Kalman Filtering on $\mathrm{SO}(3)$

Presenter
전준기

스마트미디어연구그룹

Electronics and Telecommunications Research Institute, ETRI

Jan. 31, 2017

Section 1

Kalman Filter

Figure 1: Rudolf Emil Kálmán (May 19, 1930 - July 2, 2016)

Goal

- Tracking problem
  - Find the state of a dynamical system, given the history of observations of the system
  - Filtering: find the **present** state
  - Smoothing: find **past** states
  - Prediction: find **future** states
- Control problem
  - Try to make a dynamical system into a desired state by applying certain actions, given the history of observations of the system
- States, observations, and actions are in some multi-dimensional continua

Goal

- Tracking problem
  - Find the state of a dynamical system, given the history of observations of the system
  - Filtering: find the **present** state
  - Smoothing: find **past** states
  - Prediction: find **future** states
- Control problem
  - Try to make a dynamical system into a desired state by applying certain actions, given the history of observations of the system
- States, observations, and actions are in some multi-dimensional continua

Goal

- Tracking problem
  - Find the state of a dynamical system, given the history of observations of the system
  - Filtering: find the **present** state
  - Smoothing: find **past** states
  - Prediction: find **future** states
- Control problem
  - Try to make a dynamical system into a desired state by applying certain actions, given the history of observations of the system
- States, observations, and actions are in some multi-dimensional continua

Goal

- Tracking problem
  - Find the state of a dynamical system, given the history of observations of the system
  - Filtering: find the **present** state
  - Smoothing: find **past** states
  - Prediction: find **future** states
- Control problem
  - Try to make a dynamical system into a desired state by applying certain actions, given the history of observations of the system
- States, observations, and actions are in some multi-dimensional continua

Goal

- Tracking problem
  - Find the state of a dynamical system, given the history of observations of the system
  - Filtering: find the **present** state
  - Smoothing: find **past** states
  - Prediction: find **future** states
- Control problem
  - Try to make a dynamical system into a desired state by applying certain actions, given the history of observations of the system
- States, observations, and actions are in some multi-dimensional continua

Example of System Model

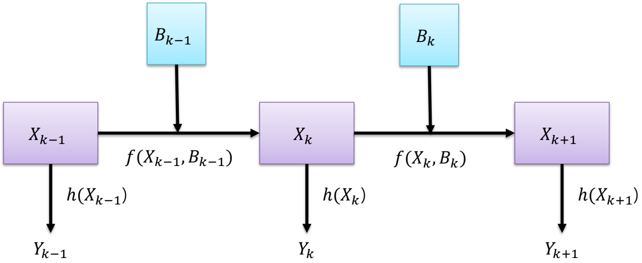


Figure 2: Discrete-Time Control System

X_k: system's state at the kth time instance
B_k: controller's action at the kth time instance
Y_k: observation of the system at the kth time instance

Approaches

- Deterministic approaches
  - "Solve" the equations
  - Minimize corresponding "cost function"

- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain "form" of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant

- Kalman's approach: parametric Bayesian

Approaches

- Deterministic approaches
  - "Solve" the equations
  - Minimize corresponding "cost function"
- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain "form" of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant
- Kalman's approach: parametric Bayesian

Approaches

- Deterministic approaches
  - "Solve" the equations
  - Minimize corresponding "cost function"
- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain "form" of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant
- Kalman's approach: parametric Bayesian

Approaches

- Deterministic approaches
  - "Solve" the equations
  - Minimize corresponding "cost function"
- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain "form" of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant
- Kalman's approach: parametric Bayesian

Approaches

- Deterministic approaches
  - "Solve" the equations
  - Minimize corresponding "cost function"
- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain "form" of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant
- Kalman's approach: parametric Bayesian

Approaches

- Deterministic approaches
  - “Solve” the equations
  - Minimize corresponding “cost function”
- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain “form” of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant
- Kalman’s approach: parametric Bayesian

Approaches

- Deterministic approaches
  - "Solve" the equations
  - Minimize corresponding "cost function"
- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain "form" of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant
- Kalman's approach: parametric Bayesian

Approaches

- Deterministic approaches
  - "Solve" the equations
  - Minimize corresponding "cost function"
- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain "form" of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant
- Kalman's approach: parametric Bayesian

Approaches

- Deterministic approaches
  - "Solve" the equations
  - Minimize corresponding "cost function"
- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain "form" of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant
- Kalman's approach: parametric Bayesian

Approaches

- Deterministic approaches
  - "Solve" the equations
  - Minimize corresponding "cost function"
- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain "form" of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant
- Kalman's approach: parametric Bayesian

Approaches

- Deterministic approaches
  - “Solve” the equations
  - Minimize corresponding “cost function”

- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain “form” of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant

- Kalman’s approach: parametric Bayesian

Approaches

- Deterministic approaches
  - "Solve" the equations
  - Minimize corresponding "cost function"
- Probabilistic approaches
  - Parametric vs. Non-parametric
    - Parametric: assumes certain "form" of probability measures
    - Non-parametric: try to find the probability measure itself
  - Bayesian vs. Non-Bayesian
    - Bayesian: parameter itself is a random variable
    - Non-Bayesian: parametr is a fixed unknown constant
- Kalman's approach: parametric Bayesian

Kalman Filter (1)

- Goal: estimate the state of a system given by a time-series $(X_k)_{k=0}^{\infty}$, from observations $(Y_k)_{k=0}^{\infty}$
- Assumptions:
  - The system evolves linearly: $X_k = FX_{k-1} + W_k$
    - F: a known, fixed linear transformation
    - W_k: **process noise**, driving the system randomly
  - The observation is derived linearly from the state: $Y_k = HX_k + U_k$
    - H: a known, fixed linear transformation
    - U_k: **measurement noise**, making the observation imprecise
  - Further assumptions:
    - X_0, W_k's, U_k's are all independent and all zero-mean Gaussian
    - W_k's are identically distributed w/ covariance Q
    - U_k's are identically distributed w/ covariance R

Kalman Filter (1)

- Goal: estimate the state of a system given by a time-series $(X_k)_{k=0}^{\infty}$, from observations $(Y_k)_{k=0}^{\infty}$
- Assumptions:
  - The system evolves linearly: $X_k = FX_{k-1} + W_k$
    - F: a known, fixed linear transformation
    - W_k: **process noise**, driving the system randomly
  - The observation is derived linearly from the state: $Y_k = HX_k + U_k$
    - H: a known, fixed linear transformation
    - U_k: **measurement noise**, making the observation imprecise
  - Further assumptions:
    - X_0, W_k's, U_k's are all independent and all zero-mean Gaussian
    - W_k's are identically distributed w/ covariance Q
    - U_k's are identically distributed w/ covariance R

Kalman Filter (1)

- Goal: estimate the state of a system given by a time-series $(X_k)_{k=0}^{\infty}$, from observations $(Y_k)_{k=0}^{\infty}$
- Assumptions:
  - The system evolves linearly: $X_k = FX_{k-1} + W_k$
    - F: a known, fixed linear transformation
    - W_k: **process noise**, driving the system randomly
  - The observation is derived linearly from the state: $Y_k = HX_k + U_k$
    - H: a known, fixed linear transformation
    - U_k: **measurement noise**, making the observation imprecise
  - Further assumptions:
    - X_0, W_k's, U_k's are all independent and all zero-mean Gaussian
    - W_k's are identically distributed w/ covariance Q
    - U_k's are identically distributed w/ covariance R

Kalman Filter (1)

- Goal: estimate the state of a system given by a time-series $(X_k)_{k=0}^{\infty}$, from observations $(Y_k)_{k=0}^{\infty}$
- Assumptions:
  - The system evolves linearly: $X_k = FX_{k-1} + W_k$
    - F: a known, fixed linear transformation
    - W_k: **process noise**, driving the system randomly
  - The observation is derived linearly from the state: $Y_k = HX_k + U_k$
    - H: a known, fixed linear transformation
    - U_k: **measurement noise**, making the observation imprecise
  - Further assumptions:
    - X_0, W_k's, U_k's are all independent and all zero-mean Gaussian
    - W_k's are identically distributed w/ covariance Q
    - U_k's are identically distributed w/ covariance R

Kalman Filter (1)

- Goal: estimate the state of a system given by a time-series $(X_k)_{k=0}^{\infty}$, from observations $(Y_k)_{k=0}^{\infty}$
- Assumptions:
  - The system evolves linearly: $X_k = FX_{k-1} + W_k$
    - F: a known, fixed linear transformation
    - W_k: **process noise**, driving the system randomly
  - The observation is derived linearly from the state: $Y_k = HX_k + U_k$
    - H: a known, fixed linear transformation
    - U_k: **measurement noise**, making the observation imprecise
  - Further assumptions:
    - X_0, W_k's, U_k's are all independent and all zero-mean Gaussian
    - W_k's are identically distributed w/ covariance Q
    - U_k's are identically distributed w/ covariance R

Example - Speed Camera

- Want to know: the *speed* of the car
- Observed: the *position* of the car

- $X_k = \begin{bmatrix} P_k \\ V_k \end{bmatrix}$
  - P_k: position at $t = k\Delta t$
  - V_k: average velocity between $t = k\Delta t$ and $t = (k+1)\Delta t$
- $\begin{bmatrix} P_k \\ V_k \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_{k-1} \\ V_{k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ A_k \Delta t \end{bmatrix}$
  - A_k: random acceleration of the car
- $Y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} P_k \\ V_k \end{bmatrix} + U_k$
  - U_k: sensor noise

Example - Speed Camera

- Want to know: the *speed* of the car
- Observed: the *position* of the car

- $X_k = \begin{bmatrix} P_k \\ V_k \end{bmatrix}$
  - P_k: position at $t = k\Delta t$
  - V_k: average velocity between $t = k\Delta t$ and $t = (k+1)\Delta t$
- $\begin{bmatrix} P_k \\ V_k \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_{k-1} \\ V_{k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ A_k \Delta t \end{bmatrix}$
  - A_k: random acceleration of the car
- $Y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} P_k \\ V_k \end{bmatrix} + U_k$
  - U_k: sensor noise

Example - Speed Camera

- Want to know: the *speed* of the car
- Observed: the *position* of the car

- $X_k = \begin{bmatrix} P_k \\ V_k \end{bmatrix}$
  - P_k: position at $t = k\Delta t$
  - V_k: average velocity between $t = k\Delta t$ and $t = (k+1)\Delta t$
- $\begin{bmatrix} P_k \\ V_k \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_{k-1} \\ V_{k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ A_k \Delta t \end{bmatrix}$
  - A_k: random acceleration of the car
- $Y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} P_k \\ V_k \end{bmatrix} + U_k$
  - U_k: sensor noise

Example - Speed Camera

- Want to know: the *speed* of the car
- Observed: the *position* of the car

- $X_k = \begin{bmatrix} P_k \\ V_k \end{bmatrix}$
  - P_k: position at $t = k\Delta t$
  - V_k: average velocity between $t = k\Delta t$ and $t = (k+1)\Delta t$
- $\begin{bmatrix} P_k \\ V_k \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_{k-1} \\ V_{k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ A_k \Delta t \end{bmatrix}$
  - A_k: random acceleration of the car
- $Y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} P_k \\ V_k \end{bmatrix} + U_k$
  - U_k: sensor noise

Kalman Filter (2)

- What is the "best" estimate of X_k given $Y_1, \cdots, Y_k$?
- Since we are doing Bayesian estimate, *average cost* is the concern

Theorem 1

Given L^2-r.v. X and a r.v. Y, a conditional expectation of X given Y is an MMSE (Minimum Mean-Square Error) estimate of X given Y; that is, for any measurable function $f : \mathcal{Y} \to \mathcal{X}$,

$$\mathrm{E}\left[\|X - \mathrm{E}[X|Y]\|^2\right] \leq \mathrm{E}\left[\|X - f(Y)\|^2\right]$$

- Suffices to find $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$

Kalman Filter (2)

- What is the "best" estimate of X_k given $Y_1, \cdots, Y_k$?
- Since we are doing Bayesian estimate, *average cost* is the concern

Theorem 1

Given L^2-r.v. X and a r.v. Y, a conditional expectation of X given Y is an MMSE (Minimum Mean-Square Error) estimate of X given Y; that is, for any measurable function $f : \mathcal{Y} \to \mathcal{X}$,

$$\mathrm{E}\left[\|X - \mathrm{E}[X|Y]\|^2\right] \leq \mathrm{E}\left[\|X - f(Y)\|^2\right]$$

- Suffices to find $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$

Kalman Filter (2)

- What is the "best" estimate of X_k given $Y_1, \cdots, Y_k$?
- Since we are doing Bayesian estimate, *average cost* is the concern

Theorem 1

Given L^2-r.v. X and a r.v. Y, a conditional expectation of X given Y is an MMSE (Minimum Mean-Square Error) estimate of X given Y; that is, for any measurable function $f : \mathcal{Y} \to \mathcal{X}$,

$$\mathrm{E}\left[\|X - \mathrm{E}[X|Y]\|^2\right] \leq \mathrm{E}\left[\|X - f(Y)\|^2\right]$$

- Suffices to find $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$

Kalman Filter (2)

- What is the "best" estimate of X_k given $Y_1, \cdots, Y_k$?
- Since we are doing Bayesian estimate, *average cost* is the concern

Theorem 1

Given L^2-r.v. X and a r.v. Y, a conditional expectation of X given Y is an MMSE (Minimum Mean-Square Error) estimate of X given Y; that is, for any measurable function $f : \mathcal{Y} \to \mathcal{X}$,

$$\mathrm{E}\left[\|X - \mathrm{E}[X|Y]\|^2\right] \leq \mathrm{E}\left[\|X - f(Y)\|^2\right]$$

- Suffices to find $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$

Kalman Filter (3)

- If $(X_1, X_2) \sim \mathrm{N}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$, then

$$(X_1 | X_2 = x_2) \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

- What's the matter?
  - Intractable to directly calculate $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$
    - Infinite (indefinitely growing) memory requirement
    - Joint distribution of $(X_k, Y_1, \cdots, Y_k)$ is too complicated
- Have to "summarize" the results before taking Y_k into account
- It turns out, $\hat{X}_{k-1|k-1} := \mathrm{E}[X_{k-1} | Y_1, \cdots, Y_{k-1}]$ and $P_{k-1|k-1} := \mathrm{E}[(X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T]$ are sufficient summaries

Kalman Filter (3)

- If $(X_1, X_2) \sim \mathrm{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix} \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$, then

$$(X_1 | X_2 = x_2) \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

- What's the matter?
  - Intractable to directly calculate $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$
    - Infinite (indefinitely growing) memory requirement
    - Joint distribution of $(X_k, Y_1, \cdots, Y_k)$ is too complicated
- Have to "summarize" the results before taking Y_k into account
- It turns out, $\hat{X}_{k-1|k-1} := \mathrm{E}[X_{k-1} | Y_1, \cdots, Y_{k-1}]$ and $P_{k-1|k-1} := \mathrm{E}[(X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T]$ are sufficient summaries

Kalman Filter (3)

- If $(X_1, X_2) \sim \mathrm{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix} \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$, then

$$(X_1 | X_2 = x_2) \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

- What's the matter?
  - Intractable to directly calculate $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$
    - Infinite (indefinitely growing) memory requirement
    - Joint distribution of $(X_k, Y_1, \cdots, Y_k)$ is too complicated
- Have to "summarize" the results before taking Y_k into account
- It turns out, $\hat{X}_{k-1|k-1} := \mathrm{E}[X_{k-1} | Y_1, \cdots, Y_{k-1}]$ and $P_{k-1|k-1} := \mathrm{E}[(X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T]$ are sufficient summaries

Kalman Filter (3)

- If $(X_1, X_2) \sim \mathrm{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix} \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$, then

$$(X_1 | X_2 = x_2) \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

- What's the matter?
  - Intractable to directly calculate $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$
    - Infinite (indefinitely growing) memory requirement
    - Joint distribution of $(X_k, Y_1, \cdots, Y_k)$ is too complicated
- Have to "summarize" the results before taking Y_k into account
- It turns out, $\hat{X}_{k-1|k-1} := \mathrm{E}[X_{k-1} | Y_1, \cdots, Y_{k-1}]$ and $P_{k-1|k-1} := \mathrm{E}[(X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T]$ are sufficient summaries

Kalman Filter (3)

- If $(X_1, X_2) \sim \mathrm{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix} \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$, then

$$(X_1 | X_2 = x_2) \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

- What's the matter?
  - Intractable to directly calculate $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$
    - Infinite (indefinitely growing) memory requirement
    - Joint distribution of $(X_k, Y_1, \cdots, Y_k)$ is too complicated
- Have to "summarize" the results before taking Y_k into account
- It turns out, $\hat{X}_{k-1|k-1} := \mathrm{E}[X_{k-1} | Y_1, \cdots, Y_{k-1}]$ and $P_{k-1|k-1} := \mathrm{E}[(X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T]$ are sufficient summaries

Kalman Filter (3)

- If $(X_1, X_2) \sim \mathrm{N}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$, then

$$(X_1 | X_2 = x_2) \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

- What's the matter?
  - Intractable to directly calculate $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$
    - Infinite (indefinitely growing) memory requirement
    - Joint distribution of $(X_k, Y_1, \cdots, Y_k)$ is too complicated
- Have to "summarize" the results before taking Y_k into account
- It turns out, $\hat{X}_{k-1|k-1} := \mathrm{E}[X_{k-1} | Y_1, \cdots, Y_{k-1}]$ and $P_{k-1|k-1} := \mathrm{E}[(X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T]$ are sufficient summaries

Kalman Filter (3)

- If $(X_1, X_2) \sim \mathrm{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix} \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$, then

$$(X_1 | X_2 = x_2) \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

- What's the matter?
  - Intractable to directly calculate $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$
    - Infinite (indefinitely growing) memory requirement
    - Joint distribution of $(X_k, Y_1, \cdots, Y_k)$ is too complicated
- Have to "summarize" the results before taking Y_k into account
- It turns out, $\hat{X}_{k-1|k-1} := \mathrm{E}[X_{k-1} | Y_1, \cdots, Y_{k-1}]$ and $P_{k-1|k-1} := \mathrm{E}[(X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T]$ are sufficient summaries

Kalman Filter (4)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

1. **Prediction phase**: compute $\hat{X}_{k|k-1}$, $P_{k|k-1}$ from $\hat{X}_{k-1|k-1}$, $P_{k-1|k-1}$ using the formulas
   - $\hat{X}_{k|k-1} = F\hat{X}_{k-1|k-1}$
   - $P_{k|k-1} = FP_{k-1|k-1}F^T + Q$
2. **Update phase**: compute $\hat{X}_{k|k}$, $P_{k|k}$ from $\hat{X}_{k|k-1}$, $P_{k|k-1}$, and Y_k using the formulas
   - $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
   - $\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Y_k - H\hat{X}_{k|k-1})$
   - $P_{k|k} = (\mathrm{I} - K_kH)P_{k|k-1}$

Kalman Filter (4)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

1. **Prediction phase**: compute $\hat{X}_{k|k-1}$, $P_{k|k-1}$ from $\hat{X}_{k-1|k-1}$, $P_{k-1|k-1}$ using the formulas
   - $\hat{X}_{k|k-1} = F\hat{X}_{k-1|k-1}$
   - $P_{k|k-1} = FP_{k-1|k-1}F^T + Q$
2. **Update phase**: compute $\hat{X}_{k|k}$, $P_{k|k}$ from $\hat{X}_{k|k-1}$, $P_{k|k-1}$, and Y_k using the formulas
   - $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
   - $\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Y_k - H\hat{X}_{k|k-1})$
   - $P_{k|k} = (\mathrm{I} - K_kH)P_{k|k-1}$

Kalman Filter (4)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

1. **Prediction phase**: compute $\hat{X}_{k|k-1}$, $P_{k|k-1}$ from $\hat{X}_{k-1|k-1}$, $P_{k-1|k-1}$ using the formulas
   - $\hat{X}_{k|k-1} = F\hat{X}_{k-1|k-1}$
   - $P_{k|k-1} = FP_{k-1|k-1}F^T + Q$
2. **Update phase**: compute $\hat{X}_{k|k}$, $P_{k|k}$ from $\hat{X}_{k|k-1}$, $P_{k|k-1}$, and Y_k using the formulas
   - $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
   - $\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Y_k - H\hat{X}_{k|k-1})$
   - $P_{k|k} = (\mathrm{I} - K_kH)P_{k|k-1}$

Good Things and Bad Things

- Good things
  - Computationally cheap; several matrix multiplications and an inversion
  - Small memory requirement; only necessary to remember $\hat{X}_{k|k}$ and $P_{k|k}$
  - Optimal w.r.t. MSE
- Bad things
  - Too limited applications
    - Without further assumptions: lose optimality
    - Linear system? Extremely rare...

Good Things and Bad Things

- Good things
  - Computationally cheap; several matrix multiplications and an inversion
  - Small memory requirement; only necessary to remember $\hat{X}_{k|k}$ and $P_{k|k}$
  - Optimal w.r.t. MSE
- Bad things
  - Too limited applications
    - Without further assumptions: lose optimality
    - Linear system? Extremely rare...

Good Things and Bad Things

- Good things
  - Computationally cheap; several matrix multiplications and an inversion
  - Small memory requirement; only necessary to remember $\hat{X}_{k|k}$ and $P_{k|k}$
  - Optimal w.r.t. MSE
- Bad things
  - Too limited applications
    - Without further assumptions: lose optimality
    - Linear system? Extremely rare...

Good Things and Bad Things

- Good things
  - Computationally cheap; several matrix multiplications and an inversion
  - Small memory requirement; only necessary to remember $\hat{X}_{k|k}$ and $P_{k|k}$
  - Optimal w.r.t. MSE
- Bad things
  - Too limited applications
    - Without further assumptions: lose optimality
    - Linear system? Extremely rare...

Good Things and Bad Things

- Good things
  - Computationally cheap; several matrix multiplications and an inversion
  - Small memory requirement; only necessary to remember $\hat{X}_{k|k}$ and $P_{k|k}$
  - Optimal w.r.t. MSE
- Bad things
  - Too limited applications
    - Without further assumptions: lose optimality
    - Linear system? Extremely rare...

Good Things and Bad Things

- Good things
  - Computationally cheap; several matrix multiplications and an inversion
  - Small memory requirement; only necessary to remember $\hat{X}_{k|k}$ and $P_{k|k}$
  - Optimal w.r.t. MSE
- Bad things
  - Too limited applications
    - Without further assumptions: lose optimality
    - Linear system? Extremely rare...

How to Go Beyond? (1)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- Filtering process:
  1. **Prediction phase**:
     - $\hat{X}_{k|k-1} = F\hat{X}_{k-1|k-1}$
     - $P_{k|k-1} = FP_{k-1|k-1}F^T + Q$
  2. **Update phase**:
     - $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
     - $\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Y_k - H\hat{X}_{k|k-1})$
     - $P_{k|k} = (\mathrm{I} - K_kH)P_{k|k-1}$
- Observation: no difference when F, H can vary over time, if we know them exactly
- Replace F to $Df_{\hat{X}_{k-1|k-1}}$, H to $Dh_{\hat{X}_{k|k-1}}$
  $\Rightarrow$ **Extended Kalman Filter (EKF)**

How to Go Beyond? (1)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- Filtering process:
  1. **Prediction phase**:
     - $\hat{X}_{k|k-1} = F\hat{X}_{k-1|k-1}$
     - $P_{k|k-1} = FP_{k-1|k-1}F^T + Q$
  2. **Update phase**:
     - $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
     - $\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Y_k - H\hat{X}_{k|k-1})$
     - $P_{k|k} = (\mathrm{I} - K_kH)P_{k|k-1}$
- Observation: no difference when F, H can vary over time, if we know them exactly
- Replace F to $Df_{\hat{X}_{k-1|k-1}}$, H to $Dh_{\hat{X}_{k|k-1}}$
  $\Rightarrow$ **Extended Kalman Filter (EKF)**

How to Go Beyond? (1)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- Filtering process:
  1. **Prediction phase**:
     - $\hat{X}_{k|k-1} = F\hat{X}_{k-1|k-1}$
     - $P_{k|k-1} = FP_{k-1|k-1}F^T + Q$
  2. **Update phase**:
     - $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
     - $\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Y_k - H\hat{X}_{k|k-1})$
     - $P_{k|k} = (\mathrm{I} - K_kH)P_{k|k-1}$
- Observation: no difference when F, H can vary over time, if we know them exactly
- Replace F to $Df_{\hat{X}_{k-1|k-1}}$, H to $Dh_{\hat{X}_{k|k-1}}$
  $\Rightarrow$ **Extended Kalman Filter (EKF)**

How to Go Beyond? (2)

- Problems of EKF
  - VERY sensitive when $\hat{X}$ approaches to singularities
  - Calculation of the Jacobian matrices could be extremely complicated
  - Cannot force constraints
- There are other alternatives:
  - Unscented Kalman Filter (UKF)
  - Particle Filter
  - Moving Horizon Filter
  - Etc...

How to Go Beyond? (2)

- Problems of EKF
  - VERY sensitive when $\hat{X}$ approaches to singularities
  - Calculation of the Jacobian matrices could be extremely complicated
  - Cannot force constraints
- There are other alternatives:
  - Unscented Kalman Filter (UKF)
  - Particle Filter
  - Moving Horizon Filter
  - Etc...

How to Go Beyond? (2)

- Problems of EKF
  - VERY sensitive when $\hat{X}$ approaches to singularities
  - Calculation of the Jacobian matrices could be extremely complicated
  - Cannot force constraints
- There are other alternatives:
  - Unscented Kalman Filter (UKF)
  - Particle Filter
  - Moving Horizon Filter
  - Etc...

How to Go Beyond? (2)

- Problems of EKF
  - VERY sensitive when $\hat{X}$ approaches to singularities
  - Calculation of the Jacobian matrices could be extremely complicated
  - Cannot force constraints

- There are other alternatives:
  - Unscented Kalman Filter (UKF)
  - Particle Filter
  - Moving Horizon Filter
  - Etc...

How to Go Beyond? (2)

- Problems of EKF
  - VERY sensitive when $\hat{X}$ approaches to singularities
  - Calculation of the Jacobian matrices could be extremely complicated
  - Cannot force constraints

- There are other alternatives:
  - Unscented Kalman Filter (UKF)
  - Particle Filter
  - Moving Horizon Filter
  - Etc...

Section 2

Attitude Tracking

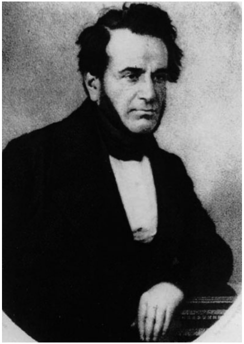


Figure 3: Olinde Rodrigues (October 6, 1795 - December 17, 1851)

Problem Definition (1)

- What "attitude" means?
  - The coordinates of three **frame vectors** with respect to the global coordinate system

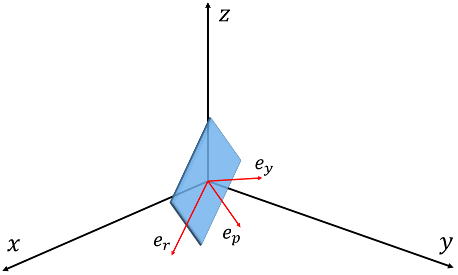
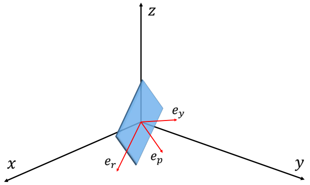


Figure 4: Frame vectors

Problem Definition (2)

- What "attitude" means?
  - (Frame vectors) = **(Rotation matrix)**
  - $r := \begin{bmatrix} e_r & e_p & e_y \end{bmatrix}^T$
  - (Rotation matrix) = **(Orthogonal matrix w/ det.=1)**

  - The set of orthogonal matrices w/ det.=1 is called the **special orthogonal group** and denoted as $\mathrm{SO}(3)$
  - In summary, we are to find an element in $\mathrm{SO}(3)$



Problem Definition (3)

- What we have?
  - **Inertial Measurement Unit (IMU)**: combination of the following three sensors:
    - **Gyroscope**: measures **angular velocity**
    - **Accelerometer**: measures **acceleration**
    - **Magnetometer**: measures **magnetic field**
  - Accelerometer measures the *gravity*
  - Magnetometr measures the *heading*; think of compass

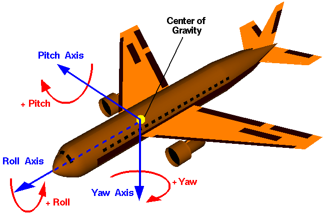


Figure 5: The output of gyroscope sensors

Problem Definition (3)

- What we have?
  - **Inertial Measurement Unit (IMU)**: combination of the following three sensors:
    - **Gyroscope**: measures **angular velocity**
    - **Accelerometer**: measures **acceleration**
    - **Magnetometer**: measures **magnetic field**
  - Accelerometer measures the *gravity*
  - Magnetometr measures the *heading*; think of compass

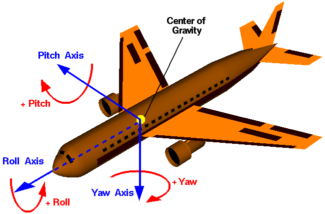


Figure 5: The output of gyroscope sensors

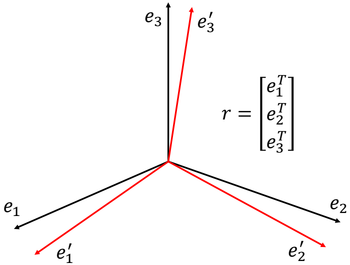
Problem Definition (4)

- Gyroscope measures change of frame vectors *with respect to the local frame vectors*

$$re'_1 \approx \begin{bmatrix}1\\0\\0\end{bmatrix} + \omega_3 \Delta t \begin{bmatrix}0\\1\\0\end{bmatrix} - \omega_2 \Delta t \begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$re'_2 \approx \begin{bmatrix}0\\1\\0\end{bmatrix} + \omega_1 \Delta t \begin{bmatrix}0\\0\\1\end{bmatrix} - \omega_3 \Delta t \begin{bmatrix}1\\0\\0\end{bmatrix}$$

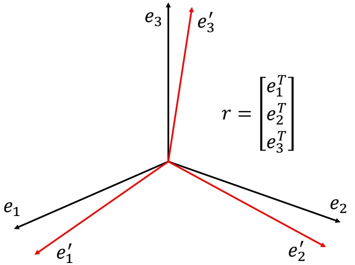
$$re'_3 \approx \begin{bmatrix}0\\0\\1\end{bmatrix} + \omega_2 \Delta t \begin{bmatrix}1\\0\\0\end{bmatrix} - \omega_1 \Delta t \begin{bmatrix}0\\1\\0\end{bmatrix}$$



Problem Definition (5)

- Gyroscope measures change of frame vectors *with respect to the local frame vectors*

$$r \begin{bmatrix} e_1' & e_2' & e_3' \end{bmatrix} \approx I + \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \Delta t \approx \exp\left([\omega]_\times \Delta t\right)$$



- Hence, $r(r')^T \approx \exp\left([\omega]_\times \Delta t\right)$, $r' \approx \exp\left(-[\omega]_\times \Delta t\right) r$

Problem Definition (6)

- Evolution equations:

$$\textbf{Attitude} \quad R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$$
$$\textbf{Angular velocity} \quad \Omega_k = \Omega_{k-1} + A_k \Delta t$$

  where A_k is a random *process noise*

- Measurement equations:

$$\textbf{Gyroscope} \quad G_k = \Omega_k + U_k$$
$$\textbf{Accelerometer} \quad A_k = R_k \mathbf{a} + V_k$$
$$\textbf{Magnetometer} \quad M_k = R_k \mathbf{m} + W_k$$

  where U_k, V_k, W_k are random *measurement noises*, $\mathbf{a}$ the constant gravity vector, and $\mathbf{m}$ the constant Earth magnetic field vector

- Valid only when
  - There is no rapid movement
  - There is no magnetic disturbance

Problem Definition (6)

- Evolution equations:

$$\begin{aligned} \textbf{Attitude} \quad R_k &= \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1} \\ \textbf{Angular velocity} \quad \Omega_k &= \Omega_{k-1} + A_k \Delta t \end{aligned}$$

  where A_k is a random *process noise*

- Measurement equations:

$$\begin{aligned} \textbf{Gyroscope} \quad G_k &= \Omega_k + U_k \\ \textbf{Accelerometer} \quad A_k &= R_k \mathbf{a} + V_k \\ \textbf{Magnetometer} \quad M_k &= R_k \mathbf{m} + W_k \end{aligned}$$

  where U_k, V_k, W_k are random *measurement noises*, $\mathbf{a}$ the constant gravity vector, and $\mathbf{m}$ the constant Earth magnetic field vector

- Valid only when
  - There is no rapid movement
  - There is no magnetic disturbance

Problem Definition (6)

- Evolution equations:

$$\begin{aligned} \textbf{Attitude} \quad & R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1} \\ \textbf{Angular velocity} \quad & \Omega_k = \Omega_{k-1} + A_k \Delta t \end{aligned}$$

  where A_k is a random *process noise*

- Measurement equations:

$$\begin{aligned} \textbf{Gyroscope} \quad & G_k = \Omega_k + U_k \\ \textbf{Accelerometer} \quad & A_k = R_k \mathbf{a} + V_k \\ \textbf{Magnetometer} \quad & M_k = R_k \mathbf{m} + W_k \end{aligned}$$

  where U_k, V_k, W_k are random *measurement noises*, $\mathbf{a}$ the constant gravity vector, and $\mathbf{m}$ the constant Earth magnetic field vector

- Valid only when
  - There is no rapid movement
  - There is no magnetic disturbance

Problem Definition (7)

- System model:
  - $R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$
  - $\Omega_k = \Omega_{k-1} + A_k \Delta t$
  - $G_k = \Omega_k + U_k$
  - $A_k = R_k \mathbf{a} + V_k$
  - $M_k = R_k \mathbf{m} + W_k$
- Goal: find R_k given Y_1^k, where $Y_k := (G_k, A_k, M_k)$
- We may assume A_k, U_k, V_k, W_k are all independent *isotropic i.i.d. Gaussian process*

Problem Definition (7)

- System model:
  - $R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$
  - $\Omega_k = \Omega_{k-1} + A_k \Delta t$
  - $G_k = \Omega_k + U_k$
  - $A_k = R_k \mathbf{a} + V_k$
  - $M_k = R_k \mathbf{m} + W_k$

- Goal: find R_k given Y_1^k, where $Y_k := (G_k, A_k, M_k)$

- We may assume A_k, U_k, V_k, W_k are all independent *isotropic i.i.d. Gaussian process*

Problem Definition (7)

- System model:
  - $R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$
  - $\Omega_k = \Omega_{k-1} + A_k \Delta t$
  - $G_k = \Omega_k + U_k$
  - $A_k = R_k \mathbf{a} + V_k$
  - $M_k = R_k \mathbf{m} + W_k$
- Goal: find R_k given Y_1^k, where $Y_k := (G_k, A_k, M_k)$
- We may assume A_k, U_k, V_k, W_k are all independent *isotropic i.i.d. Gaussian process*

State Variables Are Not in $\mathbb{R}^n$

- There is no "conditional expectation"
  - But we can find MMSE estimator *if* we know the **conditional distribution**
- There is no "Gaussian distribution"
  - Why Gaussian distribution is so nice?
    - Very stable under various kinds of transformations
      - Affine transforms
      - Conditioning
      - Etc.
    - Parametrized
    - Physically meaningful
      - Central limit theorem
      - Brownian motion, heat kernel, diffusion kernel, or related
      - Maximum entropy under energy constraint
      - Etc.

State Variables Are Not in $\mathbb{R}^n$

- There is no "conditional expectation"
  - But we can find MMSE estimator *if* we know the **conditional distribution**
- There is no "Gaussian distribution"
  - Why Gaussian distribution is so nice?
    - Very stable under various kinds of transformations
      - Affine transforms
      - Conditioning
      - Etc.
    - Parametrized
    - Physically meaningful
      - Central limit theorem
      - Brownian motion, heat kernel, diffusion kernel, or related
      - Maximum entropy under energy constraint
      - Etc.

State Variables Are Not in $\mathbb{R}^n$

- There is no "conditional expectation"
  - But we can find MMSE estimator *if* we know the **conditional distribution**
- There is no "Gaussian distribution"
  - Why Gaussian distribution is so nice?
    - Very stable under various kinds of transformations
      - Affine transforms
      - Conditioning
      - Etc.
    - Parametrized
    - Physically meaningful
      - Central limit theorem
      - Brownian motion, heat kernel, diffusion kernel, or related
      - Maximum entropy under energy constraint
      - Etc.

State Variables Are Not in $\mathbb{R}^n$

- There is no "conditional expectation"
  - But we can find MMSE estimator *if* we know the **conditional distribution**
- There is no "Gaussian distribution"
  - Why Gaussian distribution is so nice?
    - Very stable under various kinds of transformations
      - Affine transforms
      - Conditioning
      - Etc.
    - Parametrized
    - Physically meaningful
      - Central limit theorem
      - Brownian motion, heat kernel, diffusion kernel, or related
      - Maximum entropy under energy constraint
      - Etc.

State Variables Are Not in $\mathbb{R}^n$

- There is no "conditional expectation"
  - But we can find MMSE estimator *if* we know the **conditional distribution**
- There is no "Gaussian distribution"
  - Why Gaussian distribution is so nice?
    - Very stable under various kinds of transformations
      - Affine transforms
      - Conditioning
      - Etc.
    - Parametrized
    - Physically meaningful
      - Central limit theorem
      - Brownian motion, heat kernel, diffusion kernel, or related
      - Maximum entropy under energy constraint
      - Etc.

State Variables Are Not in $\mathbb{R}^n$

- There is no "conditional expectation"
  - But we can find MMSE estimator *if* we know the **conditional distribution**
- There is no "Gaussian distribution"
  - Why Gaussian distribution is so nice?
    - Very stable under various kinds of transformations
      - Affine transforms
      - Conditioning
      - Etc.
    - Parametrized
    - Physically meaningful
      - Central limit theorem
      - Brownian motion, heat kernel, diffusion kernel, or related
      - Maximum entropy under energy constraint
      - Etc.

State Variables Are Not in $\mathbb{R}^n$

- There is no "conditional expectation"
  - But we can find MMSE estimator *if* we know the **conditional distribution**
- There is no "Gaussian distribution"
  - Why Gaussian distribution is so nice?
    - Very stable under various kinds of transformations
      - Affine transforms
      - Conditioning
      - Etc.
    - Parametrized
    - Physically meaningful
      - Central limit theorem
      - Brownian motion, heat kernel, diffusion kernel, or related
      - Maximum entropy under energy constraint
      - Etc.

Drift-Diffusion Equation (1)

- What is the most natural generalization of Gaussian measures on Lie groups?

Theorem 2 (Drift-Diffusion Equation)

The solution $f : [0,\infty) \times \mathbb{R}^n \to \mathbb{C}$ to the differential equation

$$\begin{cases} \frac{\partial f(t,x)}{\partial t} = \sum_{i,j=1}^n a_{ij} \frac{\partial^2 f(x,t)}{\partial x_i \partial x_j} - \sum_{i=1}^n b_i \frac{\partial f(x,t)}{\partial x_i} \\ f(0,x) = f_0(x), \quad f_0 : \mathbb{R}^n \to \mathbb{C} \end{cases}$$

where $A := [a_{ij}]$ is positive-semidefinite is given as

$$f(t,x) = \int f_0(x-y)\, d\mu_t(y) = (f_0 * \mu_t)(x)$$

where μ_t is the measure given by its Fourier-Stieltjes transform $\hat{\mu}_t(\xi) = \exp\left(-2\pi i t \langle \xi, b\rangle - 2\pi^2 t \langle \xi, A\xi\rangle\right)$.

Drift-Diffusion Equation (1)

- What is the most natural generalization of Gaussian measures on Lie groups?

Theorem 2 (Drift-Diffusion Equation)

The solution $f : [0,\infty) \times \mathbb{R}^n \to \mathbb{C}$ *to the differential equation*

$$\begin{cases} \frac{\partial f(t,x)}{\partial t} = \sum_{i,j=1}^{n} a_{ij} \frac{\partial^2 f(x,t)}{\partial x_i \partial x_j} - \sum_{i=1}^{n} b_i \frac{\partial f(x,t)}{\partial x_i} \\ f(0,x) = f_0(x), \quad f_0 : \mathbb{R}^n \to \mathbb{C} \end{cases}$$

where $A := [a_{ij}]$ *is positive-semidefinite is given as*

$$f(t,x) = \int f_0(x-y)\, d\mu_t(y) = (f_0 * \mu_t)(x)$$

where μ_t *is the measure given by its Fourier-Stieltjes transform* $\hat{\mu}_t(\xi) = \exp\left(-2\pi i t \langle \xi, b\rangle - 2\pi^2 t \langle \xi, A\xi\rangle\right)$.

Drift-Diffusion Equation (2)

- Gaussian measures on $\mathbb{R}^n$ can be characterized as
  - Kernels of drift-diffusion equations, or
  - Measures whose Fourier transforms are of the form

$$\hat{\mu}(\xi) = \exp\left(-2\pi i \langle \xi, b\rangle - 2\pi^2 \langle \xi, A\xi\rangle\right)$$

- Equivalence of the above two characterization is not a coincidence
- Turns out, *the same is true for general Lie groups!*

Drift-Diffusion Equation (2)

- Gaussian measures on $\mathbb{R}^n$ can be characterized as
  - Kernels of drift-diffusion equations, or
  - Measures whose Fourier transforms are of the form

$$\hat{\mu}(\xi) = \exp\left(-2\pi i \langle \xi, b \rangle - 2\pi^2 \langle \xi, A\xi \rangle\right)$$

- Equivalence of the above two characterization is not a coincidence
- Turns out, *the same is true for general Lie groups!*

Drift-Diffusion Equation (2)

- Gaussian measures on $\mathbb{R}^n$ can be characterized as
  - Kernels of drift-diffusion equations, or
  - Measures whose Fourier transforms are of the form

$$\hat{\mu}(\xi) = \exp\left(-2\pi i \langle \xi, b \rangle - 2\pi^2 \langle \xi, A\xi \rangle\right)$$

- Equivalence of the above two characterization is not a coincidence
- Turns out, *the same is true for general Lie groups!*

Some Backgrounds (1)

Definition 3 (Unitary representation)

A *unitary representation* of a locally compact Hausdorff group G is a continuous group homomorphism $\xi : G \to \mathrm{U}(E)$ into the unitary group of a Hilbert space E endowed with the strong operator topology.

Definition 4 (Fourier-Stieltjes transform)

For $\mu \in M(G)$ and a unitary representation ξ of G, the *Fourier-Stieltjes transform of μ at ξ* is defined as

$$\hat{\mu}(\xi) := \int \xi(x^{-1})\, d\mu(x)$$

Some Backgrounds (2)

Definition 5 (Convolution)

For $\mu, \nu \in M(G)$, the *convolution* of μ and ν is defined as

$$\mu * \nu : A \mapsto \int \int \mathbb{1}_A(xy)\, d\mu(x)\, d\nu(y).$$

In particular, according to the embedding $L^1(G) \to M(G)$ with respect to the *right Haar measure*,

$$f * \mu : x \mapsto \int f(xy^{-1})\, d\mu(y)$$

Proposition 6

For $\mu, \nu \in M(G)$ and a unitary representation ξ of G,

$$\widehat{\mu * \nu}(\xi) = \hat{\nu}(\xi)\hat{\mu}(\xi).$$

Some Backgrounds (3)

Theorem 7 (Gelfand-Raikov)

A measure $\mu \in M(G)$ is uniquely determined by values of its Fourier-Stieltjes transform at irreducible unitary representations; that is, if $\hat{\mu}(\xi) = 0$ for all irreducible unitary representation ξ of G, then $\mu = 0$.

- However, in general, computing μ from $\hat{\mu}(\xi)$'s is very hard, even when we know the complete list of irreducible unitary representations

Some Backgrounds (3)

Theorem 7 (Gelfand-Raikov)

A measure $\mu \in M(G)$ is uniquely determined by values of its Fourier-Stieltjes transform at irreducible unitary representations; that is, if $\hat{\mu}(\xi) = 0$ for all irreducible unitary representation ξ of G, then $\mu = 0$.

- However, in general, computing μ from $\hat{\mu}(\xi)$'s is very hard, even when we know the complete list of irreducible unitary representations

Drift-Diffusion Equation (3)

Theorem 8

Let G be a Lie group and D be a left-invariant differential operator on G given as $D = -m + \frac{1}{2}\Sigma$, where $m \in \mathfrak{g}$ and $\Sigma \in U(\mathfrak{g})$ is a degree 2 symmetric positive-semidefinite element. Then the unique solution to the differential equation

$$\begin{cases} \frac{\partial f(t,x)}{\partial t} = Df(t,x) \\ f(0,x) = f_0(x), \quad f_0 : G \to \mathbb{C} \end{cases}$$

is given as

$$f(t,x) = \int f_0(xy^{-1})\, d\mu_t(y) = (f_0 * \mu_t)(x),$$

where μ_t is the unique measure such that $\hat{\mu}_t : \xi \mapsto \exp(t\xi_* D)$.

Drift-Diffusion Equation (4)

- The theorem says:
  - For any ξ, $\exp(t\xi_* D)$ is a well-defined bounded operator on the Hilber space on which ξ is defined
  - There uniquely exists a probability measure $\mu_t \in M(G)$ having $\xi \mapsto \exp(t\xi_* D)$ as its Fourier-Stieltjes transform
  - $(\mu_t)_{t\geq 0}$ is the kernel of the left-invariant drift-diffusion equation

Definition 9 (Drift-diffusion measure)

The *drift-diffusion measure* associated to D is the measure μ_1, which is the unique measure satisfying $\hat{\mu}_1 : \xi \mapsto \exp(\xi_* D)$.

Drift-Diffusion Equation (4)

- The theorem says:
  - For any ξ, $\exp(t\xi_* D)$ is a well-defined bounded operator on the Hilber space on which ξ is defined
  - There uniquely exists a probability measure $\mu_t \in M(G)$ having $\xi \mapsto \exp(t\xi_* D)$ as its Fourier-Stieltjes transform
  - $(\mu_t)_{t \geq 0}$ is the kernel of the left-invariant drift-diffusion equation

Definition 9 (Drift-diffusion measure)

The *drift-diffusion measure* associated to D is the measure μ_1, which is the unique measure satisfying $\hat{\mu}_1 : \xi \mapsto \exp(\xi_* D)$.

Drift-Diffusion Equation (4)

- The theorem says:
  - For any ξ, $\exp(t\xi_* D)$ is a well-defined bounded operator on the Hilber space on which ξ is defined
  - There uniquely exists a probability measure $\mu_t \in M(G)$ having $\xi \mapsto \exp(t\xi_* D)$ as its Fourier-Stieltjes transform
  - $(\mu_t)_{t\geq 0}$ is the kernel of the left-invariant drift-diffusion equation

Definition 9 (Drift-diffusion measure)

The *drift-diffusion measure* associated to D is the measure μ_1, which is the unique measure satisfying $\hat{\mu}_1 : \xi \mapsto \exp(\xi_* D)$.

Drift-Diffusion Equation (4)

- The theorem says:
  - For any ξ, $\exp(t\xi_* D)$ is a well-defined bounded operator on the Hilber space on which ξ is defined
  - There uniquely exists a probability measure $\mu_t \in M(G)$ having $\xi \mapsto \exp(t\xi_* D)$ as its Fourier-Stieltjes transform
  - $(\mu_t)_{t\geq 0}$ is the kernel of the left-invariant drift-diffusion equation

Definition 9 (Drift-diffusion measure)

The *drift-diffusion measure* associated to D is the measure μ_1, which is the unique measure satisfying $\hat{\mu}_1 : \xi \mapsto \exp(\xi_* D)$.

Drift-Diffusion Equation (4)

- The theorem says:
  - For any ξ, $\exp(t\xi_* D)$ is a well-defined bounded operator on the Hilber space on which ξ is defined
  - There uniquely exists a probability measure $\mu_t \in M(G)$ having $\xi \mapsto \exp(t\xi_* D)$ as its Fourier-Stieltjes transform
  - $(\mu_t)_{t\geq 0}$ is the kernel of the left-invariant drift-diffusion equation

Definition 9 (Drift-diffusion measure)

The *drift-diffusion measure* associated to D is the measure μ_1, which is the unique measure satisfying $\hat{\mu}_1 : \xi \mapsto \exp(\xi_* D)$.

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (1)

- Let's focus on the case $G = \mathrm{SO}(3)$
- Let us call a measure $\mu \in M(G)$ a *left-translated diffusion distribution* if $\hat{\mu}(\xi) = \exp\left(\frac{1}{2}\xi_*\Sigma\right)\xi(x_0^{-1})$, and denote $\mu = \mathrm{LD}(x_0, \Sigma)$
    - Σ, a symmetric positive-definite degree 2 element in $U(\mathfrak{so}(3))$ will play the role of *covariance matrix*
    - $x_0 \in G$ will play the role of *mean*
- The case when $\Sigma = \sigma^2\mathbf{1}$, that is, when it is a *Casimir element*,
    - We call μ *central* or *isotropic*
    - For $\nu = \mathrm{LD}(y_0, \Sigma)$, $\mu * \nu = \mathrm{LD}(x_0 y_0, \Sigma + \sigma^2\mathbf{1})$
    - In general, convolution of non-isotropic LD's need not an LD

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (1)

- Let's focus on the case $G = \mathrm{SO}(3)$
- Let us call a measure $\mu \in M(G)$ a *left-translated diffusion distribution* if $\hat{\mu}(\xi) = \exp\left(\frac{1}{2}\xi_*\Sigma\right)\xi(x_0^{-1})$, and denote $\mu = \mathrm{LD}(x_0, \Sigma)$
  - Σ, a symmetric positive-definite degree 2 element in $U(\mathfrak{so}(3))$ will play the role of *covariance matrix*
  - $x_0 \in G$ will play the role of *mean*
- The case when $\Sigma = \sigma^2\mathbf{1}$, that is, when it is a *Casimir element*,
  - We call μ *central* or *isotropic*
  - For $\nu = \mathrm{LD}(y_0, \Sigma)$, $\mu * \nu = \mathrm{LD}(x_0 y_0, \Sigma + \sigma^2\mathbf{1})$
  - In general, convolution of non-isotropic LD's need not an LD

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (1)

- Let's focus on the case $G = \mathrm{SO}(3)$
- Let us call a measure $\mu \in M(G)$ a *left-translated diffusion distribution* if $\hat{\mu}(\xi) = \exp\left(\frac{1}{2}\xi_* \Sigma\right) \xi(x_0^{-1})$, and denote $\mu = \mathrm{LD}(x_0, \Sigma)$
  - Σ, a symmetric positive-definite degree 2 element in $U(\mathfrak{so}(3))$ will play the role of *covariance matrix*
  - $x_0 \in G$ will play the role of *mean*
- The case when $\Sigma = \sigma^2 \mathbf{1}$, that is, when it is a *Casimir element*,
  - We call μ *central* or *isotropic*
  - For $\nu = \mathrm{LD}(y_0, \Sigma)$, $\mu * \nu = \mathrm{LD}(x_0 y_0, \Sigma + \sigma^2 \mathbf{1})$
  - In general, convolution of non-isotropic LD's need not an LD

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (1)

- Let's focus on the case $G = \mathrm{SO}(3)$
- Let us call a measure $\mu \in M(G)$ a *left-translated diffusion distribution* if $\hat{\mu}(\xi) = \exp\left(\frac{1}{2}\xi_* \Sigma\right) \xi(x_0^{-1})$, and denote $\mu = \mathrm{LD}(x_0, \Sigma)$
    - Σ, a symmetric positive-definite degree 2 element in $U(\mathfrak{so}(3))$ will play the role of *covariance matrix*
    - $x_0 \in G$ will play the role of *mean*
- The case when $\Sigma = \sigma^2 \mathbf{1}$, that is, when it is a *Casimir element*,
    - We call μ *central* or *isotropic*
    - For $\nu = \mathrm{LD}(y_0, \Sigma)$, $\mu * \nu = \mathrm{LD}(x_0 y_0, \Sigma + \sigma^2 \mathbf{1})$
    - In general, convolution of non-isotropic LD's need not an LD

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (1)

- Let's focus on the case $G = \mathrm{SO}(3)$
- Let us call a measure $\mu \in M(G)$ a *left-translated diffusion distribution* if $\hat{\mu}(\xi) = \exp\left(\frac{1}{2}\xi_*\Sigma\right)\xi(x_0^{-1})$, and denote $\mu = \mathrm{LD}(x_0, \Sigma)$
  - Σ, a symmetric positive-definite degree 2 element in $U(\mathfrak{so}(3))$ will play the role of *covariance matrix*
  - $x_0 \in G$ will play the role of *mean*
- The case when $\Sigma = \sigma^2\mathbf{1}$, that is, when it is a *Casimir element*,
  - We call μ *central* or *isotropic*
  - For $\nu = \mathrm{LD}(y_0, \Sigma)$, $\mu * \nu = \mathrm{LD}(x_0 y_0, \Sigma + \sigma^2\mathbf{1})$
  - In general, convolution of non-isotropic LD's need not an LD

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (1)

- Let's focus on the case $G = \mathrm{SO}(3)$
- Let us call a measure $\mu \in M(G)$ a *left-translated diffusion distribution* if $\hat{\mu}(\xi) = \exp\left(\frac{1}{2}\xi_*\Sigma\right)\xi(x_0^{-1})$, and denote $\mu = \mathrm{LD}(x_0, \Sigma)$
  - Σ, a symmetric positive-definite degree 2 element in $U(\mathfrak{so}(3))$ will play the role of *covariance matrix*
  - $x_0 \in G$ will play the role of *mean*
- The case when $\Sigma = \sigma^2\mathbf{1}$, that is, when it is a *Casimir element*,
  - We call μ *central* or *isotropic*
  - For $\nu = \mathrm{LD}(y_0, \Sigma)$, $\mu * \nu = \mathrm{LD}(x_0 y_0, \Sigma + \sigma^2\mathbf{1})$
  - In general, convolution of non-isotropic LD's need not an LD

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (1)

- Let's focus on the case $G = \mathrm{SO}(3)$
- Let us call a measure $\mu \in M(G)$ a *left-translated diffusion distribution* if $\hat{\mu}(\xi) = \exp\left(\frac{1}{2}\xi_* \Sigma\right) \xi(x_0^{-1})$, and denote $\mu = \mathrm{LD}(x_0, \Sigma)$
    - Σ, a symmetric positive-definite degree 2 element in $U(\mathfrak{so}(3))$ will play the role of *covariance matrix*
    - $x_0 \in G$ will play the role of *mean*
- The case when $\Sigma = \sigma^2 \mathbf{1}$, that is, when it is a *Casimir element*,
    - We call μ *central* or *isotropic*
    - For $\nu = \mathrm{LD}(y_0, \Sigma)$, $\mu * \nu = \mathrm{LD}(x_0 y_0, \Sigma + \sigma^2 \mathbf{1})$
    - In general, convolution of non-isotropic LD's need not an LD

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (1)

- Let's focus on the case $G = \mathrm{SO}(3)$
- Let us call a measure $\mu \in M(G)$ a *left-translated diffusion distribution* if $\hat{\mu}(\xi) = \exp\left(\frac{1}{2}\xi_*\Sigma\right)\xi(x_0^{-1})$, and denote $\mu = \mathrm{LD}(x_0, \Sigma)$
  - Σ, a symmetric positive-definite degree 2 element in $U(\mathfrak{so}(3))$ will play the role of *covariance matrix*
  - $x_0 \in G$ will play the role of *mean*
- The case when $\Sigma = \sigma^2\mathbf{1}$, that is, when it is a *Casimir element*,
  - We call μ *central* or *isotropic*
  - For $\nu = \mathrm{LD}(y_0, \Sigma)$, $\mu * \nu = \mathrm{LD}(x_0 y_0, \Sigma + \sigma^2\mathbf{1})$
  - In general, convolution of non-isotropic LD's need not an LD

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (2)

- A kind of central limit theorem hold [2][1]
- For *isotropic* distributions, the pdf can be calculuated numerically as

$$f(t) = \sum_{l=0}^{\infty} (2l+1) e^{-\frac{l(l+1)}{2}\sigma^2} \left(\frac{\sin\left(l + \frac{1}{2}\right) t}{\sin \frac{t}{2}} \right)$$

  where t is the distance from the center
- Is x_0 the MSE estimate when the distribution is $\mathrm{LD}(x_0, \Sigma)$?

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (2)

- A kind of central limit theorem hold [2][1]
- For *isotropic* distributions, the pdf can be calculuated numerically as

$$f(t) = \sum_{l=0}^{\infty}(2l+1)e^{-\frac{l(l+1)}{2}\sigma^2}\left(\frac{\sin\left(l+\frac{1}{2}\right)t}{\sin\frac{t}{2}}\right)$$

where t is the distance from the center

- Is x_0 the MSE estimate when the distribution is $\mathrm{LD}(x_0, \Sigma)$?

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (2)

- A kind of central limit theorem hold [2][1]
- For *isotropic* distributions, the pdf can be calculuated numerically as

$$f(t) = \sum_{l=0}^{\infty}(2l+1)e^{-\frac{l(l+1)}{2}\sigma^2}\left(\frac{\sin\left(l+\frac{1}{2}\right)t}{\sin\frac{t}{2}}\right)$$

  where t is the distance from the center
- Is x_0 the MSE estimate when the distribution is $\mathrm{LD}(x_0, \Sigma)$?

Left-Translated Diffusion Distribution on $\mathrm{SO}(3)$ (2)

- A kind of central limit theorem hold [2][1]
- For *isotropic* distributions, the pdf can be calculuated numerically as

$$f(t) = \sum_{l=0}^{\infty} (2l+1) e^{-\frac{l(l+1)}{2}\sigma^2} \left(\frac{\sin\left(l+\frac{1}{2}\right)t}{\sin\frac{t}{2}} \right)$$

  where t is the distance from the center
- Is x_0 the MSE estimate when the distribution is $\mathrm{LD}(x_0, \Sigma)$?

Update Using Gyroscope (1)

- System model:
  - $R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$
  - $\Omega_k = \Omega_{k-1} + A_k \Delta t$
  - $G_k = \Omega_k + U_k$
- Assume
  - $R_0 \perp \Omega_0$
  - $R_0 \sim \mathrm{LD}(\bar{r}_0, \Sigma_0)$
  - $\Omega_0 \sim \mathrm{N}(\bar{\omega}_0, \sigma^2_{\Omega,0}\mathbf{1})$
- $(\Omega_1 | G_1 = g_1) \sim \mathrm{N}\left(\bar{\omega}_1, \sigma^2_{\Omega,1}\mathbf{1}\right)$ where

$$\bar{\omega}_1 := \frac{\sigma_U^2 \bar{\omega}_0}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}} + \frac{\left(\sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}\right) g_1}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}}$$

$$\sigma^2_{\Omega,1} := \frac{\sigma_U^2 \left(\sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}\right)}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}}$$

Update Using Gyroscope (1)

- System model:
  - $R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$
  - $\Omega_k = \Omega_{k-1} + A_k \Delta t$
  - $G_k = \Omega_k + U_k$
- Assume
  - $R_0 \perp \Omega_0$
  - $R_0 \sim \mathrm{LD}(\bar{r}_0, \Sigma_0)$
  - $\Omega_0 \sim \mathrm{N}(\bar{\omega}_0, \sigma^2_{\Omega,0}\mathbf{1})$
- $(\Omega_1 | G_1 = g_1) \sim \mathrm{N}\left(\bar{\omega}_1, \sigma^2_{\Omega,1}\mathbf{1}\right)$ where

$$\bar{\omega}_1 := \frac{\sigma_U^2 \bar{\omega}_0}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}} + \frac{\left(\sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}\right) g_1}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}}$$

$$\sigma^2_{\Omega,1} := \frac{\sigma_U^2 \left(\sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}\right)}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}}$$

Update Using Gyroscope (1)

- System model:
  - $R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$
  - $\Omega_k = \Omega_{k-1} + A_k \Delta t$
  - $G_k = \Omega_k + U_k$
- Assume
  - $R_0 \perp \Omega_0$
  - $R_0 \sim \mathrm{LD}(\bar{r}_0, \Sigma_0)$
  - $\Omega_0 \sim \mathrm{N}(\bar{\omega}_0, \sigma^2_{\Omega,0}\mathbf{1})$
- $(\Omega_1 | G_1 = g_1) \sim \mathrm{N}\left(\bar{\omega}_1, \sigma^2_{\Omega,1}\mathbf{1}\right)$ where

$$\bar{\omega}_1 := \frac{\sigma_U^2 \bar{\omega}_0}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}} + \frac{\left(\sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}\right) g_1}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}}$$

$$\sigma^2_{\Omega,1} := \frac{\sigma_U^2 \left(\sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}\right)}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}}$$

Update Using Gyroscope (2)

- System model:
  - $R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$
  - $\Omega_k = \Omega_{k-1} + A_k \Delta t$
  - $G_k = \Omega_k + U_k$
- Similarly, $(\Omega_0 | G_1 = g_1) \sim \mathrm{N}\left(\tilde{\omega}_0, \tilde{\sigma}^2_{\Omega,0}\mathbf{1}\right)$ where

$$\tilde{\omega}_0 := \frac{\left(\sigma_A^2 \Delta t^2 + \sigma_U^2\right)\bar{\omega}_0}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}} + \frac{\sigma^2_{\Omega,0} g_1}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}}$$

$$\tilde{\sigma}^2_{\Omega,0} := \frac{\left(\sigma_A^2 \Delta t^2 + \sigma_U^2\right)\sigma^2_{\Omega,0}}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma^2_{\Omega,0}}$$

Update Using Gyroscope (3)

- System model:
  - $R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$
  - $\Omega_k = \Omega_{k-1} + A_k \Delta t$
  - $G_k = \Omega_k + U_k$
- Since $G_1 - \Omega_0 - R_0$ is a Markov chain and $R_0 \perp \Omega_0$,

$$(\text{dist. } R_1 | G_1) = (\text{dist. } \exp(-[\Omega_0]_\times \Delta t) | G_1) * (\text{dist. } R_0)$$

- Claim 1 if $\bar{\omega}, \sigma^2$ are small enough,

$$\exp_* \mathrm{N}\left(\bar{\omega}, \sigma^2 \mathbf{1}\right) \approx \mathrm{LD}\left(\exp\left([\bar{\omega}]_\times\right), \sigma^2 \mathbf{1}\right)$$

- Hence, if Δt is small enough,

$$\begin{aligned}(R_1 | G_1 = g_1) \sim & \mathrm{LD}\left(\exp\left(-[\tilde{\omega}_0]_\times \Delta t\right), \tilde{\sigma}^2_{\Omega,0} \mathbf{1}\right) * \mathrm{LD}\left(\bar{r}_0, \Sigma_0\right) \\ = & \mathrm{LD}\left(\exp\left(-[\tilde{\omega}_0]_\times \Delta t\right) \bar{r}_0, \tilde{\sigma}^2_{\Omega,0} \mathbf{1} + \Sigma_0\right)\end{aligned}$$

Update Using Gyroscope (3)

- System model:
  - $R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$
  - $\Omega_k = \Omega_{k-1} + A_k \Delta t$
  - $G_k = \Omega_k + U_k$
- Since $G_1 - \Omega_0 - R_0$ is a Markov chain and $R_0 \perp \Omega_0$,

$$(\text{dist. } R_1 | G_1) = (\text{dist. } \exp(-[\Omega_0]_\times \Delta t) | G_1) * (\text{dist. } R_0)$$

- Claim 1 if $\bar{\omega}, \sigma^2$ are small enough,

$$\exp_* \mathrm{N}\left(\bar{\omega}, \sigma^2 \mathbf{1}\right) \approx \mathrm{LD}\left(\exp\left([\bar{\omega}]_\times\right), \sigma^2 \mathbf{1}\right)$$

- Hence, if Δt is small enough,

$$\begin{aligned}(R_1 | G_1 = g_1) \sim & \mathrm{LD}\left(\exp\left(-[\tilde{\omega}_0]_\times \Delta t\right), \tilde{\sigma}^2_{\Omega,0} \mathbf{1}\right) * \mathrm{LD}\left(\bar{r}_0, \Sigma_0\right) \\ = & \mathrm{LD}\left(\exp\left(-[\tilde{\omega}_0]_\times \Delta t\right) \bar{r}_0, \tilde{\sigma}^2_{\Omega,0} \mathbf{1} + \Sigma_0\right)\end{aligned}$$

Update Using Gyroscope (3)

- System model:
  - $R_k = \exp(-[\Omega_{k-1}]_\times \Delta t) R_{k-1}$
  - $\Omega_k = \Omega_{k-1} + A_k \Delta t$
  - $G_k = \Omega_k + U_k$
- Since $G_1 - \Omega_0 - R_0$ is a Markov chain and $R_0 \perp \Omega_0$,

$$(\text{dist. } R_1 | G_1) = (\text{dist. } \exp(-[\Omega_0]_\times \Delta t) | G_1) * (\text{dist. } R_0)$$

- Claim 1 if $\bar{\omega}, \sigma^2$ are small enough,

$$\exp_* \mathrm{N}\left(\bar{\omega}, \sigma^2 \mathbf{1}\right) \approx \mathrm{LD}\left(\exp\left([\bar{\omega}]_\times\right), \sigma^2 \mathbf{1}\right)$$

- Hence, if Δt is small enough,

$$\begin{aligned}(R_1 | G_1 = g_1) \sim & \mathrm{LD}\left(\exp\left(-[\tilde{\omega}_0]_\times \Delta t\right), \tilde{\sigma}^2_{\Omega,0} \mathbf{1}\right) * \mathrm{LD}\left(\bar{r}_0, \Sigma_0\right) \\ = & \mathrm{LD}\left(\exp\left(-[\tilde{\omega}_0]_\times \Delta t\right) \bar{r}_0, \tilde{\sigma}^2_{\Omega,0} \mathbf{1} + \Sigma_0\right)\end{aligned}$$

Bingham Distribution (1)

Definition 10 (Bingham distribution)

For a symmetric 4×4 real matrix M, the *Bingham distribution* associated to M is the probability measure on $\mathrm{SU}(2) \subseteq \mathbb{R}^4$ whose pdf is of the form

$$f(q) = \frac{1}{K(M)} \exp\left(q^T M q\right).$$

- $f(q) = f(-q)$, thus the pushforward of this measure by the covering map $\pi : \mathrm{SU}(2) \to \mathrm{SO}(3)$ doubles the pdf
    - Denote this pushforward onto $\mathrm{SO}(3)$ as $\mathrm{BH}(M)$
- $\mathrm{BH}(M) = \mathrm{BH}(M + \lambda \mathbf{1})$, so we may assume the eigenvalues of M are $0 \geq \lambda_1 \geq \lambda_2 \geq \lambda_3$

Bingham Distribution (1)

Definition 10 (Bingham distribution)

For a symmetric 4×4 real matrix M, the *Bingham distribution* associated to M is the probability measure on $\mathrm{SU}(2) \subseteq \mathbb{R}^4$ whose pdf is of the form

$$f(q) = \frac{1}{K(M)} \exp\left(q^T M q\right).$$

- $f(q) = f(-q)$, thus the pushforward of this measure by the covering map $\pi : \mathrm{SU}(2) \to \mathrm{SO}(3)$ doubles the pdf
  - Denote this pushforward onto $\mathrm{SO}(3)$ as $\mathrm{BH}(M)$
- $\mathrm{BH}(M) = \mathrm{BH}(M + \lambda\mathbf{1})$, so we may assume the eigenvalues of M are $0 \geq \lambda_1 \geq \lambda_2 \geq \lambda_3$

Bingham Distribution (1)

Definition 10 (Bingham distribution)

For a symmetric 4×4 real matrix M, the *Bingham distribution* associated to M is the probability measure on $\mathrm{SU}(2) \subseteq \mathbb{R}^4$ whose pdf is of the form

$$f(q) = \frac{1}{K(M)} \exp\left(q^T M q\right).$$

- $f(q) = f(-q)$, thus the pushforward of this measure by the covering map $\pi : \mathrm{SU}(2) \to \mathrm{SO}(3)$ doubles the pdf
  - Denote this pushforward onto $\mathrm{SO}(3)$ as $\mathrm{BH}(M)$
- $\mathrm{BH}(M) = \mathrm{BH}(M + \lambda \mathbf{1})$, so we may assume the eigenvalues of M are $0 \geq \lambda_1 \geq \lambda_2 \geq \lambda_3$

Bingham Distribution (1)

Definition 10 (Bingham distribution)

For a symmetric 4×4 real matrix M, the *Bingham distribution* associated to M is the probability measure on $\mathrm{SU}(2) \subseteq \mathbb{R}^4$ whose pdf is of the form

$$f(q) = \frac{1}{K(M)} \exp\left(q^T M q\right).$$

- $f(q) = f(-q)$, thus the pushforward of this measure by the covering map $\pi : \mathrm{SU}(2) \to \mathrm{SO}(3)$ doubles the pdf
  - Denote this pushforward onto $\mathrm{SO}(3)$ as $\mathrm{BH}(M)$
- $\mathrm{BH}(M) = \mathrm{BH}(M + \lambda \mathbf{1})$, so we may assume the eigenvalues of M are $0 \geq \lambda_1 \geq \lambda_2 \geq \lambda_3$

Bingham Distribution (2)

- $f(q) = \frac{1}{K(M)} \exp\left(q^T M q\right)$
- $M = P^T \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & \lambda_3 \end{bmatrix} P$, for some $P \in \mathrm{O}(4)$
- There exists $q_0 := (s_0, v_0) \in \mathrm{SU}(2) \subseteq \mathbb{R}^{1+3}$ and $Q \in \mathrm{O}(3)$ such that

$$P = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} Q_L(q_0^{-1}) = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} s_0 & v_0^T \\ -v_0 & s_0 \mathbf{1} - [v_0]_\times \end{bmatrix}$$

- Hence, $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$

Bingham Distribution (2)

- $f(q) = \frac{1}{K(M)} \exp\left(q^T M q\right)$
- $M = P^T \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & \lambda_3 \end{bmatrix} P$, for some $P \in \mathrm{O}(4)$
- There exists $q_0 := (s_0, v_0) \in \mathrm{SU}(2) \subseteq \mathbb{R}^{1+3}$ and $Q \in \mathrm{O}(3)$ such that

$$P = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} Q_L(q_0^{-1}) = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} s_0 & v_0^T \\ -v_0 & s_0 \mathbf{1} - [v_0]_\times \end{bmatrix}$$

- Hence, $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$

Bingham Distribution (2)

- $f(q) = \frac{1}{K(M)} \exp\left(q^T M q\right)$
- $M = P^T \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & \lambda_3 \end{bmatrix} P$, for some $P \in \mathrm{O}(4)$
- There exists $q_0 := (s_0, v_0) \in \mathrm{SU}(2) \subseteq \mathbb{R}^{1+3}$ and $Q \in \mathrm{O}(3)$ such that

$$P = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} Q_L(q_0^{-1}) = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} s_0 & v_0^T \\ -v_0 & s_0 \mathbf{1} - [v_0]_\times \end{bmatrix}$$

- Hence, $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$

Bingham Distribution (2)

- $f(q) = \frac{1}{K(M)} \exp\left(q^T M q\right)$
- $M = P^T \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & \lambda_3 \end{bmatrix} P$, for some $P \in \mathrm{O}(4)$
- There exists $q_0 := (s_0, v_0) \in \mathrm{SU}(2) \subseteq \mathbb{R}^{1+3}$ and $Q \in \mathrm{O}(3)$ such that

$$P = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} Q_L(q_0^{-1}) = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} s_0 & v_0^T \\ -v_0 & s_0 \mathbf{1} - [v_0]_\times \end{bmatrix}$$

- Hence, $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q)\right)$

Bingham Distribution (3)

- $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$
- $\pi(q_0)$: unique center, at which f is maximized
- $\Sigma = Q^T \begin{bmatrix} -\frac{1}{2\lambda_1} & 0 & 0 \\ 0 & -\frac{1}{2\lambda_2} & 0 \\ 0 & 0 & -\frac{1}{2\lambda_3} \end{bmatrix} Q$ is positive-definite
- Q: principal directions
- Eigenvalues: how rapidly spreads along each principal direction
- $K(M) := \int_{\mathrm{SU}(2)} \exp\left(q^T M q\right) \, dq$ depends only on eigenvalues
- Claim 2 $\mathrm{LD}(\pi(q_0), \Sigma) \approx \mathrm{BH}(M)$
  - Numerically verified for some isotropic cases [1]

Bingham Distribution (3)

- $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$
- $\pi(q_0)$: unique center, at which f is maximized
- $\Sigma = Q^T \begin{bmatrix} -\frac{1}{2\lambda_1} & 0 & 0 \\ 0 & -\frac{1}{2\lambda_2} & 0 \\ 0 & 0 & -\frac{1}{2\lambda_3} \end{bmatrix} Q$ is positive-definite
- Q: principal directions
- Eigenvalues: how rapidly spreads along each principal direction
- $K(M) := \int_{\mathrm{SU}(2)} \exp\left(q^T M q\right) \, dq$ depends only on eigenvalues
- Claim 2 $\mathrm{LD}(\pi(q_0), \Sigma) \approx \mathrm{BH}(M)$
  - Numerically verified for some isotropic cases [1]

Bingham Distribution (3)

- $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$
- $\pi(q_0)$: unique center, at which f is maximized
- $\Sigma = Q^T \begin{bmatrix} -\frac{1}{2\lambda_1} & 0 & 0 \\ 0 & -\frac{1}{2\lambda_2} & 0 \\ 0 & 0 & -\frac{1}{2\lambda_3} \end{bmatrix} Q$ is positive-definite
- Q: principal directions
- Eigenvalues: how rapidly spreads along each principal direction
- $K(M) := \int_{\mathrm{SU}(2)} \exp\left(q^T M q\right) \, dq$ depends only on eigenvalues
- Claim 2 $\mathrm{LD}(\pi(q_0), \Sigma) \approx \mathrm{BH}(M)$
  - Numerically verified for some isotropic cases [1]

Bingham Distribution (3)

- $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$
- $\pi(q_0)$: unique center, at which f is maximized
- $\Sigma = Q^T \begin{bmatrix} -\frac{1}{2\lambda_1} & 0 & 0 \\ 0 & -\frac{1}{2\lambda_2} & 0 \\ 0 & 0 & -\frac{1}{2\lambda_3} \end{bmatrix} Q$ is positive-definite
- Q: principal directions
- Eigenvalues: how rapidly spreads along each principal direction
- $K(M) := \int_{\mathrm{SU}(2)} \exp\left(q^T M q\right) \, dq$ depends only on eigenvalues
- Claim 2 $\mathrm{LD}(\pi(q_0), \Sigma) \approx \mathrm{BH}(M)$
  - Numerically verified for some isotropic cases [1]

Bingham Distribution (3)

- $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$
- $\pi(q_0)$: unique center, at which f is maximized
- $\Sigma = Q^T \begin{bmatrix} -\frac{1}{2\lambda_1} & 0 & 0 \\ 0 & -\frac{1}{2\lambda_2} & 0 \\ 0 & 0 & -\frac{1}{2\lambda_3} \end{bmatrix} Q$ is positive-definite
- Q: principal directions
- Eigenvalues: how rapidly spreads along each principal direction
- $K(M) := \int_{\mathrm{SU}(2)} \exp\left(q^T M q\right) \, dq$ depends only on eigenvalues
- Claim 2 $\mathrm{LD}(\pi(q_0), \Sigma) \approx \mathrm{BH}(M)$
  - Numerically verified for some isotropic cases [1]

Bingham Distribution (3)

- $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$
- $\pi(q_0)$: unique center, at which f is maximized
- $\Sigma = Q^T \begin{bmatrix} -\frac{1}{2\lambda_1} & 0 & 0 \\ 0 & -\frac{1}{2\lambda_2} & 0 \\ 0 & 0 & -\frac{1}{2\lambda_3} \end{bmatrix} Q$ is positive-definite
- Q: principal directions
- Eigenvalues: how rapidly spreads along each principal direction
- $K(M) := \int_{\mathrm{SU}(2)} \exp\left(q^T M q\right) dq$ depends only on eigenvalues
- Claim 2 $\mathrm{LD}(\pi(q_0), \Sigma) \approx \mathrm{BH}(M)$
  - Numerically verified for some isotropic cases [1]

Bingham Distribution (3)

- $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$
- $\pi(q_0)$: unique center, at which f is maximized
- $\Sigma = Q^T \begin{bmatrix} -\frac{1}{2\lambda_1} & 0 & 0 \\ 0 & -\frac{1}{2\lambda_2} & 0 \\ 0 & 0 & -\frac{1}{2\lambda_3} \end{bmatrix} Q$ is positive-definite
- Q: principal directions
- Eigenvalues: how rapidly spreads along each principal direction
- $K(M) := \int_{\mathrm{SU}(2)} \exp\left(q^T M q\right) \, dq$ depends only on eigenvalues
- <u>Claim 2</u> $\mathrm{LD}(\pi(q_0), \Sigma) \approx \mathrm{BH}(M)$
  - Numerically verified for some isotropic cases [1]

Bingham Distribution (3)

- $f(q) = \frac{1}{K(M)} \exp\left((q_0^{-1} q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (q_0^{-1} q) \right)$
- $\pi(q_0)$: unique center, at which f is maximized
- $\Sigma = Q^T \begin{bmatrix} -\frac{1}{2\lambda_1} & 0 & 0 \\ 0 & -\frac{1}{2\lambda_2} & 0 \\ 0 & 0 & -\frac{1}{2\lambda_3} \end{bmatrix} Q$ is positive-definite
- Q: principal directions
- Eigenvalues: how rapidly spreads along each principal direction
- $K(M) := \int_{\mathrm{SU}(2)} \exp\left(q^T M q\right) dq$ depends only on eigenvalues
- Claim 2 $\mathrm{LD}(\pi(q_0), \Sigma) \approx \mathrm{BH}(M)$
  - Numerically verified for some isotropic cases [1]

Direction Measurement (1)

- Let $R \sim \mathrm{BH}(M)$, $V = Rh + U$ for some $h \in \mathbb{R}^3$ and $U \sim \mathrm{N}(0, \sigma^2 \mathbf{1})$
- What is the conditional distribution $R|V$?
- Bayes' rule: $f_{R|V}(r|v) = \frac{f_{V|R}(v|r) f_R(r)}{f_V(v)}$
- $(V|R = r) \sim \mathrm{N}(rh, \Sigma)$, so $f_{R|V}(r|v) \propto \exp\left(-\frac{\|v - qhq^{-1}\|^2}{2\sigma^2}\right) \exp\left(q^T M q\right) = \exp\left(q^T (M + M_1) q\right)$ where

$$M_1 := -\frac{1}{2\sigma^2} \begin{bmatrix} \|v - h\|^2 & 2(v \times h)^T \\ 2(v \times h)^T & \|v + h\|^2 \mathbf{1} - 2(vh^T + hv^T) \end{bmatrix}$$

Direction Measurement (1)

- Let $R \sim \mathrm{BH}(M)$, $V = Rh + U$ for some $h \in \mathbb{R}^3$ and $U \sim \mathrm{N}(0, \sigma^2 \mathbf{1})$
- What is the conditional distribution $R|V$?

- Bayes' rule: $f_{R|V}(r|v) = \frac{f_{V|R}(v|r) f_R(r)}{f_V(v)}$
- $(V|R = r) \sim \mathrm{N}\left(rh, \Sigma\right)$, so $f_{R|V}(r|v) \propto \exp\left(-\frac{\left\|v - qhq^{-1}\right\|^2}{2\sigma^2}\right) \exp\left(q^T M q\right) = \exp\left(q^T \left(M + M_1\right) q\right)$ where

$$M_1 := -\frac{1}{2\sigma^2} \begin{bmatrix} \|v - h\|^2 & 2(v \times h)^T \\ 2(v \times h)^T & \|v + h\|^2 \mathbf{1} - 2(vh^T + hv^T) \end{bmatrix}$$

Direction Measurement (1)

- Let $R \sim \mathrm{BH}(M)$, $V = Rh + U$ for some $h \in \mathbb{R}^3$ and $U \sim \mathrm{N}(0, \sigma^2 \mathbf{1})$
- What is the conditional distribution $R|V$?

- Bayes' rule: $f_{R|V}(r|v) = \frac{f_{V|R}(v|r) f_R(r)}{f_V(v)}$
- $(V|R = r) \sim \mathrm{N}(rh, \Sigma)$, so $f_{R|V}(r|v) \propto \exp\left(-\frac{\|v - qhq^{-1}\|^2}{2\sigma^2}\right) \exp\left(q^T M q\right) = \exp\left(q^T (M + M_1) q\right)$ where

$$M_1 := -\frac{1}{2\sigma^2} \begin{bmatrix} \|v-h\|^2 & 2(v \times h)^T \\ 2(v \times h)^T & \|v+h\|^2 \mathbf{1} - 2(vh^T + hv^T) \end{bmatrix}$$

Direction Measurement (1)

- Let $R \sim \mathrm{BH}(M)$, $V = Rh + U$ for some $h \in \mathbb{R}^3$ and $U \sim \mathrm{N}(0, \sigma^2 \mathbf{1})$
- What is the conditional distribution $R|V$?

- Bayes' rule: $f_{R|V}(r|v) = \frac{f_{V|R}(v|r) f_R(r)}{f_V(v)}$
- $(V|R = r) \sim \mathrm{N}\left(rh, \Sigma\right)$, so $f_{R|V}(r|v) \propto \exp\left(-\frac{\left\|v - qhq^{-1}\right\|^2}{2\sigma^2}\right) \exp\left(q^T M q\right) = \exp\left(q^T \left(M + M_1\right) q\right)$ where

$$M_1 := -\frac{1}{2\sigma^2} \begin{bmatrix} \|v - h\|^2 & 2(v \times h)^T \\ 2(v \times h)^T & \|v + h\|^2 \mathbf{1} - 2(vh^T + hv^T) \end{bmatrix}$$

Direction Measurement (2)

- In general, for $V_i = Rh_i + U_i$, $i = 1, \cdots, n$, define

$$M_i := -\frac{1}{2\sigma_i^2}\begin{bmatrix} \|v_i - h_i\|^2 & 2(v_i \times h_i)^T \\ 2(v_i \times h_i)^T & \|v_i + h_i\|^2 \mathbf{1} - 2(v_i h_i^T + h_i v_i^T) \end{bmatrix},$$

then

$$(R|V_1 = v_1, \cdots, V_n = v_n) \sim \mathrm{BH}\left(M + \sum_{i=1}^{n} M_i\right)$$

Update Using Accelerometer and Magnetometer

- System model:
  - $A_k = R_k \mathbf{a} + V_k$
  - $M_k = R_k \mathbf{m} + W_k$
- Update procedure:
  1. Approximate the distribution of $R_1|G_1$ as $\mathrm{BH}(M)$
  2. Calculate M_1, M_2 for A_1, M_1
  3. Then the distribution of $R_1|G_1, A_1, M_1$ is approximately $\mathrm{BH}(M + M_1 + M_2)$
  4. Approximate $\mathrm{BH}(M + M_1 + M_2)$ as $\mathrm{LD}(r, \Sigma)$

Update Using Accelerometer and Magnetometer

- System model:
  - $A_k = R_k \mathbf{a} + V_k$
  - $M_k = R_k \mathbf{m} + W_k$
- Update procedure:
  1. Approximate the distribution of $R_1 | G_1$ as $\mathrm{BH}(M)$
  2. Calculate M_1, M_2 for A_1, M_1
  3. Then the distribution of $R_1 | G_1, A_1, M_1$ is approximately $\mathrm{BH}(M + M_1 + M_2)$
  4. Approximate $\mathrm{BH}(M + M_1 + M_2)$ as $\mathrm{LD}(r, \Sigma)$

Update Using Accelerometer and Magnetometer

- System model:
  - $A_k = R_k\mathbf{a} + V_k$
  - $M_k = R_k\mathbf{m} + W_k$
- Update procedure:
  1. Approximate the distribution of $R_1|G_1$ as $\mathrm{BH}(M)$
  2. Calculate M_1, M_2 for A_1, M_1
  3. Then the distribution of $R_1|G_1, A_1, M_1$ is approximately $\mathrm{BH}(M + M_1 + M_2)$
  4. Approximate $\mathrm{BH}(M + M_1 + M_2)$ as $\mathrm{LD}(r, \Sigma)$

Update Using Accelerometer and Magnetometer

- System model:
  - $A_k = R_k\mathbf{a} + V_k$
  - $M_k = R_k\mathbf{m} + W_k$
- Update procedure:
  1. Approximate the distribution of $R_1|G_1$ as $\mathrm{BH}(M)$
  2. Calculate M_1, M_2 for A_1, M_1
  3. Then the distribution of $R_1|G_1, A_1, M_1$ is approximately $\mathrm{BH}(M + M_1 + M_2)$
  4. Approximate $\mathrm{BH}(M + M_1 + M_2)$ as $\mathrm{LD}(r, \Sigma)$

Update Using Accelerometer and Magnetometer

- System model:
  - $A_k = R_k \mathbf{a} + V_k$
  - $M_k = R_k \mathbf{m} + W_k$
- Update procedure:
  1. Approximate the distribution of $R_1 | G_1$ as $\mathrm{BH}(M)$
  2. Calculate M_1, M_2 for A_1, M_1
  3. Then the distribution of $R_1 | G_1, A_1, M_1$ is approximately $\mathrm{BH}(M + M_1 + M_2)$
  4. Approximate $\mathrm{BH}(M + M_1 + M_2)$ as $\mathrm{LD}(r, \Sigma)$

Update Using Accelerometer and Magnetometer

- System model:
  - $A_k = R_k \mathbf{a} + V_k$
  - $M_k = R_k \mathbf{m} + W_k$
- Update procedure:
  1. Approximate the distribution of $R_1 | G_1$ as $\mathrm{BH}(M)$
  2. Calculate M_1, M_2 for A_1, M_1
  3. Then the distribution of $R_1 | G_1, A_1, M_1$ is approximately $\mathrm{BH}(M + M_1 + M_2)$
  4. Approximate $\mathrm{BH}(M + M_1 + M_2)$ as $\mathrm{LD}(r, \Sigma)$

Algorithm Summary (1)

- Input: initial distribution, noise variances, constant vectors $\mathbf{a}, \mathbf{m}$
- Output: $\bar{r}$
- Initialize $\bar{r} \leftarrow \bar{r}_0$, $\Sigma \leftarrow \Sigma_0$, $\bar{\omega} \leftarrow \bar{\omega}_0$, $\sigma_\Omega^2 \leftarrow \sigma_{\Omega,0}^2$
- For each time instance,
  1. Get measurements (g, a, m)
  2. Update using g:

$$\bar{r} \leftarrow \exp\left(-\left[\frac{(\sigma_U^2 + \sigma_A^2 \Delta t^2)\bar{\omega}}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma_\Omega^2} + \frac{\sigma_\Omega^2 g}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma_\Omega^2}\right]_\times \Delta t\right) \bar{r}$$

$$\Sigma \leftarrow \Sigma + \frac{(\sigma_U^2 + \sigma_A^2 \Delta t^2)\sigma_\Omega^2 \Delta t^2}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma_\Omega^2} \mathbf{1}$$

$$\bar{\omega} \leftarrow \frac{\sigma_U^2 \bar{\omega}}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma_\Omega^2} + \frac{(\sigma_A^2 \Delta t^2 + \sigma_\Omega^2) g}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma_\Omega^2}$$

$$\sigma_\Omega^2 \leftarrow \frac{\sigma_U^2 (\sigma_A^2 \Delta t^2 + \sigma_\Omega^2)}{\sigma_U^2 + \sigma_A^2 \Delta t^2 + \sigma_\Omega^2}$$

Algorithm Summary (1)

- Input: initial distribution, noise variances, constant vectors $\mathbf{a}, \mathbf{m}$
- Output: $\bar{r}$
- For each time instance,
  3. Update using a, m:
     - Let $M_1 := \begin{bmatrix} -\frac{\|a-\mathbf{a}\|^2}{2\sigma_V^2} - \frac{\|m-\mathbf{m}\|^2}{2\sigma_W^2} & \left(\frac{\mathbf{a}\times a}{\sigma_V^2} + \frac{\mathbf{m}\times m}{\sigma_W^2}\right)^T \\ \frac{\mathbf{a}\times a}{\sigma_V^2} + \frac{\mathbf{m}\times m}{\sigma_W^2} & \begin{matrix} \frac{a\mathbf{a}^T + \mathbf{a}a^T}{\sigma_V^2} + \frac{m\mathbf{m}^T + \mathbf{m}m^T}{\sigma_W^2} \\ -\left(\frac{\|a+\mathbf{a}\|^2}{2\sigma_V^2} + \frac{\|m+\mathbf{m}\|^2}{2\sigma_W^2}\right)\mathbf{1} \end{matrix} \end{bmatrix}$
     - Find $q_0 \in \mathrm{SU}(2)$ with $\pi(q_0) = \bar{r}$
     - Let $M_2 := Q_L(q_0) \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} Q_L(q_0^{-1})$
     - Find $(\bar{r}, \Sigma)$ such that $\mathrm{LD}(\bar{r}, \Sigma) \approx \mathrm{BH}(M_1 + M_2)$

Discussions

- Why (a, m) are used only for updating R but not Ω?
- Incorrect independence assumption $R_k \perp \Omega_k$
- Consideration of gyroscope bias and accel./magnet. disturbance
- Do claims really hold?
- Extension to $\mathrm{SE}(3)$?

Discussions

- Why (a, m) are used only for updating R but not Ω?
- Incorrect independence assumption $R_k \perp \Omega_k$
- Consideration of gyroscope bias and accel./magnet. disturbance
- Do claims really hold?
- Extension to $\mathrm{SE}(3)$?

Discussions

- Why (a, m) are used only for updating R but not Ω?
- Incorrect independence assumption $R_k \perp \Omega_k$
- Consideration of gyroscope bias and accel./magnet. disturbance
- Do claims really hold?
- Extension to $\mathrm{SE}(3)$?

Discussions

- Why (a, m) are used only for updating R but not Ω?
- Incorrect independence assumption $R_k \perp \Omega_k$
- Consideration of gyroscope bias and accel./magnet. disturbance
- Do claims really hold?
- Extension to $\mathrm{SE}(3)$?

Discussions

- Why (a, m) are used only for updating R but not Ω?
- Incorrect independence assumption $R_k \perp \Omega_k$
- Consideration of gyroscope bias and accel./magnet. disturbance
- Do claims really hold?
- Extension to $\mathrm{SE}(3)$?

Discussions

- Why (a, m) are used only for updating R but not Ω?
- Incorrect independence assumption $R_k \perp \Omega_k$
- Consideration of gyroscope bias and accel./magnet. disturbance
- Do claims really hold?
- Extension to $\mathrm{SE}(3)$?

Thank you.

References

S. Matthies, J. Muller, and G. W. Vinel.
On the normal distribution in the orientation space.
Textures & Microstructures, 10:77–96, 1988.

D. I. Nikolayev and T. I. Savyolova.
Normal distribution on the rotation group $\mathrm{SO}(3)$.
Textures & Microstructures, 29:201–233, 1997.

Derivation of Kalman Filter (1)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- Denote $\hat{X}_{k|k} := \mathrm{E}[X_k | Y_1^k]$
- Define the innovation sequence

$$\begin{aligned} \tilde{Y}_k &:= Y_k - \mathrm{E}[Y_k | Y_1^{k-1}] \\ &= Y_k - H\mathrm{E}[X_k | Y_1^{k-1}] \end{aligned}$$

- From Gaussian assumption, one can show that $Y_1^{k-1} \perp \tilde{Y}_k$

Derivation of Kalman Filter (1)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$
- Denote $\hat{X}_{k|k} := \mathrm{E}[X_k|Y_1^k]$
- Define the **innovation sequence**

$$\begin{aligned}\tilde{Y}_k &:= Y_k - \mathrm{E}[Y_k|Y_1^{k-1}] \\ &= Y_k - H\mathrm{E}[X_k|Y_1^{k-1}]\end{aligned}$$

- From Gaussian assumption, one can show that $Y_1^{k-1} \perp \tilde{Y}_k$

Derivation of Kalman Filter (1)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$

- Denote $\hat{X}_{k|k} := \mathrm{E}[X_k|Y_1^k]$
- Define the **innovation sequence**

$$\begin{aligned}\tilde{Y}_k &:= Y_k - \mathrm{E}[Y_k|Y_1^{k-1}] \\ &= Y_k - H\mathrm{E}[X_k|Y_1^{k-1}]\end{aligned}$$

- From Gaussian assumption, one can show that $Y_1^{k-1} \perp \tilde{Y}_k$

Derivation of Kalman Filter (1)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- Denote $\hat{X}_{k|k} := \mathrm{E}[X_k | Y_1^k]$
- Define the **innovation sequence**

$$\begin{aligned} \tilde{Y}_k &:= Y_k - \mathrm{E}[Y_k | Y_1^{k-1}] \\ &= Y_k - H\mathrm{E}[X_k | Y_1^{k-1}] \end{aligned}$$

- From Gaussian assumption, one can show that $Y_1^{k-1} \perp \tilde{Y}_k$

Derivation of Kalman Filter (2)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$

- Denote $\hat{X}_{k|k-1} := \mathrm{E}[X_k|Y_1^{k-1}]$ so that $\tilde{Y}_k = Y_k - H\hat{X}_{k|k-1}$
- One can show that if (X, Y, Z) are jointly Gaussian and $X \perp Y$ then

$$\mathrm{E}[Z|X, Y] = \mathrm{E}[Z|X] + \mathrm{E}[Z|Y] - \mathrm{E}[Z],$$

so

$$\hat{X}_{k|k} = \mathrm{E}[X_k|Y_1^k] = \hat{X}_{k|k-1} + \mathrm{E}[X_k|\tilde{Y}_k]$$

Derivation of Kalman Filter (2)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- Denote $\hat{X}_{k|k-1} := \mathrm{E}[X_k | Y_1^{k-1}]$ so that $\tilde{Y}_k = Y_k - H\hat{X}_{k|k-1}$
- One can show that if (X, Y, Z) are jointly Gaussian and $X \perp Y$ then

$$\mathrm{E}[Z|X, Y] = \mathrm{E}[Z|X] + \mathrm{E}[Z|Y] - \mathrm{E}[Z],$$

  so

$$\hat{X}_{k|k} = \mathrm{E}[X_k | Y_1^k] = \hat{X}_{k|k-1} + \mathrm{E}[X_k | \tilde{Y}_k]$$

Derivation of Kalman Filter (3)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- $\hat{X}_{k|k-1} = \mathrm{E}[X_k | Y_1^{k-1}] = F\mathrm{E}[X_{k-1} | Y_1^{k-1}] = F\hat{X}_{k-1|k-1}$
- Need to know $\mathrm{E}[X_k | \tilde{Y}_k]$
- According to a well-known formula for jointly Gaussian r.v.s:

$$\mathrm{E}[X_k | \tilde{Y}_k] = \mathrm{E}[X_k \tilde{Y}_k^T]\mathrm{E}[\tilde{Y}_k \tilde{Y}_k^T]^{-1}\tilde{Y}_k$$

Derivation of Kalman Filter (3)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- $\hat{X}_{k|k-1} = \mathrm{E}[X_k | Y_1^{k-1}] = F\mathrm{E}[X_{k-1} | Y_1^{k-1}] = F\hat{X}_{k-1|k-1}$
- Need to know $\mathrm{E}[X_k | \tilde{Y}_k]$
- According to a well-known formula for jointly Gaussian r.v.s:

$$\mathrm{E}[X_k | \tilde{Y}_k] = \mathrm{E}[X_k \tilde{Y}_k^T]\mathrm{E}[\tilde{Y}_k \tilde{Y}_k^T]^{-1}\tilde{Y}_k$$

Derivation of Kalman Filter (3)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- $\hat{X}_{k|k-1} = \mathrm{E}[X_k | Y_1^{k-1}] = F\mathrm{E}[X_{k-1} | Y_1^{k-1}] = F\hat{X}_{k-1|k-1}$
- Need to know $\mathrm{E}[X_k | \tilde{Y}_k]$
- According to a well-known formula for jointly Gaussian r.v.s:

$$\mathrm{E}[X_k | \tilde{Y}_k] = \mathrm{E}[X_k \tilde{Y}_k^T] \mathrm{E}[\tilde{Y}_k \tilde{Y}_k^T]^{-1} \tilde{Y}_k$$

Derivation of Kalman Filter (4)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- $\tilde{Y}_k = Y_k - \mathrm{E}[Y_k | Y_1^{k-1}] = H(X_k - \hat{X}_{k|k-1}) + U_k$, so
- $\mathrm{E}[\tilde{Y}_k \tilde{Y}_k^T] = HP_{k|k-1}H^T + R$, where
  - $P_{k|k-1} := \mathrm{E}[(X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T]$
  - This is the **prediction uncertainty**
- Again from independence, $\mathrm{E}[X_k \tilde{Y}_k^T] = P_{k|k-1}H^T$

Derivation of Kalman Filter (4)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- $\tilde{Y}_k = Y_k - \mathrm{E}[Y_k | Y_1^{k-1}] = H(X_k - \hat{X}_{k|k-1}) + U_k$, so
- $\mathrm{E}[\tilde{Y}_k \tilde{Y}_k^T] = HP_{k|k-1}H^T + R$, where
  - $P_{k|k-1} := \mathrm{E}[(X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T]$
  - This is the **prediction uncertainty**
- Again from independence, $\mathrm{E}[X_k \tilde{Y}_k^T] = P_{k|k-1}H^T$

Derivation of Kalman Filter (4)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$

- $\tilde{Y}_k = Y_k - \mathrm{E}[Y_k|Y_1^{k-1}] = H(X_k - \hat{X}_{k|k-1}) + U_k$, so
- $\mathrm{E}[\tilde{Y}_k \tilde{Y}_k^T] = HP_{k|k-1}H^T + R$, where
  - $P_{k|k-1} := \mathrm{E}[(X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T]$
  - This is the **prediction uncertainty**
- Again from independence, $\mathrm{E}[X_k \tilde{Y}_k^T] = P_{k|k-1}H^T$

Derivation of Kalman Filter (5)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- Hence,

$$\mathrm{E}[X_k | \tilde{Y}_k] = K_k \tilde{Y}_k$$

  where
  - $K_k := P_{k|k-1} H^T (H P_{k|k-1} H^T + R)^{-1}$
  - This is called the **Kalman gain**
- We are to find the prediction uncertainty $P_{k|k-1} := \mathrm{E}[(X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T]$ from now on

Derivation of Kalman Filter (5)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$

- Hence,

$$\mathrm{E}[X_k|\tilde{Y}_k] = K_k\tilde{Y}_k$$

  where
  - $K_k := P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
  - This is called the **Kalman gain**
- We are to find the prediction uncertainty $P_{k|k-1} := \mathrm{E}[(X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T]$ from now on

Derivation of Kalman Filter (5)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- Hence,

$$\mathrm{E}[X_k | \tilde{Y}_k] = K_k \tilde{Y}_k$$

  where
  - $K_k := P_{k|k-1} H^T (HP_{k|k-1} H^T + R)^{-1}$
  - This is called the **Kalman gain**
- We are to find the prediction uncertainty $P_{k|k-1} := \mathrm{E}[(X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})^T]$ from now on

Derivation of Kalman Filter (6)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- $X_k - \hat{X}_{k|k-1} = F(X_{k-1} - \hat{X}_{k-1|k-1}) + W_k$, so

$$P_{k|k-1} = FP_{k-1|k-1}F^T + Q$$

where
  - $P_{k-1|k-1} := \mathrm{E}[(X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T]$
  - This is the **estimation uncertainty**
- From $\hat{X}_{k|k} = \hat{X}_{k|k-1} + \mathrm{E}[X_k | \tilde{Y}_k]$, one can derive the formula

$$P_{k|k} = (\mathrm{I} - K_k H)P_{k|k-1}$$

Derivation of Kalman Filter (6)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- $X_k - \hat{X}_{k|k-1} = F(X_{k-1} - \hat{X}_{k-1|k-1}) + W_k$, so

$$P_{k|k-1} = FP_{k-1|k-1}F^T + Q$$

where
  - $P_{k-1|k-1} := \mathrm{E}[(X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T]$
  - This is the **estimation uncertainty**
- From $\hat{X}_{k|k} = \hat{X}_{k|k-1} + \mathrm{E}[X_k|\tilde{Y}_k]$, one can derive the formula

$$P_{k|k} = (\mathrm{I} - K_k H)P_{k|k-1}$$

Derivation of Kalman Filter (6)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- $X_k - \hat{X}_{k|k-1} = F(X_{k-1} - \hat{X}_{k-1|k-1}) + W_k$, so

$$P_{k|k-1} = FP_{k-1|k-1}F^T + Q$$

  where
  - $P_{k-1|k-1} := \mathrm{E}[(X_{k-1} - \hat{X}_{k-1|k-1})(X_{k-1} - \hat{X}_{k-1|k-1})^T]$
  - This is the **estimation uncertainty**
- From $\hat{X}_{k|k} = \hat{X}_{k|k-1} + \mathrm{E}[X_k|\tilde{Y}_k]$, one can derive the formula

$$P_{k|k} = (\mathrm{I} - K_k H) P_{k|k-1}$$

Derivation of Kalman Filter (7)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$

- In summary, the overall procedure is done in two phases:
  1. **Prediction phase**: compute $\hat{X}_{k|k-1}$, $P_{k|k-1}$ from $\hat{X}_{k-1|k-1}$, $P_{k-1|k-1}$ using the formulas
     - $\hat{X}_{k|k-1} = F\hat{X}_{k-1|k-1}$
     - $P_{k|k-1} = FP_{k-1|k-1}F^T + Q$
  2. **Update phase**: compute $\hat{X}_{k|k}$, $P_{k|k}$ from $\hat{X}_{k|k-1}$, $P_{k|k-1}$, and Y_k using the formulas
     - $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
     - $\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Y_k - H\hat{X}_{k|k-1})$
     - $P_{k|k} = (\mathrm{I} - K_kH)P_{k|k-1}$

Derivation of Kalman Filter (7)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k|Y_1, \cdots, Y_k]$

- In summary, the overall procedure is done in two phases:
  1. **Prediction phase**: compute $\hat{X}_{k|k-1}$, $P_{k|k-1}$ from $\hat{X}_{k-1|k-1}$, $P_{k-1|k-1}$ using the formulas
     - $\hat{X}_{k|k-1} = F\hat{X}_{k-1|k-1}$
     - $P_{k|k-1} = FP_{k-1|k-1}F^T + Q$
  2. **Update phase**: compute $\hat{X}_{k|k}$, $P_{k|k}$ from $\hat{X}_{k|k-1}$, $P_{k|k-1}$, and Y_k using the formulas
     - $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
     - $\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Y_k - H\hat{X}_{k|k-1})$
     - $P_{k|k} = (\mathrm{I} - K_kH)P_{k|k-1}$

Derivation of Kalman Filter (7)

- System model:
  - $X_k = FX_{k-1} + W_k$
  - $Y_k = HX_k + U_k$
- To compute: $\mathrm{E}[X_k | Y_1, \cdots, Y_k]$

- In summary, the overall procedure is done in two phases:
  1. **Prediction phase**: compute $\hat{X}_{k|k-1}$, $P_{k|k-1}$ from $\hat{X}_{k-1|k-1}$, $P_{k-1|k-1}$ using the formulas
     - $\hat{X}_{k|k-1} = F\hat{X}_{k-1|k-1}$
     - $P_{k|k-1} = FP_{k-1|k-1}F^T + Q$
  2. **Update phase**: compute $\hat{X}_{k|k}$, $P_{k|k}$ from $\hat{X}_{k|k-1}$, $P_{k|k-1}$, and Y_k using the formulas
     - $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$
     - $\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k(Y_k - H\hat{X}_{k|k-1})$
     - $P_{k|k} = (\mathrm{I} - K_k H)P_{k|k-1}$