Introduction to Robotics
Localization

Erion Plaku

Department of Electrical Engineering and Computer Science
Catholic University of America
Linear Dynamical Discrete-Time System with Noise

\[ x(k + 1) = F(k)x(k) + G(k)u(k) + v(k) \]
\[ y(k) = H(k)x(k) + w(k) \]

- \( x(k) \in \mathbb{R}^n \) denotes the system state at time \( t_k = t_0 + Tk \)
  - \( t_0 \) denotes the initial time, \( T \) denotes the time step
- \( u(k) \in \mathbb{R}^m \) denotes the control input, e.g., velocity commands, torques, forces
- \( y(k) \in \mathbb{R}^p \) denotes the system output, e.g., values reported by sensors
- \( F(k) \in \mathbb{R}^{n \times n} \) encodes the system dynamics
- \( G(k) \in \mathbb{R}^{n \times m} \) describes how the inputs drive the dynamics
- \( H(k) \in \mathbb{R}^{p \times n} \) describes how states are mapped into outputs
  - assumed to be full row rank for all \( k \), although it may not be square
- \( v(k) \in \mathbb{R}^n \) denotes the process noise
  - assumed to be white Guassian noise with zero mean and covariance matrix \( V(k) \)
- \( w(k) \in \mathbb{R}^p \) denotes the measurement noise
  - assumed to be white Guassian noise with zero mean and covariance matrix \( W(k) \)
Linear Kalman Filter

Linear Dynamical Discrete-Time System with Noise

\[ x(k + 1) = F(k)x(k) + G(k)u(k) + v(k) \]
\[ y(k) = H(k)x(k) + w(k) \]

Objective: Determine the “best” estimate of \( x(k) \) given a previous estimate \( x(k - 1) \) together with the known input \( u(k) \) and output \( y(k) \)

Challenges

- Presence of the unknown and unmeasurable noise vectors \( v(k) \) and \( w(k) \)
- State cannot in general be directly determined from the outputs because \( H(k) \) may not be invertible

Approach: State estimate is constructed using the time history of the known signals \( y(k) \) and \( u(k) \) together with the known parameters \( F(k), G(k), H(k), V(k), W(k) \)
A Simple Observer

Assume that there is no noise, i.e.,

\[ x(k+1) = F(k)x(k) + G(k)u(k) \]

\[ y(k) = H(k)x(k) \]

Notation: \( \hat{x}(k_1|0\ldots k_2) \) with \( k_1 \geq k_2 \) denotes the value of the state estimate at time step \( k_1 \) given the output values \( y(0), \ldots, y(k_2) \)

Observer follows a two-step process:

1. Prediction

\[ \hat{x}(k+1|0\ldots k) = F(k)\hat{x}(k|0\ldots k) + G(k)u(k) \]

2. Update

- Given the output \( y(k+1) \), the system state is constrained to lie on the hyperplane

\[ \Omega = \{ x \in \mathbb{R}^n : y(k+1) = H(k+1)x \} \]

- Choose the next estimate \( \hat{x}(k+1|0\ldots k+1) \) to be the point in \( \Omega \) that has the shortest distance to the prediction \( \hat{x}(k+1|0\ldots k) \), i.e.,

\[ \hat{x}(k+1|0\ldots k+1) = \arg\min_{x \in \Omega} ||x, \hat{x}(k+1|0\ldots k)|| \]

Why? \( \hat{x}(k+1|0\ldots k) \) is close to the actual state, and the actual state must be in \( \Omega \)
Computing the Update

\[ \Omega = \{ x : y(k + 1) = H(k + 1)x \} \]

1. A vector \( a \in \mathbb{R}^n \) is parallel to \( \Omega \)
   \[ \iff x + a \in \Omega, \forall x \in \Omega \]
   \[ \iff H(k + 1)a = 0 \]

   Gives rise to the null-space, i.e.,

   \[ \text{NullSpace} (H(k + 1)) = \{ a \in \mathbb{R}^n : H(k + 1)a = 0 \} \]

2. A vector \( b \in \mathbb{R}^n \) is orthogonal to \( \Omega \)
   \[ \iff a^T b = 0, \forall a \in \text{NullSpace} (H(k + 1)) \]
   \[ \iff b \in \text{RowSpace} (H(k + 1)) \]

Therefore,

\[ \Delta x \text{ is orthogonal to } \Omega \iff \Delta x = H(k + 1)^T \gamma, \text{ for some } \gamma \in \mathbb{R}^p \]

\[ \Delta x = \hat{x}(k + 1 | 0 \ldots k + 1) - \hat{x}(k + 1 | 0 \ldots k) \]

\( \Delta x \) must be orthogonal to \( \Omega \)

\[ a^T \Delta x = 0 \text{ for any } a \text{ that is parallel to } \Omega \]
Let $\nu$ denote the innovation error, i.e.,

$$\nu = y(k + 1) - H(k + 1)\hat{x}(k + 1|0 \ldots k)$$

Assume for now that $\gamma$ can be written as a linear function of $\nu$, i.e.,

$$\gamma = K\nu,$$  for some $K \in \mathbb{R}^{p \times p}$

Then

$$\Delta x = H(k + 1)^T\gamma = H(k + 1)^T K\nu = H(k + 1)^T K (y(k + 1) - H(k + 1)\hat{x}(k + 1|0 \ldots k))$$

Now we need to find $K$ such that

$$y(k + 1) = H(k + 1)(\hat{x}(k + 1|0 \ldots k) + \Delta x) \implies$$

$$H(k + 1)\Delta x = y(k + 1) - H(k + 1)\hat{x}(k + 1|0 \ldots k) = \nu \implies$$

$$H(k + 1)H(k + 1)^T K\nu = \nu \implies$$

$$K = \left(H(k + 1)H(k + 1)^T\right)^{-1}$$

Does the inverse exist?
A Simple Observer: Putting it all together

Prediction

\[ \hat{x}(k + 1|0\ldots k) = F(k)\hat{x}(k|0\ldots k) + G(k)u(k) \]

Update

\[ \hat{x}(k + 1|0\ldots k + 1) = \hat{x}(k + 1|0\ldots k) + \Delta x \]
\[ = \hat{x}(k + 1|0\ldots k) + H(k + 1)^T K \nu \]

- \[ K = (H(k + 1)H(k + 1)^T)^{-1} \]
- \[ \nu = (y(k + 1) - H(k + 1)\hat{x}(k + 1|0\ldots k)) \]

What are some problems with the simple observer?

- Update is always perpendicular to \( \Omega \)
- Estimate errors in direction parallel to \( \Omega \) are never corrected
- As a result, estimate \( \hat{x} \) will not in general converge to actual state \( x \)
Before improving the simple observer . . . some statistics

Probability density function:

\[ Pr[a \leq X \leq b] = \int_{x=a}^{b} f(x) \, dx \]

Expected value for a random vector \( X : S \to \mathbb{R}^n \):

\[ E(X) = \int_{x \in \mathbb{R}^n} xf(x) \, dx \]

Variance of a scalar random variable:

\[ Var(X) = E \left( (X - E(X))^2 \right) = E(X^2) - (E(X))^2 \]

Covariance among two scalar random variables:

\[ Cov(X, Y) = E \left( (X - E(X))(Y - E(Y)) \right) = E(XY) - E(X)E(Y) \]

Covariance matrix:

\[ Cov(X) = E((X - E(X))(X - E(X))^T), \text{ i.e., } Cov_{ij}(X) = Cov(X_i, X_j) \]

Multivariate gaussian distribution with mean \( \bar{X} \) and covariance matrix \( P \):

\[ f(x; \bar{X}, P) = \frac{1}{\sqrt{(2\pi)^n |P|}} e^{-\frac{1}{2}(x - \bar{X})P^{-1}(x - \bar{X})^T} \]
Assume that there is process noise but no measurement noise, i.e.,

\[ x(k + 1) = F(k)x(k) + G(k)u(k) + v(k) \]

\[ y(k) = H(k)x(k) \]

Recall that
- \( H(k) \): assumed to be full row rank for all \( k \), although it may not be square
- \( v(k) \): assumed to be white Gaussian noise with zero mean and covariance matrix \( V(k) \)
  
  white here means \( v(k) \) is independent of \( v(k - 1) \) for all \( k \)

Objective is to generate both
- a state vector estimate \( \hat{x}(k|0\ldots k) \) and
- a covariance matrix estimate \( P(k|0\ldots k) \)

Hence
- Prediction will generate \( \hat{x}(k + 1|0\ldots k) \) and \( P(k + 1|0\ldots k) \)
- Update will generate the next estimate given \( \hat{x}(k + 1|0\ldots k + 1) \) and \( P(k + 1|0\ldots k + 1) \)
Predicted state vector:

\[ \hat{x}(k+1|0\ldots k) = F(k)\hat{x}(k|0\ldots k) + G(k)u(k) + E(v(k)) = F(k)\hat{x}(k|0\ldots k) + G(k)u(k) \]

Predicted covariance matrix:

\[
P(k + 1|0\ldots k) = E \left( (x(k + 1) - \hat{x}(k + 1|0\ldots k))(x(k + 1) - \hat{x}(k + 1|0\ldots k))^T \right)
\]

substituting \( x(k + 1) \) and \( \hat{x}(k + 1|0\ldots k) \) yields

\[
= E \left( F(k)(x(k) - \hat{x}(k|0\ldots k))(x(k) - \hat{x}(k|0\ldots k))^T F(k)^T \right) + 2F(k)(x(k) - \hat{x}(k|0\ldots k))v(k)^T + v(k)v(k)^T
\]

\[
= F(k)E \left( (x(k) - \hat{x}(k|0\ldots k))(x(k) - \hat{x}(k|0\ldots k))^T \right) F(k)^T + E \left( v(k)v(k)^T \right)
\]

\[
= F(k)P(k|0\ldots k)F(k)^T + V(k)
\]
Observing with Probability Distributions: Update Step

Chose \( \hat{x}(k + 1|0\ldots k + 1) \) to be the most likely point in the set

\[
\Omega = \{ x \in \mathbb{R}^n : y(k + 1) = H(k + 1)x \}
\]

Choose \( x \in \Omega \) that maximizes the Gaussian distribution with mean \( \hat{x}(k + 1|0\ldots k) \) and covariance matrix \( P(k + 1|0\ldots k) \), i.e.,

\[
f(x) = \exp \left( -\frac{1}{2} (x - \hat{x}(k + 1|0\ldots k)) P(k + 1|0\ldots k)^{-1} (x - \hat{x}(k + 1|0\ldots k))^T \right) \frac{1}{\sqrt{(2\pi)^n |P(k + 1|0\ldots k)|}}
\]

Choose \( x \in \Omega \) that minimizes

\[
(x - \hat{x}(k + 1|0\ldots k)) P(k + 1|0\ldots k)^{-1} (x - \hat{x}(k + 1|0\ldots k))^T
\]

Define new inner product and (Mahalanobis) distance in \( \mathbb{R}^n \) as

\[
\langle x_1, x_2 \rangle_M = x_1^T P(k + 1|0\ldots k)^{-1} x_2
\]

\[
||x||_M^2 = \langle x, x \rangle_M = x^T P(k + 1|0\ldots k)^{-1} x
\]

Let \( \Delta x = \hat{x}(k + 1|0\ldots k + 1) - \hat{x}(k + 1|0\ldots k) \).
So we want to find \( \hat{x}(k + 1|0\ldots k + 1) \) such that

1. \( ||\Delta x||_M \) is minimized
2. \( (\hat{x}(k + 1|0\ldots k) + \Delta x) \in \Omega \)
Observing with Probability Distributions: Update Step (cont.)

\[ ||\Delta x||_M \text{ is minimized} \]
\[ \implies \Delta x \text{ is orthogonal to } \Omega \text{ according to inner product } \langle \cdot, \cdot \rangle_M \]
\[ \implies \text{For all } a \in \text{NullSpace}(H(k+1)) \]
\[ aP(k+1|0\ldots k)^{-1}(\Delta x) = 0 \]
\[ \implies \Delta x \in \text{ColumnSpace}(P(k+1|0\ldots k)H(k+1)^T) \]
\[ \implies \text{For some } \gamma \in \mathbb{R}^p \]
\[ \Delta x = P(k+1|0\ldots k)H(k+1)^T \gamma \]

Let \( \nu \) denote the innovation error, i.e.,
\[ \nu = y(k+1) - H(k+1)\hat{x}(k+1|0\ldots k) \]

Assume that \( \gamma \) can be written as a linear function of \( \nu \), i.e.,
\[ \gamma = K\nu, \text{ for some } K \in \mathbb{R}^{p \times p} \]

Then \( \Delta x = P(k+1|0\ldots k)H(k+1)^T K\nu \)
\[ (\hat{x}(k+1|0 \ldots k) + \Delta x) \in \Omega \]

\[ \Rightarrow y(k+1) = H(k+1)(\hat{x}(k+1|0 \ldots k) + \Delta x) \]

\[ \Rightarrow H(k+1)\Delta x = \nu \]

\[ \Rightarrow H(k+1)P(k+1|0 \ldots k)H^T K \nu = \nu \]

(since also \( \Delta x = P(k+1|0 \ldots k)H^T K \nu \))

\[ \Rightarrow K = (H(k+1)P(k+1|0 \ldots k)H(k+1)^T)^{-1} \]

Let

\[ R = P(k+1|0 \ldots k)H(k+1)^T K \]

Then, the update for the state vector estimate is

\[ \hat{x}(k+1|0 \ldots k + 1) = \hat{x}(k+1|0 \ldots k) + \Delta x \]

\[ = \hat{x}(k+1|0 \ldots k) + P(k+1|0 \ldots k)H(k+1)^T K \nu \]

\[ = \hat{x}(k+1|0 \ldots k) + R \nu \]

Update for the covariance matrix estimate

\[ P(k+1|0 \ldots k + 1) = P(k+1|0 \ldots k) - RH(k+1)P(k+1|0 \ldots k) \]
Observing with Probability Distributions: Putting it all together

Prediction

\[ \hat{x}(k + 1|0 \ldots k) = F(k)\hat{x}(k|0 \ldots k) + G(k)u(k) \]
\[ P(k + 1|0 \ldots k) = F(k)P(k|0 \ldots k)F(k)^T + V(k) \]

Update

\[ \hat{x}(k + 1|0 \ldots k + 1) = \hat{x}(k + 1|0 \ldots k) + R\nu \]
\[ P(k + 1|0 \ldots k + 1) = P(k + 1|0 \ldots k) - RH(k + 1)P(k + 1|0 \ldots k) \]

where

\[ \nu = y(k + 1) - H(k + 1)\hat{x}(k + 1|0 \ldots k) \]
\[ R = P(k + 1|0 \ldots k)H(k + 1)^T (H(k + 1)P(k + 1|0 \ldots k)H(k + 1)^T)^{-1} \]

What are some problems with this observer?

- Since we assumed no sensor noise, the update equations can cause the covariance matrix estimate to become singular
- But if covariance matrix is singular, Gaussian distribution and Mahalanobis distance are not defined since they rely on the inverse matrix
to address these problems...the kalman filter
Linear Kalman Filter

\[
x(k+1) = F(k)x(k) + G(k)u(k) + v(k)
\]
\[
y(k) = H(k)x(k) + w(k)
\]

Recall that

- \(v(k), w(k)\): white Gaussian noise with zero mean and covariance matrix \(V(k), W(k)\)

Prediction (no changes from before)

\[
\hat{x}(k+1|0\ldots k) = F(k)\hat{x}(k|0\ldots k) + G(k)u(k)
\]
\[
P(k+1|0\ldots k) = F(k)P(k|0\ldots k)F(k)^T + V(k)
\]

Update: Changes due to the sensor noise term \(w(k)\)

- Before, we knew that the constrained the next state estimate to be in \(\Omega\), so we used the equation \(y(k+1) = H(k+1)\hat{x}(k+1|0\ldots k+1)\) to find \(\hat{x}(k+1|0\ldots k+1)\)
- Now we only know that the output is drawn from a Gaussian distribution in \(\mathbb{R}^p\) with mean \(y(k+1)\) and covariance matrix \(W(k)\)
- So will first look for the most likely output \(y^*\) given the prediction \((\hat{x}(k+1|0\ldots k), P(k+1|0\ldots k))\) together with the measured output \(y(k+1)\)
- After that, we can introduce the constraint \(y^* = H(k+1)\hat{x}(k+1|0\ldots k+1)\) and proceed as before
Project the prediction into output space

State space distribution with mean $\hat{x}(k + 1|0 \ldots k)$ and covariance matrix $P(k + 1|0 \ldots k)$ projects into a Gaussian distribution in the output space $\mathbb{R}^p$ with mean

$$\hat{y} = H(k + 1)\hat{x}(k + 1|0 \ldots k)$$

and covariance matrix

$$\hat{W} = E\left[ (\hat{y} - y(k + 1))(\hat{y} - y(k + 1))^T \right]$$

$$= E\left[ H(k + 1)(\hat{x}(k + 1|0 \ldots k) - x(k + 1))(\hat{x}(k + 1|0 \ldots k) - x(k + 1))^T H(k + 1)^T \right]$$

$$= H(k + 1)P(k + 1|0 \ldots k)H(k + 1)^T$$
y* is then the most likely point in the output space $\mathbb{R}^p$ given

- $(\hat{y}, \hat{W})$: Gaussian distribution that results from projection the state prediction
- $(y(k + 1), W(k + 1))$: Gaussian distribution that results from taking the measurement

$y^*$ will be the peak of the function that results from taking their product (since distributions $(\hat{y}, \hat{W})$ and $(y(k + 1), W(k + 1))$ are independent)

**Theorem:** The product of two Gaussians $(z_1, C_1)$ and $(z_2, C_2)$ is proportional to a third Gaussian $(z_3, C_3)$, where

$$z_3 = z_1 + C_1(C_1 + C_2)^{-1}(z_2 - z_1)$$

$$C_3 = C_1 - C_1(C_1 + C_2)^{-1}C_1$$

Then

$$y^* = \hat{y} + \hat{W}(\hat{W} + W(k + 1))^{-1}(y(k + 1) - \hat{y})$$

We can also define

$$\Omega^* = \{x \in \mathbb{R}^n : y^* = H(k + 1)x\}$$

and proceed to find $\Delta x = \hat{x}(k + 1|0 \ldots k + 1) - \hat{x}(k + 1|0 \ldots k)$ that

- minimizes $||\Delta x||_M$ and
- satisfies $\hat{x}(k + 1|0 \ldots k + 1) \in \Omega^*$
1. $||\Delta x||_M$ is minimized

$\implies \Delta x$ is orthogonal to $\Omega^*$ according to inner product $\langle \cdot, \cdot \rangle_M$

$\implies$ For some $\gamma \in \mathbb{R}^p$: $\Delta x = P(k + 1|0 \ldots k)H(k + 1)^T \gamma$

Let $\nu$ be the innovation error

$$\nu = y^* - H(k + 1)\hat{x}(k + 1|0 \ldots k)$$

Assume that $\gamma$ can be written as

$$\gamma = K\nu, \quad \text{for some } K \in \mathbb{R}^{p \times p}$$

Then $\Delta x = P(k + 1|0 \ldots k)H(k + 1)^T K\nu$

2. $(\hat{x}(k + 1|0 \ldots k) + \Delta x) \in \Omega$

$\implies y^* = H(k + 1)(\hat{x}(k + 1|0 \ldots k) + \Delta x)$

$\implies H(k + 1)\Delta x = \nu$

$\implies H(k + 1)P(k + 1|0 \ldots k)H^T K\nu = \nu$

$\implies K = (H(k + 1)P(k + 1|0 \ldots k)H(k + 1)^T)^{-1}$
Therefore, from (1) and (2),

\[ \Delta x = P(k + 1|0 \ldots k)H(k + 1)^T K \nu \]

where

- \( K = (H(k + 1)P(k + 1|0 \ldots k)H(k + 1)^T)^{-1} \)
- \( \nu = y^* - H(k + 1)\hat{x}(k + 1|0 \ldots k) \)
- \( y^* = \hat{y} + \hat{W}(\hat{W} + W(k + 1))^{-1}(y(k + 1) - \hat{y}) \)
- \( \hat{W} = H(k + 1)P(k + 1|0 \ldots k)H(k + 1)^T \)
- \( \hat{y} = H(k + 1)\hat{x}(k + 1|0 \ldots k) \)

Some simplifications:

\[
K \nu = K(y^* - H(k + 1)\hat{x}(k + 1|0 \ldots k))
\]
\[
= K\hat{W}(\hat{W} + W(k + 1))^{-1}(y(k + 1) - \hat{y})
\]
\[
= (\hat{W} + W(k + 1))^{-1}(y(k + 1) - \hat{y})
\]

Therefore, (with the shorthand notation \( H \equiv H(k + 1), P \equiv P(k + 1|0 \ldots k) \))

\[ \Delta x = PH^T(HPH^T + W(k + 1))^{-1}(y(k + 1) - H\hat{x}(k + 1|0 \ldots k)) \]
Linear Kalman Filter: Putting it all together

Prediction

\[
\hat{x}(k+1|0\ldots k) = F(k)\hat{x}(k|0\ldots k) + G(k)u(k)
\]

\[
P(k+1|0\ldots k) = F(k)P(k|0\ldots k)F(k)^T + V(k)
\]

Update

\[
\hat{x}(k+1|0\ldots k+1) = \hat{x}(k+1|0\ldots k) + \Delta x
\]

\[
= \hat{x}(k+1|0\ldots k) + PH^T(HPH^T + W(k+1))^{-1}(y(k+1) - H\hat{x}(k+1|0\ldots k))
\]

\[
P(k+1|0\ldots k+1) = E\left[(x(k+1) - \hat{x}(k+1|0\ldots k+1))(x(k+1) - \hat{x}(k+1|0\ldots k))\right]
\]

where \( H \equiv H(k+1), P \equiv P(k+1|0\ldots k) \)
Consider a mobile robot constrained to move along a straight line. 

Robot state $x = (x_r, v_r)^T$

- $x_r$: robot position
- $v_r$: robot velocity

Input control $u$: real-valued force applied to the robot. According to Newton’s law

$$\frac{dv_r}{dt} = \frac{u}{m}$$

Approximated by the discrete time equation ($T$ discretization rate (in seconds))

$$\frac{v_r(k + 1) - v_r(k)}{T} = \frac{u(k)}{m}$$

Therefore,

$$x(k + 1) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} 0 \\ \frac{T}{m} \end{bmatrix} u(k) + \nu(k)$$

where $\nu(k)$ is white Gaussian noise with zero mean and covariance matrix $V$

Suppose sensor measures velocity. Then,

$$y(k + 1) = [0, 1] x(k) + w(k)$$

where $w(k)$ is white Gaussian noise with zero mean and covariance matrix $W$