t} \mid \underline{\boldsymbol{x}_{1:t}})}_{\alpha(z_t)} \propto p(\boldsymbol{x}_t, z_t \mid \boldsymbol{x}_{1:t-1})$$

$$= p(\boldsymbol{x}_t \mid z_t)p(z_t \mid \boldsymbol{x}_{1:t-1})$$

$$= p(\boldsymbol{x}_t \mid z_t)\int p(z_t \mid z_{t-1})\underbrace{p(z_{t-1} \mid \boldsymbol{x}_{1:t-1})}_{\alpha(z_{t-1})}\,\mathrm{d}z_t$$



Derivation of $p(z_t | x_{1:t-1})$



- Predicted distribution on the state space

$$p(z_t \mid x_{1:t-1}) = \int p(z_t \mid z_{t-1})\, p(z_{t-1} \mid x_{1:t-1})\, \mathrm{d}z_t$$

Transition $p.d.f.$ $= \mathcal{N}(z_t; A z_{t-1}, \Gamma)$

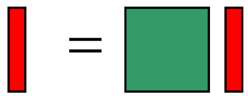
Assumed to be $\mathcal{N}(z_{t-1}; \mu_{t-1}, P_{t-1})$
(Its derivation will be given later.)

$$= \int \mathcal{N}(z_t; A z_{t-1}, \Gamma)\, \mathcal{N}(z_{t-1}; \mu_{t-1}, P_{t-1})\, \mathrm{d}z_{t-1}$$

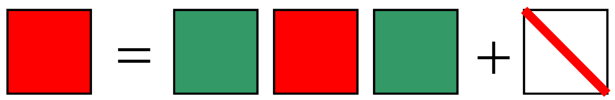
$$= \mathcal{N}(z_t; A\mu_{t-1}, A P_{t-1} A^{\mathsf{T}} + \Gamma)$$

$$= \mathcal{N}(z_t; \mu_{t|t-1}, P_{t|t-1})$$

Predicted mean : $\mu_{t|t-1} = A\mu_{t-1}$



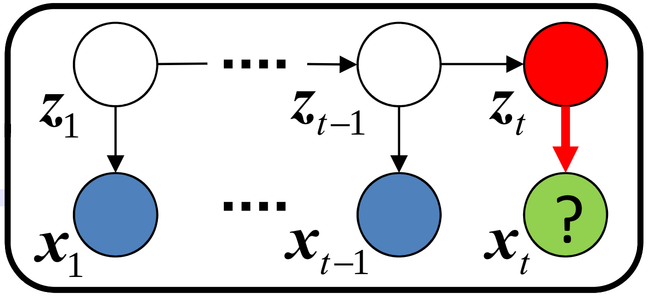
Predicted covariance : $P_{t|t-1} = A P_{t-1} A^{\mathsf{T}} + \Gamma$



$$\mu_{1|0} = \mu_0$$

$$P_{1|0} = P_0$$

Derivation of $p(\boldsymbol{x}_t | \boldsymbol{x}_{1:t-1})$



- Likelihood function
  (*i.e.*, predicted distribution on the observation space)

$$p(\boldsymbol{x}_t \mid \boldsymbol{x}_{1:t-1}) = \int p(\boldsymbol{x}_t \mid \boldsymbol{z}_t) p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:t-1}) \mathrm{d}\boldsymbol{z}_t$$

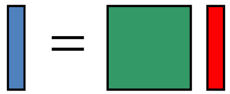
Emission $p.d.f. = \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{W}\boldsymbol{z}_t, \boldsymbol{\Sigma})$

Predicted distribution $= \mathcal{N}(\boldsymbol{z}_t; \boldsymbol{\mu}_{t|t-1}, \boldsymbol{P}_{t|t-1})$

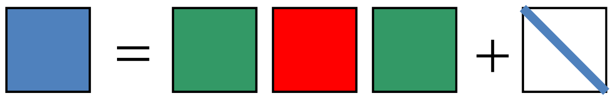
$$= \int \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{W}\boldsymbol{z}_t, \boldsymbol{\Sigma}) \mathcal{N}(\boldsymbol{z}_t; \boldsymbol{\mu}_{t|t-1}, \boldsymbol{P}_{t|t-1}) \mathrm{d}\boldsymbol{z}_t$$

$$= \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{W}\boldsymbol{\mu}_{t|t-1}, \boldsymbol{W}\boldsymbol{P}_{t|t-1}\boldsymbol{W}^{\mathsf{T}} + \boldsymbol{\Sigma})$$

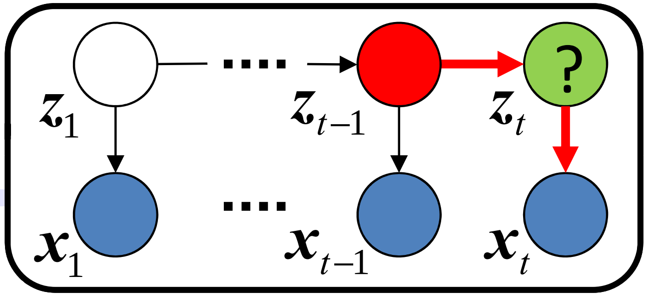
Mean : $\boldsymbol{W}\boldsymbol{\mu}_{t|t-1}$



Covariance: $\boldsymbol{W}\boldsymbol{P}_{t|t-1}\boldsymbol{W}^{\mathsf{T}} + \boldsymbol{\Sigma}$



Derivation of $\alpha(\boldsymbol{z}_t) = p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:t})$



- Updated distribution on the state space

$$\alpha(\boldsymbol{z}_t) \propto p(\boldsymbol{x}_t \mid \boldsymbol{z}_t)\, p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:t-1})$$

Posterior ∝ Likelihood x Prior

$$= p(\boldsymbol{x}_t \mid \boldsymbol{z}_t) \int \left\{ p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1})\, \alpha(\boldsymbol{z}_{t-1}) \right\} \mathrm{d}\boldsymbol{z}_{t-1}$$

$$\begin{cases} p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}) = \mathcal{N}(\boldsymbol{z}_t; \boldsymbol{A}\boldsymbol{z}_{t-1}, \boldsymbol{\Gamma}) \\ p(\boldsymbol{x}_t \mid \boldsymbol{z}_t) = \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{W}\boldsymbol{z}_t, \boldsymbol{\Sigma}) \end{cases}$$

Assuming that $\alpha(\boldsymbol{z}_t) = \mathcal{N}(\boldsymbol{z}_t; \boldsymbol{\mu}_t, \boldsymbol{P}_t)$

$$\underline{\mathcal{N}(\boldsymbol{z}_t; \boldsymbol{\mu}_t, \boldsymbol{P}_t)} \propto \underbrace{\mathcal{N}(\boldsymbol{x}_t; \boldsymbol{W}\boldsymbol{z}_t, \boldsymbol{\Sigma}) \int \underbrace{\left\{ \mathcal{N}(\boldsymbol{z}_t; \boldsymbol{A}\boldsymbol{z}_{t-1}, \boldsymbol{\Gamma})\, \underline{\mathcal{N}(\boldsymbol{z}_{t-1}; \boldsymbol{\mu}_{t-1}, \boldsymbol{P}_{t-1})} \right\}}_{\mathcal{N}(\boldsymbol{z}_{t-1}; ?\,\boldsymbol{z}_t + ?, ?)\,\mathcal{N}(\boldsymbol{z}_t; ?, ?)} \mathrm{d}\boldsymbol{z}_{t-1}}_{\mathcal{N}(\boldsymbol{z}_t; ?, ?)}$$

Kalman gain matrix : $\boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\boldsymbol{W}^{\mathsf{T}}\left(\boldsymbol{W}\boldsymbol{P}_{t|t-1}\boldsymbol{W}^{\mathsf{T}} + \boldsymbol{\Sigma}\right)^{-1}$

Updated mean : $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t|t-1} + \boldsymbol{K}_t\left(\underline{\boldsymbol{x}_t - \boldsymbol{W}\boldsymbol{\mu}_{t|t-1}}\right)$

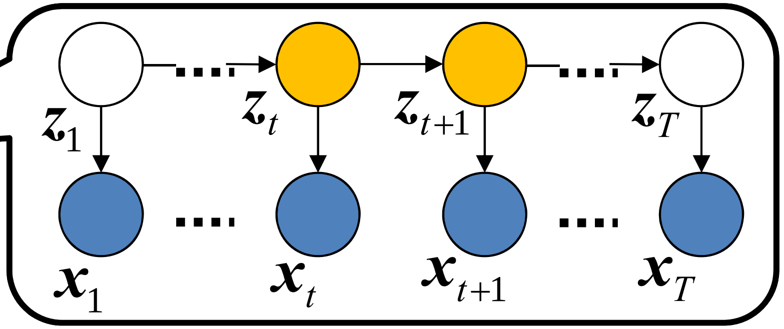
Updated covariance : $\boldsymbol{P}_t = \left(\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{W}\right)\boldsymbol{P}_{t|t-1}$

Error between predicted and observed data

Appendix

Derivation of *p.d.f.s* in Kalman Smoothing

Backward Algorithm (Kalman Smoothing)



- Joint posterior *p.d.f.* of z_t and z_{t+1}

$$
\begin{aligned}
p(z_t, z_{t+1} \mid x_{1:T}) &= p(z_{t+1} \mid x_{1:T})\,p(z_t \mid z_{t+1}, x_{1:T}) \\
&= p(z_{t+1} \mid x_{1:T})\,p(z_t \mid z_{t+1}, x_{1:t}) \\
&= p(z_{t+1} \mid x_{1:T}) \frac{p(z_t \mid x_{1:t})\,p(z_{t+1} \mid z_t, x_{1:t})}{p(z_{t+1} \mid x_{1:t})} \\
&= p(z_t \mid x_{1:t}) \frac{p(z_{t+1} \mid x_{1:T})\,p(z_{t+1} \mid z_t)}{p(z_{t+1} \mid x_{1:t})}
\end{aligned}
$$

Transition *p.d.f.*

Updated *p.d.f.* in forward algorithm

Predicted *p.d.f.* in forward algorithm

- Posterior *p.d.f.* of z_t, *i.e.*, $p(z_t \mid x_{1:T}) = \int p(z_t, z_{t+1} \mid x_{1:T})\, dz_{t+1}$

$$
\overbrace{p(z_t \mid x_{1:T})}^{\gamma(z_t)} = \overbrace{p(z_t \mid x_{1:t})}^{\alpha(z_t)} \int \frac{\overbrace{p(z_{t+1} \mid x_{1:T})}^{\gamma(z_{t+1})}\,p(z_{t+1} \mid z_t)}{p(z_{t+1} \mid x_{1:t})}\, dz_{t+1}
$$

Derivation of $\gamma(\boldsymbol{z}_t) = p(\boldsymbol{z}_t \mid \boldsymbol{x}_{1:T})$

$$\gamma(\boldsymbol{z}_t) = \alpha(\boldsymbol{z}_t) \int \frac{\gamma(\boldsymbol{z}_{t+1}) p(\boldsymbol{z}_{t+1} \mid \boldsymbol{z}_t)}{p(\boldsymbol{z}_{t+1} \mid \boldsymbol{x}_{1:t})} \mathrm{d}\boldsymbol{z}_{t+1}$$

Assuming that $\gamma(\boldsymbol{z}_t) = \mathcal{N}\left(\boldsymbol{z}_t ; \hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{P}}_t\right)$

$$\mathcal{N}\left(\boldsymbol{z}_t ; \hat{\boldsymbol{\mu}}_t, \hat{\boldsymbol{P}}_t\right) = \underbrace{\mathcal{N}(\boldsymbol{z}_t ; \boldsymbol{\mu}_t, \boldsymbol{P}_t) \int \underbrace{\frac{\mathcal{N}\left(\boldsymbol{z}_{t+1} ; \hat{\boldsymbol{\mu}}_{t+1}, \hat{\boldsymbol{P}}_{t+1}\right) \mathcal{N}(\boldsymbol{z}_{t+1} ; \boldsymbol{A}\boldsymbol{z}_t, \boldsymbol{\Gamma})}{\mathcal{N}\left(\boldsymbol{z}_{t+1} ; \boldsymbol{\mu}_{t+1|t}, \boldsymbol{P}_{t+1|t}\right)}}_{\mathcal{N}(\boldsymbol{z}_{t+1} ; ?\, \boldsymbol{z}_t + ?, ?) \mathcal{N}(\boldsymbol{z}_t ; ?, ?)} \mathrm{d}\boldsymbol{z}_{t+1}}_{\mathcal{N}(\boldsymbol{z}_t ; ?, ?)}$$

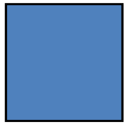
$$\boldsymbol{J}_t = \boldsymbol{P}_t \boldsymbol{A}^{\mathsf{T}} \boldsymbol{P}_{t+1|t}^{-1}$$

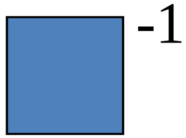
Smoothed mean : $\hat{\boldsymbol{\mu}}_t = \boldsymbol{\mu}_t + \boldsymbol{J}_t \left(\hat{\boldsymbol{\mu}}_{t+1} - \boldsymbol{\mu}_{t+1|t}\right)$

Smoothed covariance : $\hat{\boldsymbol{P}}_t = \boldsymbol{P}_t + \boldsymbol{J}_t \left(\hat{\boldsymbol{P}}_{t+1} - \boldsymbol{P}_{t+1|t}\right) \boldsymbol{J}_t^{\mathsf{T}}$

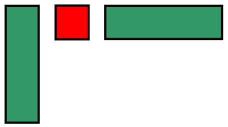
Tips: Matrix Inversion Lemma

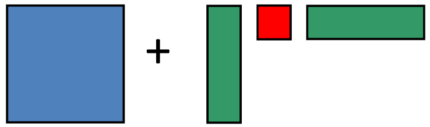
- Condition 1: Matrix $\boldsymbol{A}$ and its inverse matrix $\boldsymbol{A}^{-1}$ are given.

$\boldsymbol{A} =$ 

$\boldsymbol{A}^{-1} =$ 

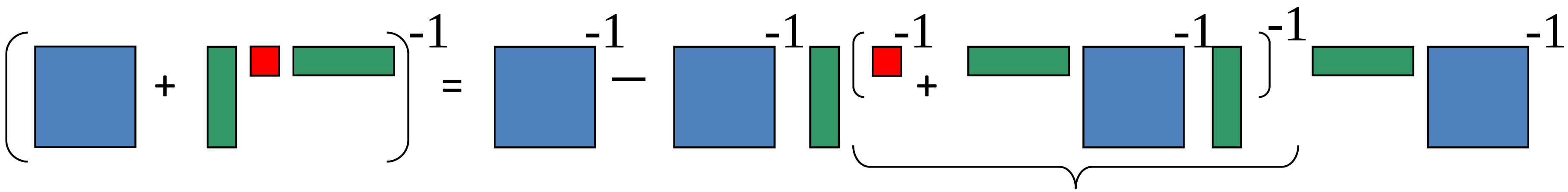
- Condition 2: Fluctuation generated on a lower-dimensional subspace $\boldsymbol{vNv}^{\mathrm{T}}$ is added to the matrix $\boldsymbol{A}$.

$\boldsymbol{vNv}^{\mathrm{T}} =$ 

$\boldsymbol{A} + \boldsymbol{vNv}^{\mathrm{T}} =$ 

- Under these conditions, an inverse matrix $\left(\boldsymbol{A} + \boldsymbol{vNv}^{\mathrm{T}}\right)^{-1}$ can be calculated as follows:

$$\left(\boldsymbol{A} + \boldsymbol{vNv}^{\mathrm{T}}\right)^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{v}\left(\boldsymbol{N}^{-1} + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{v}\right)^{-1}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}$$



Calculation of an inverse matrix on a lower dimensional space