ECE276A: Sensing & Estimation in Robotics
Lecture 13: Kalman Filtering

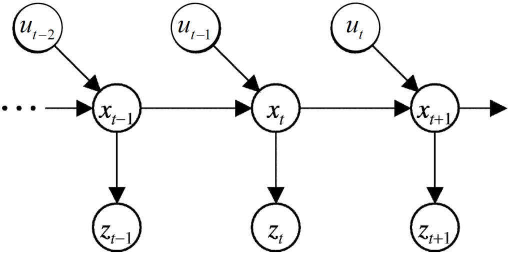
Instructor:
Nikolay Atanasov: natanasov@ucsd.edu

Teaching Assistants:
Qiaojun Feng: qif007@eng.ucsd.edu
Tianyu Wang: tiw161@eng.ucsd.edu
Ibrahim Akbar: iakbar@eng.ucsd.edu
You-Yi Jau: yjau@eng.ucsd.edu
Harshini Rajachander: hrajacha@eng.ucsd.edu

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

Filtering Problem

- The Markov assumptions are used to decompose the joint pdf of the states $x_{0:T}$, observations $z_{0:T}$, and controls $u_{0:T-1}$



- **Joint distribution**:

$$p(x_{0:T}, z_{0:T}, u_{0:T-1}) = \underbrace{p_{0|-1}(x_0)}_{\text{prior}} \prod_{t=0}^{T} \underbrace{p_h(z_t \mid x_t)}_{\text{observation model}} \prod_{t=1}^{T} \underbrace{p_f(x_t \mid x_{t-1}, u_{t-1})}_{\text{motion model}}$$

- **Filtering**: keeps track of

$$p_{t|t}(x_t) := p(x_t \mid z_{0:t}, u_{0:t-1})$$
$$p_{t+1|t}(x_{t+1}) := p(x_{t+1} \mid z_{0:t}, u_{0:t})$$

- **Smoothing**: keeps track of

$$p_{t|t}(x_{0:t}) := p(x_{0:t} \mid z_{0:t}, u_{0:t-1})$$
$$p_{t+1|t}(x_{0:t+1}) := p(x_{0:t+1} \mid z_{0:t}, u_{0:t})$$

Bayes Filter Summary

- **Prior**: $x_t \mid z_{0:t}, u_{0:t-1} \sim p_{t|t}(\cdot)$
- **Motion model**: $x_{t+1} = f(x_t, u_t, w_t) \sim p_f(\cdot \mid x_t, u_t)$
- **Observation model**: $z_t = h(x_t, v_t) \sim p_h(\cdot \mid x_t)$
- **Prediction**: $p_{t+1|t}(x) = \int p_f(x \mid s, u_t) p_{t|t}(s) ds$
- **Update**: $p_{t+1|t+1}(x) = \frac{p_h(z_{t+1}|x) p_{t+1|t}(x)}{\int p_h(z_{t+1}|s) p_{t+1|t}(s) ds}$

Kalman Filter

- A **Kalman filter** is a Bayes filter for which:
  - The prior pdf $p_{0|0}$ is Gaussian
  - The motion model is linear in the state and affected by Gaussian noise
  - The observation model is linear in the state and affected by Gaussian noise
  - The process noise w_t and measurement noise v_t are independent of each other, of the state x_t, and across time

- **Prior**: $x_t \mid z_{0:t}, u_{0:t-1} \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t})$

- **Motion Model**:

$$x_{t+1} = f(x_t, u_t, w_t) := Fx_t + Gu_t + w_t, \quad w_t \sim \mathcal{N}(0, W)$$
$$x_{t+1} \mid x_t, u_t \sim \mathcal{N}(Fx_t + Gu_t, W), \quad F \in \mathbb{R}^{d_x \times d_x}, G \in \mathbb{R}^{d_x \times d_u},\ W \in \mathbb{R}^{d_x \times d_x}$$

- **Observation Model**:

$$z_t = h(x_t, v_t) := Hx_t + v_t, \quad v_t \sim \mathcal{N}(0, V)$$
$$z_t \mid x_t \sim \mathcal{N}(Hx_t, V), \quad H \in \mathbb{R}^{d_z \times d_x},\ V \in \mathbb{R}^{d_z \times d_z}$$

Gaussian Distribution Properties

- **Stable distributon**: 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, i.e., the space of Gaussian pdfs is **closed under convolution**:

$$\int \phi(x; Fs, W)\phi(s; \mu, \Sigma)ds = \phi\left(x; F\mu, F\Sigma F^T + W\right)$$

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

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

Gaussian Marginals and Conditionals

- Consider a **joint Gaussian distribution**:

$$x := \begin{pmatrix} x_A \\ x_B \end{pmatrix} \sim \mathcal{N}\left(\underbrace{\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}}_{\mu}, \underbrace{\begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{AB}^T & \Sigma_{BB} \end{bmatrix}}_{\Sigma} \right)$$

- The **marginal distribution** is also Gaussian:

$$x_A \sim \mathcal{N}(\mu_A, \Sigma_{AA})$$

- The **conditional distribution** is also Gaussian:

$$x_B \mid x_A \sim \mathcal{N}\left(\mu_B + \Sigma_{AB}^T \Sigma_{AA}^{-1} (x_A - \mu_A), \underbrace{\Sigma_{BB} - \Sigma_{AB}^T \Sigma_{AA}^{-1} \Sigma_{AB}}_{\text{Schur complement of } \Sigma_{AA}} \right)$$

Matrix Manipulation

- The main tools for proving the previous results and for deriving the Kalman filter are:
  - **Matrix inversion lemma**:

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

  - **Matrix block inversion**:

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

  - **Square completion**:

$$\frac{1}{2}x^TAx + b^Tx + c = \frac{1}{2}\left(x + A^{-1}b\right)^TA\left(x + A^{-1}b\right) - \frac{1}{2}b^TA^{-1}b + c$$

Information Space

- Natural parameterization of a Gaussian distribution: $x \sim \mathcal{G}(\nu, \Omega)$
- **Information mean**: $\nu = \Sigma^{-1}\mu$
- **Information matrix**: $\Omega := \begin{bmatrix} \Omega_{AA} & \Omega_{AB} \\ \Omega_{AB}^T & \Omega_{BB} \end{bmatrix} = \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{AB}^T & \Sigma_{BB} \end{bmatrix}^{-1}$

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

Gaussian Marginal

- Let $\tilde{x}_A := x_A - \mu_A$ and $\tilde{x}_B := x_B - \mu_B$ and consider:

$$p(x_A) = \int \phi\left(\begin{pmatrix} x_A \\ x_B \end{pmatrix}; \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{AB}^T & \Sigma_{BB} \end{bmatrix}\right) dx_B$$

$$= \frac{1}{(2\pi)^{\frac{n+m}{2}} |\Sigma|^{1/2}} \int \exp\left(-\frac{1}{2}\left(\tilde{x}_A^T \Omega_{AA} \tilde{x}_A + 2\tilde{x}_A^T \Omega_{AB} \tilde{x}_B + \tilde{x}_B^T \Omega_{BB} \tilde{x}_B\right)\right) dx_B$$

$$\overset{Sq.Comp.}{=\!=\!=} \kappa \int \exp\left(-\frac{1}{2}\left[(\tilde{x}_B + \Omega_{BB}^{-1}\Omega_{AB}^T \tilde{x}_A)^T \Omega_{BB} (\tilde{x}_B + \Omega_{BB}^{-1}\Omega_{AB}^T \tilde{x}_A) - \tilde{x}_A^T \Omega_{AB} \Omega_{BB}^{-1} \Omega_{AB}^T \tilde{x}_A + \tilde{x}_A^T \Omega_{AA} \tilde{x}_A\right]\right) dx_B$$

$$\underset{\Sigma_{AA}^{-1} = \Omega_{AA} - \Omega_{AB}\Omega_{BB}^{-1}\Omega_{AB}^T}{\overset{\int \phi(x;\mu,\Sigma)dx = 1}{=\!=\!=\!=\!=}} \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{AA}|^{1/2}} \exp\left(-\frac{1}{2}\tilde{x}_A^T \Sigma_{AA}^{-1} \tilde{x}_A\right)$$

$$\Rightarrow \boxed{x_A \sim \mathcal{N}(\mu_A, \Sigma_{AA})}$$

Gaussian Conditional

$$p(x_A \mid x_B) = \frac{p(x_A, x_B)}{p(x_B)} = \frac{\frac{1}{\sqrt{(2\pi)^{n+m}|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}}{\frac{1}{\sqrt{(2\pi)^m|\Sigma_{BB}|}} e^{-\frac{1}{2}(x_B-\mu_B)^T\Sigma_{BB}^{-1}(x_B-\mu_B)}}$$

$$= \frac{1}{\sqrt{(2\pi)^n|\Sigma|/|\Sigma_{BB}|}} e^{-\frac{1}{2}\left((x-\mu)^T\Sigma^{-1}(x-\mu)-(x_B-\mu_B)^T\Sigma_{BB}^{-1}(x_B-\mu_B)\right)}$$

- Consider the exponent:

$$\begin{aligned}
&\tilde{x}_A^T\Omega_{AA}\tilde{x}_A + \tilde{x}_A^T\Omega_{AB}\tilde{x}_B + \tilde{x}_B^T\Omega_{AB}^T\tilde{x}_A + \tilde{x}_B^T\Omega_{BB}\tilde{x}_B - \tilde{x}_B^T\Sigma_{BB}^{-1}\tilde{x}_B \\
&\quad = \tilde{x}_A^T\Omega_{AA}\tilde{x}_A - 2\tilde{x}_A^T\Omega_{AA}\Sigma_{AB}\Sigma_{BB}^{-1}\tilde{x}_B + \tilde{x}_B^T\Sigma_{BB}^{-1}\Sigma_{AB}^T\Omega_{AA}\Sigma_{AB}\Sigma_{BB}^{-1}\tilde{x}_B \\
&\quad = (\tilde{x}_A - \Sigma_{AB}\Sigma_{BB}^{-1}\tilde{x}_B)^T\Omega_{AA}(\tilde{x}_A - \Sigma_{AB}\Sigma_{BB}^{-1}\tilde{x}_B) \\
&\quad \Rightarrow \boxed{x_A \mid x_B \sim \mathcal{N}\left(\mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(x_B - \mu_B), \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{AB}^T\right)}
\end{aligned}$$

Kalman Filter

- A **Kalman filter** is a Bayes filter for which:
  - The prior pdf $p_{0|0}$ is Gaussian
  - The motion model is linear in the state and affected by Gaussian noise
  - The observation model is linear in the state and affected by Gaussian noise
  - The process noise w_t and measurement noise v_t are independent of each other, of the state x_t, and across time
- **Prior**: $x_t \mid z_{0:t}, u_{0:t-1} \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t})$
- **Motion Model**:

$$x_{t+1} = f(x_t, u_t, w_t) := Fx_t + Gu_t + w_t, \quad w_t \sim \mathcal{N}(0, W)$$
$$x_{t+1} \mid x_t, u_t \sim \mathcal{N}(Fx_t + Gu_t, W), \quad F \in \mathbb{R}^{d_x \times d_x}, G \in \mathbb{R}^{d_x \times d_u},\ W \in \mathbb{R}^{d_x \times d_x}$$

- **Observation Model**:

$$z_t = h(x_t, v_t) := Hx_t + v_t, \quad v_t \sim \mathcal{N}(0, V)$$
$$z_t \mid x_t \sim \mathcal{N}(Hx_t, V), \quad H \in \mathbb{R}^{d_z \times d_x},\ V \in \mathbb{R}^{d_z \times d_z}$$

Kalman Filter Prediction

$$p_{t+1|t}(x) = \int p_f(x \mid s, u_t) p_{t|t}(s) ds = \int \phi(x; Fs + Gu_t, W) \phi(s; \mu_{t|t}, \Sigma_{t|t}) ds$$

$$= \kappa_{t|t} \int \exp\left\{ -\frac{1}{2}(x - Fs - Gu_t)^T W^{-1} (x - Fs - Gu_t) \right\} \times$$

$$\exp\left\{ -\frac{1}{2}(s - \mu_{t|t})^T \Sigma_{t|t}^{-1} (s - \mu_{t|t}) \right\} ds$$

$$= \kappa_{t|t} \int \exp\left\{ -\frac{1}{2} \left(s^T (F^T W^{-1} F + \Sigma_{t|t}^{-1}) s - 2(\Sigma_{t|t}^{-1} \mu_{t|t} + F^T W^{-1} (x - Gu_t))^T s + \ldots \right) \right\} ds$$

$$\overset{Sq.Comp.}{\underset{Inv.Lemma}{=\!=\!=\!=}} \phi(x; F\mu_{t|t} + Gu_t, F\Sigma_{t|t} F^T + W)$$

$$\boxed{\begin{aligned} p_{t+1|t}(x) &= \int \phi(x; Fs + Gu_t, W) \phi(s; \mu_{t|t}, \Sigma_{t|t}) ds \\ &= \phi(x; F\mu_{t|t} + Gu_t, F\Sigma_{t|t} F^T + W) \end{aligned}}$$

Kalman Filter Prediction (easy version)

- Motion model with given prior:

$$x_{t+1} = Fx_t + Gu_t + w_t, \qquad w_t \sim \mathcal{N}(0, W), \quad x_t \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t})$$

- Since w_t and x_t are independent and the Gaussian distribution is stable, we know that the distribution of x_{t+1} is Gaussian: $\mathcal{N}(\mu_{t+1|t}, \Sigma_{t+1|t})$

- We just need to compute its mean and covariance:

$$\mu_{t+1|t} = \mathbb{E}\left[Fx_t + Gu_t + w_t\right] = F\mathbb{E}[x_t] + Gu_t + \mathbb{E}[w_t] = F\mu_{t|t} + Gu_t$$

$$\begin{aligned}
\Sigma_{t+1|t} &= \mathbf{Var}\left[Fx_t + Gu_t + w_t\right] \xlongequal{\text{independence}} \mathbf{Var}\left[Fx_t\right] + 0 + \mathbf{Var}\left[w_t\right] \\
&= \mathbb{E}\left[F\left(x_t - \mu_{t|t}\right)\left(x_t - \mu_{t|t}\right)^T F^T\right] + W \\
&= F\Sigma_{t|t}F^T + W
\end{aligned}$$

Kalman Filter Update

$$p_{t+1|t+1}(x) = \frac{p(z_{t+1} \mid x) p_{t+1|t}(x)}{p(z_{t+1} \mid z_{0:t}, u_{0:t})} = \frac{\phi(z_{t+1}; Hx, V)\phi(x; \mu_{t+1|t}, \Sigma_{t+1|t})}{\int \phi(z_{t+1}; Hs, V)\phi(s; \mu_{t+1|t}, \Sigma_{t+1|t}) ds}$$

$$= \frac{\kappa_{t+1}}{\eta_{t+1}} \exp\left\{-\frac{1}{2}(z_{t+1} - Hx)^T V^{-1}(z_{t+1} - Hx)\right\} \exp\left\{-\frac{1}{2}(x - \mu_{t+1|t})^T \Sigma_{t+1|t}^{-1}(x - \mu_{t+1|t})\right\}$$

$$= \frac{\kappa_{t+1}}{\eta_{t+1}} \exp\left\{-\frac{1}{2}\left(x^T(H^T V^{-1} H + \Sigma_{t+1|t}^{-1})x + (H^T V^{-1} z_{t+1} + \Sigma_{t+1|t}^{-1}\mu_{t+1|t})^T x + \ldots\right)\right\}$$

$$\overset{Sq.Comp.}{=\!=\!=} \phi\left(x; (H^T V^{-1} H + \Sigma_{t+1|t}^{-1})^{-1}(H^T V^{-1} z_{t+1} + \Sigma_{t+1|t}^{-1}\mu_{t+1|t}), (H^T V^{-1} H + \Sigma_{t+1|t}^{-1})^{-1}\right)$$

$$\overset{Inv.Lemma}{=\!=\!=} \boxed{\phi\left(x; \mu_{t+1|t} + K_{t+1|t}(z_{t+1} - H\mu_{t+1|t}), (I - K_{t+1|t}H)\Sigma_{t+1|t}\right)}$$

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

Kalman Filter Update (easy version)

- Observation model with given prior:

$$z_{t+1} = Hx_{t+1} + v_{t+1}, \qquad v_{t+1} \sim \mathcal{N}(0, V), \quad x_{t+1} \sim \mathcal{N}(\mu_{t+1|t}, \Sigma_{t+1|t})$$

- The joint distribution of x_{t+1} and z_{t+1} is Gaussian:

$$\begin{pmatrix} x_{t+1} \\ z_{t+1} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{t+1|t} \\ H\mu_{t+1|t} \end{pmatrix}, \begin{bmatrix} \Sigma_{t+1|t} & * \\ *^T & H\Sigma_{t+1|t}H^T + V \end{bmatrix} \right)$$

$$\begin{aligned} * &= \mathbb{E}\left[\left(x_{t+1} - \mu_{t+1|t}\right) \left(z_{t+1} - H\mu_{t+1|t}\right)^T \right] \\ &= \mathbb{E}\left[\left(x_{t+1} - \mu_{t+1|t}\right) \left(\left(x_{t+1} - \mu_{t+1|t}\right)^T H^T + v_{t+1}^T \right) \right] = \Sigma_{t+1|t}H^T \end{aligned}$$

- The conditional distribution of $x_{t+1} \mid z_{t+1}$ is then also Gaussian:

$$\begin{aligned} x_{t+1} \mid z_{t+1} \sim \mathcal{N}\Big(&\mu_{t+1|t} + \Sigma_{t+1|t}H^T(H\Sigma_{t+1|t}H^T + V)^{-1}(z_{t+1} - H\mu_{t+1|t}), \\ &\Sigma_{t+1|t} - \Sigma_{t+1|t}H^T(H\Sigma_{t+1|t}H^T + V)^{-1}H\Sigma_{t+1|t} \Big) \end{aligned}$$

Kalman Filter (discrete time)

Prior: $x_t \mid z_{0:t}, u_{0:t-1} \sim \mathcal{N}(\mu_{t|t}, \Sigma_{t|t})$

Motion model: $x_{t+1} = Fx_t + Gu_t + w_t, \quad w_t \sim \mathcal{N}(0, W)$

Observation model: $z_t = Hx_t + v_t, \quad v_t \sim \mathcal{N}(0, V)$

Prediction:

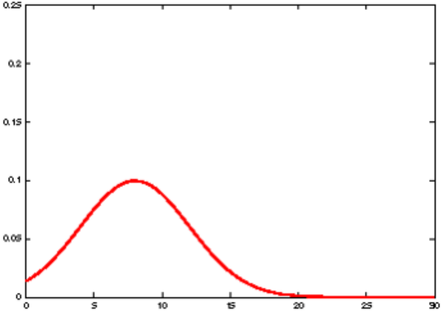
$$\mu_{t+1|t} = F\mu_{t|t} + Gu_t$$
$$\Sigma_{t+1|t} = F\Sigma_{t|t}F^T + W$$

Update:

$$\mu_{t+1|t+1} = \mu_{t+1|t} + K_{t+1|t}(z_{t+1} - H\mu_{t+1|t})$$
$$\Sigma_{t+1|t+1} = (I - K_{t+1|t}H)\Sigma_{t+1|t}$$

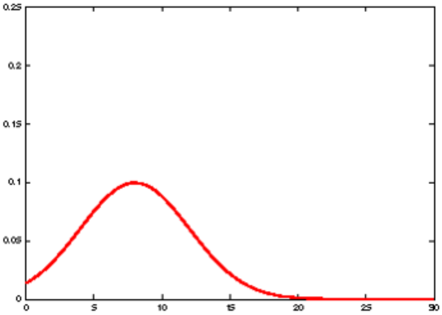
Kalman Gain: $K_{t+1|t} := \Sigma_{t+1|t}H^T \left(H\Sigma_{t+1|t}H^T + V\right)^{-1}$

Predict

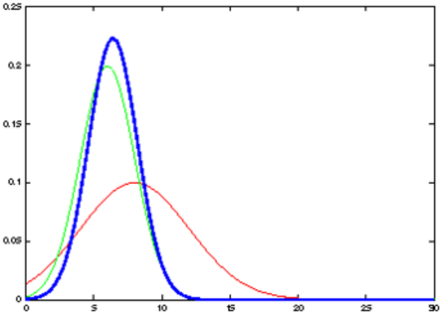


Update

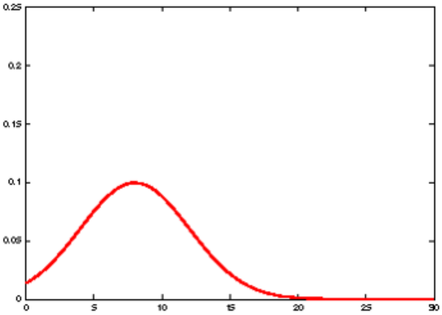
Predict

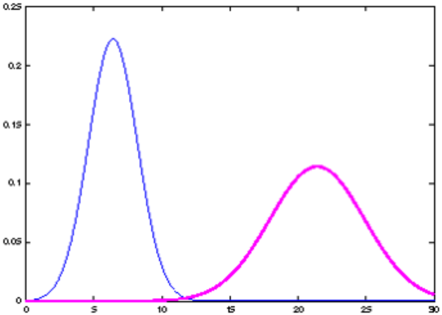


Update

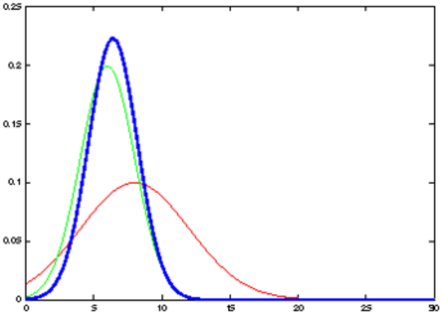


Predict

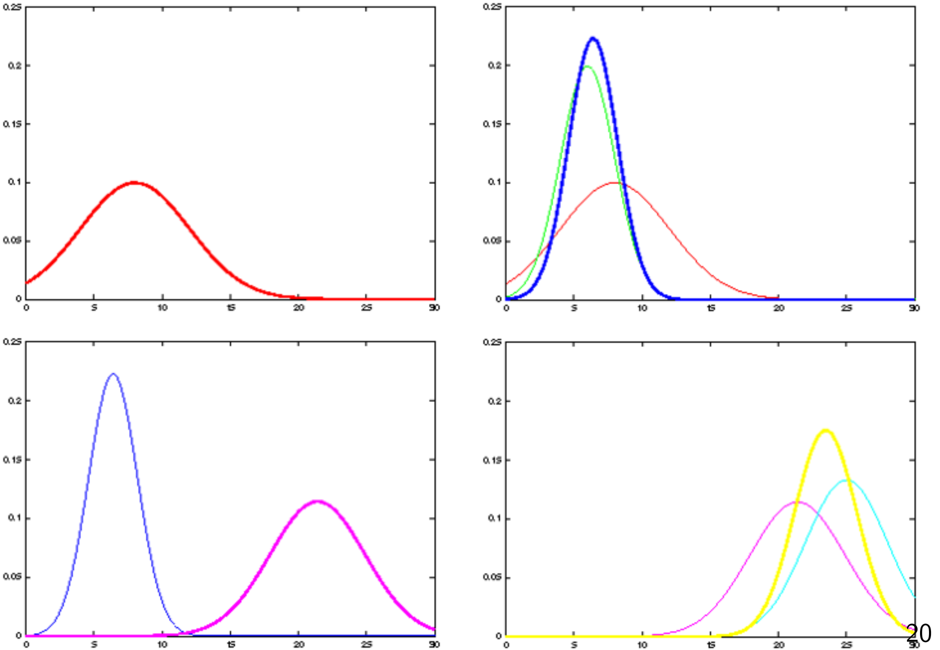




Update



Predict Update



Information Filter

- Uses the natural Gaussian parameterization $x \sim \mathcal{G}(\nu, \Omega)$ where $\nu = \Sigma^{-1}\mu$ and $\Omega = \Sigma^{-1}$
- Uses the matrix inversion lemma to convert the Kalman filter covariance equations to their information matrix counterparts

Prior: $x_t \mid z_{0:t}, u_{0:t-1} \sim \mathcal{G}(\nu_{t|t}, \Omega_{t|t})$

Motion model: $x_{t+1} = Fx_t + Gu_t + w_t, \quad w_t \sim \mathcal{G}(0, W^{-1})$

Observation model: $z_t = Hx_t + v_t, \quad v_t \sim \mathcal{G}(0, V^{-1})$

Prediction:
$$\nu_{t+1|t} = (I - C_{t|t})F^{-T}\nu_{t|t}$$
$$\Omega_{t+1|t} = (I - C_{t|t})F^{-T}\Omega_{t|t}F^{-1}(I - C_{t|t}^T) + C_{t|t}W^{-1}C_{t|t}^T$$

Information Gain: $C_{t|t} = F^{-T}\Omega_{t|t}F^{-1}\left(F^{-T}\Omega_{t|t}F^{-1} + W^{-1}\right)^{-1}$

Update:
$$\nu_{t+1|t+1} = \nu_{t+1|t} + H^T V^{-1} z_{t+1}$$
$$\Omega_{t+1|t+1} = \Omega_{t+1|t} + H^T V^{-1} H$$

Kalman-Bucy Filter (continuous time)

Prior:	$x(0) \sim \mathcal{N}(\mu(0), \Sigma(0))$
Motion model:	$\dot{x}(t) = Fx(t) + Gu(t) + w(t)$
Observation model:	$z(t) = Hx(t) + v(t)$
Mean:	$\dot{\mu}(t) = F\mu(t) + Gu(t) + K(t)\,(z(t) - H\mu(t))$
Covariance:	$\dot{\Sigma}(t) = F\Sigma(t) + \Sigma(t)F^T + W - K(t)VK^T(t)$
Kalman Gain:	$K(t) = \Sigma(t)H^T V^{-1}$

Kalman Filter Comments

- **Efficient**: polynomial in measurement and state dim: $\boxed{O(d_z^{2.376} + d_x^2)}$
- **Optimal**: under linearity, Gaussianity, and independence assumptions with respect to the mean square error (MSE):

$$\mathbb{E}\left[\|x_t - \mu_{t|t}\|_2^2\right] = \text{tr}(\Sigma_{t|t})$$

- To deal with **unknown models** we can use EM to learn the dynamics model (F, G, W) and the measurement model (H, V)
- Given data $\mathcal{D} := \{z_{0:T}, u_{0:T-1}\}$, apply EM with hidden variables $x_{0:T}$:
  - **E step**: Given initial parameter estimates $\theta^{(k)} := \{F^{(k)}, G^{(k)}, W^{(k)}, H^{(k)}, V^{(k)}\}$ calculate the likelihood of the hidden variables via the Kalman filter/smoother
  - **M step**: Optimize the parameters via MLE to obtain $\theta^{(k+1)}$ which explain the posterior distribution over $x_{0:T}$ better
- Most robotic systems are **nonlinear**!

Gaussian Mixture Filter

- **Prior**: $x_t \mid z_{0:t}, u_{0:t-1} \sim p_{t|t}(x_x) := \sum_k \alpha_{t|t}^{(k)} \phi\left(x_t; \mu_{t|t}^{(k)}, \Sigma_{t|t}^{(k)}\right)$
- **Motion model**: $x_{t+1} = Fx_t + Gu_t + w_t, \quad w_t \sim \mathcal{N}(0, W)$
- **Observation model**: $z_t = Hx_t + v_t, \quad v_t \sim \mathcal{N}(0, V)$
- **Prediction**:

$$p_{t+1|t}(x) = \int p_f(x \mid s, u_t) p_{t|t}(s) ds = \sum_k \alpha_{t|t}^{(k)} \int p_f(x \mid s, u_t) \phi\left(s; \mu_{t|t}^{(k)}, \Sigma_{t|t}^{(k)}\right) ds$$

$$= \sum_k \alpha_{t|t}^{(k)} \phi\left(x; F\mu_{t|t}^{(k)} + Gu_t, F\Sigma_{t|t}^{(k)} F^T + W\right)$$

- **Update**:

$$p_{t+1|t+1}(x) = \frac{p_h(z_{t+1} \mid x) p_{t+1|t}(x)}{p(z_{t+1} \mid z_{0:t}, u_{0:t})} = \frac{\phi\left(z_{t+1}; Hx, V\right) \sum_k \alpha_{t+1|t}^{(k)} \phi\left(x; \mu_{t+1|t}^{(k)}, \Sigma_{t+1|t}^{(k)}\right)}{\int \phi\left(z_{t+1}; Hs, V\right) \sum_j \alpha_{t+1|t}^{(j)} \phi\left(s; \mu_{t+1|t}^{(j)}, \Sigma_{t+1|t}^{(j)}\right) ds}$$

$$= \sum_k \left(\frac{\alpha_{t+1|t}^{(k)} \phi\left(z_{t+1}; Hx, V\right) \phi\left(x; \mu_{t+1|t}^{(k)}, \Sigma_{t+1|t}^{(k)}\right)}{\sum_j \alpha_{t+1|t}^{(j)} \phi\left(z_{t+1}; H\mu_{t+1|t}^{(j)}, H\Sigma_{t+1|t}^{(j)} H^T + V\right)} \times \frac{\phi\left(z_{t+1}; H\mu_{t+1|t}^{(k)}, H\Sigma_{t+1|t}^{(k)} H^T + V\right)}{\phi\left(z_{t+1}; H\mu_{t+1|t}^{(k)}, H\Sigma_{t+1|t}^{(k)} H^T + V\right)} \right)$$

$$= \sum_k \left[\frac{\alpha_{t+1|t}^{(k)} \phi\left(z_{t+1}; H\mu_{t+1|t}^{(k)}, H\Sigma_{t+1|t}^{(k)} H^T + V\right)}{\sum_j \alpha_{t+1|t}^{(j)} \phi\left(z_{t+1}; H\mu_{t+1|t}^{(j)}, H\Sigma_{t+1|t}^{(j)} H^T + V\right)} \right] \phi\left(x; \mu_{t+1|t}^{(k)} + K_{t+1|t}^{(k)}(z_{t+1} - H\mu_{t+1|t}^{(k)}), (I - K_{t+1|t}^{(k)} H)\Sigma_{t+1|t}^{(k)}\right)$$

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

Gaussian Mixture Filter

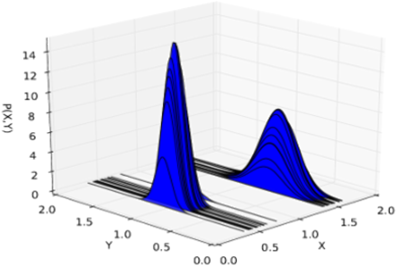
- **pdf**: $x_t \mid z_{0:t}, u_{0:t-1} \sim p_{t|t}(x) := \sum_k \alpha_{t|t}^{(k)} \phi\left(x_t; \mu_{t|t}^{(k)}, \Sigma_{t|t}^{(k)}\right)$
- **mean**: $\mu_{t|t} := \mathbb{E}\left[x_t \mid z_{0:t}, u_{0:t-1}\right] = \int x p_{t|t}(x) dx = \sum_k \alpha_{t|t}^{(k)} \mu_{t|t}^{(k)}$
- **cov**: $\Sigma_{t|t} := \mathbb{E}\left[x_t x_t^T \mid z_{0:t}, u_{0:t-1}\right] - \mu_{t|t}\mu_{t|t}^T$
  $= \int x x^T p_{t|t}(x) dx - \mu_{t|t}\mu_{t|t}^T = \sum_k \alpha_{t|t}^{(k)} \left(\Sigma_{t|t}^{(k)} + \mu_{t|t}^{(k)} (\mu_{t|t}^{(k)})^T\right) - \mu_{t|t}\mu_{t|t}^T$
- The GMF is just a **bank of Kalman filters**; sometimes called **Gaussian Sum filter**

Gaussian Mixture Filter

- If the motion or observation models are nonlinear, we can apply the EKF or UKF tricks to get a nonlinear GMF
- Additional operations are needed when strong nonlinearities are present in the motion or observation models:
  - **Refinement**: introduces additional components to reduce the linearization error
  - **Pruning**: approximates the overall distribution with a smaller number of components (e.g., using KL divergence as a measure of accuracy)
- More details:
  - Nonlinear Gaussian Filtering: Theory, Algorithms, and Applications: Huber
  - Bayesian Filtering and Smoothing: Särkkä

Rao-Blackwellized Particle Filter

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



$$
\begin{aligned}
x^n_{t+1} &= f^n_t(x^n_t) + A^n_t(x^n_t)x^l_t + G^n_t(x^n_t)w^n_t \\
x^l_{t+1} &= f^l_t(x^n_t) + A^l_t(x^n_t)x^l_t + G^l_t(x^n_t)w^l_t \\
z_t &= h_t(x^n_t) + C_t(x^n_t)x^l_t + v_t
\end{aligned}
$$

Nonlinear states: x^n_t

Linear states: x^l_t

- **Idea**: exploit linear-Gaussian sub-structure to handle high dim. problems

$$
p\left(x^l_t, x^n_{0:t} \mid z_{0:t}, u_{0:t-1}\right) = \underbrace{p\left(x^l_t \mid z_{0:t}, u_{0:t-1}, x^n_{0:t}\right)}_{\text{Kalman Filter}} \underbrace{p\left(x^n_{0:t} \mid z_{0:t}, u_{0:t-1}\right)}_{\text{Particle Filter}}
$$

$$
= \sum_{k=1}^{N_{t|t}} \alpha^{(k)}_{t|t} \delta\left(x^n_{0:t}; m^{(k)}_{t|t}\right) \phi\left(x^l_t; \mu^{(k)}_{t|t}, \Sigma^{(k)}_{t|t}\right)
$$

- The RBPF is a combination of the particle filter and the Kalman filter, in which each particle has a Kalman filter associated to it