Kalman Filters

(Notes are From Principles of Adaptive Filters)

and Self-Learning Systems by A. Zaknich

The Kalman filter is used for estimating or predicting the next stage of a system based on a moving average of measurements driven by white noise, which is completely unpredictable. It needs a model of the relationship between inputs and outputs to provide feedback signals, but it can follow changing noise statistics quite well. ...

Kalman filters were first used in control applications in aerospace and they are still used primarily for control and tracking applications related to vessels, spacecraft, radar, and target trajectories. ...

The Kalman filter represents the most widely applied and useful result to emerge from the state variable approach of ‘modern control theory’.

Kalman filters are not completely unlike Wiener filters in some of the theory and math. They can be 'anticipated from a special case of the causal IIR Wiener filter for estimating a process x[n] from noisy measurements.'
There are essentially two parts to the Kalman Filter:

1. A model describing how the system behaves. That is, difference equations or state equations.
2. A measurement or observation that indicates what the present state of the system is.

Both of these usually will contain noise that is inherent. That is, the system’s response will not be perfect and will contain random fluctuations from a number of potential sources. Any measurement taken of the actual state of the system will also contain noise as part of the measurement process.

Mathematically, we can start with the model of the system being an AR(1) difference equation (autoregressive system with one history term)

\[ x[n] = a_1 x[n-1] + w[n] \]

The present state of the system depends on the previous state plus some noise in the system response.

We will also observe or measure the state of the system.
This is given as:
\[ y[n] = x[n] + v[n] \]

where \( v[n] \) and \( w[n] \) are zero mean white noise processes that are uncorrelated with each other. That is, the random fluctuations in how the system behaves has nothing to do with the randomness in our measurements of the system.

We can extend this idea to an AR[p] process:
\[ x[n] = \sum_{k=1}^{p} a_k x[n-k] + w[n] \]

which can be measured as
\[ y[n] = x[n] + v[n] \]

We can rewrite these in matrix form as
\[
\begin{bmatrix}
\hat{x}[n]
\end{bmatrix} = \begin{bmatrix}
\mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{A} & \cdots & \mathbf{0} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}
\end{bmatrix}
\begin{bmatrix}
\hat{x}[n-1] \\
\hat{x}[n-2] \\
\vdots \\
\hat{x}[n-p+1]
\end{bmatrix} + \begin{bmatrix}
\mathbf{0} \\
\mathbf{0} \\
\vdots \\
\mathbf{w}[n]
\end{bmatrix}
\]

\[ \hat{x}[n] = \mathbf{A} \hat{x}[n-1] + \mathbf{w}[n] \]

where
\[ \hat{x}[n] = \begin{bmatrix}
x[n] \\
x[n-1] \\
\vdots \\
x[n-p+1]
\end{bmatrix} \]

\[ \hat{x}[n] = \begin{bmatrix}
x[n] \\
x[n-1] \\
\vdots \\
x[n-p+1]
\end{bmatrix} \]

a column vector
\( \hat{A} \) is a \( p \times p \) state transition matrix
\[
\hat{\omega}[n] = [\hat{w}[n], 0, 0, \ldots, 0]^T \\
\text{a column vector zero-mean white noise process}
\]

Writing this out as a set of equations perhaps makes it more obvious what is occurring.
\[
x[n] = a_1 x[n-1] + a_2 x[n-2] + \ldots + a_{p-1} x[n-p+1] + a_p x[n-p] + \omega[n]
\]
\[
\begin{align*}
x[n-1] &= x[n-1] \\
x[n-2] &= x[n-2] \\
&\vdots \\
x[n-p] &= x[n-p]
\end{align*}
\]

Also
\[
y[n] = [1, 0, 0, \ldots] x[n] + \nu[n]
\]
\[
= c^T x[n] + \nu[n]
\]

where \( c \) is a unit vector of length \( p \).

If \( x[n] \) and \( y[n] \) are jointly wide-sense stationary processes then the “\( a \)” coefficients in the difference equation are constant and we can estimate the state of the system as
\[
\hat{x}[n] = \hat{A} \hat{x}[n-1] + K [y[n] - c^T \hat{A} \hat{x}[n-1]]
\]

Our previous guess \( \hat{x}[n] \) like the learning coefficient of the LMS algorithm

The current measurement - the last = “error” guess

\( K \) Current measurement - our last guess
If the process is nonstationary then we can adjust the state equation to:

\[ \vec{x}[n] = \tilde{A}(n-1) \vec{x}[n-1] + \vec{w}[n] \]

where \( \tilde{A}(n-1) \) is a time-varying \( p \times p \) state transition matrix.

Let's quickly review some notation and details of the noise.

In general our data could be complex so we need to use the Hermitian transpose of a complex vector:

\[ \vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} \Rightarrow \vec{w}^H = (\vec{w}^*)^T = (\vec{w}^T)^* = [w_1^* \ w_2^* \ w_3^* \ \cdots \ w_p^*] \]

The zero mean white noise process represented by vector \( \vec{w}[n] \) has an expectation value \( E \{ \vec{w}[n] \vec{w}^H[k] \} \) of

\[ E \{ \vec{w}[n] \vec{w}^H[k] \} = \begin{cases} Q_w(n) & : k = n \\ 0 & : k \neq n \end{cases} \]

Okay... so what the heck does that mean?
First let's figure out \( \mathbf{w}[n] \mathbf{w}^H[k] \)

\[
\mathbf{w}[n] \mathbf{w}^H[k] = \begin{bmatrix}
\mathbf{w}^* \\
\mathbf{w}_{n-1}^* \\
\mathbf{w}_{n-2}^* \\
\vdots \\
\mathbf{w}_{n-p+1}^*
\end{bmatrix}
\begin{bmatrix}
\mathbf{w}_k^* & \mathbf{w}_{k-1}^* & \mathbf{w}_{k-2}^* & \cdots & \mathbf{w}_{k-p+1}^*
\end{bmatrix}
\]

The values of noise at time step \( k, k-1, n, n-1, \text{etc.} \)

This is an outer product that produces the matrix

\[
\mathbf{w}[n] \mathbf{w}^H[k] = \begin{bmatrix}
\mathbf{w}_n \mathbf{w}_k^* & \mathbf{w}_n \mathbf{w}_{k-1}^* & \mathbf{w}_n \mathbf{w}_{k-2}^* & \cdots & \mathbf{w}_n \mathbf{w}_{k-p+1}^* \\
\mathbf{w}_{n-1} \mathbf{w}_k^* & \mathbf{w}_{n-1} \mathbf{w}_{k-1}^* & \mathbf{w}_{n-1} \mathbf{w}_{k-2}^* & \cdots & \mathbf{w}_{n-1} \mathbf{w}_{k-p+1}^* \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{w}_{n-p+1} \mathbf{w}_k^* & \cdots & \mathbf{w}_{n-p+1} \mathbf{w}_{k-p+1}^*
\end{bmatrix}
\]

Since \( \mathbf{w}_n \) and \( \mathbf{w}_k \) are random noise values they will be arbitrarily different when \( n \neq k \) and so this matrix will be a random jumble of positive and negative numbers.

The expectation value of those numbers will be zero (the typical value will be centered on zero).

Thus \( \mathbb{E}\{\mathbf{w}[n] \mathbf{w}^H[k]\} = 0 \) for \( n \neq k \)

Note what happens when \( n = k \) though
For \( n = k \):
\[
\begin{bmatrix}
  w_n w_n^* & w_n w_{n-1}^* & \cdots & \cdots \\
  w_{n-1} w_n^* & w_{n-1} w_{n-2}^* & \cdots & \cdots \\
  \vdots & \vdots & \ddots & \vdots \\
  w_{n-p+1} w_n^* & w_{n-p+1} w_{n-p+2}^* & \cdots & 1
\end{bmatrix}
\]

The off-diagonal terms where you have \( w_{n-i} w_{n-j}^* \) with \( i \neq j \) will again be a truly random mix of positive and negative values since the noise values at two different time steps are uncorrelated with each other.

However, the diagonal terms are guaranteed to be all positive so the expectation values of these matrix elements are

\[
E\{\hat{\hat{\omega}}[n] \hat{\omega}^+[n]\} = \begin{bmatrix}
  \|w_n\|^2 & 0 & \cdots & 0 \\
  0 & \|w_{n-1}\|^2 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & \|w_{n-p+1}\|^2
\end{bmatrix}
\]

Conclusion: \( U_m \) yes the noise is random.

We can now rewrite our measurement equation as:
\[
\hat{y}[n] = \hat{C}(n) \hat{x}[n] + \hat{\nu}[n]
\]
when \( \hat{C}(n) \) is a time-varying \( q \times p \) matrix and we have

\[
E\{\hat{\nu}[n] \hat{\nu}^+[k]\} = \begin{cases}
  \Phi_{\nu}(n) : k = n \\
  0 & : k \neq n
\end{cases}
\]
And thus we can update our estimate/guess equation from:
\[ \hat{\mathbf{x}}[n] = \mathbf{A} \hat{\mathbf{x}}[n-1] + \mathbf{K}[\mathbf{y}[n] - \mathbf{C}^T \hat{\mathbf{x}}[n-1]] \]

to:
\[ \hat{\mathbf{x}}[n] = \hat{\mathbf{A}}(n-1) \hat{\mathbf{x}}[n-1] + \mathbf{K}(n)[\hat{\mathbf{y}}[n] - \mathbf{C}(n)\hat{\mathbf{A}}(n-1)\hat{\mathbf{x}}[n-1]] \]

where \( \mathbf{K}(n) \) is the appropriate Kalman gain matrix.

Now we will use even more complicated notation and try not to strain your frontal lobe.

Let us define \( \hat{\mathbf{x}}[n | n] \) as the best linear estimate of \( \hat{\mathbf{x}}[n] \) at time \( n \) given all the observations \( \hat{\mathbf{y}}[i] \) for \( i = 1, 2, 3, \ldots, n \).

\( \hat{\mathbf{x}}[n | n-1] \) is then the best estimate given all the observations up to and including time \( n-1 \).

The corresponding state estimation errors are then:
\[ \hat{\mathbf{e}}[n | n] = \mathbf{x}[n] - \hat{\mathbf{x}}[n | n] \]
\[ \hat{\mathbf{e}}[n | n-1] = \mathbf{x}[n] - \hat{\mathbf{x}}[n | n-1] \]

Let's expand this notation to hopefully make it more clear as to what is really going on in these equations.
For instance, let’s assume an AR(3) process. Then at time step 4 we have:

\[
\hat{X}[4 | 4] = \hat{X}[4] - \hat{X}[4 | 4]
\]

\[
= \begin{bmatrix} x[4] \\ y[3] \\ x[2] \end{bmatrix} - \left\{ \begin{bmatrix} \hat{X}[4] \\ \hat{X}[3] \\ \hat{X}[2] \end{bmatrix} - \begin{bmatrix} A \hat{X}[n-1] + K_y[n] - C^T \hat{X}[n-1] \\ K_1 \\ K_2 \end{bmatrix} \right\}
\]

\[
= \begin{bmatrix} x[4] \\ y[3] \\ x[2] \end{bmatrix} - \left\{ \begin{bmatrix} a_1, a_2, a_3 \\ 1, 0, 0 \\ 0, 1, 0 \end{bmatrix} \begin{bmatrix} \hat{X}[3] \\ \hat{X}[2] \\ \hat{X}[1] \end{bmatrix} + \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} \begin{bmatrix} y[4] - 100 \end{bmatrix} \right\}
\]

\[
= \begin{bmatrix} x[4] \\ y[3] \\ x[2] \end{bmatrix} - \left\{ \begin{bmatrix} a_1 \hat{X}[3] + a_2 \hat{X}[2] + a_3 \hat{X}[1] \\ \hat{X}[3] \\ \hat{X}[2] \end{bmatrix} + \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} \begin{bmatrix} y[4] - 100 \end{bmatrix} \right\}
\]

\[
= \begin{bmatrix} x[4] \\ y[3] \\ x[2] \end{bmatrix} - \left\{ \begin{bmatrix} q_1, q_2, q_3 \\ 1, 0, 0 \\ 0, 1, 0 \end{bmatrix} \begin{bmatrix} \hat{X}[3] \\ \hat{X}[2] \\ \hat{X}[1] \end{bmatrix} + \begin{bmatrix} q_1 \hat{X}[3] + q_2 \hat{X}[2] + q_3 \hat{X}[1] \end{bmatrix} \right\}
\]

Let’s call this \( d_n \), the difference between the measured value and our estimate.

Then we have
\[
\hat{e}^T[4|4] = \begin{bmatrix} x[4] \\ x[3] \\ x[2] \end{bmatrix} - \begin{bmatrix} a_1 \hat{y}[3] + a_2 \hat{x}[z] + a_3 \hat{x}[1] + K_1 d_{me} \\ \hat{x}[3] + K_2 d_{me} \\ \hat{x}[2] + K_3 d_{me} \end{bmatrix}
\]

\[
= \begin{bmatrix} x[4] - (a_1 \hat{y}[3] + a_2 \hat{x}[z] + a_3 \hat{x}[1] + K_1 d_{me}) \\ x[3] - (\hat{x}[3] + K_2 d_{me}) \\ x[2] - (\hat{x}[2] + K_3 d_{me}) \end{bmatrix}
\]

The corresponding covariance errors are then defined as

\[
\hat{P}(n|n) = E\left\{\hat{e}[n|n]\hat{e}^H[n|n]\right\}
\]

\[
\bar{P}(n|n-1) = E\left\{\hat{e}[n|n-1]\hat{e}^H[n|n-1]\right\}
\]

Short review here: Covariance = Correlation when the mean is zero

\[
C_x(\tau) = E\left\{(x(t) - \bar{x})(x(t+\tau) - \bar{x})\right\}
\]

\[
C_x(\tau) = \Gamma_x(\tau) - \bar{x}^2
\]

\[
C_x(\tau) = \Gamma_x(\tau) \quad \text{for} \quad \bar{x} = 0
\]

So what is covariance error thing then?

Let's expand it out again to see what's really really going on in the matrix.
\[ P(n | n) = E \{ \hat{e}[n | n] \hat{e}^T[n | n] \} \]

\[ = E \left\{ \left[ \begin{array}{c} \hat{x}[4] - (a, \hat{x}[3] + a_2 \hat{x}[2] + a_3 \hat{x}[1] + k_f \phi) \\ \hat{x}[3] - (\hat{x}[3] + k_d d) \\ \hat{x}[2] - (\hat{x}[2] + k_3 d) \end{array} \right] \right\} \]

\[ = E \left\{ \left[ \begin{array}{c} \hat{x}[4] - (a, \hat{x}[3] + a_2 \hat{x}[2] + a_3 \hat{x}[1] + k_f \phi) \\ \hat{x}[3] - (\hat{x}[3] + k_d d) \\ \hat{x}[2] - (\hat{x}[2] + k_3 d) \end{array} \right]^2 \right\} \]

The expectation values of the cross-terms would be zero so this becomes:

\[ = \left[ \begin{array}{ccc} e_z^2 & 0 & 0 \\ 0 & e_z^2 & 0 \\ 0 & 0 & e_z^2 \end{array} \right] \]

So, given an estimate \( \hat{x}[0|0] \) of the state \( \hat{x}[0] \) and if \( \hat{P}(0|0) \) is known, then when measurement \( \hat{y}[1] \) becomes available the goal is to update \( \hat{x}[0|0] \) and find \( \hat{x}[1|1] \) that minimizes the mean square error given by:

\[ \xi(1) = E \{ \| \hat{e}[1|1] \|^2 \} = tr \{ \hat{P}(1|1) \} = \sum_{i=0}^{e-1} E \{ |e_k[1|1]|^2 \} \]
This estimation is repeated for the next observation $\hat{y}[2]$ and so on. For each $n > 0$, given $\hat{x}[n-1|n-1]$ and $P(n-1|n-1)$ when new observation $y[n]$ becomes available, the problem is to find the minimum mean square estimate $\hat{x}[n|n]$ of the state vector $\hat{x}[n]$. The problem is solved in two steps:

1. Given $\hat{x}[n-1|n-1]$ find $\hat{x}[n|n-1]$ which is the best estimate of $\hat{x}[n]$ without observation $y[n]$.

2. Given $\hat{x}[n|n-1]$ and $y[n]$ estimate $\hat{x}[n]$.  

Step 1. In this step all we know is the evolution of $\hat{x}[n]$ according to the state equation:

$$\hat{x}[n] = \hat{A}(n-1)\hat{x}[n-1] + \hat{W}[n]$$

(This is what the system is supposed to do)

Since $\hat{w}[n]$ is an unknown zero mean white noise process we can predict the state of the system as

$$\hat{x}[n|n-1] = \hat{A}(n-1)\hat{x}[n-1|n-1] + 0$$

Just use the state equation to estimate a new value $\hat{x}[n|n-1]$.
This has an unbiased (i.e., $E\{\hat{\varepsilon}[n\mid n-1]\} = 0$) estimation error $\tilde{e}[n\mid n-1]$ defined by

$$
\tilde{e}[n\mid n-1] = \hat{x}[n] - \hat{x}[n\mid n-1]
$$

Actual system state

Our guess using only the state equation

$$
= \hat{A}(n-1)\hat{x}[n-1] + \tilde{w}[n] - \hat{A}(n-1)\hat{x}[n\mid n-1](n-1)]
$$

Actual

Estimate

$$
= \hat{A}(n-1)\hat{x}[n-1] - \hat{A}(n-1)\hat{x}[n\mid n-1](n-1)] + \tilde{w}[n]
$$

$$
= \hat{A}(n-1)\{\hat{x}[n-1] - \hat{x}[n\mid n-1](n-1)]\} + \tilde{w}[n]
$$

$$
= \hat{A}(n-1)\tilde{e}[n\mid n-1] + \tilde{w}[n]
$$

Since $\tilde{w}[n]$ (the noise present at time $n$) is uncorrelated with estimation errors $\tilde{e}[n\mid n-1]$ and $\tilde{e}[n\mid n-1]$ we can define covariance error matrix $\hat{P}(n\mid n-1)$ as

$$
\hat{P}(n\mid n-1) = E\{\tilde{e}[n\mid n-1]\tilde{e}^H[n\mid n-1]\} \quad \text{and we have } \tilde{e}[n\mid n-1]
$$

from above so:

(I think this matrix math is correct, but don't quote me for sure. The end result is correct but don't consider this a rigorous math proof... yet.)

$$
E\{\tilde{e}[n\mid n-1]\tilde{e}^H[n\mid n-1]\} =
E\{[\hat{A}(n-1)\tilde{e}[n-1\mid n-1]+\tilde{w}[n]]\tilde{e}^H[n\mid n-1]\hat{A}(n-1) + \tilde{w}^H[n]\}
$$
This would give us (in terms of matrix dimensions)

\[
[ \text{exp}(\rho x^1) + (\rho x^1) ] \left[ \text{exp}(\rho x^1) + (1 x^1) \right] = \\
[ p x 1 ] [ 1 x p ] = p x p \text{ matrix for the expectation result.}
\]

So continuing on...

\[
E \{ \hat{A}(n-1) \hat{\varepsilon}[n-1] \hat{\varepsilon}^H[n-1] \hat{A}^H(n-1) + \hat{\mu}[n] \hat{\varepsilon}^H[n-1] \hat{A}(n-1) + \hat{A}(n-1) \hat{\varepsilon}[n-1] \hat{\varepsilon}^H[n] + \hat{\mu}[n] \hat{\mu}^H[n] \}
\]

Each of these terms is p x p

Recall though that \( \hat{\mu}[n] \) is uncorrelated with estimation error \( \hat{\varepsilon}[n-1] \) so terms 2 and 3 would have expectation values of \( \Phi \) and so we have:

\[
E \{ \hat{\mu}[n] \hat{\mu}^H[n] \} = \hat{Q} \mu[n]
\]

\( \hat{A} \) is a set of coefficients that do not contain randomness so they do not affect the expectation operator

\[
= \hat{A}(n-1) E \{ \hat{\varepsilon}[n-1] \hat{\varepsilon}^H[n-1] \hat{A}^H(n-1) + \hat{Q} \mu[n] \}
\]

\[
= \hat{A}(n-1) \hat{P}(n-1 | n-1) \hat{A}^H(n-1) + \hat{Q} \mu[n]
\]

Thus

\[
\hat{P}(n | n-1) = \hat{A}(n-1) \hat{P}(n-1 | n-1) \hat{A}^H(n-1) + \hat{Q} \mu[n]
\]
Step 2: The new measurement $\hat{y}[n]$ is incorporated into the estimate $\hat{x}[n|n-1]$. A linear estimate of $\hat{x}[n]$ can be defined by

$$\hat{x}[n|n] = \hat{x}'(n) \hat{x}[n|n-1] + \hat{K}(n) \hat{y}[n]$$

- Weighted guess without measurement (What the model tells us we should see)
- Weighted piece of the measurement (What we see)

We are "blending" or averaging these together to make our new state estimate. Our new state estimate $\hat{x}[n|n]$ must be an unbiased result (that's what we want anyway), so $E\{\hat{e}[n|n]\} = 0$ and it must minimize the mean square error (i.e. $E\{|\hat{e}[n|n]|^2\} = \xi_{\text{min}}(n)$)

OK, so what is the error?

$$\hat{e}[n|n] = \hat{x}[n] - \hat{x}[n|n]$$

$$= \hat{x}[n] - \hat{K}'(n) \hat{x}[n|n-1] - \hat{K}(n) \hat{y}[n]$$

$$= \hat{x}[n] - \hat{K}'(n) [\hat{x}(n) - \hat{e}[n|n-1]] - \hat{K}(n) [\hat{e}(n) \hat{x}[n] + \hat{y}[n]]$$

$$= \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$
Since $E\{\tilde{v}[n]\} = 0$ and $E\{\tilde{e}[n|n-1]\} = 0$ terms
2 and 3 above will contribute $\emptyset$ to the
expectation value of $\tilde{e}[n|n]$

Thus the only way $\hat{x}[n|n]$ can be unbiased
for any $\tilde{x}[n]$, (that is, $E[\tilde{e}[n|n]] = 0$) is if the
term
$$[I - \hat{K}'(n) - \hat{K}(n)C(n)] = 0$$
and thus
$$\hat{K}'(n) = I - \hat{K}(n)C(n)$$ — The amount we
weight what the
model tells us we should
see is not independent
of how much we
weight our measurement
(what we did see).

Substitute this above to eliminate $\hat{K}'$

$$\hat{x}[n|n] = [I - \hat{K}(n)C(n)]\hat{x}[n|n-1] + \hat{K}(n)\tilde{y}[n]$$

$$\hat{x}[n|n] = \hat{x}[n|n-1] + \hat{K}(n)[\tilde{y}[n] - \tilde{z}(n)\hat{x}[n|n-1]]$$

Estimate with = Estimate without + Weighted error between measured
measurement measurement value and our estimate
(what the system
model tells us)

* Why not just use $\tilde{y}[n]$ as our estimate of the
system’s position?

Because $\tilde{y}[n]$ is a noisy measurement.
The system isn’t actually at $\tilde{y}[n]$
* Why not (only) use system state equations to predict the state. That is, why not just use $\hat{x}[n|n-1]$?

Because the system doesn't respond perfectly according to the state equations, there is noise in the system response,

* We are blending the two noisy things together to get a better estimate than we could obtain by using either one individually.

The error for our estimate guess $\hat{x}[n|n]$ is:

$$\hat{e}[n|n] = x[n] - \hat{x}[n|n]$$

$$= x[n] - \hat{K}(n)^* \hat{x}[n|n-1] - \hat{K}(n) y[n]$$

$$= x[n] - \hat{K}(n)^* \hat{x}[n|n-1] - \hat{K}(n)^* [\hat{C}(n) \hat{x}[n] + \hat{v}[n]]$$

$$= x[n] - \hat{K}(n)^* \hat{x}[n|n-1] - \hat{K}(n)^* \hat{C}(n) \hat{x}[n] - \hat{K}(n)^* \hat{v}[n]$$

$$= x[n] - \left[\{ \hat{I} - \hat{K}(n)^* \hat{C}(n) \} \hat{x}[n|n-1] - \hat{K}(n)^* \hat{C}(n) \hat{x}[n] - \hat{K}(n)^* \hat{v}[n]\right]$$

$$= \left[\{ \hat{I} - \hat{K}(n)^* \hat{C}(n) \} \hat{x}[n] + \left[ \hat{I} - \hat{K}(n)^* \hat{C}(n) \right] \hat{x}[n|n-1] - \hat{K}(n)^* \hat{v}[n]\right]$$

$$= \left[\{ \hat{I} - \hat{K}(n)^* \hat{C}(n) \} \hat{x}[n] - \hat{K}(n)^* \hat{x}[n|n-1]\right] - \hat{K}(n)^* \hat{v}[n]$$

* $\hat{e}[n|n] = \left[\{ \hat{I} - \hat{K}(n)^* \hat{C}(n) \} \hat{x}[n|n-1] - \hat{K}(n)^* \hat{v}[n]\right]$*

Similar to before, because $\hat{v}[n]$ is a white noise process (and uncorrelated with $\hat{e}[n|n]$) it is therefore uncorrelated with $x[n]$ and $\hat{x}[n|n-1]$ (the actual state of the system and our estimate of the system that does not use the most recent measurement).
It is therefore also uncorrelated with $\varepsilon^{[n|n-1]}$ (that is $\text{E}\{\varepsilon^{[n|n-1]}\varepsilon^{[n|n]}\} = 0$) so we can write the error covariance matrix for $\varepsilon^{[n|n]}$ as

$$
\hat{\Sigma}^{[n|n]} = \text{E}\{\varepsilon^{[n|n]}\hat{\Sigma}^{[n|n]}\} = \text{E}\{(\hat{1} - \hat{K}(n)\hat{C}(n))\varepsilon^{[n|n-1]}(\hat{1} - \hat{K}(n)\hat{C}(n))^H - \hat{V}(n)\hat{K}(n)\}
$$

$$
= \text{E}\{\hat{1} - \hat{K}(n)\hat{C}(n)\varepsilon^{[n|n-1]}(\hat{1} - \hat{K}(n)\hat{C}(n))^H - \hat{K}(n)\hat{V}(n)\hat{K}(n)^H\}
$$

$$
= \text{E}\{\hat{1} - \hat{K}(n)\hat{C}(n)\varepsilon^{[n|n-1]}(\hat{1} - \hat{K}(n)\hat{C}(n))^H + \hat{K}(n)\hat{V}(n)\hat{K}(n)^H\}
$$

$$
= \text{E}\{\hat{1} - \hat{K}(n)\hat{C}(n)\varepsilon^{[n|n-1]}(\hat{1} - \hat{K}(n)\hat{C}(n))^H\} + \text{E}\{\hat{K}(n)\hat{V}(n)\hat{K}(n)^H\}
$$

$$
= [\hat{1} - \hat{K}(n)\hat{C}(n)]\text{E}\{\varepsilon^{[n|n-1]}(\hat{1} - \hat{K}(n)\hat{C}(n))^H\} + \hat{K}(n)\text{E}\{\hat{V}(n)\hat{V}(n)^H\} \hat{K}(n)^H
$$

$$
\hat{\Sigma}^{[n|n]} = [\hat{1} - \hat{K}(n)\hat{C}(n)]\hat{\Sigma}^{[n|n-1]}(\hat{1} - \hat{K}(n)\hat{C}(n))^H + \hat{K}(n)\hat{Q}(n)\hat{K}(n)^H
$$

The Kalman gain $\hat{K}(n)$ which then minimizes the mean square error $\ell(n) = \text{tr}\{\hat{\Sigma}(n|n)\}$ can be found by differentiating with respect to $\hat{K}(n)$ and then setting the result equal to zero (find the bottom of the bowl).

Recall (or note) that

$$
\frac{d}{dk} \text{tr}\{\hat{K}\hat{A}\} = \hat{A}^H \quad \text{and} \quad \frac{d}{dk} \text{tr}\{\hat{K}\hat{A}^H\hat{K}\} = 2\hat{K}\hat{A}^T
$$

Thus,

$$
\frac{d}{dk} \text{tr}\{\hat{\Sigma}(n|n)\} = \frac{d}{dk} \text{tr}\{[\hat{1} - \hat{K}(n)\hat{C}(n)]\hat{\Sigma}^{[n|n-1]}(\hat{1} - \hat{K}(n)\hat{C}(n))^H + \hat{K}(n)\hat{Q}(n)\hat{K}(n)^H\} = 0
$$
\[
\begin{align*}
&= \mathcal{Z} \left[ \mathbf{I} - \mathbf{K}(n) \mathbf{C}(n) \right] \mathbf{P}(n|n-1) \mathbf{C}^H(n) + \mathbf{Z} \mathbf{K}(n) \mathbf{Q}_v(n) \\
&\quad \uparrow \quad \text{"dynamics" of} \\
&= -\mathcal{Z} \left[ \mathbf{I} - \mathbf{K}(n) \mathbf{C}(n) \right] \mathbf{P}(n|n-1) \mathbf{C}^H(n) + \mathbf{Z} \mathbf{K}(n) \mathbf{Q}_v(n) \\
&\quad \text{Setting this equal to zero we get} \\
&\left[ \mathbf{I} - \mathbf{K}(n) \mathbf{C}(n) \right] \mathbf{P}(n|n-1) \mathbf{C}^H(n) - \mathbf{K}(n) \mathbf{Q}_v(n) = 0 \\
&\mathbf{P}(n|n-1) \mathbf{C}^H(n) - \mathbf{K}(n) \mathbf{C}(n) \mathbf{P}(n|n-1) \mathbf{C}^H(n) - \mathbf{K}(n) \mathbf{Q}_v(n) = 0 \\
&\mathbf{K}(n) \mathbf{C}(n) \mathbf{P}(n|n-1) \mathbf{C}^H(n) + \mathbf{K}(n) \mathbf{Q}_v(n) = \mathbf{P}(n|n-1) \mathbf{C}^H(n) \\
&\mathbf{K}(n) \left[ \mathbf{C}(n) \mathbf{P}(n|n-1) \mathbf{C}^H(n) + \mathbf{Q}_v(n) \right] = \mathbf{P}(n|n-1) \mathbf{C}^H(n) \\
&\mathbf{K}(n) = \mathbf{P}(n|n-1) \mathbf{C}^H(n) \left[ \mathbf{C}(n) \mathbf{P}(n|n-1) \mathbf{C}^H(n) + \mathbf{Q}_v(n) \right]^{-1} \\
&\text{This is how to find the gains} \\
&\text{Now reorganize the} \quad \mathbf{P}(n|n) \quad \text{equation} \\
&\mathbf{P}(n|n) = \left[ \mathbf{I} - \mathbf{K}(n) \mathbf{C}(n) \right] \mathbf{P}(n|n-1) \left[ \mathbf{I} - \mathbf{K}(n) \mathbf{C}(n) \right]^H + \mathbf{K}(n) \mathbf{Q}_v(n) \mathbf{K}^H(n) \\
&\quad \text{Multiply this part out with first part} \\
&= \left[ \mathbf{I} - \mathbf{K}(n) \mathbf{C}(n) \right] \mathbf{P}(n|n-1) \mathbf{C}^H(n) - \left[ \mathbf{I} - \mathbf{K}(n) \mathbf{C}(n) \right] \mathbf{P}(n|n-1) \mathbf{C}^H(n) \mathbf{K}(n) + \mathbf{K}(n) \mathbf{Q}_v(n) \mathbf{K}^H(n) \\
&= \left[ \mathbf{I} - \mathbf{K}(n) \mathbf{C}(n) \right] \mathbf{P}(n|n-1) - \left\{ \left[ \mathbf{I} - \mathbf{K}(n) \mathbf{C}(n) \right] \mathbf{P}(n|n-1) \mathbf{C}^H(n) - \mathbf{K}(n) \mathbf{Q}_v(n) \right\} \mathbf{K}(n) \\
&\quad \text{But this is equal to zero} \\
\end{align*}
\]
\[ P(n|n) = (I - K(n) \theta(n)) P(n|n-1) \]

We now have a recursive set of equations to compute the Kalman Filter behavior. It must be initialized for \( n = 0 \).

Choose an initial guess as \( \hat{x}[0|0] = E\{ \hat{x}[0] \} \) and initial covariance matrix as \( P(0|0) = E\{ \hat{x}[0] \hat{x}^T[0] \} \).

This will ensure "unbiased" results.

The Kalman gain \( K(n) \) and error covariance matrix \( P(n|n) \) can be computed offline prior to filtering since they are not dependent on \( \hat{x}[n] \).

| Summary of Kalman Filter estimator: |

Given

State vector: \( \hat{x}[n] \)
Observation vector: \( \hat{y}[n] \)
Covariance matrix of process noise: \( \tilde{Q}[n] = \tilde{Q}[n] \)
Covariance matrix of measurement noise: \( \tilde{R}[n] = \tilde{R}[n] \)
State Transition Matrix from \( n-1 \) to \( n \): \( \tilde{A}(n-1,n) \)
Measurement Matrix:
State Equation: \[ \hat{x}[n] = \tilde{A}(n-1,n) \hat{x}[n-1] + \tilde{w}[n] \]
Observation Equation: \[ \hat{y}[n] = \tilde{C}(n) \hat{x}[n] + \tilde{v}[n] \]
Initialization

Initial estimate: \[ \hat{X}[0|0] = E\{x[0]\} \]

Error covariance matrix

for estimate \( \hat{X}[0|0] \): \[ \hat{P}(0|0) = E\{\hat{x}[0]\hat{x}^H[0]\} \]

Computation

For \( n \geq 1, 2, \ldots \) compute all the following equations in sequence:

\[ \hat{X}[n|n-1] = A(n-1,n) \hat{X}[n-1|n-1] \]

\[ \hat{P}(n|n-1) = A(n-1,n) \hat{P}(n-1|n-1) A^H(n-1,n) + Q(n) \]

Filter Gain \( K(n) = \hat{P}(n|n-1) C^H(n) \left[ C(n)\hat{P}(n|n-1) C^H(n) + Q(n) \right]^{-1} \)

Estimator \( \hat{X}[n|n] = \hat{X}[n|n-1] + K(n) \left[ y(n) - C(n)\hat{X}[n|n-1] \right] \)

Error Covariance Matrix \( \hat{P}(n|n) = [I - K(n)C(n)] \hat{P}(n|n-1) \)

These can be rearranged, or rederived to instead predict the state of the system at \( n+1 \)

State Equation: \( \hat{X}[n] = A(n-1,n) \hat{X}[n-1] + \tilde{w}[n] \)

Observation Equation: \( y[n] = C(n)\hat{X}[n] + \tilde{v}[n] \)

Initialization

Initial Estimate: \( \hat{X}[0|0] = E\{x[0]\} \)

Error covariance matrix

for estimate \( \hat{X}[0|0] \): \[ \hat{P}(0|0) = E\{\hat{x}[0]\hat{x}^H[0]\} \]
Computation

For \( n=1,2,\ldots \), compute all the following equations in sequence.

\[
\hat{x}[n|n-1] = \hat{A}(n-1,n) \hat{x}[n-1|n-1]
\]

\[
\bar{P}(n|n-1) = \hat{A}(n-1,n) \bar{P}(n-1|n-1) \hat{A}^H(n-1,n) + \bar{Q}_n(n)
\]

Predictor Gain:

\[
\hat{G}(n) = \hat{A}(n-1,n) \bar{P}(n|n-1) \hat{C}^H(n)[\hat{C}(n) \bar{P}(n|n-1) \hat{C}^H(n) + \bar{Q}_n(n)]^{-1}
\]

Predictor: (best prediction at \( n+1 \) given all observations up to \( n \))

\[
\hat{x}[n+1|n] = \hat{A}(n-1,n) \hat{x}[n|n-1] + \hat{G}(n)[y[n] - \hat{C}(n) \hat{x}[n|n-1]]
\]

Error Covariance Matrix:

\[
\bar{P}[n+1|n] = [\hat{A}(n-1,n) - \hat{G}(n) \hat{C}(n)] \bar{P}(n|n-1) \hat{A}^H(n-1,n) + \bar{Q}_n(n)
\]

Example 1: Use the Kalman filter to estimate the following AR(1) process

\[
X[n] = 0.8 X[n-1] + w[n]
\]

\[
y[n] = X[n] + v[n]
\]

where \( w[n] \) and \( v[n] \) are uncorrelated white noise processes with respective variances of \( \sigma_w^2 = 0.36 \) and \( \sigma_v^2 = 1 \).

Here, \( \hat{A}(n) = 0.8 \) and \( \hat{C}(n) = 1 \) and the Kalman
The state equation is:
\[ \hat{x}[n] = 0.8 \hat{x}[n-1] + K(n) [y(n) - 0.8 \hat{x}[n]] \]

These are scalar equations this time
\[ P(n \mid n-1) = A(n-1) P(n-1 \mid n-1) A^H(n-1) + \Omega_w(n) \text{ becomes} \]
\[ P(n \mid n-1) = (0.8) P(n-1 \mid n-1) (0.8) + 0.36 \]
* \[ P(n \mid n-1) = (0.8)^2 P(n-1 \mid n-1) + 0.36 \]
\[ K(n) = P(n \mid n-1) C^H(n) [C(n) P(n \mid n-1) C^H(n) + \Omega_v(n)]^{-1} \text{ becomes} \]
* \[ K(n) = \frac{P(n \mid n-1)}{P(n \mid n-1) + 1} \]
\[ P(n \mid n) = [I - K(n) C(n)] P(n \mid n-1) \text{ becomes} \]
* \[ P(n \mid n) = [1 - K(n)] P(n \mid n-1) \]

With \( \hat{x}[0] = E\{x[0]\} = 0 \) and \( P(0 \mid 0) = E\{|x[0]|^2\} = 1 \)

The first few values of Kalman gain and error covariances become
| n   | P(n|n-1) | K(n) | P(n/n)  |
|-----|---------|------|---------|
| 1   | 1.0     | 0.5  | 0.5     |
| 2   | 0.68    | 0.4048 | 0.4048 |
| 3   | 0.6190 | 0.3824 | 0.3824 |
| 4   | 0.6047 | 0.3768 | 0.3768 |
| 5   | 0.6012 | 0.3755 | 0.3755 |
| 6   | 0.6003 | 0.3751 | 0.3751 |

Thus, the Kalman filter reaches a steady state of

\[
\hat{x}[n] = 0.8 \hat{x}[n-1] + 0.375(y[n] - 0.8 \hat{x}[n-1])
\]

**Example 2: Radar Tracking of a Plane**

State variables are

- \(x_1[n]\) = \(p(n)\), aircraft radial range
- \(x_2[n]\) = \(\dot{p}(n)\), aircraft radial velocity
- \(x_3[n]\) = \(\theta(n)\), aircraft bearing
- \(x_4[n]\) = \(\dot{\theta}(n)\), aircraft bearing rate or angular velocity

The state equation is

\[
\hat{X}[n] = \hat{A}(n-1,n) \hat{X}[n-1] + \hat{W}[n-1]
\]

* \(\hat{A}, \hat{W}\)
Expanding this out we see

\[
\begin{bmatrix}
X_1[n] \\
X_2[n] \\
X_3[n] \\
X_4[n]
\end{bmatrix} =
\begin{bmatrix}
1 & \Delta T & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & \Delta T \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
X_1[n-1] \\
X_2[n-2] \\
X_3[n-3] \\
X_4[n-4]
\end{bmatrix} +
\begin{bmatrix}
0 \\
U_1[n-1] \\
0 \\
U_2[n-1]
\end{bmatrix}
\]

\(\Delta T\) is the time interval between steps/measurements.

The noise terms \(U_1[n-1]\) and \(U_2[n-1]\) represent the change in radial velocity and bearing rate interval \(\Delta T\). They are \(\Delta T\) times the radial and angular acceleration, are random with zero means, and uncorrelated with each other from one time step to another.

\[
x_1[n] = x_1[n-1] + \Delta T x_2[n-1] + 0 \quad \Rightarrow \quad \dot{x}_1[n] = \dot{x}_1[n-1] + \Delta T \ddot{x}_2[n-1] + 0
\]
\[
x_2[n] = x_2[n-1] + u_1[n-1] \quad \Rightarrow \quad \dot{x}_2[n] = \dot{x}_2[n-1] + u_1[n-1]
\]
\[
x_3[n] = x_3[n-1] + \Delta T x_4[n-1] + 0 \quad \Rightarrow \quad \dot{x}_3[n] = \dot{x}_3[n-1] + \Delta T \ddot{x}_4[n-1] + 0
\]
\[
x_4[n] = x_4[n-1] + u_2[n-1] \quad \Rightarrow \quad \dot{x}_4[n] = \dot{x}_4[n-1] + u_2[n-1]
\]

Process noise is incorporated into the acceleration terms.

The radar sensors provide noisy estimates of the range and bearing \(x_1[n] = \hat{r}(n), x_3[n] = \hat{\theta}(n)\). Measurements are:

\[
y_1[n] = x_1[n] + v_1[n] \quad \Rightarrow \quad \hat{y}_1[n] = \hat{C}(n) \hat{x}[n] + \hat{v}[n]
\]
\[
y_2[n] = x_3[n] + v_2[n] \quad \Rightarrow \quad \hat{y}_2[n] = \hat{C}(n) \hat{x}[n] + \hat{v}[n]
\]

Expand:

\[
\begin{bmatrix}
y_1[n] \\
y_2[n]
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
x_1[n] \\
x_2[n] \\
x_3[n] \\
x_4[n]
\end{bmatrix} +
\begin{bmatrix}
v_1[n] \\
v_2[n]
\end{bmatrix}
\]

The radar sensors provide noisy estimates of the range and bearing \(x_1[n] = \hat{r}(n), x_3[n] = \hat{\theta}(n)\). Measurements are:

\[
y_1[n] = x_1[n] + v_1[n] \quad \Rightarrow \quad \hat{y}_1[n] = \hat{C}(n) \hat{x}[n] + \hat{v}[n]
\]
\[
y_2[n] = x_3[n] + v_2[n] \quad \Rightarrow \quad \hat{y}_2[n] = \hat{C}(n) \hat{x}[n] + \hat{v}[n]
\]

Expand:

\[
\begin{bmatrix}
y_1[n] \\
y_2[n]
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
x_1[n] \\
x_2[n] \\
x_3[n] \\
x_4[n]
\end{bmatrix} +
\begin{bmatrix}
v_1[n] \\
v_2[n]
\end{bmatrix}
\]

The radar sensors provide noisy estimates of the range and bearing \(x_1[n] = \hat{r}(n), x_3[n] = \hat{\theta}(n)\). Measurements are:

\[
y_1[n] = x_1[n] + v_1[n] \quad \Rightarrow \quad \hat{y}_1[n] = \hat{C}(n) \hat{x}[n] + \hat{v}[n]
\]
\[
y_2[n] = x_3[n] + v_2[n] \quad \Rightarrow \quad \hat{y}_2[n] = \hat{C}(n) \hat{x}[n] + \hat{v}[n]
\]

Expand:

\[
\begin{bmatrix}
y_1[n] \\
y_2[n]
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
x_1[n] \\
x_2[n] \\
x_3[n] \\
x_4[n]
\end{bmatrix} +
\begin{bmatrix}
v_1[n] \\
v_2[n]
\end{bmatrix}
\]
\[ \hat{\nu}[n] \] noise components are Gaussian with zero means and respective components variances \( \sigma_f^2(n) \) and \( \sigma_\theta^2(n) \).

The covariance matrices \( \bar{R}(n) = \bar{Q}_v(n) \) for the system and \( \bar{Q}(n) = \bar{Q}_w(n) \) for the measurement are

\[ \bar{Q}_v(n) = R(n) = E\{ \hat{\nu}[n] \hat{\nu}^H[n] \} = \begin{bmatrix} \sigma_f^2 & 0 \\ 0 & \sigma_\theta^2 \end{bmatrix} \]

\[ \hat{\nu}[n] \hat{\nu}^H[n] = \begin{bmatrix} v_1[n] \\ v_2[n] \end{bmatrix} \begin{bmatrix} v_1[n] & v_2[n] \end{bmatrix} = \begin{bmatrix} v_1^2[n] & v_1[n]v_2[n] \\ v_2[n]v_1[n] & v_2^2[n] \end{bmatrix} \]

\[ E\{ \hat{\nu}[n] \hat{\nu}^H[n] \} = E\left\{ \begin{bmatrix} v_1^2[n] & v_1[n]v_2[n] \\ v_2[n]v_1[n] & v_2^2[n] \end{bmatrix} \right\} = \begin{bmatrix} E\{v_1^2[n]\} & 0 \\ 0 & E\{v_2^2[n]\} \end{bmatrix} \]

\[ = \begin{bmatrix} \sigma_f^2 & 0 \\ 0 & \sigma_\theta^2 \end{bmatrix} \]

\[ \bar{Q}_w(n) = Q(n) = E\{ \hat{\omega}[n] \hat{\omega}^H[n] \} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_1^{-2}(n) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_2^{-2}(n) \end{bmatrix} \]

where \( \sigma_1^{-2}(n) = E\{U_1^2[n]\} \) and \( \sigma_2^{-2}(n) = E\{U_2^2[n]\} \)

To start the Kalman processing the gain matrix \( \bar{K}(n) \) is initialized by specifying the error covariance matrix \( \bar{P}(n|n) \) in some way the reference author [Digital and Kalman Filtering by S. M. Bozic 1994] suggests using two measurements of range and bearing at times \( n=1 \) and \( n=2 \).
\[
\hat{X}_{[2]} = \begin{bmatrix}
\hat{x}_1[2] = \hat{p}(2) = y_1[2] \\
\hat{x}_2[2] = \hat{v}(2) = (1/\Delta T) [y_1[2] - y_1[1]] \\
\hat{x}_3[2] = \hat{v}(2) = y_2[2] \\
\hat{x}_4[2] = \hat{v}(2) = (1/\Delta T) [y_2[2] - y_2[1]] \\
\end{bmatrix}
\]

By definition, \( \bar{P}(2 | z) = E \left\{ \left[ \bar{x}[z] - \hat{x}[z] \right] \left[ \bar{x}[z] - \hat{x}[z] \right]^T \right\} \)

Thus we have:

\[
\bar{x}[z] - \hat{x}[z] = \begin{bmatrix}
x_1[2] - \hat{x}_1[2] \\
x_2[2] - \hat{x}_2[2] \\
x_3[2] - \hat{x}_3[2] \\
x_4[2] - \hat{x}_4[2] \\
\end{bmatrix} = \begin{bmatrix}
\end{bmatrix}
\]

\[
\bar{x}[z] - \hat{x}[z] = \begin{bmatrix}
x_1[2] - (x_1[2] + v_1[2]) \\
\left( x_2[2] - (x_1[2] + v_1[2]) - (x_1[1] - v_1[1]) \right)/\Delta T \\
\left( x_4[2] - (x_3[2] + v_3[2]) - (x_3[1] - v_3[1]) \right)/\Delta T \\
\end{bmatrix}
\]

Look at the term \( x_2[2] - (x_1[2] + v_1[2]) - (x_1[1] - v_1[1]) \)

Rewrite \( x_2[2] = x_1[1] + u_1[1] \). Then this term becomes

\[
x_2[1] + u_1[1] - \left( \frac{(x_1[2] + v_1[2]) - (x_1[1] + v_1[1])}{\Delta T} \right)
\]
Regroup in parentheses

\[ X_2[1] + u_1[1] - \left( \frac{X_1[2] - X_1[1]}{\Delta T} + \frac{u_1[2] - u_1[1]}{\Delta T} \right) \]

This is

\[ X_2[1] \]

\[ X_1[2] = X_1[1] + \Delta T \cdot X_2[1] \]

\[ \frac{X_1[2] - X_1[1]}{\Delta T} = X_2[1] \quad \text{So it will cancel out.} \]

You're left with

\[ u_1[1] - \left( \frac{u_1[2] - u_1[1]}{\Delta T} \right) \]

Thus

\[
\begin{bmatrix}
X_1[2] - X_1[1]
\end{bmatrix} =
\begin{bmatrix}
- u_1[2] \\
U_1[1] - \frac{u_1[2] - u_1[1]}{\Delta T} \\
- v_2[2] \\
U_2[1] - \frac{v_1[2] - v_2[1]}{\Delta T}
\end{bmatrix}
\]

Since \( u \) and \( v \) are independent we get:

\[
\bar{p}(z|z) = E \left\{ [\bar{X}[2] - \hat{X}[2]] [\bar{X}[2] - \hat{X}[2]]^T \right\}
\]

\[
= \begin{bmatrix}
\sigma^2_g & \sigma^2_\phi & 0 & 0 \\
\sigma^2_\phi & \frac{2\sigma^2_\phi + \sigma^2_1}{T^2} & 0 & 0 \\
0 & \sigma^2_\phi & \frac{\sigma^2_\phi}{T} & 0 \\
0 & 0 & \sigma^2_\phi & \frac{\sigma^2_\phi}{T}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\sigma^2_g & \sigma^2_\phi & 0 & 0 \\
\sigma^2_\phi & \frac{2\sigma^2_\phi + \sigma^2_1}{T^2} & 0 & 0 \\
0 & \sigma^2_\phi & \frac{\sigma^2_\phi}{T} & 0 \\
0 & 0 & \sigma^2_\phi & \frac{\sigma^2_\phi}{T}
\end{bmatrix}
\]
Let $R = 160 \text{ km, } \Delta T = 15 \text{ seconds, maximum acceleration is } M = 2.1 \text{ m/s}^2$

Let $\sigma_g = 1000 \text{ m and } \sigma_\theta = 0.017 \text{ radians}$

The noise variances in $Q$ are $\sigma_1^2 = 330$

$$\sigma_2^2 = 1.3 \times 10^{-8}$$

$$\mathbf{P}(2 \mid 2) = \begin{bmatrix} 10^6 & 6.7 \times 10^4 & 0 & 0 \\ 6.7 \times 10^4 & 0.9 \times 10^4 & 0 & 0 \\ 0 & 0 & 2.9 \times 10^{-4} & 1.9 \times 10^{-5} \\ 0 & 0 & 1.9 \times 10^{-5} & 2.6 \times 10^{-6} \end{bmatrix}$$

The predictor gain is $\mathbf{G}(3)$

$$\mathbf{G}(3) = \mathbf{A}(2,3) \mathbf{P}(3 \mid 2) \mathbf{C}^H(3) \left[ \mathbf{C}(3) \mathbf{P}(3 \mid 2) \mathbf{C}^H(3) + \mathbf{R}(3) \right]^{-1}$$

where $\mathbf{P}(3 \mid 2)$ can be calculated as $\mathbf{P}(n \mid n-1)$

so $\mathbf{P}(3 \mid 2) = \left[ \mathbf{A}(1,2) - \mathbf{G}(2) \mathbf{C}(2) \right] \mathbf{P}(2 \mid 1) \mathbf{A}^H(1,2) + \mathbf{Q}(n)$

The problem at the beginning though is that we don't have or know $\mathbf{G}(2)$ or $\mathbf{P}(2 \mid 1)$.

However our Kalman predictor equations give us another way to find the necessary initial values.*

*Use the $\mathbf{P}(n+1 \mid n)$ equation $\Rightarrow \mathbf{P}(3 \mid 2)$.

Thus $\mathbf{P}(3 \mid 2)$ is $\mathbf{A}(2,3) \mathbf{P}(2 \mid 2) \mathbf{A}^H(2,3) + \mathbf{Q}(n)$.
\[ \mathbf{P}(3 \mid 2) = \begin{bmatrix} 1 & 15 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 15 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 10^6 & 6.7 \times 10^4 & 0 & 0 \\ 6.7 \times 10^4 & 0.9 \times 10^4 & 0 & 0 \\ 0 & 0 & 2.9 \times 10^5 & 1.9 \times 10^5 \\ 0 & 0 & 1.9 \times 10^{-5} & 2.6 \times 10^{-6} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 330 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.3 \times 10^{-8} \end{bmatrix} \]

\[ \mathbf{P}(3 \mid 2) = \begin{bmatrix} 5 \times 10^6 & 2 \times 10^5 & 0 & 0 \\ 2 \times 10^5 & 9.3 \times 10^3 & 0 & 0 \\ 0 & 0 & 14.5 \times 10^{-4} & 5.8 \times 10^{-5} \\ 0 & 0 & 5.8 \times 10^{-5} & 2.6 \times 10^{-6} \end{bmatrix} \]

Then \[ \mathbf{G}(3) = A(2,3) \mathbf{P}(3 \mid 2) \mathbf{C}^H(3) \left[ \mathbf{C}(3) \mathbf{P}(3 \mid 2) \mathbf{C}^H(3) + \mathbf{R}(3) \right]^{-1} \]

\[ = \begin{bmatrix} 1 & 15 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 15 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 \times 10^6 & 2 \times 10^5 & 0 & 0 \\ 2 \times 10^5 & 9.3 \times 10^3 & 0 & 0 \\ 0 & 0 & 14.5 \times 10^{-4} & 5.8 \times 10^{-5} \\ 0 & 0 & 5.8 \times 10^{-5} & 2.6 \times 10^{-6} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}^{-1} \]

\[ \mathbf{G}(3) = \begin{bmatrix} 1.33 & 0 \\ 3.3 \times 10^2 & 0 \\ 0 & 1.33 \\ 0 & 3.3 \times 10^2 \end{bmatrix} \]

\[ \mathbf{G}(3) \text{ can now be used to find } \mathbf{G}[n+1|n] \text{ and the process can now iterate from here.} \]