Kalman Filter

EE468/CE468: Mobile Robotics

Dr. Basit Memon

Electrical and Computer Engineering
Habib University

October 18, 2023

Table of Contents

Table of Contents

Combining two noisy measurements, x_1 and x_2, of x

- The best linear estimate ($\hat{x} = a_1 x_1 + a_2 x_2$) is:

$$\hat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2$$

- This can also be written as:

$$\hat{x} = \overbrace{x_1}^{\text{Previous Measurement}} + \underbrace{\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}}_{\text{Kalman Gain}} \overbrace{(x_2 - x_1)}^{\text{Measurement Difference}}$$

- The obtained estimator is a special case of Discrete Kalman Filter.

Kalman Filter is the best linear unbiased estimator.

- Kalman Filter is used to estimate state when state is continuously changing according to some dynamics, usually.
- Instead of two noisy measurements of state, a KF uses one noisy prediction for the state and one measurement for the state.
- KF estimates the state at each time step.
- It uses the estimated state to generate prediction for the next time step.

Discrete Kalman Filter Algorithm in 1D, i.e. x is scalar

$$x_k = a\,x_{k-1} + b\,u_k + w_k$$
$$w_k \sim \mathcal{N}(0, r^2)$$
$$z_k = c\,x_k + n_k$$
$$n_k \sim \mathcal{N}(0, \sigma_2^2)$$

- State dynamics: how the state is determined from the previous state and control.
- The state evolution is also noisy (w_k). Noise has known covariance.
- State is not being directly measured, but a function of the state cx_k is measured.
- Measurement is also noisy with known covariance.

Evolution of state of robot system

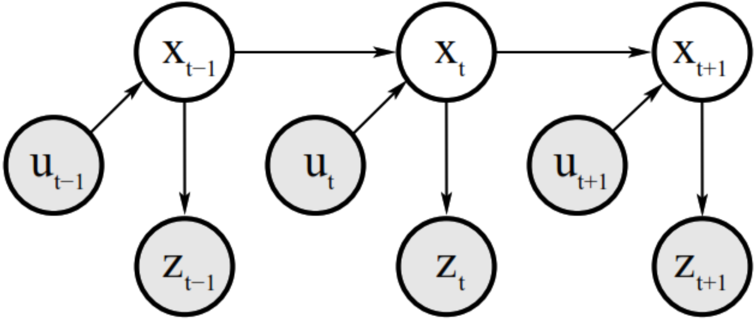


Figure: Relationship between state, action, measurements [2]

Estimate for state using prediction and measurement:

$$x_k = a\,x_{k-1} + b\,u_k + w_k$$
$$z_k = c\,x_k + n_k$$
$$w_k \sim \mathcal{N}(0, r^2)$$
$$n_k \sim \mathcal{N}(0, \sigma_2^2)$$

- We want to estimate state x_k at time k. Label this estimate as $\hat{x}_k$.
- We have two sources of information for this:
  - Prediction for the state x_k based on our estimate for the state at time $k-1$, i.e. $\hat{x}_{k-1}$. Let's label this prediction as $\bar{x}_k$.
  - A measurement z_k, which also yields a value for the state at time k, i.e. z_k/c.
- *All bar terms will be predictions based on past values and hat terms will be our final estimates.*

Combine two candidates based on our previous BLUE equation

$$\hat{x}_k = \underbrace{\bar{x}_k}_{1} + \underbrace{K}_{2} \underbrace{\left(\frac{z_k}{c} - \bar{x}_k\right)}_{3}$$

$$x_k = a\,x_{k-1} + b\,u_k + w_k$$
$$z_k = c\,x_k + n_k$$
$$w_t \sim \mathcal{N}(0, r^2)$$
$$n_t \sim \mathcal{N}(0, \sigma_2^2)$$

To find K, we will need variances:

$$Var\left[\frac{z_k}{c}\right] = Var\left[\frac{cx_k + v_k}{c}\right] = Var\left[x_k + \frac{v_k}{c}\right]$$

Since x_k and v_k are independent:

$$Var\left[\frac{z_k}{c}\right] = Var[x_k] + Var\left[\frac{v_k}{c}\right] = 0 + \frac{\sigma_2^2}{c^2}$$

Explanation of different terms in equations

1. Prediction for x_k
2. Kalman Gain: Which do you prefer - the measurement or the prior?
3. Error between actual measurement and expected measurement.

- Thus, estimate for state at time k is:

$$\begin{aligned}\hat{x}_k &= \bar{x}_k + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2/c^2}\left(\frac{z_k}{c} - \bar{x}_k\right) \\ &= \underbrace{\bar{x}_k}_{1} + \underbrace{\frac{c\sigma_1^2}{c^2\sigma_1^2 + \sigma_2^2}}_{2}\underbrace{(z_k - c\bar{x}_k)}_{3}\end{aligned}$$

Uncertainty associated with this estimate is the error variance.

1. Variance of the prediction $\bar{x}_k$.
2. Variance of the state estimate due to measurement.
3. Kalman Gain: Which do you prefer - the measurement or the prior?

- The error variance is $\Sigma_k = E[(x - \hat{x})^2]$:

$$\Sigma_k = \overbrace{\sigma_1^2}^{1} - \frac{\sigma_1^4}{\sigma_1^2 + \underbrace{\sigma_2^2/c^2}_{2}}$$

$$= \underbrace{\sigma_1^2}_{1} - \underbrace{\frac{c\sigma_1^2}{c^2\sigma_1^2 + \sigma_2^2}}_{3} c\sigma_1^2$$

Even poor measurements will increase the precision of estimate.

- $$\sigma^2 = \sigma_1^2 \boxed{-} \frac{c\sigma_1^2}{c^2\sigma_1^2 + \sigma_2^2} c\sigma_1^2$$

No matter what, variance of posterior will be reduced, i.e. confidence in estimate increases even with noisy measurements.

Where does the prediction $\bar{x}_k$ come from?

- Assume that the estimate for x_{k-1} computed at the previous time step was $\hat{x}_{k-1}$. Then,

$$\bar{x}_k = a\,\hat{x}_{k-1} + b\,u_k$$

What is the associated prediction error variance, σ_1^2?

$$\sigma_1^2 = \bar{\Sigma}_k = E\left[(x_k - \bar{x}_k)^2\right]$$
$$= E\left[(ax_{k-1} + bu_k + w_k - a\hat{x}_{k-1} - bu_k)^2\right]$$
$$= E\left[a^2(x_{k-1} - \hat{x}_{k-1})^2 + w_k^2 + 2a(x_{k-1} - \hat{x}_{k-1})\,w_k\right]$$
$$= a^2 E\left[(x_{k-1} - \hat{x}_{k-1})^2\right] + E\left[w_k^2\right] + 2aE\left[x_{k-1} - \hat{x}_{k-1}\right]E\left[w_k\right],$$

where the last term is due to the uncorrelatedness of $\hat{x}_{k-1}$ and w_k

$$\sigma_1^2 = \bar{\Sigma}_k = a^2\,\Sigma_{k-1} + r^2$$

where last term is zero as $E[x_{k-1} - \hat{x}_{k-1}] = 0$ because of estimator being unbiased.

$$x_k = a\,x_{k-1} + b\,u_k + w_k$$
$$z_k = c\,x_k + n_k$$
$$w_k \sim \mathcal{N}(0, r^2)$$
$$n_k \sim \mathcal{N}(0, \sigma_2^2)$$

1. $\bar{x}_k = a\,\hat{x}_{k-1} + b\,u_k$
2. $\bar{\Sigma}_k = a^2\,\Sigma_{k-1} + r^2$
3. $K_k = \frac{c\,\bar{\Sigma}_k}{c^2\,\bar{\Sigma}_K + \sigma_2^2}$
4. $\hat{x}_k = \bar{x}_k + K_k\,(z_k - c\,\bar{x}_k)$
5. $\Sigma_k = (1 - c\,K_k)\,\bar{\Sigma}_k$

What are the underlying assumptions for KF?

- Motion model is linear with additive zero-mean uncorrelated ($E[w_k w_{k+1}] = E[w_k]E[w_{k+1}]$) Gaussian process noise:

$$x_k = A_k x_{k-1} + B_k u_k + w_k, \quad \text{where, } w_k \sim \mathcal{N}(0, R_k).$$

- Measurement model is linear with additive independent Gaussian noise:

$$z_k = C_k x_k + n_k, \quad \text{where, } n_k \sim \mathcal{N}(0, Q_k).$$

- Initial belief, $bel(x_0)$, must be normally distributed.

- Markov assumption.

Why are there so many assumptions?

- Assumptions ensure that all PDFs, including posterior, are Gaussian at each step, and we only need to store and process their mean and covariance.

- **Optimal filter under these assumptions.**

- Motion model is linear with additive uncorrelated Gaussian noise:

$$x_k = A_k x_{k-1} + B_k u_k + w_k, \quad \text{where, } w_k \sim \mathcal{N}(0, R_k).$$

- Measurement model is linear with additive independent Gaussian noise:

$$z_k = C_k x_k + n_k, \quad \text{where, } n_t \sim \mathcal{N}(0, Q_k).$$

- Initial belief, $bel(x_0)$, must be normally distributed.

- Markov assumption.

More insights into the KF

- Use the provided MATLAB livescript file to play around with KF.

Table of Contents

Kalman Filter Algorithm [1, 5.3]

$$
\begin{aligned}
x_k &= A_k\, x_{k-1} + B_k\, u_k + w_k \\
z_k &= C_k\, x_k + n_k \\
w_k &\sim \mathcal{N}(0, R_k) \\
n_k &\sim \mathcal{N}(0, Q_k)
\end{aligned}
$$

1. $\bar{x}_k = A_k\, \hat{x}_{k-1} + B_k\, u_k$
2. $\bar{\Sigma}_k = A_k\, \Sigma_{k-1}\, A_k^T + R_k$
3. $K_k = \bar{\Sigma}_k\, C_k^T\, (C_k\, \bar{\Sigma}_k\, C_k^T + Q_k)^{-1}$
4. $\hat{x}_k = \bar{x}_k + K_k\, (z_k - C_k\, \bar{x}_k)$
5. $\Sigma_k = (I - K_k\, C_k)\, \bar{\Sigma}_k$

$$A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$$

$$Q_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Delta t = 1$$

- We have a robot moving in a straight line. It's state vector is $x_k = (p, v)^T$, where p is position and v is velocity.
- Suppose we don't have any control over this robot and we predict zero acceleration in near future.

$$p_k = p_{k-1} + v_{k-1}\Delta t + w_p(k)$$
$$v_k = v_{k-1} + w_v(k)$$

Written in standard form:

$$x_k = A\,x_k + w_k$$
$$w_k \sim \mathcal{N}(0, Q_k)$$

Kalman Filter Example in 2D: **Measurement Model**

- We observe the position only, i.e. $z_k = p_k + n_k$.

- Written in standard form,

$$
\begin{aligned}
z_k &= C_k \, x_k + n_k \\
C_k &= \begin{bmatrix} 1 & 0 \end{bmatrix} \\
n_k &\sim \mathcal{N}(0, R_k) \\
R_k &= 1
\end{aligned}
$$

Kalman Filter Example in 2D: **Prior**

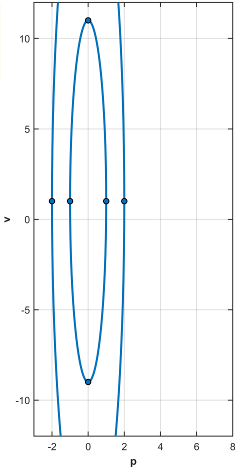


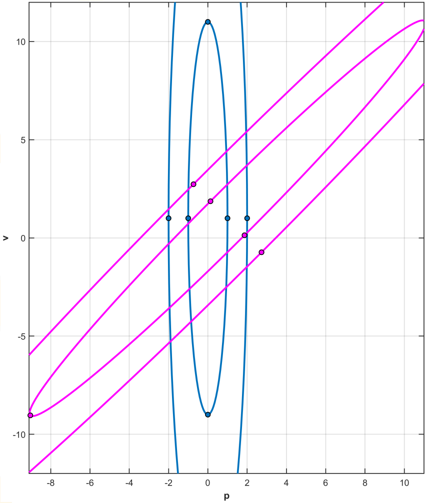
Figure: 1σ and 2σ ellipses for $x_{t|t}$

- At time k, the belief of state is given by $(\hat{x}_k, \hat{\Sigma}_k)$, such that

$$\hat{x}_k = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \hat{\Sigma}_k = \begin{bmatrix} 1^2 & 0 \\ 0 & 10^2 \end{bmatrix}$$

- We're certain with high probability (~ 0.997) that position is within $3\sigma_p = 3$ range of the mean position, 0.

- Less certain about velocity. Confident with probability (~ 0.997) that velocity is within $3\sigma_v = 30$ range of the mean velocity, 1.



- Find $\bar{x}_{k+1}$ and $\bar{\Sigma}_{k+1}$.
-

$$\bar{x}_{k+1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\bar{\Sigma}_{k+1} = \begin{bmatrix} 102 & 100 \\ 100 & 101 \end{bmatrix}$$

Kalman Filter Example in 2D: **Update Step**

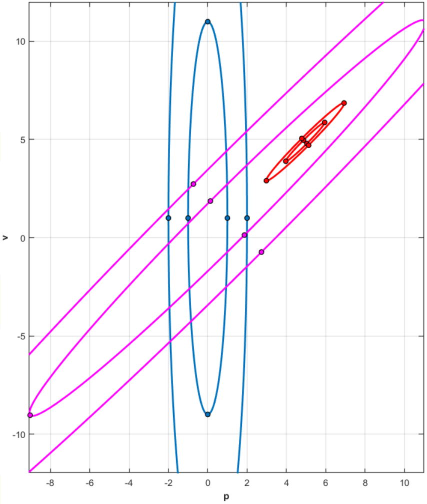


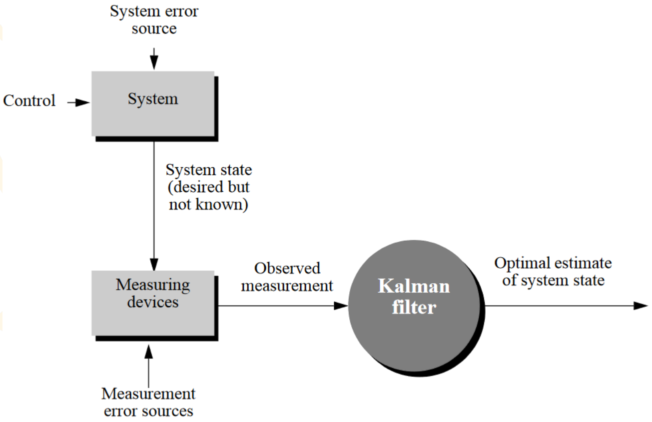
Figure: 1σ and 2σ ellipses for $x_{t+1|t+1}$

- Find $\hat{x}_{k+1}$ and Σ_{k+1}, when $z_{k+1} = 5$.
-

$$\hat{x}_{k+1} = \begin{bmatrix} 4.96 \\ 4.88 \end{bmatrix}$$

$$\Sigma_{k+1} = \begin{bmatrix} 0.99 & 0.97 \\ 0.97 & 0.98 \end{bmatrix}$$

More insights into the KF



Reading:
http://www.bzarg.com/p/how-a-kalman-filter-works-in-pictures/

Table of Contents

Linearity is not realistic!

Most robotic systems are nonlinear!

$$\xcancel{x_k = A_k x_{k-1} + B_k u_k + w_k} \qquad \xcancel{z_k = C_k x_k + n_k}$$

$$x_k = g(u_k, x_{k-1}) + w_k \qquad z_k = h(x_k) + n_k$$

But, the KF relied on linear models! What happens to KF?

Underlying Belief distribution is no longer Gaussian !

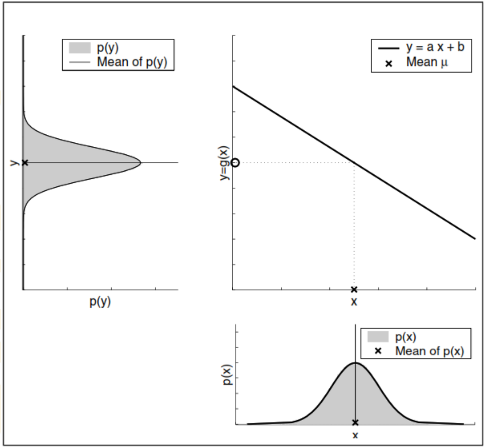


Figure: Linear Transformation

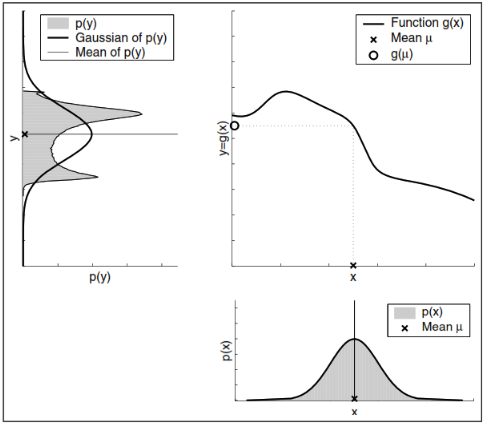


Figure: Nonlinear Transformation. Closed form doesn't exist.

How did we even draw posterior?

- Take a lot of samples of $p(x)$. Pass them through $g(x)$. Draw histogram of results.
- Find mean and variance of histogram distribution. Solid line is Gaussian corresponding to this mean and variance.
- EKF approximates true belief by Gaussian.

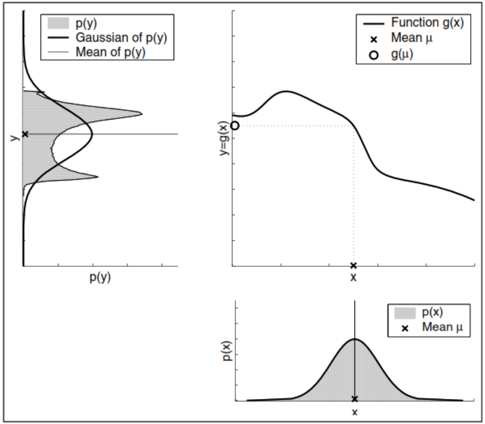


Figure: Nonlinear Transformation. Closed form doesn't exist.

How to effectively compute mean and covariance?

- Method on last slide loses computational advantage of KF.
- **Idea:** Linearization of models **about mean** using Taylor approximation.

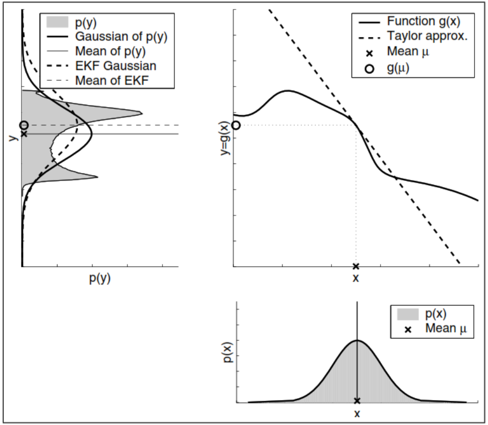


Figure: Linearization about mean

Linearizations of Prediction and Measurement

$$x_k = g(u_k, x_{k-1}) + w_k$$
$$z_k = h(x_k) + n_k$$

- **Prediction**

$$g(x_{k-1}, u_k) \approx g(\hat{x}_{k-1}, u_k) + \underbrace{\frac{\partial g(\hat{x}_{k-1}, u_t)}{\partial x_{k-1}}}_{\text{Jacobian}} (x_{k-1} - \hat{x}_{k-1})$$

$$g(x_{k-1}, u_k) \approx g(\hat{x}_{k-1}, u_k) + \overbrace{G_k}\,(x_{k-1} - \hat{x}_{k-1})$$

- **Measurement**

$$h(x_k) \approx h(\bar{x}_k) + \underbrace{\frac{\partial h(\bar{x}_k)}{\partial x_k}}_{\text{Jacobian}} (x_k - \bar{x}_k)$$

$$h(x_k) \approx h(\bar{x}_k) + \overbrace{H_k}\,(x_k - \bar{x}_k)$$

Kalman Filter

1. $\bar{x}_k = A_k \, \hat{x}_{k-1} + B_k \, u_k$
2. $\bar{\Sigma}_k = A_k \, \Sigma_{k-1} \, A_k^T + R_k$
3. $K_k = \bar{\Sigma}_k \, C_k^T \, (C_k \, \bar{\Sigma}_k \, C_k^T + Q_k)^{-1}$
4. $\hat{x}_k = \bar{x}_k + K_k \, (z_k - C_k \, \bar{x}_k)$
5. $\Sigma_k = (I - K_k \, C_k) \, \bar{\Sigma}_k$

Extended Kalman Filter

1. $\bar{x}_k = g(\hat{x}_{k-1}, u_k)$
2. $\bar{\Sigma}_k = G_k \, \Sigma_{k-1} \, G_k^T + R_k$
3. $K_k = \bar{\Sigma}_k \, H_k^T \, (H_k \, \bar{\Sigma}_t \, H_k^T + Q_k)^{-1}$
4. $\hat{x}_k = \bar{x}_k + K_k \, (z_k - h(\bar{x}_k))$
5. $\Sigma_k = (I - K_k \, H_k) \, \bar{\Sigma}_k$

$$G_k = \frac{\partial g(\hat{x}_{k-1}, u_k)}{\partial x_{k-1}}$$

$$H_k = \frac{\partial h(\bar{x}_k)}{\partial x_k}$$

- Given a vector-valued function

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}$$

- Jacobian matrix, given $x \in \mathbb{R}^n$, is:

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Filter Divergence: Keep uncertainty of state estimate small.

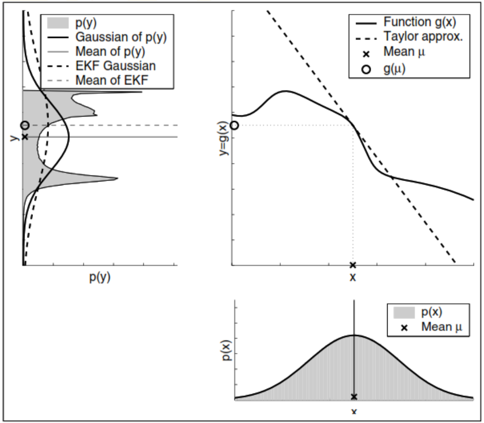


Figure: Mean is same, but variance of x is larger.

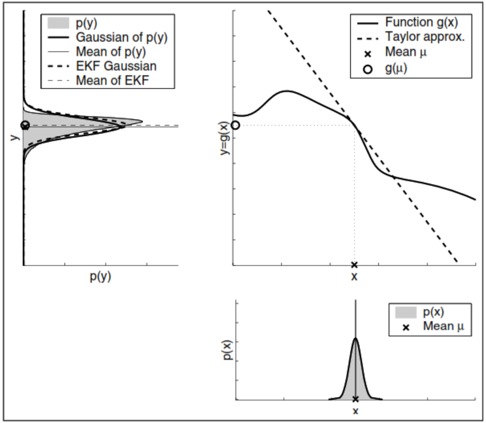


Figure: Small variance of x results in EKF mean and variance aligning with the one computed from large number of samples.

Final Comments

- There exist better ways for dealing with nonlinearities, such as unscented Kalman filter.
- EKF is not an optimal estimator.

Figure: I was on every Apollo lunar mission.

- Could be argued that sensor fusion is key to robust localization, and we mostly don't need the complete posterior PDF.
- KF doesn't approximate posterior PDF, but computes exact one by making simplifying assumptions about robotic system.
- Applied in economics, control, weather forecasting, satellites, robotics, etc.
- Typically used for position tracking or local localization.

Table of Contents

[1] **Alonzo Kelly.**
Mobile robotics: mathematics, models, and methods.
Cambridge University Press, 2013.

[2] **Sebastian Thrun, Wolfram Burgard, and Dieter Fox.**
Probabilistic robotics.
2006.