PRML 13.3-13.3.2

12/7/2018

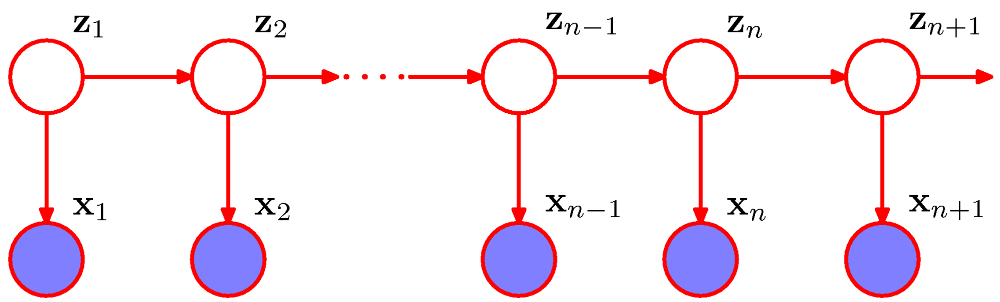
Kazunori Sakai

Outline

HMM LDS

discrete continuous

maximum likelihood



Contents

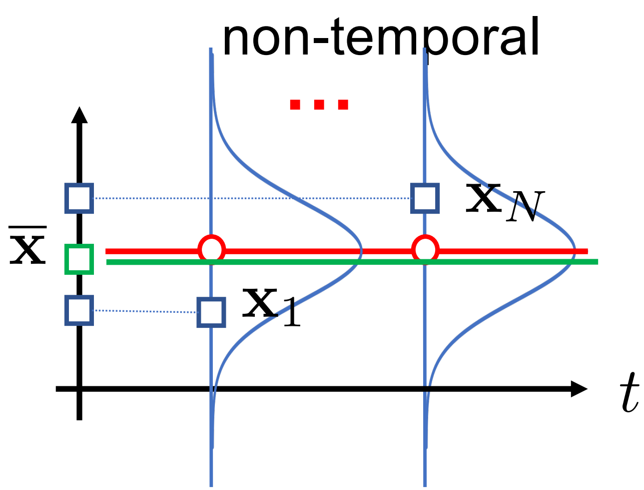
13.3 Linear Dynamical Systems

Motivation behind LDS

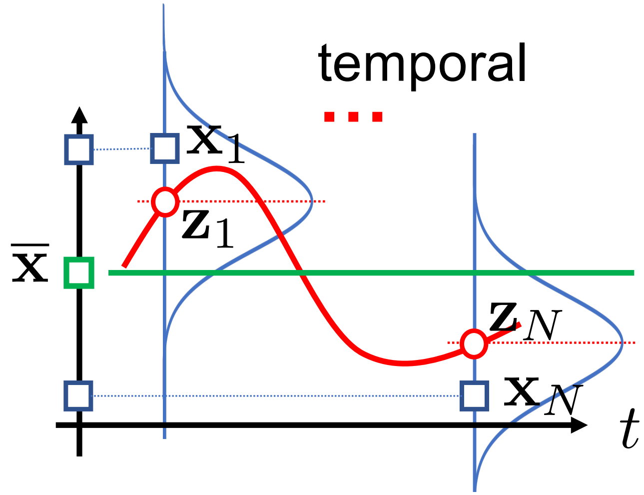
- **Estimating z with observable x.**

  non-observable: $\mathbf{z}$

  observable: $\mathbf{x} = \mathbf{z} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I}\right)$



- non-temporal
  - given a single measurement : $\mathbf{z} = \mathbf{x}$
  - given lots of measurements : $\mathbf{z} = \frac{1}{N}\sum_n^N \mathbf{x}_n$



- temporal
  - $\bar{\mathbf{x}}$ will be a new source of error.

Motivation behind LDS

- temporal

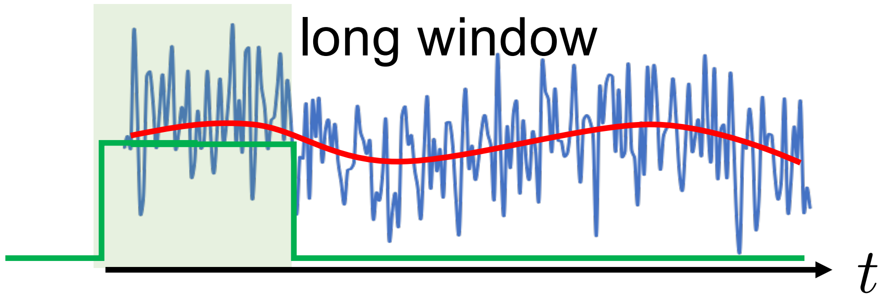
$$\mathbf{z}_N = \sum_{n=N-L}^{N} \mathbf{x}_n$$

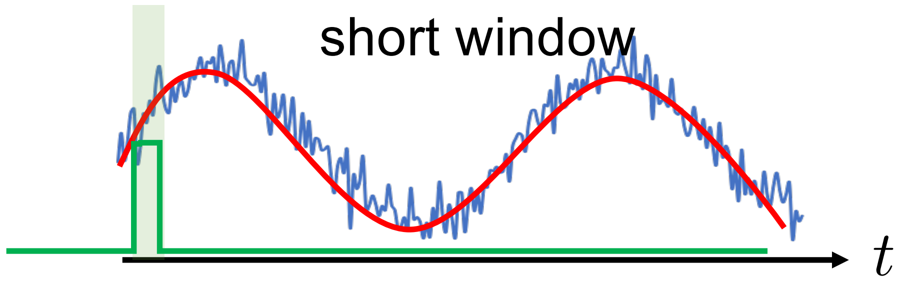
  - change slowly, noise level is high.

    long window.

  - change quickly, noise level is low.

    short window.

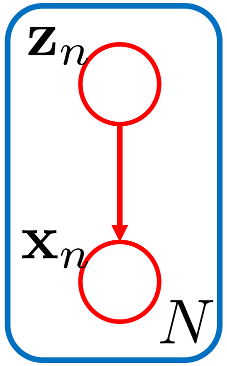
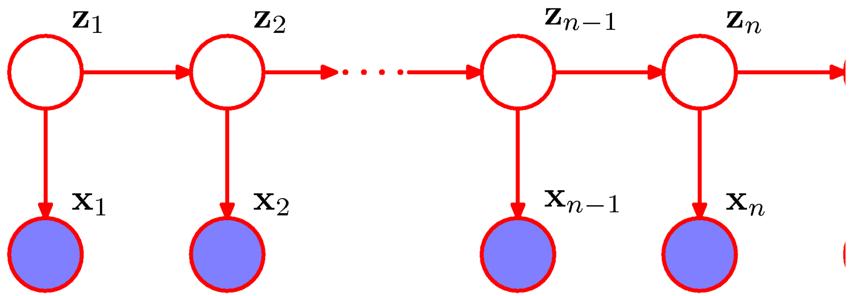




- More recent observations will contribute more.

- **How to form a weighted average?**

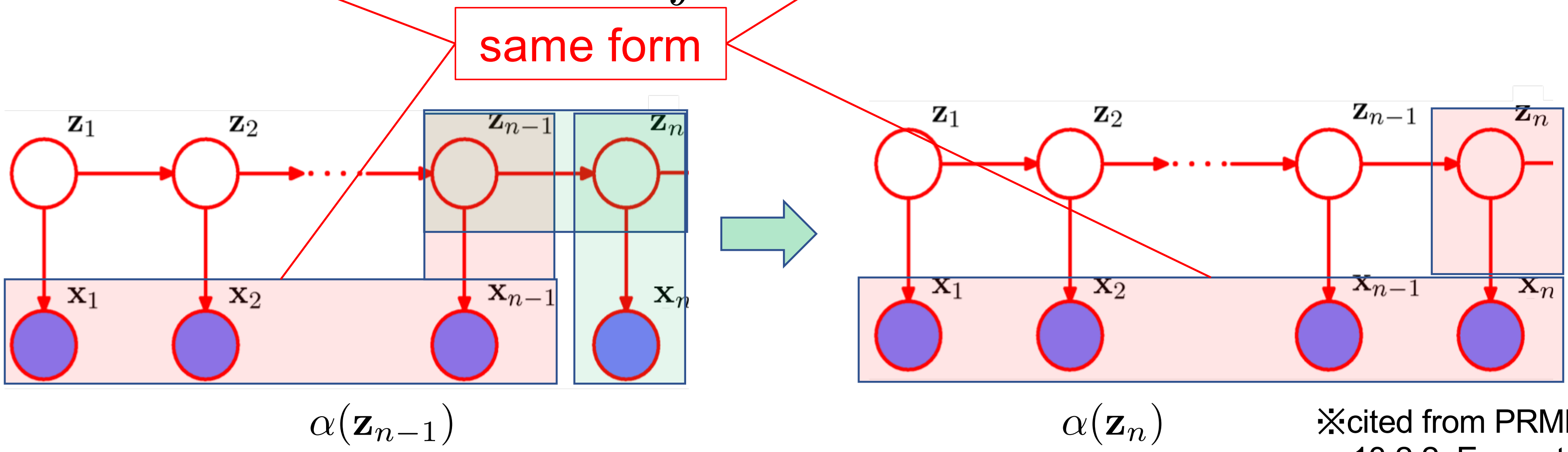
From HMM to LDS

	non-temporal	temporal
		
discrete	mixture models	HMM
continuous	continuous latent variable models	LDS

Requirements

- Efficient algorithm for inference.

$$c_n \widehat{\alpha}\left(\mathbf{z}_n\right) = p\left(\mathbf{x}_n|\mathbf{z}_n\right) \int \widehat{\alpha}\left(\mathbf{z}_{n-1}\right) p\left(\mathbf{z}_n|\mathbf{z}_{n-1}\right) \mathrm{d}\mathbf{z}_{n-1}$$

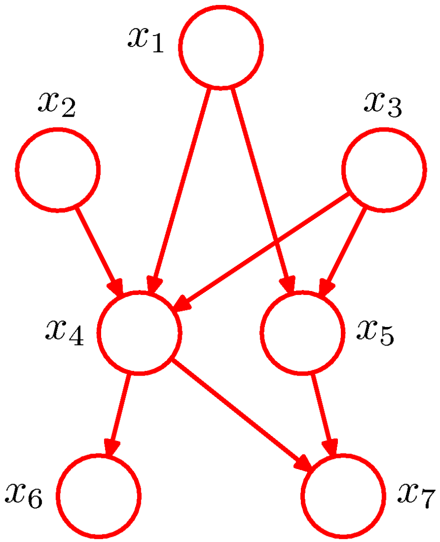


※cited from PRML13.2.1-13.2.2_Enomoto #20

- distributions belonging to the exponential family.

Linear-Gaussian Models

PRML 8.1.4
PRML 2.3



$$\mathbf{x} = \{x_1, x_2, \ldots, x_7\}$$

$$p\left(x_i | \mathrm{pa}_i\right) = \mathcal{N}\left(x_i \,\middle|\, \sum_{j \in \mathrm{pa}_i} w_{ij} x_j + b_i, v_i\right)$$

$$\ln p(\mathbf{x}) = \sum_{i=1}^{D} \ln p\left(x_i | \mathrm{pa}_i\right)$$

$$= -\sum_{i=1}^{D} \frac{1}{2v_i}\left(x_i - \sum_{j \in \mathrm{pa}_i} w_{ij} x_j - b_i\right)^2 + \text{ const}$$

Joint distribution $p\left(\mathbf{x}\right)$ is Gaussian distribution.

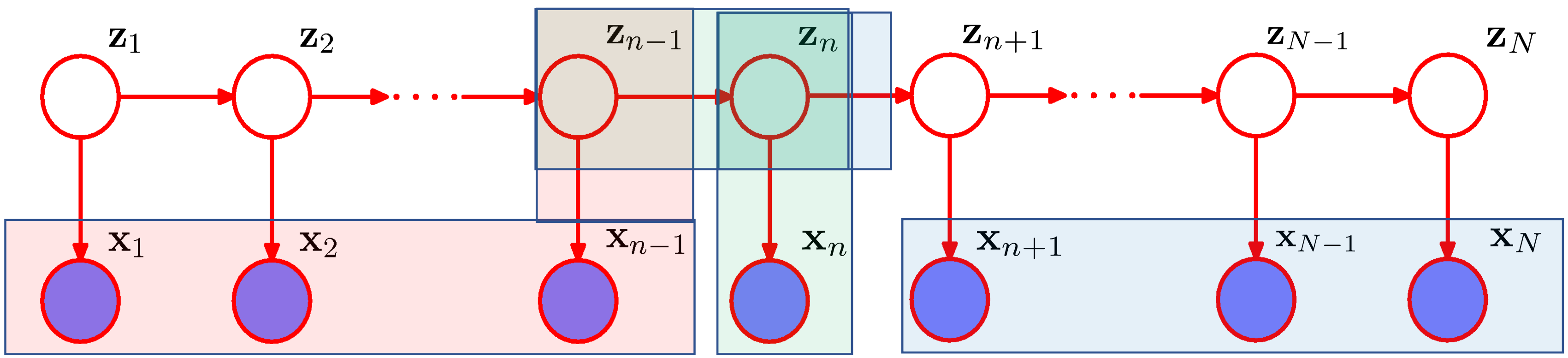
Linear-Gaussian Models

$$\widehat{\alpha}(\mathbf{z}_n) = p(\mathbf{z}_n | \mathbf{x}_1, \ldots, \mathbf{x}_n)$$

$$\widehat{\beta}(\mathbf{z}_n) = \frac{p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_n)}{p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{x}_1, \ldots, \mathbf{x}_n)}$$

$$\gamma(\mathbf{z}_n) = \widehat{\alpha}(\mathbf{z}_n)\,\widehat{\beta}(\mathbf{z}_n)$$

Gaussian



Gaussian Mixture Models 

The emission densities $p(\mathbf{x}_n|\mathbf{z}_n)$ comprise a mixture of K Gaussians.

$$p\left(\mathbf{x}_n|\mathbf{z}_n\right) = \sum^{K} \pi_k \mathcal{N}\left(\mathbf{x}_n|\mathbf{A}_k\mathbf{z}_n, \mathbf{\Sigma}_k\right)$$

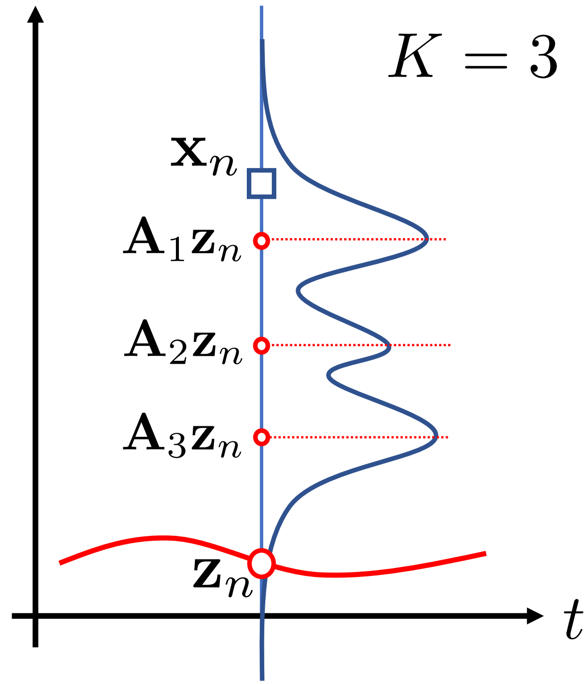
$$= \sum_{\mathbf{s}} p\left(\mathbf{s}\right) p\left(\mathbf{x}_n|\mathbf{z}_n, \mathbf{s}\right)$$

where

$$p\left(\mathbf{s}\right) = \prod^{K} \pi_k^{s_k}$$

$$p\left(\mathbf{x}_n|\mathbf{z}_n, \mathbf{s}\right) = \prod^{K} \mathcal{N}\left(\mathbf{x}_n|\mathbf{A}_k\mathbf{z}_n, \mathbf{\Sigma}_k\right)^{s_k}$$

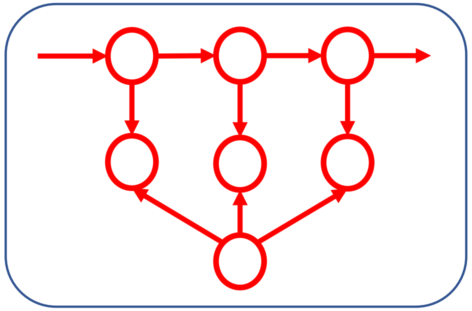
To measure z using three noisy sensors.

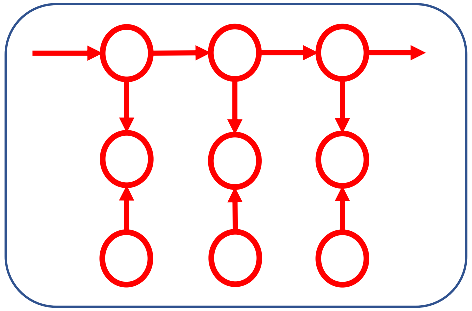


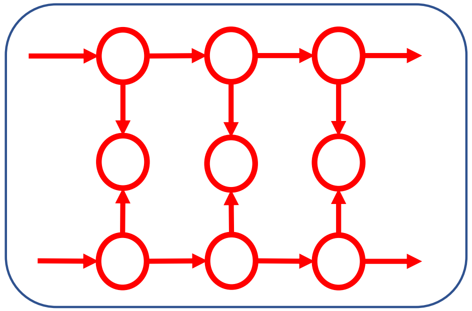
$$p\left(\mathbf{z}_n|\mathbf{x}_n\right) \propto p\left(\mathbf{x}_n|\mathbf{z}_n\right) p\left(\mathbf{z}_n\right)$$

GMM — GMM — Gaussian?

Gaussian Mixture Models 







Even if $\widehat{\alpha}(\mathbf{z}_1)$ is Gaussian,

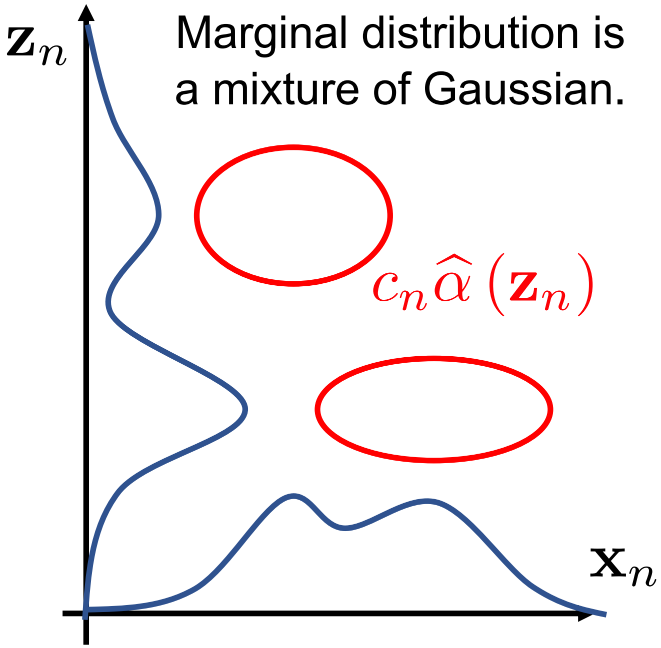
K-Gaussians　K-Gaussians　Gaussian

$$c_2\widehat{\alpha}(\mathbf{z}_2) = p(\mathbf{x}_2|\mathbf{z}_2)\int \widehat{\alpha}(\mathbf{z}_1)\, p(\mathbf{z}_2|\mathbf{z}_1)\, \mathrm{d}\mathbf{z}_1$$

$$c_3\widehat{\alpha}(\mathbf{z}_3) = p(\mathbf{x}_3|\mathbf{z}_3)\int \widehat{\alpha}(\mathbf{z}_2)\, p(\mathbf{z}_3|\mathbf{z}_2)\, \mathrm{d}\mathbf{z}_2$$

K^2-Gaussians　K-Gaussians　K-Gaussians

- Exact inference will not be practical.



The most probable latent sequence

	$x = 0$	$x = 1$
$y = 0$	0.3	0.4
$y = 1$	0.3	0.0

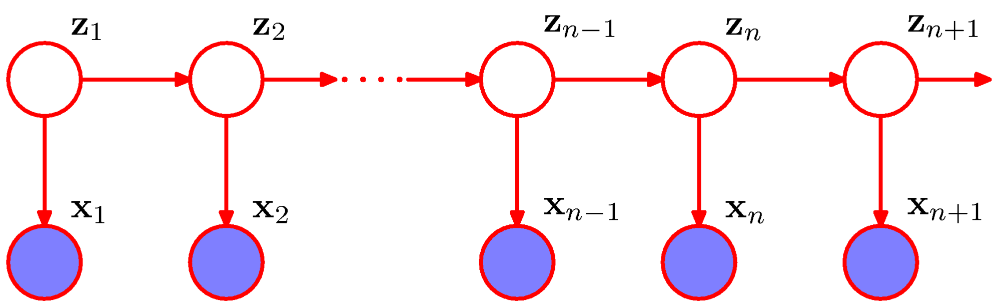
Defining the Model

- The transition distributions.

$$p\left(\mathbf{z}_n|\mathbf{z}_{n-1}\right) = \mathcal{N}\left(\mathbf{z}_n|\mathbf{A}\mathbf{z}_{n-1}, \mathbf{\Gamma}\right)$$

- The emission distributions.

$$p\left(\mathbf{x}_n|\mathbf{z}_n\right) = \mathcal{N}\left(\mathbf{x}_n|\mathbf{C}\mathbf{z}_n, \mathbf{\Sigma}\right)$$



- The initial latent variable.

$$p\left(\mathbf{z}_1\right) = \mathcal{N}\left(\mathbf{z}_1|\boldsymbol{\mu}_0, \mathbf{P}_0\right)$$

- The parameters.

$$\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{\Gamma}, \mathbf{C}, \mathbf{\Sigma}, \boldsymbol{\mu}_0, \mathbf{P}_0\}$$

Equivalent expression

$$p(\mathbf{z}_n|\mathbf{z}_{n-1}) = \mathcal{N}(\mathbf{z}_n|\mathbf{A}\mathbf{z}_{n-1}, \mathbf{\Gamma})$$
$$p(\mathbf{x}_n|\mathbf{z}_n) = \mathcal{N}(\mathbf{x}_n|\mathbf{C}\mathbf{z}_n, \mathbf{\Sigma})$$
$$p(\mathbf{z}_1) = \mathcal{N}(\mathbf{z}_1|\boldsymbol{\mu}_0, \mathbf{P}_0)$$

$$\mathbf{z}_n = \mathbf{A}\mathbf{z}_{n-1} + \mathbf{w}_n$$
$$\mathbf{x}_n = \mathbf{C}\mathbf{z}_n + \mathbf{v}_n$$
$$\mathbf{z}_1 = \boldsymbol{\mu}_0 + \mathbf{u}$$

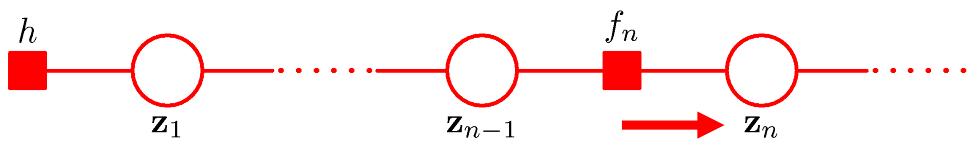
$$\mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0}, \mathbf{\Gamma})$$
$$\mathbf{v} \sim \mathcal{N}(\mathbf{v}|\mathbf{0}, \mathbf{\Sigma})$$
$$\mathbf{u} \sim \mathcal{N}(\mathbf{u}|\mathbf{0}, \mathbf{V}_0)$$

13.3.1 Inference in LDS

The forward equations

Given: $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{\Gamma}, \mathbf{C}, \mathbf{\Sigma}, \boldsymbol{\mu}_0, \mathbf{P}_0\}$

Objective: $p(\mathbf{z}_n | \mathbf{x}_1, \ldots, \mathbf{x}_n), \; p(\mathbf{z}_n | \mathbf{x}_1, \ldots, \mathbf{x}_{n-1}), \; p(\mathbf{x}_n | \mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$



$\widehat{\alpha}(\mathbf{z}_n) = ?$

$h(\mathbf{z}_1) = p(\mathbf{z}_1)\, p(\mathbf{x}_1 | \mathbf{z}_1)$: Gaussian

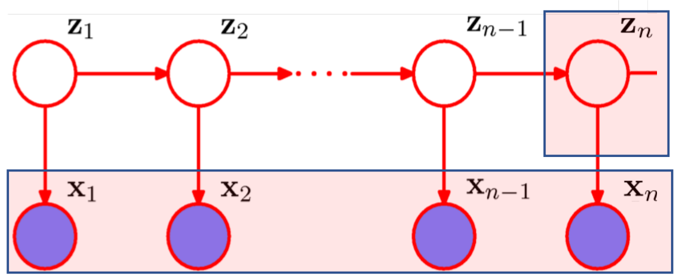
$f_n(\mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{z}_n | \mathbf{z}_{n-1})\, p(\mathbf{x}_n | \mathbf{z}_n)$: Gaussian

$$\mu_{f_n \to \mathbf{z}_n}(\mathbf{z}_n) = \int f_n(\mathbf{z}_{n-1}, \mathbf{z}_n)\, \mu_{f_{n-1} \to \mathbf{z}_{n-1}}(\mathbf{z}_{n-1}) \mathrm{d}\mathbf{z}_{n-1}$$

Form of $\widehat{\alpha}(\mathbf{z}_n)$ must be Gaussian.

The forward equations

$$\widehat{\alpha}(\mathbf{z}_n) = \mathcal{N}(\mathbf{z}_n | \boldsymbol{\mu}_n, \mathbf{V}_n)$$



- Determine $\boldsymbol{\mu}_n, \mathbf{V}_n$

$$c_n \widehat{\alpha}(\mathbf{z}_n) = p(\mathbf{x}_n | \mathbf{z}_n) \int \widehat{\alpha}(\mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1}) \, \mathrm{d}\mathbf{z}_{n-1}$$

$$p(\mathbf{z}_n | \mathbf{z}_{n-1}) = \mathcal{N}(\mathbf{z}_n | \mathbf{A}\mathbf{z}_{n-1}, \boldsymbol{\Gamma})$$

$$p(\mathbf{x}_n | \mathbf{z}_n) = \mathcal{N}(\mathbf{x}_n | \mathbf{C}\mathbf{z}_n, \boldsymbol{\Sigma})$$

Assuming $\boldsymbol{\mu}_{n-1}, \mathbf{V}_{n-1}$ are known.

$$c_n \mathcal{N}(\mathbf{z}_n | \boldsymbol{\mu}_n, \mathbf{V}_n) = \mathcal{N}(\mathbf{x}_n | \mathbf{C}\mathbf{z}_n, \boldsymbol{\Sigma}) \times \int \mathcal{N}(\mathbf{z}_n | \mathbf{A}\mathbf{z}_{n-1}, \boldsymbol{\Gamma}) \mathcal{N}(\mathbf{z}_{n-1} | \boldsymbol{\mu}_{n-1}, \mathbf{V}_{n-1}) \, \mathrm{d}\mathbf{z}_{n-1}$$

The forward equations

$$\int \mathcal{N}\left(\mathbf{z}_n | \mathbf{A}\mathbf{z}_{n-1}, \mathbf{\Gamma}\right) \mathcal{N}\left(\mathbf{z}_{n-1} | \boldsymbol{\mu}_{n-1}, \mathbf{V}_{n-1}\right) \mathrm{d}\mathbf{z}_{n-1}$$

$$= \mathcal{N}\left(\mathbf{z}_n | \mathbf{A}\boldsymbol{\mu}_{n-1}, \mathbf{P}_{n-1}\right)$$

where $\quad \mathbf{P}_{n-1} = \mathbf{A}\mathbf{V}_{n-1}\mathbf{A}^{\mathrm{T}} + \mathbf{\Gamma}$

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{\Lambda}^{-1})$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{L}^{-1})$$

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{M})$$

where $\quad \mathbf{M} = \mathbf{L}^{-1} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}}$

$p\left(\mathbf{x}_n | \mathbf{x}_1, \ldots, \mathbf{x}_{n-1}\right)$ $\quad p\left(\mathbf{z}_n | \mathbf{x}_1, \ldots, \mathbf{x}_n\right)$ $\quad p\left(\mathbf{x}_n | \mathbf{z}_n\right)$ $\quad p\left(\mathbf{z}_n | \mathbf{x}_1, \ldots, \mathbf{x}_{n-1}\right)$

$$c_n \mathcal{N}\left(\mathbf{z}_n | \boldsymbol{\mu}_n, \mathbf{V}_n\right) = \mathcal{N}\left(\mathbf{x}_n | \mathbf{C}\mathbf{z}_n, \mathbf{\Sigma}\right) \times \mathcal{N}\left(\mathbf{z}_n | \mathbf{A}\boldsymbol{\mu}_{n-1}, \mathbf{P}_{n-1}\right)$$

$$c_n = \mathcal{N}\left(\mathbf{x}_n | \mathbf{C}\mathbf{A}\boldsymbol{\mu}_{n-1}, \mathbf{C}\mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}} + \mathbf{\Sigma}\right)$$

The forward equations

$\boxed{p\left(\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}\right)}$ $\boxed{p\left(\mathbf{z}_n|\mathbf{x}_1,\ldots,\mathbf{x}_n\right)}$ $\boxed{p\left(\mathbf{x}_n|\mathbf{z}_n\right)}$ $\boxed{p\left(\mathbf{z}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}\right)}$

$$c_n \mathcal{N}\left(\mathbf{z}_n|\boldsymbol{\mu}_n, \mathbf{V}_n\right) = \mathcal{N}\left(\mathbf{x}_n|\mathbf{C}\mathbf{z}_n, \boldsymbol{\Sigma}\right) \times \mathcal{N}\left(\mathbf{z}_n|\mathbf{A}\boldsymbol{\mu}_{n-1}, \mathbf{P}_{n-1}\right)$$

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1})$$
$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}+\mathbf{b}, \mathbf{L}^{-1})$$
$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\left\{\mathbf{A}^{\mathrm{T}}\mathbf{L}(\mathbf{y}-\mathbf{b})+\boldsymbol{\Lambda}\boldsymbol{\mu}\right\}, \boldsymbol{\Sigma})$$
$$\text{where} \quad \boldsymbol{\Sigma} = \left(\boldsymbol{\Lambda}+\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A}\right)^{-1}$$

$$\boldsymbol{\mu}_n = \mathbf{V}_n\left(\mathbf{C}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}_n + \mathbf{P}_{n-1}^{-1}\mathbf{A}\boldsymbol{\mu}_{n-1}\right)$$
$$\mathbf{V}_n = \left(\mathbf{P}_{n-1}^{-1} + \mathbf{C}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{C}\right)^{-1}$$

$$\text{(C.5)} \quad \left(\mathbf{P}^{-1}+\mathbf{B}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{R}^{-1} = \mathbf{P}\mathbf{B}^{\mathrm{T}}\left(\mathbf{B}\mathbf{P}\mathbf{B}^{\mathrm{T}}+\mathbf{R}\right)^{-1}$$
$$\text{(C.7)} \quad \left(\mathbf{A}+\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}+\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1}$$

The forward equations

- Using (C.7)

$$\mathbf{V}_n = \left(\mathbf{P}_{n-1}^{-1} + \mathbf{C}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{C}\right)^{-1}$$
$$= \mathbf{P}_{n-1} - \mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}}\left(\boldsymbol{\Sigma} + \mathbf{C}\mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}}\right)^{-1}\mathbf{C}\mathbf{P}_{n-1}$$
$$= \left(\mathbf{I} - \mathbf{K}_n\mathbf{C}\right)\mathbf{P}_{n-1}$$

where $\mathbf{K}_n = \mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}}\left(\mathbf{C}\mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}} + \boldsymbol{\Sigma}\right)^{-1}$: *Kalman gain matrix*

- Using (C.5)

$$\mathbf{V}_n\mathbf{C}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}_n = \left(\mathbf{P}_{n-1}^{-1} + \mathbf{C}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}_n$$
$$= \mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}}\left(\boldsymbol{\Sigma} + \mathbf{C}\mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}}\right)^{-1}\mathbf{x}_n$$
$$= \mathbf{K}_n\mathbf{x}_n$$

$$\boldsymbol{\mu}_n = \mathbf{K}_n\mathbf{x}_n + \mathbf{V}_n\mathbf{P}_{n-1}^{-1}\mathbf{A}\boldsymbol{\mu}_{n-1} = \mathbf{K}_n\mathbf{x}_n + \left(\mathbf{I} - \mathbf{K}_n\mathbf{C}\right)\mathbf{P}_{n-1}\mathbf{P}_{n-1}^{-1}\mathbf{A}\boldsymbol{\mu}_{n-1}$$
$$= \mathbf{A}\boldsymbol{\mu}_{n-1} + \mathbf{K}_n\left(\mathbf{x}_n - \mathbf{C}\mathbf{A}\boldsymbol{\mu}_{n-1}\right)$$

Kalman Filter equations

observed — prediction for $\mathbf{x}_n$

$$\begin{cases} \boldsymbol{\mu}_n = \mathbf{A}\boldsymbol{\mu}_{n-1} + \mathbf{K}_n \underbrace{\left(\mathbf{x}_n - \mathbf{C}\mathbf{A}\boldsymbol{\mu}_{n-1}\right)}_{\text{observable space}} \\ \mathbf{V}_n = \left(\mathbf{I} - \mathbf{K}_n\mathbf{C}\right)\mathbf{P}_{n-1} \end{cases}$$

$\mathbf{V}_n$: unaffected by observations

$(\mathbf{I} - \mathbf{K}_n\mathbf{C})$: make the variance small

$\mathbf{A}\mathbf{V}_{n-1}\mathbf{A}^{\mathrm{T}} + \mathbf{\Gamma}$: make the variance large

where
$$\mathbf{P}_{n-1} = \mathbf{A}\mathbf{V}_{n-1}\mathbf{A}^{\mathrm{T}} + \mathbf{\Gamma}$$

$$\mathbf{K}_n = \mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}}\left(\mathbf{C}\mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}} + \mathbf{\Sigma}\right)^{-1}$$

Example for understanding Kalman gain

$$\mathbf{A} = \mathbf{C} = \mathbf{I},\ \boldsymbol{\Gamma} = \sigma_{\mathbf{z}}^2 \mathbf{I},\ \boldsymbol{\Sigma} = \sigma_{\mathbf{x}}^2 \mathbf{I},\ \mathbf{V}_{n-1} = \sigma_{\mathbf{z}_{n-1}}^2 \mathbf{I}$$

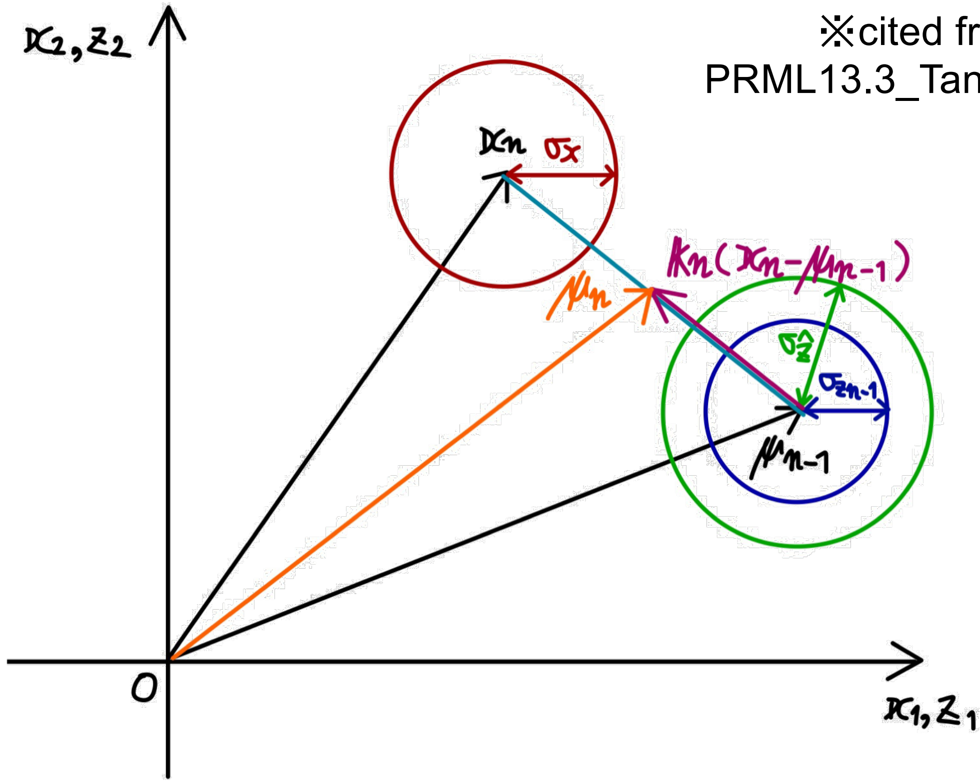
$$\mathbf{P}_{n-1} = \left(\sigma_{\mathbf{z}}^2 + \sigma_{\mathbf{z}_{n-1}}^2\right) \mathbf{I}$$

$$\mathbf{K}_n = \frac{\sigma_{\mathbf{z}}^2 + \sigma_{\mathbf{z}_{n-1}}^2}{\sigma_{\mathbf{z}}^2 + \sigma_{\mathbf{z}_{n-1}}^2 + \sigma_{\mathbf{x}}^2} \mathbf{I}$$

$$\boldsymbol{\mu}_n = \boldsymbol{\mu}_{n-1} + \mathbf{K}_n \left(\mathbf{x}_n - \boldsymbol{\mu}_{n-1}\right)$$

$$\mathbf{V}_n = \frac{\sigma_{\mathbf{x}}^2 \left(\sigma_{\mathbf{z}}^2 + \sigma_{\mathbf{z}_{n-1}}^2\right)}{\sigma_{\mathbf{z}}^2 + \sigma_{\mathbf{z}_{n-1}}^2 + \sigma_{\mathbf{x}}^2} \mathbf{I}$$

※cited from
PRML13.3_Taniyama #9



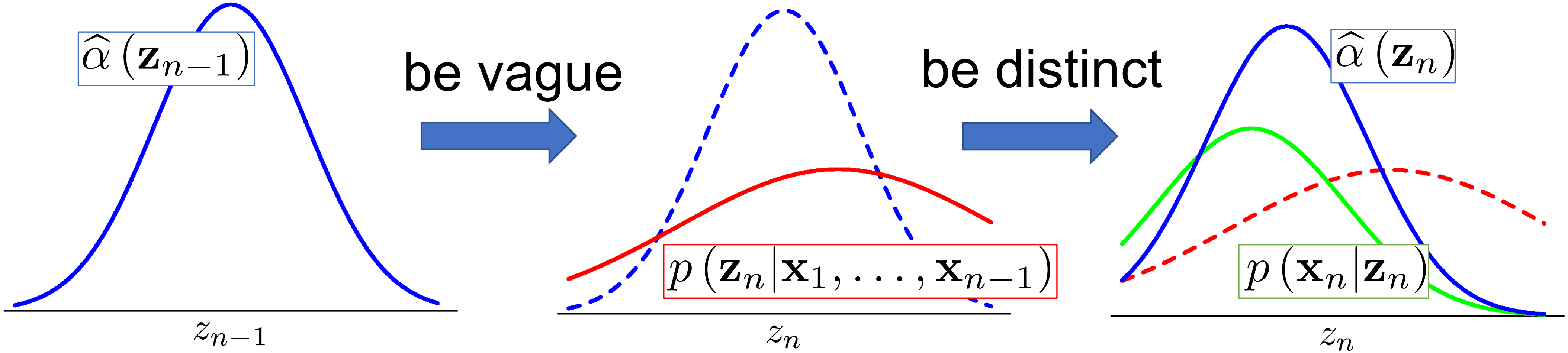
The initial conditions

$$c_1 \widehat{\alpha}\left(\mathbf{z}_1\right) = p\left(\mathbf{z}_1\right) p\left(\mathbf{x}_1 | \mathbf{z}_1\right)$$

$$\begin{aligned}
\boldsymbol{\mu}_1 &= \boldsymbol{\mu}_0 + \mathbf{K}_1\left(\mathbf{x}_1 - \mathbf{C}\boldsymbol{\mu}_0\right) \\
\mathbf{V}_1 &= \left(\mathbf{I} - \mathbf{K}_1\mathbf{C}\right)\mathbf{P}_0 \\
c_1 &= \mathcal{N}\left(\mathbf{x}_1 | \mathbf{C}\boldsymbol{\mu}_0, \mathbf{C}\mathbf{P}_0\mathbf{C}^{\mathrm{T}} + \boldsymbol{\Sigma}\right)
\end{aligned}$$

$$\mathbf{K}_1 = \mathbf{P}_0\mathbf{C}^{\mathrm{T}}\left(\mathbf{C}\mathbf{P}_0\mathbf{C}^{\mathrm{T}} + \boldsymbol{\Sigma}\right)^{-1}$$

Correction by observation



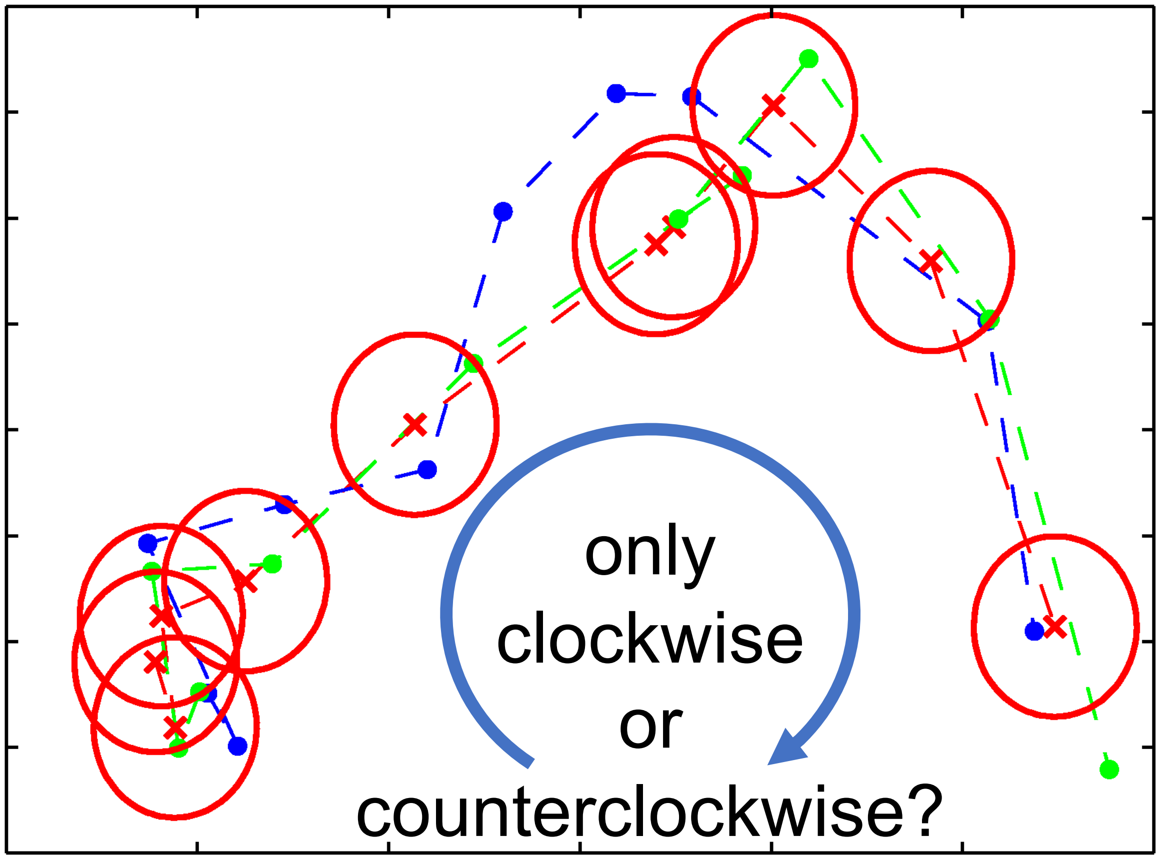
posterior

$$c_n\widehat{\alpha}(\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n)\int \widehat{\alpha}(\mathbf{z}_{n-1})\, p(\mathbf{z}_n|\mathbf{z}_{n-1})\,\mathrm{d}\mathbf{z}_{n-1}$$

$$= p(\mathbf{x}_n|\mathbf{z}_n)\, p(\mathbf{z}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1})$$

likelihood prior

Tracking



true values

observed values

predicted values

the means of posterior

$$\boldsymbol{\mu}_n = \mathbf{A}\boldsymbol{\mu}_{n-1} + \mathbf{K}_n \left(\mathbf{x}_n - \mathbf{C}\mathbf{A}\boldsymbol{\mu}_{n-1} \right)$$

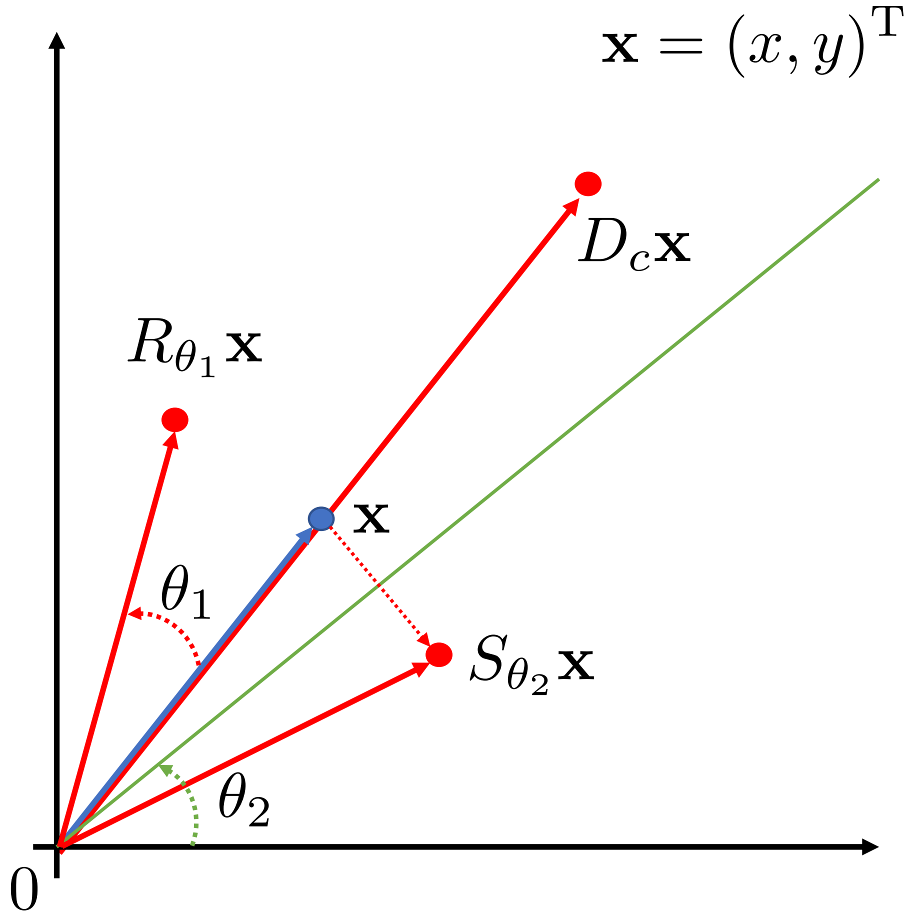
Tracking – linear transformation

- Examples of linear transformation $T:\mathbb{R}^2 \to \mathbb{R}^2$

$$D_c\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} cx \\ cy \end{pmatrix}$$

$$R_\theta\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}$$

$$S_\theta\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos 2\theta + y\sin 2\theta \\ x\sin 2\theta - y\cos 2\theta \end{pmatrix}$$



Tracking – linear transformation

- If we assuming the following two latent variables for the tracking,

  $\mathbf{x}_n = (x_n, y_n)^{\mathrm{T}} \qquad \mathbf{z}_n = (z_{x_n}, z_{y_n})^{\mathrm{T}}$

- This model is restricted on two dimensional linear transformation.

- In practice, introducing additional variables is common.
  - For example, velocity and acceleration.

$$\mathbf{x}_n = (x_n, y_n)^{\mathrm{T}}$$

$$\mathbf{z}_n = (z_{x_n}, z_{y_n}, v_n, a_n)^{\mathrm{T}}$$

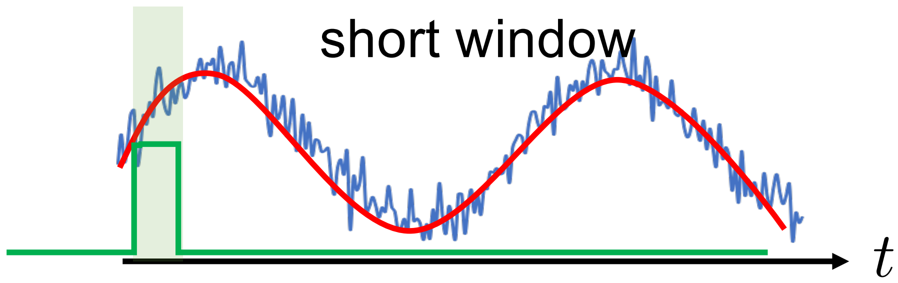
No noise

$$\mathbf{\Sigma} = \mathbf{0}, \ \ \mathbf{C} = \mathbf{I}$$

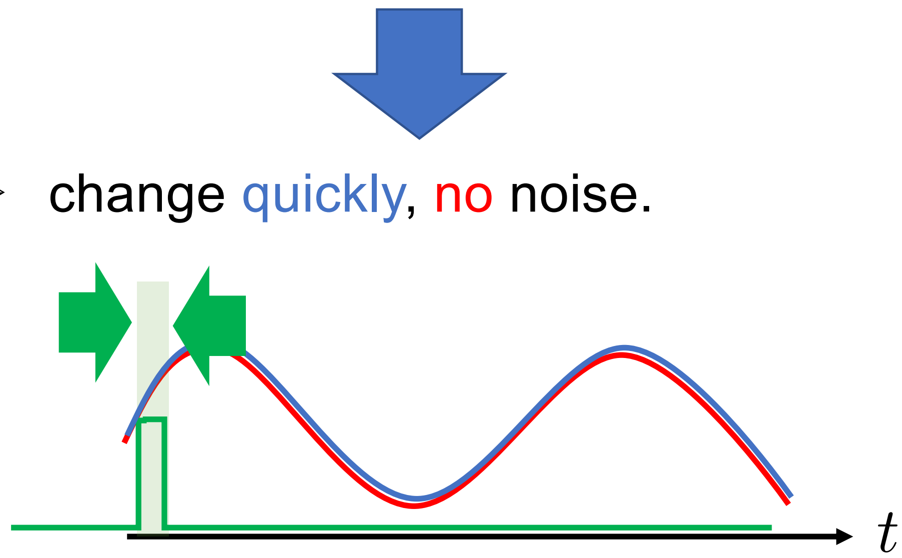
$$\mathbf{K}_n = \mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}}\left(\mathbf{C}\mathbf{P}_{n-1}\mathbf{C}^{\mathrm{T}} + \mathbf{\Sigma}\right)^{-1}$$
$$= \mathbf{I}$$

$$\boldsymbol{\mu}_n = \mathbf{A}\boldsymbol{\mu}_{n-1} + \mathbf{K}_n\left(\mathbf{x}_n - \mathbf{C}\mathbf{A}\boldsymbol{\mu}_{n-1}\right)$$
$$= \mathbf{A}\boldsymbol{\mu}_{n-1} + \mathbf{x}_n - \mathbf{A}\boldsymbol{\mu}_{n-1}$$
$$= \mathbf{x}_n$$

- change quickly, noise level is low.



- change quickly, no noise.



No change

$$\mathbf{\Gamma} = \mathbf{0},\ \mathbf{A} = \mathbf{I},\ \mathbf{C} = \mathbf{I},\ \mathbf{P}_0 \to \infty$$

$$\mathbf{K}_1 = \mathbf{P}_0\mathbf{C}^{\mathrm{T}}\left(\mathbf{\Sigma} + \mathbf{C}\mathbf{P}_0\mathbf{C}^{\mathrm{T}}\right)^{-1} = \mathbf{P}_0\left(\mathbf{\Sigma} + \mathbf{P}_0\right)^{-1}$$

$$\lim_{\mathbf{P}_0 \to \infty} \mathbf{K}_1 = \mathbf{I}$$

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \mathbf{K}_1\left(\mathbf{x}_1 - \mathbf{C}\mathbf{A}\boldsymbol{\mu}_0\right) = \boldsymbol{\mu}_0 + \mathbf{K}_1\left(\mathbf{x}_1 - \boldsymbol{\mu}_0\right)$$

$$\lim_{\mathbf{P}_0 \to \infty} \boldsymbol{\mu}_1 = \mathbf{x}_1$$

$$\mathbf{V}_1 = \left(\mathbf{P}_0^{-1} + \mathbf{C}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{C}\right)^{-1} = \left(\mathbf{P}_0^{-1} + \mathbf{\Sigma}^{-1}\right)^{-1}$$

$$\lim_{\mathbf{P}_0 \to \infty} \mathbf{V}_1 = \mathbf{\Sigma}$$

$$\mathbf{P}_{n-1} = \mathbf{A}\mathbf{V}_{n-1}\mathbf{A}^{\mathrm{T}} + \mathbf{\Gamma}$$

assume that for n

$$\boldsymbol{\mu}_n = \overline{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$$

$$\mathbf{V}_n = \frac{1}{n}\mathbf{\Sigma}$$

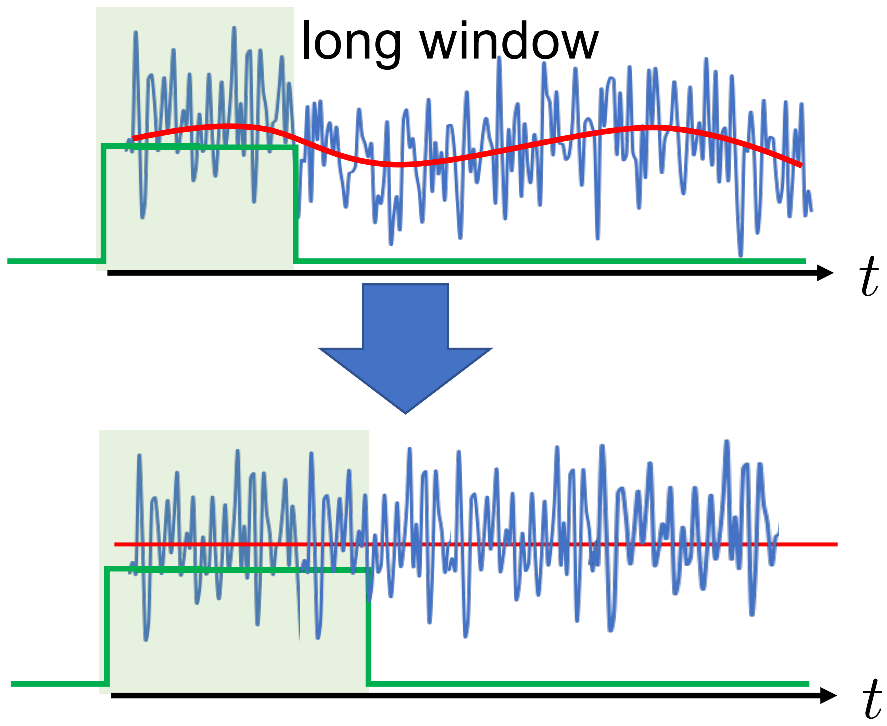
$$\mathbf{K}_n = \frac{1}{n}\mathbf{I}$$

No change

$$\mathbf{K}_{n+1} = \mathbf{P}_n \mathbf{C}^{\mathrm{T}} \left(\mathbf{C} \mathbf{P}_n \mathbf{C}^{\mathrm{T}} + \boldsymbol{\Sigma} \right)^{-1}$$
$$= \frac{1}{n} \boldsymbol{\Sigma} \left(\frac{1}{n} \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \right)^{-1} = \frac{1}{n+1} \mathbf{I}$$

$$\mathbf{V}_{n+1} = \left(\mathbf{I} - \mathbf{K}_{n+1} \mathbf{C} \right) \mathbf{P}_n$$
$$= \left(\mathbf{I} - \frac{1}{n+1} \mathbf{I} \right) \frac{1}{n} \boldsymbol{\Sigma} = \frac{1}{n+1} \boldsymbol{\Sigma}$$

$$\boldsymbol{\mu}_{n+1} = \boldsymbol{\mu}_n + \mathbf{K}_{n+1} \left(\mathbf{x}_{n+1} - \boldsymbol{\mu}_n \right)$$
$$= \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i + \frac{1}{n+1} \left(\mathbf{x}_n - \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \right) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbf{x}_i$$



The backward equations

Given: $\boldsymbol{\theta} = \{\mathbf{A}, \boldsymbol{\Gamma}, \mathbf{C}, \boldsymbol{\Sigma}, \boldsymbol{\mu}_0, \mathbf{P}_0\}, \ \widehat{\alpha}(\mathbf{z}_1), \ldots, \widehat{\alpha}(\mathbf{z}_N)$

Objective: $p(\mathbf{z}_n | \mathbf{x}_1, \ldots, \mathbf{x}_N)$

- $p(\mathbf{z}_n | \mathbf{x}_1, \ldots, \mathbf{x}_N)$ also must be Gaussian.

$$\gamma(\mathbf{z}_n) = \widehat{\alpha}(\mathbf{z}_n)\, \widehat{\beta}(\mathbf{z}_n) = \mathcal{N}\left(\mathbf{z}_n | \widehat{\boldsymbol{\mu}}_n, \widehat{\mathbf{V}}_n\right)$$

- Determine $\widehat{\boldsymbol{\mu}}_n, \widehat{\mathbf{V}}_n$

$$c_{n+1} \widehat{\beta}(\mathbf{z}_n) = \int \widehat{\beta}(\mathbf{z}_{n+1})\, p(\mathbf{x}_{n+1} | \mathbf{z}_{n+1})\, p(\mathbf{z}_{n+1} | \mathbf{z}_n)\, \mathrm{d}\mathbf{z}_{n+1}$$

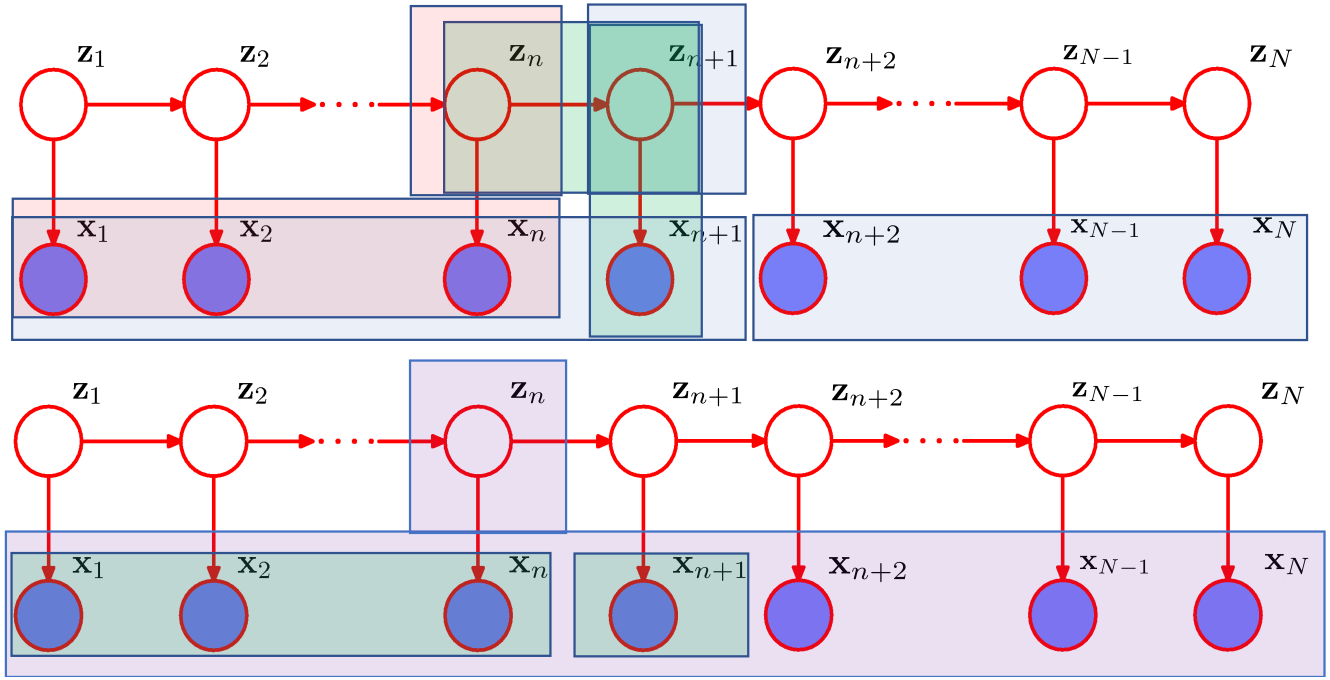
$$\widehat{\alpha}(\mathbf{z}_n) = \mathcal{N}(\mathbf{z}_n | \boldsymbol{\mu}_n, \mathbf{V}_n)$$

$$c_{n+1} \gamma(\mathbf{z}_n) = \widehat{\alpha}(\mathbf{z}_n) \int \widehat{\beta}(\mathbf{z}_{n+1})\, p(\mathbf{x}_{n+1} | \mathbf{z}_{n+1})\, p(\mathbf{z}_{n+1} | \mathbf{z}_n)\, \mathrm{d}\mathbf{z}_{n+1}$$

The backward equations

$$c_{n+1}\gamma(\mathbf{z}_n) = \widehat{\alpha}(\mathbf{z}_n) \int \widehat{\beta}(\mathbf{z}_{n+1})\, p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})\, p(\mathbf{z}_{n+1}|\mathbf{z}_n)\, \mathrm{d}\mathbf{z}_{n+1}$$

$$\widehat{\beta}(\mathbf{z}_{n+1}) = \frac{p(\mathbf{x}_{n+2},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1})}{p(\mathbf{x}_{n+2},\ldots,\mathbf{x}_N|\mathbf{x}_1,\ldots,\mathbf{x}_{n+1})}$$



$$c_{n+1}\gamma(\mathbf{z}_n)$$

The backward equations

$$
\begin{aligned}
c_{n+1}\gamma(\mathbf{z}_n) &= \int \widehat{\beta}(\mathbf{z}_{n+1})\, p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})\, p(\mathbf{z}_{n+1}|\mathbf{z}_n)\, \widehat{\alpha}(\mathbf{z}_n)\, \mathrm{d}\mathbf{z}_{n+1} \\
&= \int \widehat{\beta}(\mathbf{z}_{n+1})\, p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})\, p(\mathbf{z}_{n+1}, \mathbf{z}_n|\mathbf{x}_1, \ldots, \mathbf{x}_n)\, \mathrm{d}\mathbf{z}_{n+1} \\
&= \int \widehat{\beta}(\mathbf{z}_{n+1})\, p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})\, p(\mathbf{z}_{n+1}|\mathbf{x}_1, \ldots, \mathbf{x}_n)\, p(\mathbf{z}_n|\mathbf{z}_{n+1}, \mathbf{x}_1, \ldots, \mathbf{x}_n)\, \mathrm{d}\mathbf{z}_{n+1} \\
&= \int \widehat{\beta}(\mathbf{z}_{n+1})\, c_{n+1}\widehat{\alpha}(\mathbf{z}_{n+1})\, p(\mathbf{z}_n|\mathbf{z}_{n+1}, \mathbf{x}_1, \ldots, \mathbf{x}_n)\, \mathrm{d}\mathbf{z}_{n+1} \\
&= c_{n+1} \int \gamma(\mathbf{z}_{n+1})\, p(\mathbf{z}_n|\mathbf{z}_{n+1}, \mathbf{x}_1, \ldots, \mathbf{x}_n)\, \mathrm{d}\mathbf{z}_{n+1}
\end{aligned}
$$

$$
p(\mathbf{z}_n|\mathbf{z}_{n+1}, \mathbf{x}_1, \ldots, \mathbf{x}_n) = \mathcal{N}(\mathbf{z}_n|\mathbf{m}_n, \mathbf{M}_n)
$$

- Determine $\mathbf{m}_n, \mathbf{M}_n$

The backward equations

$$\mathbf{m}_n = \mathbf{M}_n \left(\mathbf{A}^{\mathrm{T}}\mathbf{\Gamma}^{-1}\mathbf{z}_{n+1} + \mathbf{V}_n^{-1}\boldsymbol{\mu}_n\right)$$

$$\begin{aligned}
\mathbf{M}_n &= \left(\mathbf{A}^{\mathrm{T}}\mathbf{\Gamma}^{-1}\mathbf{A} + \mathbf{V}_n\right)^{-1} \\
&= \mathbf{V}_n - \mathbf{V}_n\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{V}_{n-1}\mathbf{A}^{\mathrm{T}} + \mathbf{\Gamma}\right)^{-1}\mathbf{A}\mathbf{V}_n \\
&= \mathbf{V}_n - \mathbf{V}_n\mathbf{A}^{\mathrm{T}}\mathbf{P}_n^{-1}\mathbf{A}\mathbf{V}_n \\
&= \left(\mathbf{I} - \mathbf{J}_n\mathbf{A}\right)\mathbf{V}_n \qquad \text{where} \qquad \mathbf{J}_n = \mathbf{V}_n\mathbf{A}^{\mathrm{T}}\mathbf{P}_n^{-1}
\end{aligned}$$

$$\mathcal{N}\left(\mathbf{z}_n|\widehat{\boldsymbol{\mu}}_n, \widehat{\mathbf{V}}_n\right) = \int \mathcal{N}\left(\mathbf{z}_{n+1}|\widehat{\boldsymbol{\mu}}_{n+1}, \widehat{\mathbf{V}}_{n+1}\right)\mathcal{N}\left(\mathbf{z}_n|\mathbf{m}_n, \mathbf{M}_n\right)\mathrm{d}\mathbf{z}_{n+1}$$

- Determine $\widehat{\boldsymbol{\mu}}_n, \widehat{\mathbf{V}}_n$

Kalman Smoother equations

$$
\begin{aligned}
\hat{\boldsymbol{\mu}}_n &= \mathbf{M}_n \left(\mathbf{A}^{\mathrm{T}} \boldsymbol{\Gamma}^{-1} \widehat{\boldsymbol{\mu}}_{n+1} + \mathbf{V}_n^{-1} \boldsymbol{\mu}_n \right) \\
&= \mathbf{J}_n \widehat{\boldsymbol{\mu}}_{n+1} + \left(\mathbf{I} - \mathbf{J}_n \mathbf{A} \right) \mathbf{V}_n \mathbf{V}_n^{-1} \boldsymbol{\mu}_n \\
&= \boldsymbol{\mu}_n + \mathbf{J}_n \left(\widehat{\boldsymbol{\mu}}_{n+1} - \mathbf{A} \boldsymbol{\mu}_n \right)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\mathbf{V}}_n &= \mathbf{M}_n \mathbf{A}^{\mathrm{T}} \boldsymbol{\Gamma}^{-1} \widehat{\mathbf{V}}_{n+1} \boldsymbol{\Gamma}^{-1} \mathbf{A} \mathbf{M}_n + \mathbf{M}_n \\
&= \mathbf{V}_n + \mathbf{J}_n \left(\widehat{\mathbf{V}}_{n+1} - \mathbf{P}_n \right) \mathbf{J}_n^{\mathrm{T}}
\end{aligned}
$$

where $\qquad \mathbf{J}_n = \mathbf{V}_n \mathbf{A}^{\mathrm{T}} \mathbf{P}_n^{-1}$

The pairwise posterior marginals

- For the EM algorithm.

$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = (c_n)^{-1} \widehat{\alpha}(\mathbf{z}_{n-1})\, p(\mathbf{x}_n|\mathbf{z}_n)\, p(\mathbf{z}_n|\mathbf{z}_{-1})\, \widehat{\beta}(\mathbf{z}_n)$$

$$= \frac{\mathcal{N}(\mathbf{z}_{n-1}|\boldsymbol{\mu}_{n-1}, \mathbf{V}_{n-1})\, \mathcal{N}(\mathbf{z}_n|\mathbf{A}\mathbf{z}_{n-1}, \boldsymbol{\Gamma})\, \mathcal{N}(\mathbf{x}_n|\mathbf{C}\mathbf{z}_n, \boldsymbol{\Sigma})\, \mathcal{N}\left(\mathbf{z}_n|\widehat{\boldsymbol{\mu}}_n, \widehat{\mathbf{V}}_n\right)}{c_n \widehat{\alpha}(\mathbf{z}_n)}$$

$$\begin{pmatrix} \mathrm{E}[\mathbf{z}_{n-1}] \\ \mathrm{E}[\mathbf{z}_n] \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\mu}}_{n-1} \\ \hat{\boldsymbol{\mu}}_n \end{pmatrix}$$

$$\begin{pmatrix} \mathrm{cov}[\mathbf{z}_{n-1}, \mathbf{z}_{n-1}] & \mathrm{cov}[\mathbf{z}_{n-1}, \mathbf{z}_n] \\ \mathrm{cov}[\mathbf{z}_n, \mathbf{z}_{n-1}] & \mathrm{cov}[\mathbf{z}_n, \mathbf{z}_n] \end{pmatrix} = \begin{pmatrix} \widehat{\mathbf{V}}_{n-1} & \mathbf{J}_{n-1}\widehat{\mathbf{V}}_n \\ \widehat{\mathbf{V}}_n \mathbf{J}_{n-1}^{\mathrm{T}} & \widehat{\mathbf{V}}_n \end{pmatrix}$$

13.3.2 Learning in LDS

Maximum Likelihood

- Determine $\boldsymbol{\theta} = \{\mathbf{A}, \boldsymbol{\Gamma}, \mathbf{C}, \boldsymbol{\Sigma}, \boldsymbol{\mu}_0, \mathbf{P}_0\}$ by using the EM.

- The following expectations are required.

$$\mathbb{E}\left[\mathbf{z}_n\right] = \widehat{\boldsymbol{\mu}}_n$$

$$\mathbb{E}\left[\mathbf{z}_n\mathbf{z}_{n-1}^{\mathrm{T}}\right] = \mathbf{J}_{n-1}\widehat{\mathbf{V}}_n + \widehat{\boldsymbol{\mu}}_n\widehat{\boldsymbol{\mu}}_{n-1}^{\mathrm{T}}$$

$$\mathbb{E}\left[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}\right] = \widehat{\mathbf{V}}_n + \widehat{\boldsymbol{\mu}}_n\widehat{\boldsymbol{\mu}}_n^{\mathrm{T}}$$

The complete-data likelihood

$$\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) = \ln p\left(\mathbf{z}_1|\boldsymbol{\mu}_0, \mathbf{P}_0\right) + \sum_{n=2}^{N} \ln p\left(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}, \boldsymbol{\Gamma}\right) + \sum_{n=1}^{N} \ln p\left(\mathbf{x}_n|\mathbf{z}_n, \mathbf{C}, \boldsymbol{\Sigma}\right)$$

$$Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}\right) = \mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})\right]$$

The M step

$$Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}\right) = -\frac{1}{2}\ln|\mathbf{P}_0| - \mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\frac{1}{2}\left(\mathbf{z}_1 - \boldsymbol{\mu}_0\right)^{\mathrm{T}}\mathbf{P}_0^{-1}\left(\mathbf{z}_1 - \boldsymbol{\mu}_0\right)\right] + \text{const}$$

$$\boldsymbol{\mu}_0^{\text{new}} = \mathbb{E}\left[\mathbf{z}_1\right]$$

$$\mathbf{P}_0^{\text{new}} = \mathbb{E}\left[\mathbf{z}_1\mathbf{z}_1^{\mathrm{T}}\right] - \mathbb{E}\left[\mathbf{z}_1\right]\mathbb{E}\left[\mathbf{z}_1^{\mathrm{T}}\right]$$

The M step

$$Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}\right) = -\frac{N-1}{2} \ln |\boldsymbol{\Gamma}|$$

$$-\mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}} \left[\frac{1}{2} \sum_{n=2}^{N} \left(\mathbf{z}_n - \mathbf{A}\mathbf{z}_{n-1}\right)^{\mathrm{T}} \boldsymbol{\Gamma}^{-1} \left(\mathbf{z}_n - \mathbf{A}\mathbf{z}_{n-1}\right) \right] + \text{ const}$$

$$\mathbf{A}^{\text{ new}} = \left(\sum_{n=2}^{N} \mathbb{E}\left[\mathbf{z}_n \mathbf{z}_{n-1}^{\mathrm{T}}\right] \right) \left(\sum_{n=2}^{N} \mathbb{E}\left[\mathbf{z}_{n-1} \mathbf{z}_{n-1}^{\mathrm{T}}\right] \right)^{-1}$$

$$\boldsymbol{\Gamma}^{\text{ new}} = \frac{1}{N-1} \sum_{n=2}^{N} \left\{ \mathbb{E}\left[\mathbf{z}_n \mathbf{z}_n^{\mathrm{T}}\right] - \mathbf{A}^{\text{ new}}\, \mathbb{E}\left[\mathbf{z}_{n-1} \mathbf{z}_n^{\mathrm{T}}\right] \right.$$

$$\left. - \mathbb{E}\left[\mathbf{z}_n \mathbf{z}_{n-1}^{\mathrm{T}}\right] \mathbf{A}^{\text{ new}} + \mathbf{A}^{\text{new}} \mathbb{E}\left[\mathbf{z}_{n-1} \mathbf{z}_{n-1}^{\mathrm{T}}\right] \left(\mathbf{A}^{\text{ new}}\right)^{\mathrm{T}} \right\}$$

The M step

$$Q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}\right) = -\frac{N}{2}\ln|\boldsymbol{\Sigma}| - \mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}^{\text{old}}}\left[\frac{1}{2}\sum_{n=1}^{N}\left(\mathbf{x}_n - \mathbf{C}\mathbf{z}_n\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}_n - \mathbf{C}\mathbf{z}_n\right)\right] + \text{ const.}$$

$$\mathbf{C}^{\text{new}} = \left(\sum_{n=1}^{N}\mathbf{x}_n\mathbb{E}\left[\mathbf{z}_n^{\mathrm{T}}\right]\right)\left(\sum_{n=1}^{N}\mathbb{E}\left[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}\right]\right)^{-1}$$

$$\boldsymbol{\Sigma}^{\text{new}} = \frac{1}{N}\sum_{n=1}^{N}\left\{\mathbf{x}_n\mathbf{x}_n^{\mathrm{T}} - \mathbf{C}^{\text{new}}\mathbb{E}\left[\mathbf{z}_n\right]\mathbf{x}_n^{\mathrm{T}} - \mathbf{x}_n\mathbb{E}\left[\mathbf{z}_n^{\mathrm{T}}\right]\mathbf{C}^{\text{new}} + \mathbf{C}^{\text{new}}\mathbb{E}\left[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}\right]\mathbf{C}^{\text{new}}\right\}$$

The whole of the algorithm

1. Initialize $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{\Gamma}, \mathbf{C}, \mathbf{\Sigma}, \boldsymbol{\mu}_0, \mathbf{P}_0\}$
2. The E step. (Kalman Filter and Smoother)
3. The M step.
4. Return step 2.