Mathematics for Robotics (ROB-GY 6013 Section A)

- **Week 13:**
  - Kalman Filter

Thinking

- **Batch:**
  - Everything, all at once
  - Complete information is already there
    - No sense of probability concepts, covariance
- **Recursive:**
  - You receive new pieces of information as time progresses
    - For a discrete-time system, the index k is time
    - The matrix A_k is a function of time

Thinking

- **Dynamics:**
  - Prediction based on equation that describes system behavior/state evolution

$$x_{k+1} = A_k x_k + G_k w_k$$

- **Measurement:**
  - Sensor feedback

$$y_k = C_k x_k + v_k$$

Thinking

- **Estimation** is about finding a **probability distribution** or **density**
  - Normal density is defined by mean and covariance matrix
  - At each step of the Kalman Filter we are searching for some conditional density by computing an updated mean and covariance matrix

Model

- Linear **time-varying** discrete-time system with **“white” Gaussian noise**

$$x_{k+1} = A_k x_k + G_k w_k \qquad\qquad y_k = C_k x_k + v_k$$

$x \in \mathbb{R}^n$, $w \in \mathbb{R}^p$, $y \in \mathbb{R}^m$, $v \in \mathbb{R}^m$.

- Initial condition: x_0
- x_0, and, for $k \geq 0$, w_k, v_k are independent Gaussian random vectors.

$$\operatorname{cov}\left(\begin{bmatrix} w_k \\ v_k \\ x_0 \end{bmatrix}, \begin{bmatrix} w_l \\ v_l \\ x_0 \end{bmatrix} \right) = \begin{bmatrix} R_k \delta_{kl} & 0 & 0 \\ 0 & Q_k \delta_{kl} & 0 \\ 0 & 0 & P_0 \end{bmatrix}, \quad \delta_{kl} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

Stochastic Assumptions

- For all $k \geq 0,\ l \geq 0$, x_0, w_k, v_l are jointly Gaussian.
- w_k is a 0-mean white noise process: $\mathcal{E}\{w_k\} = 0$, and $\operatorname{cov}(w_k, w_l) = R_k \delta_{kl}$
- v_k is a 0-mean white noise process: $\mathcal{E}\{v_k\} = 0$, and $\operatorname{cov}(v_k, v_l) = Q_k \delta_{kl}$
- Uncorrelated noise processes: $\operatorname{cov}(w_k, v_l) = 0$
- The initial condition x_0 is uncorrelated with all other noise sequences.
- We denote the mean and covariance of x_0 by

$$\bar{x}_0 = \mathcal{E}\{x_0\} \ \text{ and } \ P_0 = \operatorname{cov}(x_0) = \operatorname{cov}(x_0, x_0) = \mathcal{E}\{(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^\top\}$$

Properties of x_k and y_k

$$x_{k+1} = A_k x_k + G_k w_k \qquad\qquad y_k = C_k x_k + v_k$$

- For all $k \geq 1$, x_k is a linear combination of x_0 and $w_0, \ldots, w_{k-1}$.

- For all $k \geq 1$, y_k is a linear combination of x_0 and $w_0, \ldots, w_{k-1}$, and $v_0, \ldots, v_k$.

Properties of x_k and y_k

$$x_{k+1} = A_k x_k + G_k w_k \qquad\qquad y_k = C_k x_k + v_k$$

- For all $k \geq 1$, x_k is a linear combination of x_0 and $w_0, \ldots, w_{k-1}$.
  - In particular, x_k is uncorrelated with w_k.
- For all $k \geq 1$, y_k is a linear combination of x_0 and $w_0, \ldots, w_{k-1}$, and $v_0, \ldots, v_k$.
  - In particular, y_k is uncorrelated with w_k.
- For all $k \geq 0$, v_k is uncorrelated with x_k.

Basic Kalman Filter: Terms

- Update estimates of x and P at each time instant
- $k \mid k$ vs. $k{+}1 \mid k$

Basic Kalman Filter: Terms

- Update estimates of x and P at each time instant
- $k \mid k$ vs. $k{+}1 \mid k$

$$\widehat{x}_{k|k} := \mathcal{E}\{x_k | y_0, \cdots, y_k\}$$

$$P_{k|k} := \mathcal{E}\{(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^\top | y_0, \cdots, y_k\}$$

Basic Kalman Filter: Terms

- Update estimates of x and P at each time instant
- $k \mid k$ vs. $k{+}1 \mid k$

$$\widehat{x}_{k|k} := \mathcal{E}\{x_k | y_0, \cdots, y_k\}$$

$$P_{k|k} := \mathcal{E}\{(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^\top | y_0, \cdots, y_k\}$$

$$\widehat{x}_{k+1|k} := \mathcal{E}\{x_{k+1} | y_0, \cdots, y_k\}$$

$$P_{k+1|k} := \mathcal{E}\{(x_{k+1} - \widehat{x}_{k+1|k})(x_{k+1} - \widehat{x}_{k+1|k})^\top | y_0, \cdots, y_k\}$$

Basic Kalman Filter: Initial Conditions

- **Initial conditions:** $\widehat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}$, and $P_{0|-1} := P_0 = \mathrm{cov}(x_0)$

$$\widehat{x}_{k|k} := \mathcal{E}\{x_k | y_0, \cdots, y_k\}$$
$$P_{k|k} := \mathcal{E}\{(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^\top | y_0, \cdots, y_k\}$$

$$\widehat{x}_{k+1|k} := \mathcal{E}\{x_{k+1} | y_0, \cdots, y_k\}$$
$$P_{k+1|k} := \mathcal{E}\{(x_{k+1} - \widehat{x}_{k+1|k})(x_{k+1} - \widehat{x}_{k+1|k})^\top | y_0, \cdots, y_k\}$$

Basic Kalman Filter: Naïve MVE

- **Initial conditions:** $\widehat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}, \text{ and } P_{0|-1} := P_0 = \mathrm{cov}(x_0)$

$$\widehat{x}_{k|k} := \mathcal{E}\{x_k | y_0, \cdots, y_k\}$$
$$P_{k|k} := \mathcal{E}\{(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^\top | y_0, \cdots, y_k\}$$

- **Idea: Just use this for batch computation with MVE! These are conditional densities with multivariate normal random vectors**

Basic Kalman Filter: Naïve MVE

- **Initial conditions:** $\widehat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}$, and $P_{0|-1} := P_0 = \operatorname{cov}(x_0)$

$$\widehat{x}_{k|k} := \mathcal{E}\{x_k | y_0, \cdots, y_k\}$$
$$P_{k|k} := \mathcal{E}\{(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^\top | y_0, \cdots, y_k\}$$

- **BAD Idea: Just use this for batch computation with MVE! These are conditional densities with multivariate normal random vectors**
  - **Huge number of measurements** $Y_k = (y_k, y_{k-1}, \ldots, y_0)$.
  - **Must run batch computation for *each*** x_k

Note on Notation

- **Matrix:** $Y_k = (y_k, y_{k-1}, \dots, y_0)$
  - $Y_k = (y_k, Y_{k-1})$

Note on Notation

- **Matrix:** $Y_k = (y_k, y_{k-1}, \dots, y_0)$
    - $Y_k = (y_k, Y_{k-1})$
- **Mean and Covariance of the conditional random vector:** $x_k \mid Y_k$
- **Mean and Covariance of the conditional random vector:** $x_{k+1} \mid Y_k$

Note on Notation

- **Matrix:** $Y_k = (y_k, y_{k-1}, \dots, y_0)$
  - $Y_k = (y_k, Y_{k-1})$
- **Mean and Covariance of the conditional random vector:** $x_k \mid Y_k$

$$\widehat{x}_{k|k} := \mathcal{E}\{x_k | Y_k\}$$
$$P_{k|k} := \mathcal{E}\{(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^\top | Y_k\}$$

- **Mean and Covariance of the conditional random vector:** $x_{k+1} \mid Y_k$

$$\widehat{x}_{k+1|k} := \mathcal{E}\{x_{k+1} | Y_k\}$$
$$P_{k+1|k} := \mathcal{E}\{(x_{k+1} - \widehat{x}_{k+1|k})(x_{k+1} - \widehat{x}_{k+1|k})^\top | Y_k\}$$

Note on Notation

- **Matrix:** $Y_k = (y_k, y_{k-1}, \ldots, y_0)$
  - $Y_k = (y_k, Y_{k-1})$
- **Mean and Covariance of the conditional random vector:** $x_k \mid Y_k$
  - Density of: $N(\widehat{x}_{k|k}, P_{k|k})$
- **Mean and Covariance of the conditional random vector:** $x_{k+1} \mid Y_k$
  - Density of: $N(\widehat{x}_{k+1|k}, P_{k+1|k})$

Basic Kalman Filter

Initial Conditions:

$$\widehat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}, \text{ and } P_{0|-1} := P_0 = \mathrm{cov}(x_0)$$

For $k \geq 0$

Measurement Update Step:

$$\begin{aligned} K_k &= P_{k|k-1}C_k^\top \left(C_k P_{k|k-1} C_k^\top + Q_k\right)^{-1} \quad \text{(Kalman Gain)} \\ \widehat{x}_{k|k} &= \widehat{x}_{k|k-1} + K_k\left(y_k - C_k \widehat{x}_{k|k-1}\right) \\ P_{k|k} &= P_{k|k-1} - K_k C_k P_{k|k-1} \end{aligned}$$

Time Update or Prediction Step:

$$\begin{aligned} \widehat{x}_{k+1|k} &= A_k \widehat{x}_{k|k} \\ P_{k+1|k} &= A_k P_{k|k} A_k^\top + G_k R_k G_k^\top \end{aligned}$$

End of For Loop (Just stated this way to emphasize the recursive nature of the filter)

Review

Key Fact 1: Conditional Distributions of Gaussian Random Vectors

- Mean $$\mu_{1|2} := \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$
- Covariance $$\Sigma_{1|2} := \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$
- Proof: http://fourier.eng.hmc.edu/e161/lectures/gaussianprocess/node7.html

Key Fact 2: Conditional Independence

- Suppose we have 3 vectors X_1, X_2 and X_3 that are **jointly normally distributed** and X_1 and X_3 are each **independent** of X_2. The covariance matrix has the form

$$\Sigma = \begin{bmatrix} \Sigma_{11} & 0 & \Sigma_{13} \\ 0 & \Sigma_{22} & 0 \\ \Sigma_{13}^\top & 0 & \Sigma_{33} \end{bmatrix}$$

Key Fact 2: Conditional Independence

- Suppose we have 3 vectors X_1, X_2 and X_3 that are **jointly normally distributed** and X_1 and X_3 are each **independent** of X_2. The covariance matrix has the form

$$\Sigma = \left[\begin{array}{cc|c} \Sigma_{11} & 0 & \Sigma_{13} \\ 0 & \Sigma_{22} & 0 \\ \hline \Sigma_{13}^\top & 0 & \Sigma_{33} \end{array}\right]$$

X_1 and X_2 are **conditionally independent** given X_3

Key Fact 2: Conditional Independence

- Suppose we have 3 vectors X_1, X_2 and X_3 that are **jointly normally distributed** and X_1 and X_3 are each **independent** of X_2.

$$\Sigma = \left[\begin{array}{cc|c} \Sigma_{11} & 0 & \Sigma_{13} \\ 0 & \Sigma_{22} & 0 \\ \hline \Sigma_{13}^\top & 0 & \Sigma_{33} \end{array}\right]$$

X_1 and X_2 are **conditionally independent** given X_3

- **Using Key Fact 1 for covariance:**

$$\left.\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right| X_3$$

$$\mathrm{cov}\left(\begin{bmatrix} X_{1|X_3} \\ X_{2|X_3} \end{bmatrix}, \begin{bmatrix} X_{1|X_3} \\ X_{2|X_3} \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} - \begin{bmatrix} \Sigma_{13} \\ 0 \end{bmatrix} \Sigma_{33}^{-1} \begin{bmatrix} \Sigma_{13}^\top & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{11} - \Sigma_{13}\Sigma_{33}^{-1}\Sigma_{13}^\top & 0 \\ 0 & \Sigma_{22} \end{bmatrix}$$

Key Fact 3: Covariance of a Sum of Independent Normal Random Variables

- **Linear Combination:** $Y = AX_1 + BX_2$
- **Mean:** $\mu_Y = A\mu_1 + B\mu_2$
- **Covariance:** $\mathrm{cov}(Y, Y) = A\Sigma_{11}A^T + B\Sigma_{22}B^T.$

Key Fact 4

- Suppose that X, Y, and Z are jointly distributed random vectors with density f_{XYZ}.

$$(X|Z)\big|(Y|Z) \sim \frac{f_{(X|Z)(Y|Z)}}{f_{(Y|Z)}} = \frac{f_{XYZ}}{f_{YZ}} \sim X\Big|\begin{bmatrix} Y \\ Z \end{bmatrix}$$

Key Fact 4

- Suppose that X, Y, and Z are jointly distributed random vectors with density f_{XYZ}.

$$(X|Z)\big|(Y|Z) \sim \frac{f_{(X|Z)(Y|Z)}}{f_{(Y|Z)}} = \frac{f_{XYZ}}{f_{YZ}} \sim X\Big|\begin{bmatrix} Y \\ Z \end{bmatrix}$$

- **Proof:**

$$(X|Z)\big|(Y|Z) \sim \frac{f_{(X|Z)(Y|Z)}}{f_{(Y|Z)}} = \frac{f_{\left[\begin{array}{c} X \\ Y \end{array}\right]|Z}}{f_{Y|Z}} = \frac{\frac{f_{XYZ}}{f_Z}}{\frac{f_{YZ}}{f_Z}} = \frac{f_{XYZ}}{f_{YZ}} \sim X\Big|\begin{bmatrix} Y \\ Z \end{bmatrix}$$

Key Fact 4

- Suppose that X, Y, and Z are jointly distributed random vectors with density f_{XYZ}.

$$(X|Z)\big|(Y|Z) \sim \frac{f_{(X|Z)(Y|Z)}}{f_{(Y|Z)}} = \frac{f_{XYZ}}{f_{YZ}} \sim X\Big|\begin{bmatrix} Y \\ Z \end{bmatrix}$$

- **Proof:**

$$(X|Z)\big|(Y|Z) \sim \frac{f_{(X|Z)(Y|Z)}}{f_{(Y|Z)}} = \frac{f_{\left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right]|Z}}{f_{Y|Z}} = \frac{\frac{f_{XYZ}}{f_Z}}{\frac{f_{YZ}}{f_Z}} = \frac{f_{XYZ}}{f_{YZ}} \sim X\Big|\begin{bmatrix} Y \\ Z \end{bmatrix}$$

- **Does not require random vectors to be normal (Gaussian)!**

Key Fact 4

- Suppose that X, Y, and Z are jointly distributed random vectors with density f_{XYZ}.

$$(X|Z)\big|(Y|Z) \sim \frac{f_{(X|Z)(Y|Z)}}{f_{(Y|Z)}} = \frac{f_{XYZ}}{f_{YZ}} \sim X\Big|\begin{bmatrix} Y \\ Z \end{bmatrix}$$

- **Proof:**

$$(X|Z)\big|(Y|Z) \sim \frac{f_{(X|Z)(Y|Z)}}{f_{(Y|Z)}} = \frac{f_{\left[\begin{array}{c} X \\ Y \end{array}\right]|Z}}{f_{Y|Z}} = \frac{\frac{f_{XYZ}}{f_Z}}{\frac{f_{YZ}}{f_Z}} = \frac{f_{XYZ}}{f_{YZ}} \sim X\Big|\begin{bmatrix} Y \\ Z \end{bmatrix}$$

- **Does not require random vectors to be normal (Gaussian)!**
- **Key to recursion!**

Basic Kalman Filter
(refer to Textbook section 5.7 for derivation)

Initial Conditions:

$$\widehat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}, \text{ and } P_{0|-1} := P_0 = \mathrm{cov}(x_0)$$

For $k \geq 0$

Measurement Update Step:

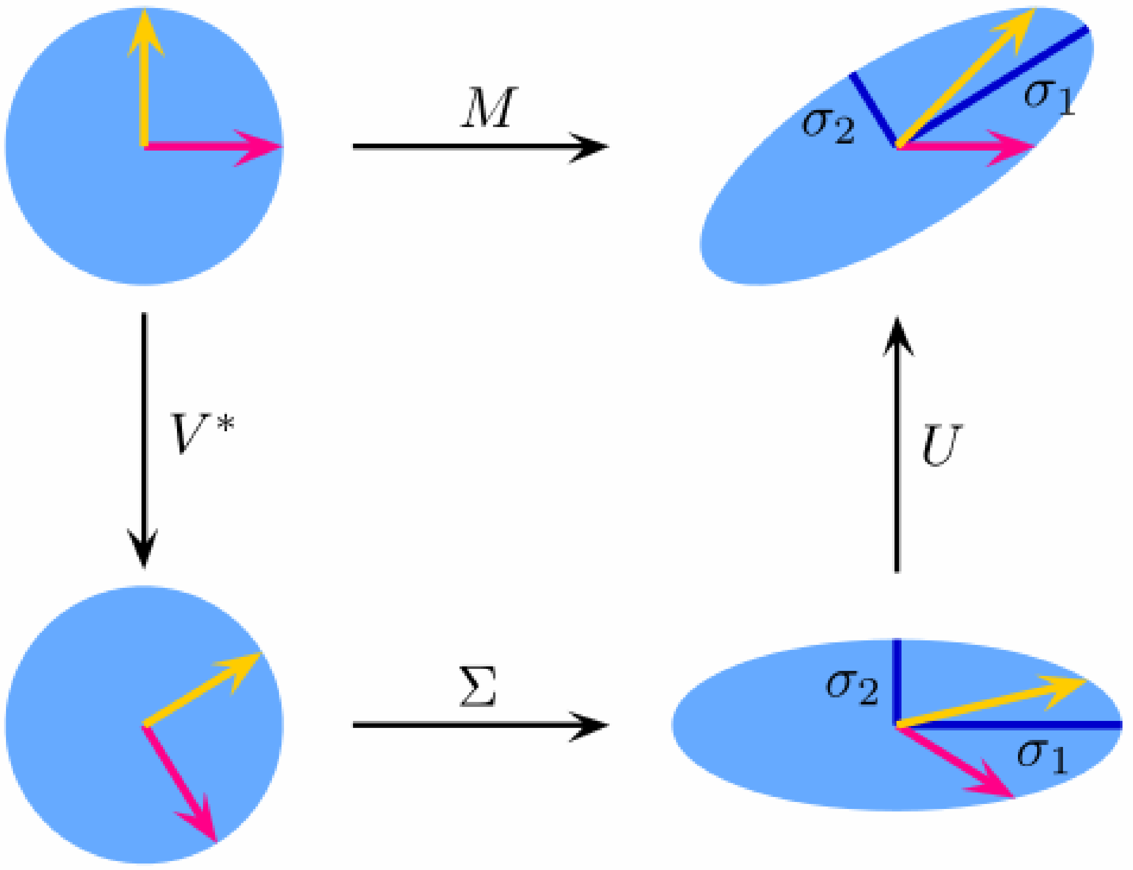
$$K_k = P_{k|k-1} C_k^\top \left(C_k P_{k|k-1} C_k^\top + Q_k\right)^{-1} \quad \text{(Kalman Gain)}$$
$$\widehat{x}_{k|k} = \widehat{x}_{k|k-1} + K_k \left(y_k - C_k \widehat{x}_{k|k-1}\right)$$
$$P_{k|k} = P_{k|k-1} - K_k C_k P_{k|k-1}$$

Time Update or Prediction Step:

$$\widehat{x}_{k+1|k} = A_k \widehat{x}_{k|k}$$
$$P_{k+1|k} = A_k P_{k|k} A_k^\top + G_k R_k G_k^\top$$

End of For Loop (Just stated this way to emphasize the recursive nature of the filter)

Singular Value Decomposition (SVD)



$$M = U \cdot \Sigma \cdot V^*$$

Plan next week

- Cover matrix factorizations
  - QR factorization
  - Finish SVD
  - LU factorizations
- Newton-Raphson
- A taste of linear programming and optimization