An Introduction to Kalman Filter
Overview

• A conceptual view (scalar problem)
• Kalman Filter Formula
• Limitation of traditional Kalman Filter and imperfect solution
• Homework (if you are interested)
Conceptual View (One-dimension)

Model
Estimate my speed, I estimate I am at position $\chi^f = 500\text{km}$ from my origin

Somewhat uncertain,
Expressed with standard deviation
$\sigma_1 = 50\text{km}$

Where am I? $\chi^a$

Observation
I also have a very old GPS, it tell me my position at $y=600\text{km}$, with a standard deviation
$\sigma_2 = 20\text{km}$

Combine the both information to get a best *estimation* of my location?

Somewhere between, but more close to GPS location, why?
Mathematical Formulation

\[ x^a = ky + (1 - k)x^f \]

\( x^a \) is our best estimation with uncertainty

\( k \) is the unknown coefficients

We want to minimize the uncertainty, i.e. standard deviation \( \sigma \)
\[
\sigma^2 = E[(x^a - x^t)^2] = E[(ky + (1 - k)x^f - x^t)^2]
\]
\[
= E[(k(y - x^t) + (1 - k)(x^f - x^t))^2]
\]
\[
= k^2 E[(y - x^t)^2] + (1 - k)^2 E[(x^f - x^t)^2]
\]
\[
= k^2 \sigma_2^2 + (1 - k)^2 \sigma_1^2 = (\sigma_1^2 + \sigma_2^2)k^2 - 2\sigma_1^2 k + \sigma_1^2
\]
\[
\frac{d\sigma^2}{dk} = 0 \quad \text{Give you the minimal } \sigma
\]
\[
\frac{d\sigma^2}{dk} = 2(\sigma_1^2 + \sigma_2^2)k - 2\sigma_1^2 = 0
\]
\[
=> k = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}
\]

Here we assume the
\[
\text{cov}(\sigma_1, \sigma_2) = 0
\]
Mathematical Formulation

\[ \sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \]

i.e.

\[ \frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \]

Uncertainty was reduced!

\[ x^a = (1 - k)x^f + ky = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x^f + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} y \]

\[ x^a = x^f + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} (y - x^f) \]

Observation

analysis model  Gain  innovation
Statistical View

Uncertainty

Your prediction!
You prediction!

GPS measurement!
*Best Estimation when combining your prediction and measurement*

- Corrected mean is the new optimal estimate of position
- New uncertainty is smaller than either of the previous two variances
Flow chart of the process---doing previously repeatedly

\[ x^a(i) = x^f(i) + K[y(i) - x^f(i)] \]

\[ x^a(i + \Delta T) = x^f(i + \Delta T) + K[y(i + \Delta T) - x^f(i + \Delta T)] \]

- \( x^f \): forecast;
- \( x^a \): improved estimation;
- \( y \): measurement;
- \( K \): Gain;
- \( i, i + \Delta T \): time step
- \( \Delta T \): assimilation interval
- \( M_{i \rightarrow i + \Delta T} \): model integration

\[
k = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}
\]
Without and With assimilation
Long-term prediction without assimilation?

Larger and larger error and uncertainty

\[
\frac{1}{\sigma_x(t_2)} f_{x(t)|z(t_1),z(t_2)}(x|z_1,z_2)
\]

In general case: not scalar

Given the linear dynamical system:

\[ x_k = M_{k-1} x_{k-1} + B_{k-1} u_{k-1} + v_{k-1} \]
\[ y_k = H_k x_k + w_k \]

- \( x_k \) is the \( n \)-dimensional state vector (unknown)
- \( u_k \) is the \( m \)-dimensional input vector (known)
- \( y_k \) is the \( p \)-dimensional output vector (known, measured)
- \( M_k, B_k, H_k \) are appropriately dimensioned system matrices (known)
- \( v_k, w_k \) are zero-mean, white Gaussian noise with (known)
  - covariance matrices \( Q(k), R(k) \)

the Kalman Filter is a recursion that provides the "best" estimate of the state vector \( x \).
In general case: not a scalar

• Kalman Filter

Step 1. Model prediction: $x^f_k$ is estimate based on

\[ x^f_k = Mx^a_{k-1} + Bu_k + v_k \]

Noise ($v$) with covariance $Q$
this is your estimation of the error propagation.

\[ P^f_k = MP^a_{k-1}M^T + Q \]

Step 2. Calculate Kalman Gain:

\[ K = \frac{P^f_k H^T}{(HP^f_k H^T + R)} \]

step 3: correction of model state by KF analysis

\[ x^a_k = x^f_k + K(y_k - Hx^f_k) \]

this is your new estimation of the error covariance, reduced from $P^f_k$ to $P^a_k$
In general case: not a scalar

\( x^f \): forecast \([\text{N} \times 1]\); a state vector (a scalar before: your prediction of your location 1-D)

\( x^a \): (analysis) \([\text{N} \times 1]\); a state vector (a scalar before: 1-D location)

\( y \): observation \([\text{N}_0 \times 1]\); a observationa vector (a scalar before: GPS measurement location 1-D)

\( K \): Kalman gain \([\text{N} \times \text{N}_0]\); a matrix (a scalar before: \( \) )

\[
K = \frac{\sigma^2}{\sigma^2_1 + \sigma^2_2}
\]

\[
x^a = x^f + \frac{1}{\sigma^2_1 + \sigma^2_2} O^{-1} (y - x^f)
\]

\( H \) is the observation operator, interpolate the \( x^f \) to \( y \), this is because the observation size are usually smaller than your model state vector.
If we are sure about measurements:
- Measurement error covariance (R) decreases to zero
- K decreases and weights residual more heavily than prediction

If we are sure about prediction
- Prediction error covariance P^f decreases to zero
- K increases and weights prediction more heavily than residual

Scalar case
\[ x^a = x^f + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} (y - x^f) \]
\[ K = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \]

General case
\[ x^a = x^f + \frac{P^f H^T}{H P^f H^T + R} (y - H x^f) \]
\[ K = \frac{P^f H^T}{H P^f H^T + R} \]

If we are sure about measurements:
\[ \lim_{R_k \to 0} K_k = H^{-1}. \]

If we are sure about prediction
\[ \lim_{P_k \to 0} K_k = 0. \]
Flow chart of the process---doing previously repeatedly

Measurements $y$ \( t = i \)

Filter

- \[ x^a(i) = x^f(i) + K[y(i) - H x^f(i)] \]
- \[ x^a(i + \Delta T) = x^f(i + \Delta T) + K[y(i + \Delta T) - H x^f(i + \Delta T)] \]

Your estimation

- \[ x^f(i + \Delta T) = M_{i \rightarrow i + \Delta T}[x^a(i)] \]
- \[ x^f(i + 2\Delta T) = M_{i + \Delta T \rightarrow i + 2\Delta T}[x^a(i + \Delta T)] \]

\( x^f \): forecast [N×1] ; model prediction
\( X^a \): KF analysis condition [N×1] ;
\( y \): observation [N_o×1] ; from field measurements
\( K \): Kalman gain
\( \Delta T \): assimilation interval
\( M_{i \rightarrow i + \Delta T} \): model integration from time $i$ to $i + \Delta T$
Summary

• Recursive data processing algorithm
• Generates optimal estimate of desired quantities given the set of measurements
• Optimal?
  – For linear system and white Gaussian errors, Kalman filter is “best” estimate based on all previous measurements
  – For non-linear system optimality is ‘qualified’
• Recursive?
  – Doesn’t need to store all previous measurements and reprocess all data each time step
Limitation

• Weak nonlinear system (Extended Kalman Filter)

• Computation loads
  \[ K = \frac{P^f H^T}{HP^f H^T + R} \]
  – P=O(1e6) x O(1e6) matrix
  – Reduced Rank Kalman Filter (project to leading error subspace O(1e2) from EOF analysis and doing KF in model error subspace then project back)
  – Ensemble Kalman Filter (Represents error statistics \( P^f \) using an ensemble of model states.)

• (see Chen etc. 2009 for coastal ocean idealized case)
Homework (if you are interested)

• Consider you are in a room: your estimation is the temperature is constant (you can have your first guess with any temperature)

• You have a thermometer, with a known variance (uncertainty) $\sigma_1^2 = 1$

• We know the true room temperature is 10 °C with some perturbation, variance $\sigma_2^2 = 0.25$

• Using matlab/others to construct a KF model, show your model states analysis and error variance convergence in KF

• Assuming all error distribution is Gaussian
Analysis state is improved and error variance converged.
Overestimate the model error variance by a factor of 10
Underestimate the model error variance by a factor of 10
References


- Greg Welch and Gary Bishop: An Introduction to the Kalman Filter

- Buehner, M., and P. Malanotte-Rizzoli, Reduced-rank Kalman filters applied to an idealized model of the wind-driven ocean circulation, *JGR*

- Evensen 2003, Ocean Dynamics, Vol 53, No 4 The Ensemble Kalman Filter: Theoretical formulation and practical implementation