Kalman Filter

SOMA BISWAS

DEPARTMENT OF ELECTRICAL ENGINEERING
IISC, BANGALORE
Rudolf Emil Kalman

- Born 1930 in Hungary
- BS and MS from MIT
- PhD 1957 from Columbia
- Filter developed in 1960-61
- KF: optimal recursive data processing algorithm
- Optimal: Under the assumptions made
- Recursive: does not require all previous data to be stored & reprocessed every time a new measurement is taken – practical for implementation purpose
Sailors a sea – do not know your location
Star sighting to establish your position
At time t1 you establish your position to be z1 (you measure)
Measurement inaccuracy – precision: σz1 (std)
Conditional probability of x(t1) given observed z1.

Based on this conditional pdf the best estimate of your position is
\[ \hat{x}(t_1) = z_1 \]
and the variance of the error in the estimate is
\[ \sigma_x^2(t_1) = \sigma_{z1}^2 \]
A Static Example – Contd.

- You have a friend that is a trained sailor.
- At time instant $t_2$ ($t_2 = t_1$) (the boat did not move) this trained sailor measures $z_2$ with a variance $\sigma_{z_2}^2$.
- As this second sailor has larger skills, assume that the variance in his measurement is smaller than the measurement in yours.

Based on this conditional pdf the best estimate of the position given by the trained sailor $\hat{x}(t_2) = z_2$.

and the variance of the error in the estimate is $\sigma_x^2(t_2) = \sigma_{z_2}^2$.

FIG. 1.5 Conditional density of position based on measurement $z_2$ alone.
At this point we have two measurements, with different uncertainty

How to combine these data?

What do we want to know?

- Conditional position at time $t_2$ ($t_2 = t_1$) given both $z_1$ and $z_2$

$$f_{x(t_2)|z(t_1),z(t_2)}(x|z_1,z_2)$$

is Gaussian

$$\mu = \frac{\sigma^2_{z_2}}{\sigma^2_{z_1} + \sigma^2_{z_2}} Z_1 + \frac{\sigma^2_{z_1}}{\sigma^2_{z_1} + \sigma^2_{z_2}} Z_2$$

$$\frac{1}{\sigma^2} = \frac{1}{\sigma^2_{z_1}} + \frac{1}{\sigma^2_{z_2}}$$

The uncertainty in the position estimate decreased by combining the two pieces of information.

**FIG. 1.6** Conditional density of position based on data $z_1$ and $z_2$. 
A Static Example – Contd.

- Which is the best estimate at time $t_2$ given $z_1$ and $z_2$?
  - The mean (also the maximum) of the conditional pdf

\[
\hat{x}(t_2) = \frac{\sigma^2_{z_2}}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_1 + \frac{\sigma^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_2
\]

\[
\hat{x}(t_2) = z_1 + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} (z_2 - z_1)
\]

In the previous time instant
\[
\hat{x}(t_1) = z_1
\]

\[
\hat{x}(t_2) = \hat{x}(t_1) + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} (z_2 - \hat{x}(t_1))
\]

This is the same as the Kalman filter implementation

The best prediction at time $t_2$, is given by the previous estimate plus an error multiplied by a gain

Interpret the error
A Static Example – Contd.

- Which is the best estimate at time $t_2$ given $z_1$ and $z_2$?
  - The mean (also the maximum) of the conditional pdf

\[ \hat{x}(t_2) = \hat{x}(t_1) + K(t_2)(z_2 - \hat{x}(t_1)) \]

- Which is the uncertainty of the estimate at time $t_2$ given $z_1$ and $z_2$?
  - The variance of the conditional pdf

\[ \frac{1}{\sigma^2} = \frac{1}{\sigma^2_{z_1}} + \frac{1}{\sigma^2_{z_2}} \]

\[ K(t_2) = \frac{\sigma^2_{z_1}}{\sigma^2_{z_1} + \sigma^2_{z_2}} \]

\[ \sigma^2_x(t_2) = \sigma^2_x(t_1) - K(t_2)\sigma^2_x(t_1) \]
Introduce Dynamics

- From now on observations are done at different time instants and the boat travels between the time instants where measurements are taken.

\[
\frac{dx}{dt} = u + w
\]

Motion model:
- \( u \) = velocity
- \( w \) = noise term representing uncertainty in the actual knowledge of velocity, assumed zero mean

- At time \( t_2 \) we already have \( \hat{x}(t_2) \) and \( \sigma_x^2(t_2) \).
- Now the vehicle is travelling.
- Before doing a measurement at time instant \( t_3 \) (i.e., at \( t_3^- \)) which is the best prediction that we can do about the position and associated uncertainty?

\[
\hat{x}(t_3^-) = \hat{x}(t_2) + u(t_3 - t_2)
\]

\[
\sigma_x^2(t_3^-) = \sigma_x^2(t_2) + \sigma_w^2(t_3 - t_2)
\]

Gaussian with zero mean & variance \( \sigma_w^2 \).
A measurement $z_3$ is done at time instant $t_3$ with an assumed variance $\sigma_{z_3}^2$.
Which is now the best estimate $\hat{x}(t_3)$?

Once again we have two Gaussian probability density functions:
- One associated with information up to $t_3^-$
- One provided by the measurement itself

Combining both:

\[
\hat{x}(t_3) = \hat{x}(t_3^-) + K(t_3)(z_3 - \hat{x}(t_3^-))
\]

\[
\sigma_x^2(t_3) = \sigma_x^2(t_3^-) - K(t_3)\sigma_x^2(t_3^-)
\]

\[
K(t_3) = \frac{\sigma_x^2(t_3^-)}{\sigma_x^2(t_3^-) + \sigma_{z_3}^2} \quad \text{Kalman Gain}
\]
Consider a dynamic process described by an $n$-th order difference equation

\[ y_{i+1} = a_0, i y_i + \ldots + a_{n-1}, i y_{i-n+1} + u_i, \ i \geq 0 \]

**State Space Model**

**Process Model**

\[ \dot{x}_{i+1} = A \dot{x}_i + Gu_i \]

**Measurement Model**

\[ \hat{y}_i = H \hat{x}_i \]

Equation (3.1) represents the way a new state $\dot{x}_{i+1}$ is modeled as a linear combination of both the previous state $\dot{x}_i$ and some process noise $u_i$. Equation (3.2) describes the way the process measurements or observations $\hat{y}_i$ are derived from the internal state $\dot{x}_i$. These two

**Process and Measurement Noise**

\[ x_k = A x_{k-1} + Bu_k + w_{k-1} \]

\[ z_k = H x_k + v_k \]

$w_k$ and $v_k$
Kalman Filter – 2-step Process

- **Predict:** computes estimates of the current state variables, along with their uncertainties.
- **Correct:** estimates are updated using a weighted average with the noisy measurement.
- More weight being given to estimates with higher certainty.
- Algorithm recursive nature - > Can run in **real time** using only present input measurements and previously calculated state.

- The KF does not require all previous data to be kept in storage and reprocessed every time a new measurement is taken.

\[
\begin{align*}
  z(0) & \quad z(1) & \quad z(2) & \quad \ldots & \quad z(k) \\
  \downarrow & & & & & & \downarrow \\
  \hat{x}(k) & & & & & & &
\end{align*}
\]

\[
\begin{align*}
  z(0) & \quad z(1) & \quad z(2) & \quad \ldots & \quad z(k) & \quad z(k+1) \\
  \downarrow & & & & & & \downarrow & & \downarrow \\
  \hat{x}(k) & \quad \hat{x}(k+1) & & & & &
\end{align*}
\]

To evaluate \( \hat{x}(k+1) \) the KF only requires \( \hat{x}(k) \) and \( z(k+1) \).
Goal: Estimate the state \( x \in \mathbb{R}^n \)

Discrete-time controlled process governed by the linear stochastic difference equation

\[
x_k = Ax_{k-1} + Bu_k + w_{k-1}
\]

Measurement

\[
z \in \mathbb{R}^m \quad z_k = Hx_k + v_k
\]

Process and measurement noise

\( w_k \) and \( v_k \)

Assumed to be independent (of each other), white, and with normal probability distributions

\[
p(w) \sim N(0, Q) \\
p(v) \sim N(0, R)
\]

A: relates state at previous time step to state at current step
B: relates control input to the state \( x \).
H: relates the state to the measurement
A,H,Q,R: assumed constant (can vary with time)
State Estimates

- \( \hat{x}_k \in \mathbb{R}^n \): *a priori* state estimate at step \( k \) given knowledge of the process prior to step \( k \)

- \( \hat{x}_k \in \mathbb{R}^n \): *a posteriori* state estimate at step \( k \) given measurement \( z_k \)

- *a priori* and *a posteriori* estimate errors

The *a priori* estimate error covariance is then

\[
P_k^- = E[e_k^- e_k^-^T]
\]

and the *a posteriori* estimate error covariance is

\[
P_k = E[e_k e_k^T]
\]
Kalman Filter equation that computes an *a posteriori* state estimate $\hat{x}_k$ as a linear combination of an *a priori* estimate $\hat{x}_k^-$ and a weighted difference between an actual measurement $z_k$ and a measurement prediction $H\hat{x}_k^-$ as shown below in equation (4.7). Some justification for

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-)$$

- $(z_k - H\hat{x}_k^-)$: Measurement *innovation* or *residual*
- Reflects discrepancy between predicted and actual measurement
- $K$: *gain* or *blending factor* that minimizes the *a posteriori* error covariance equation

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1}$$

$$= \frac{P_k^- H^T}{H P_k^- H^T + R}$$
Kalman Filter

Looking at equation (4.8) we see that as the measurement error covariance $R$ approaches zero, the gain $K$ weights the residual more heavily. Specifically,

$$\lim_{R_k \to 0} K_k = H^{-1}.$$ 

On the other hand, as the \textit{a priori} estimate error covariance $P_k^-$ approaches zero, the gain $K$ weights the residual less heavily. Specifically,

$$\lim_{P_k^- \to 0} K_k = 0.$$ 

- Measurement error covariance approaches zero, the actual measurement is “trusted” more and more,
- As the \textit{a priori} estimate error covariance approaches zero the actual measurement is trusted less and less, while the predicted measurement is trusted more and more.
Filter estimates the state at some time and then obtains feedback in the form of (noisy) measurements.

**Time update (Predictor) equations**: project forward (in time) current state and error covariance estimates to obtain *a priori* estimates for the next time.

**Measurement update (Corrector) equations**: feedback—i.e. incorporate a new measurement into the *a priori* estimate to obtain an improved *a posteriori* estimate.

*Figure 4.1*: The ongoing discrete Kalman filter cycle. The *time update* projects the current state estimate ahead in time. The *measurement update* adjusts the projected estimate by an actual measurement at that time.
Kalman Filter Update Equations

**Time Update ("Predict")**

1. Project the state ahead
   \[ \hat{x}_k^- = A\hat{x}_{k-1} + Bu_k \]
2. Project the error covariance ahead
   \[ P_k^- = AP_{k-1}A^T + Q \]

**Measurement Update ("Correct")**

1. Compute the Kalman gain
   \[ K_k = P_k^-H^T(HP_k^-H^T + R)^{-1} \]
2. Update estimate with measurement \( z_k \)
   \[ \hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \]
3. Update the error covariance
   \[ P_k = (I - K_kH)P_k^- \]

Initial estimates for \( \hat{x}_{k-1} \) and \( P_{k-1} \)
Practical implementation of the Kalman Filter is often difficult due to the inability in getting a good estimate of the noise covariance matrices $Q_k$ and $R_k$.

Extensive research has been done in this field to estimate these covariances from data.

It is known from the theory that the Kalman filter is optimal in case that a) the model perfectly matches the real system, b) the entering noise is white and c) the covariances of the noise are exactly known.
Limitations

- Kalman Filter Assumption: *linear* stochastic difference equation
- What happens if the process to be estimated and (or) the measurement relationship to the process is non-linear?
- A Kalman filter that linearizes about the current mean and covariance is referred to as an *extended Kalman filter* or EKF.
**EKF: Extended Kalman Filter**

- **Assumption**: Process is governed by the *non-linear* stochastic difference equation

\[ x_k = f(x_{k-1}, u_k, w_{k-1}) \]

- \( x \in \mathbb{R}^n \): State vector
- Non-linear function
- Driving function
- Zero-mean process noise

\[ z_k = h(x_k, v_k) \]

- \( z \in \mathbb{R}^m \): Measurement
- \( w_k \) and \( v_k \): Process and Measurement Noise
Extended Kalman Filter

Discrete-time predict and update equations

Predict

Predicted state estimate
\[ \hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}, u_{k-1}) \]
Predicted covariance estimate
\[ P_{k|k-1} = F_{k-1} P_{k-1|k-1} F_{k-1}^T + Q_{k-1} \]

Update

Innovation or measurement residual
\[ \tilde{y}_k = z_k - h(\hat{x}_{k|k-1}) \]
Innovation (or residual) covariance
\[ S_k = H_k P_{k|k-1} H_k^T + R_k \]
Near-optimal Kalman gain
\[ K_k = P_{k|k-1} H_k^T S_k^{-1} \]
Updated state estimate
\[ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k \]
Updated covariance estimate
\[ P_{k|k} = (I - K_k H_k) P_{k|k-1} \]

where the state transition and observation matrices are defined to be the following Jacobians

\[ F_{k-1} = \frac{\partial f}{\partial x} \bigg|_{\hat{x}_{k-1|k-1}, u_{k-1}} \]
\[ H_k = \frac{\partial h}{\partial x} \bigg|_{\hat{x}_{k|k-1}} \]

Like Taylor series, linearize the estimation around current estimate using the partial derivatives of process and measurement functions to compute estimates even when equation is non-linear.
Kalman Filter

Predict

Predicted (a priori) state estimate
\( \hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + B_k u_{k-1} \)
Predicted (a priori) estimate covariance
\( P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \)

Update

Innovation or measurement residual
\( \tilde{y}_k = z_k - H_k \hat{x}_{k|k-1} \)
Innovation (or residual) covariance
\( S_k = H_k P_{k|k-1} H_k^T + R_k \)
Optimal Kalman gain
\( K_k = P_{k|k-1} H_k^T S_k^{-1} \)
Updated (a posteriori) state estimate
\( \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k \)
Updated (a posteriori) estimate covariance
\( P_{k|k} = (I - K_k H_k) P_{k|k-1} \)
Unscented Kalman Filter

- probability distribution that is characterized only in terms of a finite set of statistics
- **Unscented Transform** (UT) is a mathematical function used to estimate the result of applying a given nonlinear transformation to a probability distribution
- Application: nonlinear projection of mean and covariance estimates in the context of nonlinear extensions of the Kalman Filter

- state of a system in the form of a mean vector \( m \) and an associated error covariance matrix \( M \) (this is the first 2 moments of an underlying, but otherwise unknown probability distribution)

- 2-d position of an object of interest might be represented by a mean position vector \([x,y]\), with an uncertainty given by a 2x2 covariance matrix

- Linear Transformation \( T \) – Output \( Tm \) and \( TMT^T \)
Unscented Kalman Filter

- Not possible to determine m and M resulting from a non-linear transformation
- They can only be approximated

- Approach 1: Linearize the non-linear function (EKF)
- UT: Given mean and covariance can be exactly coded in a set of points (sigma points), which if treated as elements of a discrete prob. Distribution has mean m and M equal to the given m and M
- Distribution can be propagated exactly by applying the non-linear function to each point
- M and M of the transformed set of points = the desired transformed estimate.

- Advantage: Non-linear function is fully used, as opposed to EKF which replaces it with a linear one
- Can be applied with any function, whereas linearization may not be possible for functions that are not differentiable
- Easy to implement
Unscented Transform

- N+1 sigma points necessary and sufficient to define a discrete distribution (n=dimensions)
- 2d
  \[ s_1 = [0, \sqrt{2}]^T, \quad s_2 = \left[ -\frac{3}{2}, -\sqrt{1/2} \right]^T, \quad s_3 = \left[ \frac{3}{2}, \sqrt{1/2} \right]^T \]

- Mean = 0, Covariance S = I
- Given any 2d (x,X), desired sigma points:
- Multiply each point by matrix square root of X and add x
- Eg – Conversion from Cartesian to polar co-ordinate system
- Suppose in Cartesian co-ordinates

\[ m = [12.3, 7.6]^T, \quad M = \begin{bmatrix} 1.44 & 0 \\ 0 & 2.89 \end{bmatrix} \]

- Transformation to polar co-ordinates

\[ r = \sqrt{x^2 + y^2}, \quad \theta = \arctan(y/x) \]
Unscented Transform

Multiplying each of the canonical simplex sigma points (given above) by $M^{1/2} = \begin{bmatrix} 1.2 & 0 \\ 0 & 1.7 \end{bmatrix}$ and adding the mean, $m$, gives:

- $m_1 = [0, 2.40] + [12.3, 7.6] = [12.3, 10.0]$
- $m_2 = [-1.47, -1.20] + [12.3, 7.6] = [10.8, 6.40]$
- $m_3 = [1.47, -1.20] + [12.3, 7.6] = [13.8, 6.40]$

Applying the transformation function $f()$ to each of the above points gives:

- $m^+_1 = f(12.3, 10.0) = [15.85, 0.68]$
- $m^+_2 = f(10.8, 6.40) = [12.58, 0.53]$
- $m^+_3 = f(13.8, 6.40) = [15.18, 0.44]$

The mean of these three transformed points, $m_{UT} = \frac{1}{3} \sum_{i=1}^{3} m^+_i$, is the UT estimate of the mean in polar coordinates:

$\quad m_{UT} = [14.539, 0.551]$

The UT estimate of the covariance is:

$\quad M_{UT} = \frac{1}{3} \sum_{i=1}^{3} (m^+_i - m_{UT})^2$
Unscented Transform

In filtering, cumulative effect of small errors can lead to unrecoverable estimates.

Effect of errors worse: when covariance is underestimated -> this causes filter to be overconfident in the accuracy of the mean.

Result suggests that linearization has likely produced an underestimate of the actual error in its mean.

The Unscented Transform, especially as part of the UKF, has largely replaced the EKF in many nonlinear filtering and control applications, including for underwater,\textsuperscript{[9]} ground and air navigation,\textsuperscript{[10]} and spacecraft.\textsuperscript{[11]}
Kalman Filter – Foreground Segmentation

- Segmenting Foreground Objects from a Dynamic Textured Background via a Robust Kalman Filter – Zhong & Sclaroff, ICCV 2003

- Time-varying backgrounds: waves on water, clouds moving, trees moving, etc
- Detection & segmentation of foreground objects in video, when the background is a dynamic texture.
- Eg- ships on the sea, people riding an escalator, etc.
- Assumption: foreground objects are distinctive in spatial or temporal statistics.
- Dynamic texture modeled by: Autoregressive Moving Average Model (ARMA).

\[ X_{t+1} = AX_t + v_t, \quad X_0 = x_0; v_t \sim N(0, Q) \]
\[ Y_t = CX_t + w_t, \quad w_t \sim N(0, R) \]
Kalman Filter – Foreground Segmentation

- Foreground objects can be considered as outliers of the dynamic texture model
- A robust Kalman filter algorithm proposed - iteratively estimates the intrinsic appearance of the dynamic texture, as well as the foreground objects.

Figure 4: The third example of foreground objects detection in real data. (a) Sequence of waving river. (b) Waving river with floating bobo. (c) Detected foreground object region.
Goal: Using known camera motion to estimate depth from image sequences
Applications: robot navigation and manipulation.
Requirement: depth estimation algorithm that operates in an on-line, incremental fashion.

This requires a representation that records uncertainty in depth estimates and a mechanism that integrates new measurements with existing depth estimates to reduce the uncertainty over time.

Kalman filter provides this mechanism.

Kalman Filter-based Algorithms for Estimating Depth from Image Sequences, Matthies and kanade, IJCV 1989
Kalman Filter – Depth from Image Sequence
Kalman Filter – Tracking

- Face Tracking: video communication and human computer interaction.
- Video phone: reduce communication bandwidth by locating and transmitting only the fraction of a video frame which contains the speaker’s face.
- HCI: face tracking to direct attention to a user to correctly recognize the user’s facial expressions, gestures, or speech