

# ECE276A: Sensing & Estimation in Robotics

## Lecture 9: Kalman Filter

Nikolay Atanasov  
natanasov@ucsd.edu

**UC San Diego**  
**JACOBS SCHOOL OF ENGINEERING**  
Electrical and Computer Engineering

# Bayes Filter

- ▶ **Markov Assumptions:**

- ▶ **Motion model:** given  $\mathbf{x}_t, \mathbf{u}_t$ , the state  $\mathbf{x}_{t+1}$  is independent of the history  $\mathbf{x}_{0:t-1}, \mathbf{z}_{0:t-1}, \mathbf{u}_{0:t-1}$ :

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) \sim p_f(\cdot | \mathbf{x}_t, \mathbf{u}_t)$$

- ▶ **Observation model:** given  $\mathbf{x}_t$ , the observation  $\mathbf{z}_t$  is independent of the history  $\mathbf{x}_{0:t-1}, \mathbf{z}_{0:t-1}, \mathbf{u}_{0:t-1}$ :

$$\mathbf{z}_t = h(\mathbf{x}_t, \mathbf{v}_t) \sim p_h(\cdot | \mathbf{x}_t)$$

- ▶ **Prior:**  $p_{t|t}(\mathbf{x}_t) := p(\mathbf{x}_t | \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1})$

- ▶ **Prediction:**  $p_{t+1|t}(\mathbf{x}) = \int p_f(\mathbf{x} | \mathbf{s}, \mathbf{u}_t) p_{t|t}(\mathbf{s}) d\mathbf{s}$

- ▶ **Update:**  $p_{t+1|t+1}(\mathbf{x}) = \frac{p_h(\mathbf{z}_{t+1} | \mathbf{x}) p_{t+1|t}(\mathbf{x})}{\int p_h(\mathbf{z}_{t+1} | \mathbf{s}) p_{t+1|t}(\mathbf{s}) d\mathbf{s}}$

# Kalman Filter

▶ **Kalman filter:** Bayes filter with additional **assumptions:**

- ▶ The prior pdf  $p_{t|t}$  is Gaussian
- ▶ The motion model is linear in the state  $\mathbf{x}_t$  with Gaussian noise  $\mathbf{w}_t$
- ▶ The observation model is linear in the state  $\mathbf{x}_t$  with Gaussian noise  $\mathbf{v}_t$
- ▶ The motion noise  $\mathbf{w}_t$  and observation noise  $\mathbf{v}_t$  are independent of each other, of the state  $\mathbf{x}_t$ , and across time

▶ **Prior:**  $\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim \mathcal{N}(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t})$

▶ **Motion Model:**

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) = F\mathbf{x}_t + G\mathbf{u}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, W)$$

$$\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_t \sim \mathcal{N}(F\mathbf{x}_t + G\mathbf{u}_t, W), \quad F \in \mathbb{R}^{d_x \times d_x}, G \in \mathbb{R}^{d_x \times d_u}, W \in \mathbb{R}^{d_x \times d_x}$$

▶ **Observation Model:**

$$\mathbf{z}_t = h(\mathbf{x}_t, \mathbf{v}_t) = H\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(0, V)$$

$$\mathbf{z}_t \mid \mathbf{x}_t \sim \mathcal{N}(H\mathbf{x}_t, V), \quad H \in \mathbb{R}^{d_z \times d_x}, V \in \mathbb{R}^{d_z \times d_z}$$

## Matrix Manipulation Lemmas

- ▶ The following results are necessary for deriving the Kalman filter:

- ▶ **Matrix inversion lemma:**

$$(A + BDC)^{-1} = A^{-1} - A^{-1}B \left( D^{-1} + CA^{-1}B \right)^{-1} CA^{-1}$$

- ▶ **Matrix block inversion:**

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & \underbrace{D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}}_{(D - CA^{-1}B)^{-1}} \end{bmatrix}$$

- ▶ **Schur complement:**  $\det \left( \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right) = \det(A) \det(S_A) = \det(D) \det(S_D)$

- ▶ The Schur complement of block  $A$  is  $S_A = D - CA^{-1}B$
- ▶ The Schur complement of block  $D$  is  $S_D = A - BD^{-1}C$

- ▶ **Square completion:** for quadratic form with invertible  $A$ :

$$\frac{1}{2} \mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x} + \mathbf{c} = \frac{1}{2} \left( \mathbf{x} + A^{-1} \mathbf{b} \right)^T A \left( \mathbf{x} + A^{-1} \mathbf{b} \right) - \frac{1}{2} \mathbf{b}^T A^{-1} \mathbf{b} + \mathbf{c}$$

# Gaussian Distribution

▶ **Gaussian random vector**  $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$

▶ parameters:

▶ **mean**  $\boldsymbol{\mu} \in \mathbb{R}^d$

▶ **covariance**  $\Sigma \in \mathbb{S}_{>0}^d$  (symmetric positive definite matrix)

▶ pdf:  $\phi(\mathbf{x}; \boldsymbol{\mu}, \Sigma) := \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$

▶ expectation:  $\mathbb{E}[X] = \int \mathbf{x} \phi(\mathbf{x}; \boldsymbol{\mu}, \Sigma) d\mathbf{x} = \boldsymbol{\mu}$

▶ (co)variance:  $\text{Var}[X] = \text{Cov}[X, X] = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top\right] = \Sigma$

# Information Form of the Gaussian Distribution

- ▶ Alternative Gaussian distribution parametrization:  $X \sim \mathcal{G}(\boldsymbol{\nu}, \Omega)$ 
  - ▶ parameters:
    - ▶ **information mean**  $\boldsymbol{\nu} := \Sigma^{-1}\boldsymbol{\mu} \in \mathbb{R}^d$
    - ▶ **information matrix**:  $\Omega := \Sigma^{-1} \in \mathbb{S}_{\succ 0}^d$
  - ▶ pdf: obtained using square completion:

$$\begin{aligned}\phi(\mathbf{x}; \boldsymbol{\nu}, \Omega) &= \sqrt{\frac{\det(\Omega)}{(2\pi)^d}} \exp\left(-\frac{1}{2} \left(\mathbf{x}^\top \Omega \mathbf{x} - 2\boldsymbol{\nu}^\top \mathbf{x} + \boldsymbol{\nu}^\top \Omega^{-1} \boldsymbol{\nu}\right)\right) \\ &\propto \exp\left(-\frac{1}{2} \mathbf{x}^\top \Omega \mathbf{x} + \boldsymbol{\nu}^\top \mathbf{x}\right)\end{aligned}$$

- ▶ expectation:  $\mathbb{E}[X] = \Omega^{-1}\boldsymbol{\nu}$
- ▶ (co)variance:  $\text{Var}[X] = \Omega^{-1}$

## Covariance and Information Matrix Relationship

▶ Let  $\Omega = \begin{bmatrix} \Omega_{AA} & \Omega_{AB} \\ \Omega_{AB}^\top & \Omega_{BB} \end{bmatrix} = \Sigma^{-1} = \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{AB}^\top & \Sigma_{BB} \end{bmatrix}^{-1}$

- ▶ The blocks of  $\Omega$  and  $\Sigma$  are related by the matrix block inversion lemma:

$$\Omega_{AA} = (\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{AB}^\top)^{-1}$$

$$\Omega_{AB} = -\Omega_{AA}\Sigma_{AB}\Sigma_{BB}^{-1}$$

$$\begin{aligned} \Omega_{BB} &= \Sigma_{BB}^{-1} + \Sigma_{BB}^{-1}\Sigma_{AB}^\top\Omega_{AA}\Sigma_{AB}\Sigma_{BB}^{-1} \\ &= (\Sigma_{BB} - \Sigma_{AB}^\top\Sigma_{AA}^{-1}\Sigma_{AB})^{-1} \end{aligned}$$

- ▶ The determinants of  $\Omega$  and  $\Sigma$  are related via Schur complementation:

$$\det(\Sigma) = \det(\Sigma_{AA}) \det(\Omega_{BB}^{-1}) = \det(\Sigma_{BB}) \det(\Omega_{AA}^{-1})$$

$$\det(\Omega) = \det(\Omega_{AA}) \det(\Sigma_{BB}^{-1}) = \det(\Omega_{BB}) \det(\Sigma_{AA}^{-1})$$

## Gaussian Marginals and Conditionals

- ▶ Consider random vectors  $\mathbf{x}_A$  and  $\mathbf{x}_B$  with **joint Gaussian distribution**:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_A \\ \mathbf{x}_B \end{pmatrix} \sim \mathcal{N} \left( \underbrace{\begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}}_{\boldsymbol{\mu}}, \underbrace{\begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{AB}^\top & \Sigma_{BB} \end{bmatrix}}_{\Sigma} \right)$$

- ▶ The **marginal distributions** are also Gaussian:

$$\mathbf{x}_A \sim \mathcal{N}(\boldsymbol{\mu}_A, \Sigma_{AA}) \quad \mathbf{x}_B \sim \mathcal{N}(\boldsymbol{\mu}_B, \Sigma_{BB})$$

- ▶ The **conditional distributions** are also Gaussian:

$$\mathbf{x}_B \mid \mathbf{x}_A \sim \mathcal{N} \left( \boldsymbol{\mu}_B + \Sigma_{AB}^\top \Sigma_{AA}^{-1} (\mathbf{x}_A - \boldsymbol{\mu}_A), \underbrace{\Sigma_{BB} - \Sigma_{AB}^\top \Sigma_{AA}^{-1} \Sigma_{AB}}_{\text{Schur complement of } \Sigma_{AA}} \right)$$
$$\mathbf{x}_A \mid \mathbf{x}_B \sim \mathcal{N} \left( \boldsymbol{\mu}_A + \Sigma_{AB} \Sigma_{BB}^{-1} (\mathbf{x}_B - \boldsymbol{\mu}_B), \underbrace{\Sigma_{AA} - \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{AB}^\top}_{\text{Schur complement of } \Sigma_{BB}} \right)$$



## Proof: Gaussian Marginal is Gaussian

▶ Let  $\tilde{\mathbf{x}}_A := \mathbf{x}_A - \boldsymbol{\mu}_A \in \mathbb{R}^n$  and  $\tilde{\mathbf{x}}_B := \mathbf{x}_B - \boldsymbol{\mu}_B \in \mathbb{R}^m$  and consider:

$$\begin{aligned}
 p(\mathbf{x}_A) &= \int \phi \left( \begin{pmatrix} \mathbf{x}_A \\ \mathbf{x}_B \end{pmatrix}; \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{AB}^\top & \Sigma_{BB} \end{bmatrix} \right) d\mathbf{x}_B \\
 &= \frac{1}{(2\pi)^{\frac{n+m}{2}} \det(\Sigma)^{1/2}} \int \exp \left( -\frac{1}{2} (\tilde{\mathbf{x}}_A^\top \Omega_{AA} \tilde{\mathbf{x}}_A + 2\tilde{\mathbf{x}}_A^\top \Omega_{AB} \tilde{\mathbf{x}}_B + \tilde{\mathbf{x}}_B^\top \Omega_{BB} \tilde{\mathbf{x}}_B) \right) d\tilde{\mathbf{x}}_B \\
 &\stackrel{\text{square}}{\stackrel{\text{completion}}{=}} \frac{1}{(2\pi)^{\frac{n+m}{2}} \det(\Sigma)^{1/2}} \int \exp \left( -\frac{1}{2} [(\tilde{\mathbf{x}}_B + \Omega_{BB}^{-1} \Omega_{AB}^\top \tilde{\mathbf{x}}_A)^\top \Omega_{BB} (\tilde{\mathbf{x}}_B + \Omega_{BB}^{-1} \Omega_{AB}^\top \tilde{\mathbf{x}}_A) \right. \\
 &\quad \left. - \tilde{\mathbf{x}}_A^\top \Omega_{AB} \Omega_{BB}^{-1} \Omega_{AB}^\top \tilde{\mathbf{x}}_A + \tilde{\mathbf{x}}_A^\top \Omega_{AA} \tilde{\mathbf{x}}_A] \right) d\tilde{\mathbf{x}}_B
 \end{aligned}$$

▶ Note that:

- ▶  $\Omega_{AA} - \Omega_{AB} \Omega_{BB}^{-1} \Omega_{AB}^\top = \Sigma_{AA}^{-1}$
- ▶  $\det(\Sigma) = \det(\Sigma_{AA}) \det(\Omega_{BB}^{-1})$
- ▶  $\int \phi(\tilde{\mathbf{x}}_B; -\Omega_{BB}^{-1} \Omega_{AB}^\top \tilde{\mathbf{x}}_A, \Omega_{BB}^{-1}) d\tilde{\mathbf{x}}_B = 1$

$$\begin{aligned}
 p(\mathbf{x}_A) &= \frac{1}{(2\pi)^{\frac{n}{2}} \det(\Sigma_{AA})^{1/2}} \exp \left( -\frac{1}{2} \tilde{\mathbf{x}}_A^\top \Sigma_{AA}^{-1} \tilde{\mathbf{x}}_A \right) \int \phi(\tilde{\mathbf{x}}_B; -\Omega_{BB}^{-1} \Omega_{AB}^\top \tilde{\mathbf{x}}_A, \Omega_{BB}^{-1}) d\tilde{\mathbf{x}}_B \\
 &\Rightarrow \mathbf{x}_A \sim \mathcal{N}(\boldsymbol{\mu}_A, \Sigma_{AA})
 \end{aligned}$$

## Proof: Gaussian Conditional is Gaussian

$$\begin{aligned}\rho(\mathbf{x}_A | \mathbf{x}_B) &= \frac{\rho(\mathbf{x}_A, \mathbf{x}_B)}{\rho(\mathbf{x}_B)} = \frac{\frac{1}{\sqrt{(2\pi)^{n+m} \det(\Sigma)}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}}{\frac{1}{\sqrt{(2\pi)^m \det(\Sigma_{BB})}} e^{-\frac{1}{2}(\mathbf{x}_B-\boldsymbol{\mu}_B)^\top \Sigma_{BB}^{-1}(\mathbf{x}_B-\boldsymbol{\mu}_B)}} \\ &= \frac{1}{\sqrt{(2\pi)^n \det(\Sigma) / \det(\Sigma_{BB})}} e^{-\frac{1}{2}((\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) - (\mathbf{x}_B-\boldsymbol{\mu}_B)^\top \Sigma_{BB}^{-1}(\mathbf{x}_B-\boldsymbol{\mu}_B))}\end{aligned}$$

► Consider the exponent:

$$\begin{aligned}&\tilde{\mathbf{x}}_A^\top \Omega_{AA} \tilde{\mathbf{x}}_A + \tilde{\mathbf{x}}_A^\top \Omega_{AB} \tilde{\mathbf{x}}_B + \tilde{\mathbf{x}}_B^\top \Omega_{AB}^\top \tilde{\mathbf{x}}_A + \tilde{\mathbf{x}}_B^\top \Omega_{BB} \tilde{\mathbf{x}}_B - \tilde{\mathbf{x}}_B^\top \Sigma_{BB}^{-1} \tilde{\mathbf{x}}_B \\ &= \tilde{\mathbf{x}}_A^\top \Omega_{AA} \tilde{\mathbf{x}}_A - 2\tilde{\mathbf{x}}_A^\top \Omega_{AA} \Sigma_{AB} \Sigma_{BB}^{-1} \tilde{\mathbf{x}}_B + \tilde{\mathbf{x}}_B^\top \Sigma_{BB}^{-1} \Sigma_{AB}^\top \Omega_{AA} \Sigma_{AB} \Sigma_{BB}^{-1} \tilde{\mathbf{x}}_B \\ &= (\tilde{\mathbf{x}}_A - \Sigma_{AB} \Sigma_{BB}^{-1} \tilde{\mathbf{x}}_B)^\top \Omega_{AA} (\tilde{\mathbf{x}}_A - \Sigma_{AB} \Sigma_{BB}^{-1} \tilde{\mathbf{x}}_B) \\ &\Rightarrow \mathbf{x}_A | \mathbf{x}_B \sim \mathcal{N}(\boldsymbol{\mu}_A + \Sigma_{AB} \Sigma_{BB}^{-1} (\mathbf{x}_B - \boldsymbol{\mu}_B), \Sigma_{AA} - \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{AB}^\top)\end{aligned}$$

## The Gaussian Distribution is Stable

- ▶ **Stable distribution:** a linear combination  $aX_1 + bX_2$  of two independent copies of a random variable  $X$  has the same distribution as  $cX + d$  up to location  $d$  and scale  $c > 0$  parameters
- ▶ The Gaussian distribution is **stable**
- ▶ The space of Gaussian pdfs is **closed under convolution:**

$$\int \phi(\mathbf{x}; F\mathbf{s}, W)\phi(\mathbf{s}; \boldsymbol{\mu}, \Sigma) d\mathbf{s} = \phi(\mathbf{x}; F\boldsymbol{\mu}, F\Sigma F^\top + W)$$

- ▶ The space of Gaussian pdfs is **closed under geometric averages** (up to scaling):

$$\prod_k \phi^{\alpha_k}(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k) \propto \phi\left(\mathbf{x}; \left(\sum_k \alpha_k \Sigma_k^{-1}\right)^{-1} \left(\sum_k \Sigma_k^{-1} \boldsymbol{\mu}_k\right), \left(\sum_k \alpha_k \Sigma_k^{-1}\right)^{-1}\right)$$

# Kalman Filter

- ▶ **Kalman filter:** Bayes filter with additional **assumptions**:
  - ▶ The prior pdf  $p_{t|t}$  is Gaussian
  - ▶ The motion model is linear in the state  $\mathbf{x}_t$  with Gaussian noise  $\mathbf{w}_t$
  - ▶ The observation model is linear in the state  $\mathbf{x}_t$  with Gaussian noise  $\mathbf{v}_t$
  - ▶ The motion noise  $\mathbf{w}_t$  and observation noise  $\mathbf{v}_t$  are independent of each other, of the state  $\mathbf{x}_t$ , and across time

- ▶ **Prior:**  $\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim \mathcal{N}(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t})$

- ▶ **Motion Model:**

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) := F\mathbf{x}_t + G\mathbf{u}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, W)$$

$$\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_t \sim \mathcal{N}(F\mathbf{x}_t + G\mathbf{u}_t, W), \quad F \in \mathbb{R}^{d_x \times d_x}, G \in \mathbb{R}^{d_x \times d_u}, W \in \mathbb{R}^{d_x \times d_x}$$

- ▶ **Observation Model:**

$$\mathbf{z}_t = h(\mathbf{x}_t, \mathbf{v}_t) := H\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(0, V)$$

$$\mathbf{z}_t \mid \mathbf{x}_t \sim \mathcal{N}(H\mathbf{x}_t, V), \quad H \in \mathbb{R}^{d_z \times d_x}, V \in \mathbb{R}^{d_z \times d_z}$$

## Kalman Filter Prediction

- ▶ Motion model with given prior:

$$\mathbf{x}_{t+1} = F\mathbf{x}_t + G\mathbf{u}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, W), \quad \mathbf{x}_t \sim \mathcal{N}(\boldsymbol{\mu}_{t|t}, \Sigma_{t|t})$$

$$\begin{aligned} p_{t+1|t}(\mathbf{x}) &= \int p_f(\mathbf{x} | \mathbf{s}, \mathbf{u}_t) p_{t|t}(\mathbf{s}) d\mathbf{s} = \int \phi(\mathbf{x}; F\mathbf{s} + G\mathbf{u}_t, W) \phi(\mathbf{s}; \boldsymbol{\mu}_{t|t}, \Sigma_{t|t}) d\mathbf{s} \\ &= \underbrace{\frac{1}{(2\pi)^{d_x} \sqrt{\det(W) \det(\Sigma_{t|t})}}}_{\kappa_{t|t}} \int \exp \left\{ -\frac{1}{2} (\mathbf{x} - F\mathbf{s} - G\mathbf{u}_t)^\top W^{-1} (\mathbf{x} - F\mathbf{s} - G\mathbf{u}_t) \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} (\mathbf{s} - \boldsymbol{\mu}_{t|t})^\top \Sigma_{t|t}^{-1} (\mathbf{s} - \boldsymbol{\mu}_{t|t}) \right\} d\mathbf{s} \\ &= \kappa_{t|t} \int \exp \left\{ -\frac{1}{2} \left( \mathbf{s}^\top (F^\top W^{-1} F + \Sigma_{t|t}^{-1}) \mathbf{s} - 2(\Sigma_{t|t}^{-1} \boldsymbol{\mu}_{t|t} + F^\top W^{-1} (\mathbf{x} - G\mathbf{u}_t))^\top \mathbf{s} + \dots \right) \right\} d\mathbf{s} \\ &\stackrel{\text{sq.comp.}}{\stackrel{\text{inv.lemma}}{=}} \phi(\mathbf{x}; F\boldsymbol{\mu}_{t|t} + G\mathbf{u}_t, F\Sigma_{t|t}F^\top + W) \end{aligned}$$

- ▶ Kalman filter prediction step:

$$\begin{aligned} p_{t+1|t}(\mathbf{x}) &= \int \phi(\mathbf{x}; F\mathbf{s} + G\mathbf{u}_t, W) \phi(\mathbf{s}; \boldsymbol{\mu}_{t|t}, \Sigma_{t|t}) d\mathbf{s} \\ &= \phi(\mathbf{x}; F\boldsymbol{\mu}_{t|t} + G\mathbf{u}_t, F\Sigma_{t|t}F^\top + W) \end{aligned}$$

## Kalman Filter Prediction (Easy Version)

- ▶ Motion model with given prior:

$$\mathbf{x}_{t+1} = F\mathbf{x}_t + G\mathbf{u}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, W), \quad \mathbf{x}_t \sim \mathcal{N}(\boldsymbol{\mu}_{t|t}, \Sigma_{t|t})$$

- ▶ Since  $\mathbf{w}_t$  and  $\mathbf{x}_t$  are independent and the Gaussian distribution is stable, we know that the distribution of  $\mathbf{x}_{t+1}$  is Gaussian:  $\mathcal{N}(\boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t})$
- ▶ We just need to compute the predicted mean  $\boldsymbol{\mu}_{t+1|t}$  and covariance  $\Sigma_{t+1|t}$ :

$$\boldsymbol{\mu}_{t+1|t} = \mathbb{E}[F\mathbf{x}_t + G\mathbf{u}_t + \mathbf{w}_t] = F\mathbb{E}[\mathbf{x}_t] + G\mathbf{u}_t + \mathbb{E}[\mathbf{w}_t] = F\boldsymbol{\mu}_{t|t} + G\mathbf{u}_t$$

$$\begin{aligned}\Sigma_{t+1|t} &= \text{Var}[F\mathbf{x}_t + G\mathbf{u}_t + \mathbf{w}_t] \stackrel{\text{independence}}{=} \text{Var}[F\mathbf{x}_t] + 0 + \text{Var}[\mathbf{w}_t] \\ &= \mathbb{E}\left[F\left(\mathbf{x}_t - \boldsymbol{\mu}_{t|t}\right)\left(\mathbf{x}_t - \boldsymbol{\mu}_{t|t}\right)^\top F^\top\right] + W \\ &= F\Sigma_{t|t}F^\top + W\end{aligned}$$

## Kalman Filter Update

- Observation model with given prior:

$$\mathbf{z}_{t+1} = H\mathbf{x}_{t+1} + \mathbf{v}_{t+1}, \quad \mathbf{v}_{t+1} \sim \mathcal{N}(0, V), \quad \mathbf{x}_{t+1} \sim \mathcal{N}(\boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t})$$

$$\begin{aligned} p_{t+1|t+1}(\mathbf{x}) &= \frac{p_h(\mathbf{z}_{t+1} | \mathbf{x})p_{t+1|t}(\mathbf{x})}{p(\mathbf{z}_{t+1} | \mathbf{z}_{0:t}, \mathbf{u}_{0:t})} = \frac{\phi(\mathbf{z}_{t+1}; H\mathbf{x}, V)\phi(\mathbf{x}; \boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t})}{\int \phi(\mathbf{z}_{t+1}; H\mathbf{s}, V)\phi(\mathbf{s}; \boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t})d\mathbf{s}} \\ &= \frac{\phi(\mathbf{z}_{t+1}; H\mathbf{x}, V)\phi(\mathbf{x}; \boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t})}{\phi(\mathbf{z}_{t+1}; H\boldsymbol{\mu}_{t+1|t}, H\Sigma_{t+1|t}H^\top + V)} \\ &= \frac{\kappa_{t+1}}{\eta_{t+1}} \exp \left\{ -\frac{1}{2}(\mathbf{z}_{t+1} - H\mathbf{x})^\top V^{-1}(\mathbf{z}_{t+1} - H\mathbf{x}) \right\} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{t+1|t})^\top \Sigma_{t+1|t}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{t+1|t}) \right\} \\ &= \frac{\kappa_{t+1}}{\eta_{t+1}} \exp \left\{ -\frac{1}{2} \left( \mathbf{x}^\top (H^\top V^{-1}H + \Sigma_{t+1|t}^{-1})\mathbf{x} + 2(H^\top V^{-1}\mathbf{z}_{t+1} + \Sigma_{t+1|t}^{-1}\boldsymbol{\mu}_{t+1|t})^\top \mathbf{x} + \dots \right) \right\} \\ &\stackrel{\text{sq. comp.}}{=} \phi \left( \mathbf{x}; (H^\top V^{-1}H + \Sigma_{t+1|t}^{-1})^{-1}(H^\top V^{-1}\mathbf{z}_{t+1} + \Sigma_{t+1|t}^{-1}\boldsymbol{\mu}_{t+1|t}), (H^\top V^{-1}H + \Sigma_{t+1|t}^{-1})^{-1} \right) \\ &\stackrel{\text{inversion lemma}}{=} \phi \left( \mathbf{x}; \boldsymbol{\mu}_{t+1|t} + K_{t+1|t}(\mathbf{z}_{t+1} - H\boldsymbol{\mu}_{t+1|t}), (I - K_{t+1|t}H)\Sigma_{t+1|t} \right) \end{aligned}$$

- Kalman gain:  $K_{t+1|t} := \Sigma_{t+1|t}H^\top (H\Sigma_{t+1|t}H^\top + V)^{-1}$

## Kalman Filter Update (Easy Version)

- ▶ Observation model with given prior:

$$\mathbf{z}_{t+1} = H\mathbf{x}_{t+1} + \mathbf{v}_{t+1}, \quad \mathbf{v}_{t+1} \sim \mathcal{N}(0, V), \quad \mathbf{x}_{t+1} \sim \mathcal{N}(\boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t})$$

- ▶ The joint distribution of  $\mathbf{x}_{t+1}$  and  $\mathbf{z}_{t+1}$  is Gaussian:

$$\begin{pmatrix} \mathbf{x}_{t+1} \\ \mathbf{z}_{t+1} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \boldsymbol{\mu}_{t+1|t} \\ H\boldsymbol{\mu}_{t+1|t} \end{pmatrix}, \begin{bmatrix} \Sigma_{t+1|t} & C_{t+1|t} \\ C_{t+1|t}^\top & H\Sigma_{t+1|t}H^\top + V \end{bmatrix} \right)$$

$$\begin{aligned} C_{t+1|t} &= \mathbb{E} \left[ \left( \mathbf{x}_{t+1} - \boldsymbol{\mu}_{t+1|t} \right) \left( \mathbf{z}_{t+1} - H\boldsymbol{\mu}_{t+1|t} \right)^\top \right] \\ &= \mathbb{E} \left[ \left( \mathbf{x}_{t+1} - \boldsymbol{\mu}_{t+1|t} \right) \left( \left( \mathbf{x}_{t+1} - \boldsymbol{\mu}_{t+1|t} \right)^\top H^\top + \mathbf{v}_{t+1}^\top \right) \right] = \Sigma_{t+1|t}H^\top \end{aligned}$$

- ▶ The conditional distribution of  $\mathbf{x}_{t+1} \mid \mathbf{z}_{t+1}$  is then also Gaussian:

$$\begin{aligned} \mathbf{x}_{t+1} \mid \mathbf{z}_{t+1} &\sim \mathcal{N} \left( \boldsymbol{\mu}_{t+1|t} + \Sigma_{t+1|t}H^\top (H\Sigma_{t+1|t}H^\top + V)^{-1} (\mathbf{z}_{t+1} - H\boldsymbol{\mu}_{t+1|t}), \right. \\ &\quad \left. \Sigma_{t+1|t} - \Sigma_{t+1|t}H^\top (H\Sigma_{t+1|t}H^\top + V)^{-1} H\Sigma_{t+1|t} \right) \end{aligned}$$



## Kalman Filter Summary

Motion model:  $\mathbf{x}_{t+1} = F\mathbf{x}_t + G\mathbf{u}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, W)$

Observation model:  $\mathbf{z}_t = H\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(0, V)$

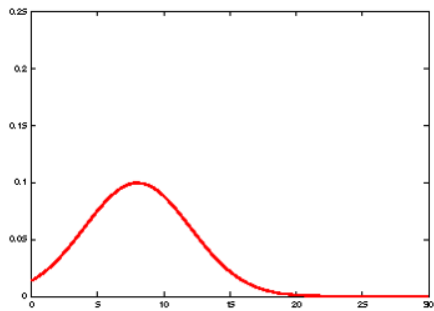
Prior:  $\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim \mathcal{N}(\boldsymbol{\mu}_{t|t}, \Sigma_{t|t})$

Prediction: 
$$\begin{aligned}\boldsymbol{\mu}_{t+1|t} &= F\boldsymbol{\mu}_{t|t} + G\mathbf{u}_t \\ \Sigma_{t+1|t} &= F\Sigma_{t|t}F^\top + W\end{aligned}$$

Update: 
$$\begin{aligned}\boldsymbol{\mu}_{t+1|t+1} &= \boldsymbol{\mu}_{t+1|t} + K_{t+1|t}(\mathbf{z}_{t+1} - H\boldsymbol{\mu}_{t+1|t}) \\ \Sigma_{t+1|t+1} &= (I - K_{t+1|t}H)\Sigma_{t+1|t}\end{aligned}$$

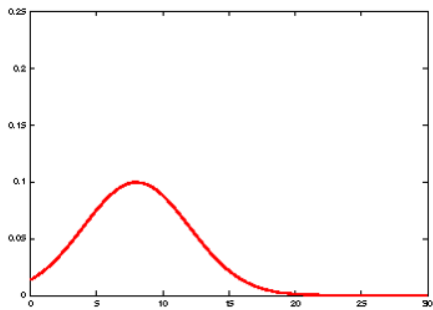
Kalman gain:  $K_{t+1|t} = \Sigma_{t+1|t}H^\top (H\Sigma_{t+1|t}H^\top + V)^{-1}$

# Predict

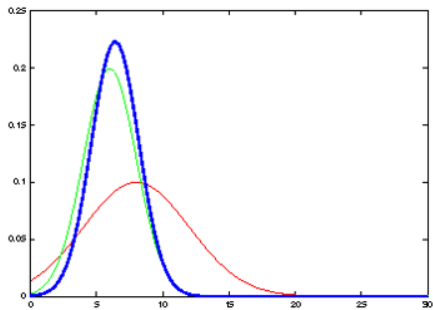


# Update

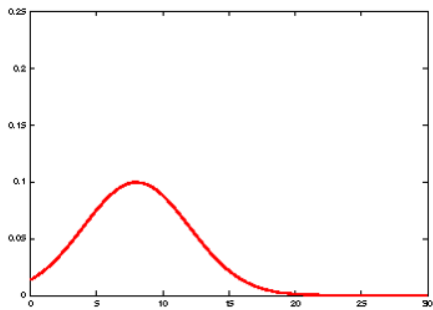
### Predict



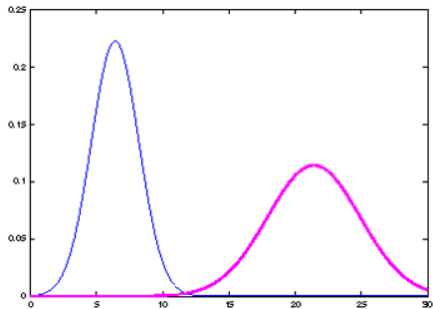
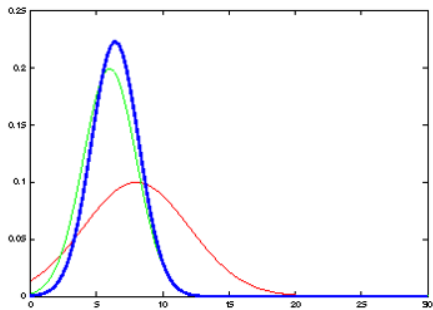
### Update



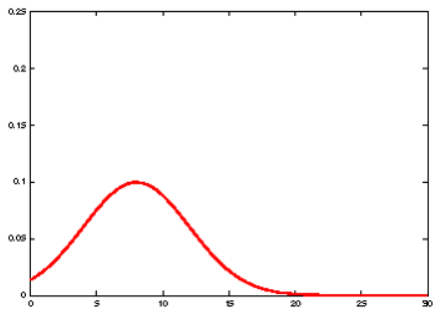
### Predict



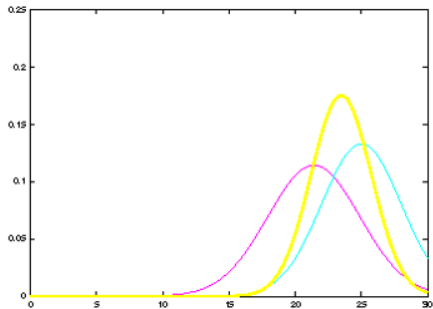
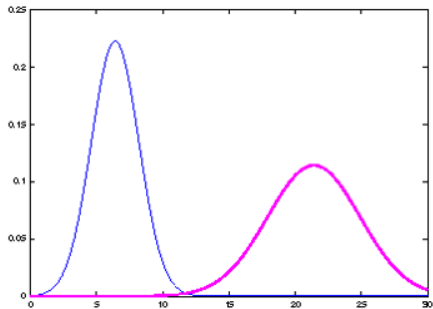
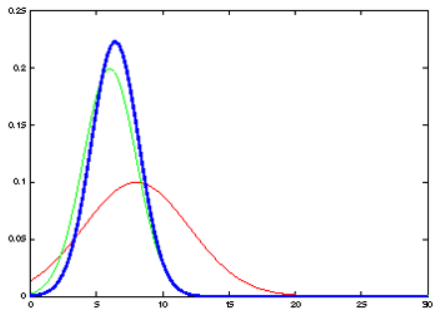
### Update



### Predict



### Update



## Kalman Filter Comments

- ▶ The normalization factor in the update step (Bayes rule) is Gaussian:

$$\begin{aligned} p(\mathbf{z}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t}) &= \int p_h(\mathbf{z}_{t+1} \mid \mathbf{x}_{t+1}) p(\mathbf{x}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t}) d\mathbf{x}_{t+1} \\ &= \int \phi(\mathbf{z}_{t+1}; H\mathbf{x}_{t+1}, V) \phi(\mathbf{x}_{t+1}; \boldsymbol{\mu}_{t+1|t}, \Sigma_{t+1|t}) d\mathbf{x}_{t+1} \\ &= \phi(\mathbf{z}_{t+1}; H\boldsymbol{\mu}_{t+1|t}, H\Sigma_{t+1|t}H^\top + V) \end{aligned}$$

- ▶ The term  $\mathbf{r}_{t+1} = \mathbf{z}_{t+1} - H\boldsymbol{\mu}_{t+1|t}$  used to correct the mean in the update step is called **innovation**
- ▶ The innovation  $\mathbf{r}_{t+1}$  conditioned on the past information  $\mathbf{z}_{0:t}, \mathbf{u}_{0:t}$  has the same covariance as  $p(\mathbf{z}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t})$ :

$$\begin{aligned} \mathbb{E}[\mathbf{r}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t}] &= H\boldsymbol{\mu}_{t+1|t} - H\boldsymbol{\mu}_{t+1|t} = 0 \\ \mathbb{E}[\mathbf{r}_{t+1}\mathbf{r}_{t+1}^\top \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t}] &= H\Sigma_{t+1|t}H^\top + V \end{aligned}$$

- ▶ The **Kalman gain**  $K_{t+1|t}$  scales the innovation  $\mathbf{r}_{t+1}$  by its covariance and determines how much to trust this correction to the mean in the update step

## Kalman Filter Comments

- ▶ The KF has **polynomial complexity** in measurement and state dimensions

$$O(d_z^{2.376} + d_x^2)$$

- ▶ Under the linear, Gaussian, and independence assumptions, the KF is an **optimal filter** with respect to the mean square error (MSE) criterion:

$$\mathbb{E} \left[ \|\mathbf{x}_t - \boldsymbol{\mu}_{t|t}\|_2^2 \right] = \text{tr}(\Sigma_{t|t})$$

- ▶ **Expectation Maximization (EM)** can be used to simultaneously estimate the state distribution  $p(\mathbf{x}_{0:t})$  and learn the motion model parameters  $(F, G, W)$  and the observation model parameters  $(H, V)$
- ▶ Given data  $\mathcal{D} := \{\mathbf{z}_{0:t}, \mathbf{u}_{0:t-1}\}$ , apply EM with hidden variables  $\mathbf{x}_{0:t}$ :
  - ▶ **E step**: Given initial parameter estimates  $\boldsymbol{\theta}^{(k)} = \{F^{(k)}, G^{(k)}, W^{(k)}, H^{(k)}, V^{(k)}\}$  calculate the likelihood of the hidden variables via the Kalman filter/smoother
  - ▶ **M step**: Optimize the parameters via maximum likelihood estimation (MLE) to obtain  $\boldsymbol{\theta}^{(k+1)}$  which explains the posterior distribution of  $\mathbf{x}_{0:t}$  better

## Information Filter

- **Information filter:** converts the Kalman filter equations to information form  $\mathbf{x} \sim \mathcal{G}(\boldsymbol{\nu}, \Omega)$  with  $\boldsymbol{\nu} = \Sigma^{-1}\boldsymbol{\mu}$  and  $\Omega = \Sigma^{-1}$  via the matrix inversion lemma

Motion model:  $\mathbf{x}_{t+1} = F\mathbf{x}_t + G\mathbf{u}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{G}(0, W^{-1})$

Observation model:  $\mathbf{z}_t = H\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{G}(0, V^{-1})$

Prior:  $\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim \mathcal{G}(\boldsymbol{\nu}_{t|t}, \Omega_{t|t})$

Prediction: 
$$\boldsymbol{\nu}_{t+1|t} = \Omega_{t+1|t} \left( F\Omega_{t|t}^{-1}\boldsymbol{\nu}_{t|t} + G\mathbf{u}_t \right)$$
$$\Omega_{t+1|t} = \left( F\Omega_{t|t}^{-1}F^\top + W \right)^{-1}$$

Update: 
$$\boldsymbol{\nu}_{t+1|t+1} = \boldsymbol{\nu}_{t+1|t} + H^\top V^{-1}\mathbf{z}_{t+1}$$
$$\Omega_{t+1|t+1} = \Omega_{t+1|t} + H^\top V^{-1}H$$



## Kalman-Bucy Filter (Continuous-Time KF)

Motion model:  $\dot{\mathbf{x}}(t) = F\mathbf{x}(t) + G\mathbf{u}(t) + \mathbf{w}(t)$

Observation model:  $\mathbf{z}(t) = H\mathbf{x}(t) + \mathbf{v}(t)$

Prior:  $\mathbf{x}(0) \sim \mathcal{N}(\boldsymbol{\mu}(0), \Sigma(0))$

Mean:  $\dot{\boldsymbol{\mu}}(t) = F\boldsymbol{\mu}(t) + G\mathbf{u}(t) + K(t)(\mathbf{z}(t) - H\boldsymbol{\mu}(t))$

Covariance:  $\dot{\Sigma}(t) = F\Sigma(t) + \Sigma(t)F^\top + W - K(t)VK^\top(t)$

Kalman Gain:  $K(t) = \Sigma(t)H^\top V^{-1}$

## Gaussian Mixture Filter

► **Motion model:**  $\mathbf{x}_{t+1} = F\mathbf{x}_t + G\mathbf{u}_t + \mathbf{w}_t$ ,  $\mathbf{w}_t \sim \mathcal{N}(0, W)$

► **Observation model:**  $\mathbf{z}_t = H\mathbf{x}_t + \mathbf{v}_t$ ,  $\mathbf{v}_t \sim \mathcal{N}(0, V)$

► **Prior:**  $\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim p_{t|t}(\mathbf{x}_t) := \sum_k \alpha_{t|t}^{(k)} \phi(\mathbf{x}_t; \boldsymbol{\mu}_{t|t}^{(k)}, \Sigma_{t|t}^{(k)})$

► **Prediction:**

$$\begin{aligned} p_{t+1|t}(\mathbf{x}) &= \int p_f(\mathbf{x} \mid \mathbf{s}, \mathbf{u}_t) p_{t|t}(\mathbf{s}) d\mathbf{s} = \sum_k \alpha_{t|t}^{(k)} \int \phi(\mathbf{x}; F\mathbf{s} + G\mathbf{u}_t, W) \phi(\mathbf{s}; \boldsymbol{\mu}_{t|t}^{(k)}, \Sigma_{t|t}^{(k)}) d\mathbf{s} \\ &= \sum_k \alpha_{t|t}^{(k)} \phi(\mathbf{x}; F\boldsymbol{\mu}_{t|t}^{(k)} + G\mathbf{u}_t, F\Sigma_{t|t}^{(k)}F^\top + W) \end{aligned}$$

► **Update:**

$$\begin{aligned} p_{t+1|t+1}(\mathbf{x}) &= \frac{p_h(\mathbf{z}_{t+1} \mid \mathbf{x}) p_{t+1|t}(\mathbf{x})}{p(\mathbf{z}_{t+1} \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t})} = \frac{\phi(\mathbf{z}_{t+1}; H\mathbf{x}, V) \sum_k \alpha_{t+1|t}^{(k)} \phi(\mathbf{x}; \boldsymbol{\mu}_{t+1|t}^{(k)}, \Sigma_{t+1|t}^{(k)})}{\int \phi(\mathbf{z}_{t+1}; H\mathbf{s}, V) \sum_j \alpha_{t+1|t}^{(j)} \phi(\mathbf{s}; \boldsymbol{\mu}_{t+1|t}^{(j)}, \Sigma_{t+1|t}^{(j)}) d\mathbf{s}} \\ &= \sum_k \left( \frac{\alpha_{t+1|t}^{(k)} \phi(\mathbf{z}_{t+1}; H\mathbf{x}, V) \phi(\mathbf{x}; \boldsymbol{\mu}_{t+1|t}^{(k)}, \Sigma_{t+1|t}^{(k)})}{\sum_j \alpha_{t+1|t}^{(j)} \phi(\mathbf{z}_{t+1}; H\boldsymbol{\mu}_{t+1|t}^{(j)}, H\Sigma_{t+1|t}^{(j)}H^\top + V)} \frac{\phi(\mathbf{z}_{t+1}; H\boldsymbol{\mu}_{t+1|t}^{(k)}, H\Sigma_{t+1|t}^{(k)}H^\top + V)}{\phi(\mathbf{z}_{t+1}; H\boldsymbol{\mu}_{t+1|t}^{(k)}, H\Sigma_{t+1|t}^{(k)}H^\top + V)} \right) \\ &= \sum_k \left[ \frac{\alpha_{t+1|t}^{(k)} \phi(\mathbf{z}_{t+1}; H\boldsymbol{\mu}_{t+1|t}^{(k)}, H\Sigma_{t+1|t}^{(k)}H^\top + V)}{\sum_j \alpha_{t+1|t}^{(j)} \phi(\mathbf{z}_{t+1}; H\boldsymbol{\mu}_{t+1|t}^{(j)}, H\Sigma_{t+1|t}^{(j)}H^\top + V)} \right] \phi(\mathbf{x}; \boldsymbol{\mu}_{t+1|t}^{(k)} + K_{t+1|t}^{(k)}(\mathbf{z}_{t+1} - H\boldsymbol{\mu}_{t+1|t}^{(k)}), (I - K_{t+1|t}^{(k)}H)\Sigma_{t+1|t}^{(k)}) \end{aligned}$$

► **Kalman Gain:**  $K_{t+1|t}^{(k)} := \Sigma_{t+1|t}^{(k)} H^\top (H\Sigma_{t+1|t}^{(k)} H^\top + V)^{-1}$

## Gaussian Mixture Filter

- ▶ The GMF is a **bank of Kalman filters** with weights  $\alpha^{(k)}$  scaled by the observation model in the update step
- ▶ **pdf:**  $\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim p_{t|t}(\mathbf{x}) := \sum_k \alpha_{t|t}^{(k)} \phi(\mathbf{x}_t; \boldsymbol{\mu}_{t|t}^{(k)}, \boldsymbol{\Sigma}_{t|t}^{(k)})$
- ▶ **mean:**  $\boldsymbol{\mu}_{t|t} := \mathbb{E}[\mathbf{x}_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1}] = \int \mathbf{x} p_{t|t}(\mathbf{x}) d\mathbf{x} = \sum_k \alpha_{t|t}^{(k)} \boldsymbol{\mu}_{t|t}^{(k)}$
- ▶ **covariance:**  $\boldsymbol{\Sigma}_{t|t} := \mathbb{E}[\mathbf{x}_t \mathbf{x}_t^\top \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1}] - \boldsymbol{\mu}_{t|t} \boldsymbol{\mu}_{t|t}^\top$   
 $= \int \mathbf{x} \mathbf{x}^\top p_{t|t}(\mathbf{x}) d\mathbf{x} - \boldsymbol{\mu}_{t|t} \boldsymbol{\mu}_{t|t}^\top = \sum_k \alpha_{t|t}^{(k)} \left( \boldsymbol{\Sigma}_{t|t}^{(k)} + \boldsymbol{\mu}_{t|t}^{(k)} (\boldsymbol{\mu}_{t|t}^{(k)})^\top \right) - \boldsymbol{\mu}_{t|t} \boldsymbol{\mu}_{t|t}^\top$

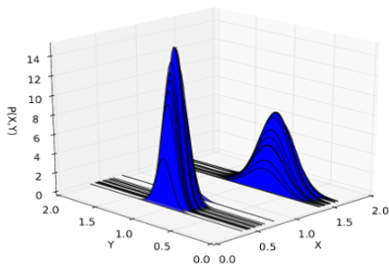
## Rao-Blackwellized Particle Filter

- ▶ The Rao-Blackwellized (**marginalized**) particle filter is applicable to conditionally linear-Gaussian models:

$$\mathbf{x}_{t+1}^n = f_t^n(\mathbf{x}_t^n) + A_t^n(\mathbf{x}_t^n)\mathbf{x}_t^l + G_t^n(\mathbf{x}_t^n)\mathbf{w}_t^n$$

$$\mathbf{x}_{t+1}^l = f_t^l(\mathbf{x}_t^n) + A_t^l(\mathbf{x}_t^n)\mathbf{x}_t^l + G_t^l(\mathbf{x}_t^n)\mathbf{w}_t^l$$

$$\mathbf{z}_t = h_t(\mathbf{x}_t^n) + C_t(\mathbf{x}_t^n)\mathbf{x}_t^l + \mathbf{v}_t$$



Nonlinear states:  $\mathbf{x}_t^n$

Linear states:  $\mathbf{x}_t^l$

- ▶ **RBPF**: exploits linear-Gaussian structure to handle high-dim problems:

$$p(\mathbf{x}_t^l, \mathbf{x}_{0:t}^n | \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1}) = \underbrace{p(\mathbf{x}_t^l | \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1}, \mathbf{x}_{0:t}^n)}_{\text{Kalman Filter}} \underbrace{p(\mathbf{x}_{0:t}^n | \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1})}_{\text{Particle Filter}}$$

$$= \phi(\mathbf{x}_t^l; \boldsymbol{\mu}_{t|t}^{(k)}, \boldsymbol{\Sigma}_{t|t}^{(k)}) \sum_{k=1}^{N_{t|t}} \alpha_{t|t}^{(k)} \delta(\mathbf{x}_{0:t}^n - \mathbf{m}_{t|t}^{(k)})$$