EE210A: Adaptation and Learning
Professor Ali H. Sayed
KALMAN FILTERING

Sections in order: 7.1-7.7
7.1 INNOVATIONS PROCESS

Consider two zero-mean random variables \( \{x, y\} \). We already know from Thm. 3.1 that the linear least-mean-squares estimator of \( x \) given \( y \) is \( \hat{x} = K_o y \), where \( K_o \) is any solution to the normal equations

\[
K_o R_y = R_{xy}
\]  

(7.1)

In the sequel we assume that \( R_y \) is positive-definite so that \( K_o \) is uniquely defined as \( K_o = R_{xy} R_y^{-1} \).

Usually, the variable \( y \) is vector-valued, say \( y = \text{col}\{y_0, y_1, \ldots, y_N\} \), where each \( y_i \) is also possibly a vector. Now assume that we could somehow replace \( y \) by another vector \( e \) of similar dimensions, say

\[
e = Ay
\]  

(7.2)

for some lower triangular invertible matrix \( A \).
Assume further that the transformation $A$ could be chosen such that the entries of $e$, denoted by $e = \text{col}\{e_0, e_1, \ldots, e_N\}$, are uncorrelated with each other, i.e.,

$$E e_i e_j^* \triangleq R_{e,i} \delta_{ij}$$

where $\delta_{ij}$ denotes the Kronecker delta function that is unity when $i = j$ and zero otherwise, and $R_{e,i}$ denotes the covariance matrix of $e_i$. Then the covariance matrix of $e$ will be block diagonal,

$$R_e \triangleq E ee^* = \text{diag}\{R_{e,0}, R_{e,1}, \ldots, R_{e,N}\}$$

and, in addition, the problem of estimating $x$ from $y$ would be equivalent to the problem of estimating $x$ from $e$. To see this, let $\hat{x}|_e$ denote the linear least-mean-squares estimator of $x$ given $e$, i.e.,

$$\hat{x}|_e = R_{xe} R_e^{-1} e$$  \hfill (7.3)
Likewise, let $\hat{x}|_y$ denote the estimator of $x$ given $y$,

$$\hat{x}|_y = R_{xy}R_y^{-1}y$$

(7.4)

Then since

$$R_e = \mathbb{E} ee^* = A (\mathbb{E} yy^*) A^* = AR_yA^*$$

and

$$R_{xe} = \mathbb{E} xe^* = (\mathbb{E} xy^*) A^* = R_{xy}A^*$$

we find that

$$\hat{x}|_e = R_{xe}R_e^{-1}e = R_{xy}A^* (AR_yA^*)^{-1} e = R_{xy}R_y^{-1}A^{-1}e = R_{xy}R_y^{-1}y$$

That is,

$$\hat{x}|_e = \hat{x}|_y$$

(7.5)
The key advantage of working with $e$ instead of $y$ is that $R_e$ in (7.3) is block-diagonal and, hence, the estimator $\hat{x}_{|e}$ can be evaluated as the combined sum of individual estimators. Specifically, expression (7.3) gives

$$\hat{x}_{|e} = \sum_{i=0}^{N} \left( \mathbb{E} x e_i^* \right) R_{e,i}^{-1} e_i = \sum_{i=0}^{N} \hat{x}_{|e_i}$$

This result shows that we can estimate $x$ from $y$ by estimating $x$ individually from each $e_i$ and then combining the resulting estimators. In particular, if we replace the notations $\hat{x}_{|e}$ and $\hat{x}_{|y}$ by the more suggestive notation $\hat{x}_{|N}$, in order to indicate that the estimator of $x$ is based on the observations $y_0$ through $y_N$, then the above expression shows that

$$\hat{x}_{|N} = \sum_{i=0}^{N} \hat{x}_{|e_i} = \hat{x}_{|e_N} + \sum_{i=0}^{N-1} \hat{x}_{|e_i}$$

where the last sum on the right-hand side is simply the estimator of $x$ using the observations $y_0$ through $y_{N-1}$. 
It follows that

\[ \hat{x}_N = \hat{x}_{N-1} + \hat{x}_{eN} \]

i.e.,

\[ \hat{x}_N = \hat{x}_{N-1} + (\mathbf{E} \mathbf{e}_N^*) R_{e,N}^{-1} e_N \]  \hspace{1cm} (7.6)

This is a useful recursive formula; it shows how the estimator of \( \hat{x} \) can be updated recursively by adding the contribution of the most recent variable \( e_N \).

The question now is how to generate the variables \( \{e_i\} \) from the \( \{y_i\} \). One possible transformation is the so-called Gram-Schmidt procedure. Let \( \hat{y}_{i|i-1} \) denote the estimator of \( y_i \) that is based on the observations up to time \( i - 1 \), i.e., on \( \{y_0, y_1, \ldots, y_{i-1}\} \). The same argument that led to (7.5) shows that \( \hat{y}_{i|i-1} \) can be alternatively calculated by estimating \( y_i \) from \( \{e_0, \ldots, e_{i-1}\} \). Then we can construct \( e_i \) as

\[ e_i \overset{\Delta}{=} y_i - \hat{y}_{i|i-1} \] \hspace{1cm} (7.7)

That is, we can choose \( e_i \) as the estimation error that results from estimating \( y_i \) from the observations \( \{y_0, y_1 \ldots, y_{i-1}\} \).
In order to verify that the resulting \( \{e_i\} \) are uncorrelated with each other, we recall that, by virtue of the orthogonality condition of linear least-mean-squares estimation (cf. Thm. 4.1),

\[
e_i \perp \{y_0, y_1, \ldots, y_{i-1}\}
\]

That is, \( e_i \) is uncorrelated with the observations \( \{y_0, y_1, \ldots, y_{i-1}\} \). It then follows that \( e_i \) should be uncorrelated with any \( e_j \) for \( j < i \) since, by definition, \( e_j \) is a linear combination of the observations \( \{y_0, y_1, \ldots, y_j\} \) and, moreover,

\[
\{y_0, y_1, \ldots, y_j\} \subset \{y_0, y_1, \ldots, y_{i-1}\} \quad \text{for} \quad j < i
\]

By the same token, \( e_i \) is uncorrelated with any \( e_j \) for \( j > i \).
It is instructive to see what choice of a transformation $A$ in (7.2) corresponds to the use of the Gram-Schmidt procedure. Assume, for illustration purposes, that $N = 2$. Then writing (7.7) for $i = 0, 1, 2$ we get

$$
\begin{bmatrix}
  e_0 \\
  e_1 \\
  e_2
\end{bmatrix} = \begin{bmatrix}
  I \\
  -(E y_1 y_0^*)(E y_0 y_0^*)^{-1} I \\
  \times \\
  \times \\
  I
\end{bmatrix} \begin{bmatrix}
  y_0 \\
  y_1 \\
  y_2
\end{bmatrix}
$$

where the entries $\times$ arise from the calculation

$$
[ \times \times ] = (E y_2 \begin{bmatrix}
  y_0^* \\
  y_1^*
\end{bmatrix}) \left( E \begin{bmatrix}
  y_0 \\
  y_1
\end{bmatrix} \begin{bmatrix}
  y_0 \\
  y_1
\end{bmatrix}^* \right)^{-1}
$$

We thus find that $A$ is a lower triangular transformation with unit entries along its diagonal. The lower triangularity of $A$ is relevant since it translates into a causal relationship between the $\{e_i\}$ and the $\{y_i\}$. By causality we mean that each $e_i$ can be computed from $\{y_j, j \leq i\}$ and, similarly, each $y_i$ can be recovered from $\{e_j, j \leq i\}$. We also see from the construction (7.7) that we can regard $e_i$ as the “new information” in $y_i$ given $\{y_0, \ldots, y_{i-1}\}$. Therefore, it is customary to refer to the $\{e_i\}$ as the innovations process associated with the $\{y_i\}$.
7.2 STATE-SPACE MODEL

As we now proceed to show, the Kalman filter is an efficient procedure for determining the innovations when the observation process \( \{y_i\} \) arises from a finite-dimensional linear state-space model.

What we mean by a state-space model for \( \{y_i\} \) is the following. We assume that \( y_i \) satisfies an equation of the form

\[
y_i = H_i x_i + v_i, \quad i \geq 0
\]  

(7.8)

in terms of an \( n \times 1 \) so-called state-vector \( x_i \), which in turn obeys a recursion of the form

\[
x_{i+1} = F_i x_i + G_i n_i, \quad i \geq 0
\]  

(7.9)

The processes \( v_i \) and \( n_i \) are assumed to be \( p \times 1 \) and \( m \times 1 \) zero-mean white noise processes, respectively, with covariances and cross-covariances denoted by

\[
E \left[ \begin{bmatrix} n_i \\ v_i \end{bmatrix} \right] \left[ \begin{bmatrix} n_j \\ v_j \end{bmatrix} \right]^* \triangleq \begin{bmatrix} Q_i & S_i \\ S_i^* & R_i \end{bmatrix} \delta_{ij}
\]
whereas the initial state $x_0$ is assumed to have zero mean, covariance matrix $\Pi_0$, and to be uncorrelated with $\{n_i\}$ and $\{v_i\}$, i.e.,

$$
E x_0^* x_0 = \Pi_0, \quad E n_i x_0^* = 0, \quad \text{and} \quad E v_i x_0^* = 0 \quad \text{for all} \quad i \geq 0
$$

The assumptions on $\{x_0, n_i, v_i\}$ can be compactly restated as

$$
E \left[ \begin{array}{c} n_i \\ v_i \\ x_0 \\
\end{array} \right] \left[ \begin{array}{c} n_j \\ v_j \\ x_0 \\
\end{array} \right]^* = \left[ \begin{array}{ccc} Q_i & S_i \\ S_i^* & R_i \\
\end{array} \right] \delta_{ij} \left[ \begin{array}{c} 0 \\ 0 \\
\end{array} \right] \left[ \begin{array}{ccc} 0 \\ 0 \\ \Pi_0 \\
\end{array} \right]
$$

(7.10)

It is also assumed that the matrices

$$
F_i \ (n \times n), \quad G_i \ (n \times m), \quad H_i \ (p \times n), \quad Q_i \ (m \times m), \quad R_i \ (p \times p), \quad S_i \ (m \times p), \quad \Pi_0 \ (n \times n)
$$

are known a priori. The process $v_i$ is called measurement noise and the process $n_i$ is called process noise. We now examine how the innovations $\{e_i\}$ of a process $\{y_i\}$ satisfying a state-space model of the form (7.8)–(7.10) can be evaluated.
7.3 RECURSION FOR THE STATE ESTIMATOR

Let \( \{ \hat{y}_{i|i-1}, \hat{x}_{i|i-1}, \hat{v}_{i|i-1} \} \) denote the estimators of the variables \( \{ y_i, x_i, v_i \} \) from the observations \( \{ y_0, y_1, \ldots, y_{i-1} \} \), respectively. Then using \( y_i = H_i x_i + v_i \), and appealing to linearity, we have

\[
\hat{y}_{i|i-1} = H_i \hat{x}_{i|i-1} + \hat{v}_{i|i-1}
\]  

(7.11)

Now the assumptions on our state-space model imply that

\[ v_i \perp y_j \quad \text{for} \quad j \leq i - 1 \]

i.e., \( v_i \) is uncorrelated with the observations \( \{ y_j, j \leq i - 1 \} \), so that

\[ \hat{v}_{i|i-1} = 0 \]

This is because from the model (7.8)–(7.9), \( y_j \) is a linear combination of the variables \( \{ v_j, n_{j-1}, \ldots, n_0, x_0 \} \), all of which are uncorrelated with \( v_i \) for \( j \leq i - 1 \).
Consequently,

\[ e_i = y_i - \hat{y}_{i|i-1} = y_i - H_i \hat{x}_{i|i-1} \]  

(7.12)

Therefore, the problem of finding the innovations reduces to one of finding \( \hat{x}_{i|i-1} \). For this purpose, we can use (7.6) to write

\[
\hat{x}_{i+1|i} = \hat{x}_{i+1|i-1} + (E x_{i+1} e_i^*) R_{e,i}^{-1} e_i
\]

\[
= \hat{x}_{i+1|i-1} + (E x_{i+1} e_i^*) R_{e,i}^{-1} (y_i - H_i \hat{x}_{i|i-1})
\]  

(7.13)

where

\[
R_{e,i} \triangleq E e_i e_i^*
\]  

(7.14)

But since \( x_{i+1} \) obeys the state equation \( x_{i+1} = F_i x_i + G_i n_i \), we also obtain, again by linearity, that

\[
\hat{x}_{i+1|i-1} = F_i \hat{x}_{i|i-1} + G_i \hat{n}_{i|i-1} = F_i \hat{x}_{i|i-1} + 0
\]  

(7.15)

since \( n_i \perp y_j, j \leq i - 1 \).
By combining Eqs. (7.12)–(7.15) we arrive at the following recursive equations for determining the innovations:

\[
\begin{align*}
e_i &= y_i - H_i \hat{x}_{i|i-1} \\
\hat{x}_{i+1|i} &= F_i \hat{x}_{i|i-1} + K_{p,i} e_i, \quad i \geq 0
\end{align*}
\]  

(7.16)

with initial conditions

\[
\hat{x}_{0|-1} = 0, \quad e_0 = y_0
\]  

(7.17)

and where we have defined the gain matrix

\[
K_{p,i} \triangleq (E x_{i+1} e_i^*) R_{e,i}^{-1}
\]  

(7.18)

The subscript “p” indicates that \( K_{p,i} \) is used to update a predicted estimator of the state vector. By combining the equations in (7.16) we also find that

\[
\hat{x}_{i+1|i} = F_{p,i} \hat{x}_{i|i-1} + K_{p,i} y_i, \quad F_{p,i} \triangleq F_i - K_{p,i} H_i, \quad \hat{x}_{0|-1} = 0, \quad i \geq 0
\]  

(7.19)

which shows that in finding the innovations, we actually also have a complete recursion for the state-estimator \( \{ \hat{x}_{i|i-1} \} \).
7.4 COMPUTING THE GAIN MATRIX

We still need to evaluate $K_{p,i}$ and $R_{e,i}$. To do so, we introduce the state-estimation error $\tilde{x}_{i|i-1} = x_i - \hat{x}_{i|i-1}$, and denote its covariance matrix by

$$P_{i|i-1} \triangleq \mathbb{E} \tilde{x}_{i|i-1} \tilde{x}_{i|i-1}^*$$  \hspace{1cm} (7.20)

Then, as we are going to see, the $\{K_{p,i}, R_{e,i}\}$ can be expressed in terms of $P_{i|i-1}$ and, in addition, the evaluation of $P_{i|i-1}$ will require propagating a so-called Riccati recursion.

To see this, note first that

$$e_i = y_i - H_i \hat{x}_{i|i-1} = H_i x_i - H_i \hat{x}_{i|i-1} + v_i = H_i \tilde{x}_{i|i-1} + v_i$$  \hspace{1cm} (7.21)

Moreover, $v_i \perp \tilde{x}_{i|i-1}$. This is because $\tilde{x}_{i|i-1}$ is a linear combination of the variables $\{v_0, \ldots, v_{i-1}, x_0, n_0, \ldots, n_{i-1}\}$, all of which are uncorrelated with $v_i$. This claim follows from the definition $\tilde{x}_{i|i-1} = x_i - \hat{x}_{i|i-1}$ and from the fact that $\hat{x}_{i|i-1}$ is a linear combination of $\{y_0, \ldots, y_{i-1}\}$ and $x_i$ is a linear combination of $\{x_0, n_0, \ldots, n_{i-1}\}$. Therefore, we get

$$R_{e,i} = \mathbb{E} e_i e_i^* = R_i + H_i P_{i|i-1} H_i^*$$  \hspace{1cm} (7.22)
Likewise, since

\[ E x_{i+1} e_i^* = F_i (E x_i e_i^*) + G_i (E n_i e_i^*) \] (7.23)

with the terms \( E x_i e_i^* \) and \( E n_i e_i^* \) given by

\[
\begin{align*}
E x_i e_i^* &= E \left( \tilde{x}_{i|i-1} + \tilde{x}_{i|i-1} \right) e_i^* \\
&= E \tilde{x}_{i|i-1} e_i^*, \quad \text{since} \ e_i \perp \tilde{x}_{i|i-1} \\
&= E \tilde{x}_{i|i-1} (H_i \tilde{x}_{i|i-1} + v_i)^* \\
&= E \tilde{x}_{i|i-1} (H_i \tilde{x}_{i|i-1} + 0), \quad \text{since} \ v_i \perp \tilde{x}_{i|i-1} \\
&= P_{i|i-1} H_i^*
\end{align*}
\]

and

\[
\begin{align*}
E n_i e_i^* &= E n_i (H_i \tilde{x}_{i|i-1} + v_i)^* \\
&= 0 + E n_i v_i^*, \quad \text{since} \ n_i \perp \tilde{x}_{i|i-1} \\
&= S_i
\end{align*}
\]

we get

\[
K_{p,i} = (E x_{i+1} e_i^*) R_{e,i}^{-1} = (F_i P_{i|i-1} H_i^* + G_i S_i) R_{e,i}^{-1}
\] (7.24)
Since $n_i \perp x_i$, it can be easily seen from $x_{i+1} = F_i x_i + G_i n_i$ that the covariance matrix of $x_i$ obeys the recursion

$$\Pi_{i+1} = F_i \Pi_i F_i^* + G_i Q_i G_i^*, \quad \Pi_i \triangleq \mathbb{E} x_i x_i^*$$

Likewise, since $e_i \perp \hat{x}_{i|i-1}$, it can be seen from $\hat{x}_{i+1|i} = F_i \hat{x}_{i|i-1} + K_{p,i} e_i$ that the covariance matrix of $\hat{x}_{i|i-1}$ satisfies the recursion

$$\Sigma_{i+1} = F_i \Sigma_i F_i^* + K_{p,i} R_{e,i} K_{p,i}^*, \quad \Sigma_i \triangleq \mathbb{E} \hat{x}_{i|i-1} \hat{x}_{i|i-1}^*$$

with initial condition $\Sigma_0 = 0$. Now the orthogonal decomposition

$$x_i = \hat{x}_{i|i-1} + \tilde{x}_{i|i-1} \quad \text{with} \quad \hat{x}_{i|i-1} \perp \tilde{x}_{i|i-1}$$

shows that $\Pi_i = \Sigma_i + P_{i|i-1}$. It is then immediate to conclude that the matrix $P_{i+1|i} = \Pi_{i+1} - \Sigma_{i+1}$ satisfies the recursion

$$P_{i+1|i} = F_i P_{i|i-1} F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*, \quad P_{0\mid-1} = \Pi_0$$
In summary, we arrive at the following statement of the Kalman filter, also known as the covariance form of the filter.

Algorithm 7.1 (The Kalman filter) Given observations \( \{ y_i \} \) that satisfy the state-space model (7.8)–(7.10), the innovations process \( \{ e_i \} \) can be recursively computed as follows. Start with \( \hat{x}_{0|1} = 0 \), \( P_{0|1} = \Pi_0 \), and repeat for \( i \geq 0 \):

\[
R_{e,i} = R_i + H_i P_{i|i-1} H_i^*
\]
\[
K_{p,i} = (F_i P_{i|i-1} H_i^* + G_i S_i) R_{e,i}^{-1}
\]
\[
e_i = y_i - H_i \hat{x}_{i|i-1}
\]
\[
\hat{x}_{i+1|i} = F_i \hat{x}_{i|i-1} + K_{p,i} e_i
\]
\[
P_{i+1|i} = F_i P_{i|i-1} F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^*
\]
Algorithm 7.2 (Time- and measurement-update forms) Given observations $\{y_i\}$ that satisfy the state-space model (7.8), (7.9), and (7.10), the innovations process $\{e_i\}$ can be recursively computed as follows. Start with $\hat{x}_{0|-1} = 0$, $P_{0|-1} = \Pi_0$, and repeat for $i \geq 0$:

$$R_{e,i} = R_i + H_i P_{i|-1} H_i^*$$

$$K_{f,i} = P_{i|-1} H_i^* R_{e,i}^{-1}$$

$$e_i = y_i - H_i \hat{x}_{i|-1}$$

$$\hat{x}_{i|i} = \hat{x}_{i|-1} + K_{f,i} e_i$$

$$\hat{x}_{i+1|i} = F_i \hat{x}_{i|i} + G_i S_i R_{e,i}^{-1} e_i$$

$$P_{i|i} = P_{i|-1} - P_{i|-1} H_i^* R_{e,i}^{-1} H_i P_{i|-1}$$

$$P_{i+1|i} = F_i P_{i|i} F_i^* + G_i (Q_i - S_i R_{e,i}^{-1} S_i^*) G_i^* - F_i K_{f,i} S_i^* G_i^* - G_i S_i K_{f,i}^* F_i^*$$
EXAMPLE

(From this point onwards, material is from handout on Kalman Filtering)

Consider the first-order model

\[ x(n + 1) = \frac{1}{4} x(n) + u(n) \]

\[ y(n) = x(n) + v(n), \quad n \geq 0 \]  (36.68)

Comparing with the standard model (36.29)–(36.30), we find that all variables are now scalars (and we writing \( x(n) \) instead of \( x_n \) to emphasize that the variable is scalar). The model coefficients are time-invariant and given by

\[ F = \frac{1}{4}, \quad G = 1, \quad H = \frac{1}{2}, \quad Q = 1, \quad R = \frac{1}{2}, \quad S = 0, \quad \Pi_0 = 1 \]  (36.69)
Writing down the Kalman recursions we get the following. Start with $\hat{x}(0|1) = 0, p(0|1) = 1$, and repeat for $n \geq 0$:

\[
\begin{align*}
  r_e(n) &= \frac{1}{2} + \frac{1}{4} p(n|n-1) \\
  k_p(n) &= \frac{1}{8} p(n|n-1) \\
  e(n) &= y(n) - \frac{1}{2} \hat{x}(n|n-1) \\
  \hat{x}(n+1|n) &= \frac{1}{4} \hat{x}(n|n-1) + k_p(n)e(n) \\
  &= \left[ \frac{1}{4} - \frac{k_p(n)}{2} \right] \hat{x}(n|n-1) + k_p(n)y(n) \\
  p(n+1|n) &= \frac{1}{16} p(n|n-1) + 1 - \frac{1}{64} p^2(n|n-1) \\
  &= \frac{1}{2} + \frac{1}{4} p(n|n-1)
\end{align*}
\]
In particular, these recursions allow us to evaluate the predictors of \( y(n) \) given all prior observations from time 0 up to and including time \( n - 1 \):

\[
\hat{y}(n|n-1) = \frac{1}{2} \hat{x}(n|n-1)
\]

Substituting into (36.70d) we get

\[
2\hat{y}(n+1|n) = \left[ \frac{1}{2} - k_p(n) \right] \hat{y}(n|n-1) + k_p(n)y(n)
\]

or, delaying by one time unit,

\[
\hat{y}(n|n-1) = \left[ \frac{1}{4} - \frac{k_p(n-1)}{2} \right] \hat{y}(n-1|n-2) + \frac{k_p(n-1)}{2} y(n-1)
\]

This is a first-order difference recursion with time-variant coefficients: the input is \( y(n-1) \) and the output is \( \hat{y}(n|n-1) \).
EXAMPLE: TRACKING A TARGET
EXAMPLE: TRACKING A TARGET

We consider a simplified model and assume the target is moving within the plane. The target is launched from location \((x_o, y_o)\) at an angle \(\theta\) with the horizontal axis at an initial speed \(v\). The initial velocity components along the horizontal and vertical directions are therefore

\[
v_x(0) = v \cos \theta, \quad v_y(0) = v \sin \theta
\]

(36.197)

The motion of the object is governed by Newton’s equations; the acceleration along the vertical direction is downwards and its magnitude is given by \(g \approx 10 \text{ m/s}^2\)

\[
v_x(t) = v \cos \theta, \quad t \geq 0
\]
\[
v_y(t) = v \sin \theta - gt, \quad t \geq 0
\]
\[
\frac{dx(t)}{dt} = v_x(t), \quad \frac{dy(t)}{dt} = v_y(t)
\]
EXAMPLE: TRACKING A TARGET

We sample the equations of motion every $T$ units of time and write

\[ v_x(n) \overset{\Delta}{=} v_x(nT) = v \cos \theta \]
\[ v_y(n) \overset{\Delta}{=} v_y(nT) = v \sin \theta - ngT \]
\[ x(n+1) = x(n) + Tv_x(n) \]
\[ y(n+1) = y(n) + Tv_y(n) \]
EXAMPLE: TRACKING A TARGET

\[
\begin{bmatrix}
  x(n+1) \\
  y(n+1) \\
  v_x(n+1) \\
  v_y(n+1)
\end{bmatrix}_{x_{n+1}} =
\begin{bmatrix}
  1 & 0 & T & 0 \\
  0 & 1 & 0 & T \\
  0 & 0 & 1 & 0 \\
  0 & 0 & 0 & 1
\end{bmatrix}_{F}
\begin{bmatrix}
  x(n) \\
  y(n) \\
  v_x(n) \\
  v_y(n)
\end{bmatrix}_{x_n}
- \begin{bmatrix}
  0 \\
  0 \\
  0 \\
  1
\end{bmatrix}_{d_n}^{gT}
\]

\[
z_n = \begin{bmatrix}
  1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0
\end{bmatrix}_{H}
\begin{bmatrix}
  x(n) \\
  y(n) \\
  v_x(n) \\
  v_y(n)
\end{bmatrix}
+ w_n
\]
EXAMPLE: TRACKING A TARGET

We start with $\hat{x}_{0|1} = 0$, $P_{0|1} = \Pi_0$

\[
R_{e,n} = R + HP_{n|n-1}H^*
\]
\[
K_{p,n} = FP_{n|n-1}H^*R_{e,n}^{-1}
\]
\[
e_n = z_n - H\hat{x}_{n|n-1}
\]
\[
\hat{x}_{n+1|n} = F\hat{x}_{n|n-1} + K_{p,n}e_n + d_n
\]
\[
P_{n+1|n} = FP_{n|n-1}F^* - K_{p,n}R_{e,n}K_{p,n}^*
\]
EXAMPLE: TRACKING A TARGET

\[ \Pi_0 = I, \quad R = \begin{bmatrix} 0.3 \\ 0.3 \end{bmatrix}, \quad (x_0, y_0) = (1, 30), \quad v = 15, \quad T = 0.01, \quad \theta = 60^\circ \]
Whitening Filter:

\[
\begin{align*}
\hat{x}_{n+1|n} &= [F_n - K_{p,n}H_n] \hat{x}_{n|n-1} + K_{p,n} y_n \\
e_n &= -H_n \hat{x}_{n|n-1} + y_n
\end{align*}
\]

This model has \( y_n \) as input and \( e_n \) as output. Since \( e_n \) is an uncorrelated sequence with variance matrix \( R_{e,n} \), we call the above model a whitening filter: it shows how to decorrelate the input sequence \( y_n \).

Modeling Filter:

\[
\begin{align*}
\hat{x}_{n+1|n} &= F_n \hat{x}_{n|n-1} + K_{p,n} e_n \\
y_n &= H_n \hat{x}_{n|n-1} + e_n
\end{align*}
\]

This model has \( e_n \) as input and \( y_n \) as output. Since \( e_n \) is an uncorrelated sequence, we therefore say that the above filter is a modeling filter: it shows how to generate the sequence \( y_n \) from the uncorrelated input sequence, \( e_n \).
Assume now that the state-space model is time-invariant.

\[
    x_{n+1} = Fx_n + Gu_n, \quad n > -\infty \\
    y_n = Hx_n + v_n
\]

with

\[
    E \begin{bmatrix} u_n \\ v_n \\ x_0 \\ 1 \end{bmatrix} \begin{bmatrix} u_m \\ v_m \\ x_0 \end{bmatrix}^* = \begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \delta_{nm} \begin{bmatrix} 0 \\ 0 \\ 0 \\ \Pi_0 \end{bmatrix}
\]

It follows from the above state-space model that the transfer matrix function from \( u_n \) to \( y_n \) is given by

\[
    H_{uy}(z) = H(zI - F)^{-1}G
\]

while the transfer function from \( v_n \) to \( y_n \) is \( H_{vy}(z) = 1 \). Assuming \( F \) is a stable matrix (meaning all its eigenvalues are inside the unit circle), then \( H_{uy}(z) \) will be a BIBO stable mapping. Since the processes \( \{u_n, v_n\} \) are wide-sense stationary, it follows that \( \{y_n\} \) will also be wide sense stationary process.
Solutions of DARE
Now, the Kalman recursions that correspond to the above model are given by:

\[
\begin{align*}
R_{e,n} & = R + HP_{n|n-1}H^* \\
K_{p,n} & = (FP_{n|n-1}H^* + GS)R_{e,n}^{-1} \\
e_n & = y_n - H\hat{x}_{n|n-1} \\
\hat{x}_{n+1|n} & = F\hat{x}_{n|n-1} + K_{p,n}e_n \\
P_{n+1|n} & = FP_{n|n-1}F^* + GQG^* - K_{p,n}R_{e,n}K_{p,n}^*
\end{align*}
\]
There are conditions under which the Riccati recursion can be shown to converge to a unique positive-definite matrix $P$. Discussion of these technical conditions is beyond the scope of this text. It suffices to consider here the case when $S = 0$ and $F$ is a stable matrix. In this case, the Riccati recursion can be shown to converge, as $n \to \infty$, to a unique positive-definite matrix $P$ that satisfies the so-called Discrete Algebraic Riccati Equation (DARE):

$$P = FPF^* + GQG^* - K_p R_e K_p^*$$

where

$$R_e = R + HPH^*$$
$$K_p = FPH^* R_e^{-1}$$

Moreover, the resulting closed-loop matrix $F - K_p H$ will also be stable (i.e., will have all its eigenvalues inside the unit circle).
Let us now focus on the case in which the sequences \( \{y_n, v_n\} \) are scalar sequences, replaced by \( \{y(n), v(n)\} \); the discussion can be easily extended to vector processes but it is sufficient for our purposes to study scalar-valued output processes. In this situation, the \( H \) matrix becomes a row vector, say \( h^T \), and the covariance matrix \( R \) becomes a scalar, say \( r \). Additionally, the innovations process \( e_n \) becomes scalar-valued, say, \( e(n) \), with variance \( r_e \) in steady-state, and the gain matrix \( K_p \) becomes a column vector, \( k_p \). Writing down the resulting modeling filter (36.84a)–(36.84b) in steady-state, as \( n \to \infty \), we have

\[
\hat{x}_{n+1|n} = F \hat{x}_{n|n-1} + k_p e(n)
\]

\[
y(n) = h^T \hat{x}_{n|n-1} + e(n)
\]
The transfer function from $e(n)$ to $y(n)$ is the following causal modeling filter $L(z)$:

$$L(z) = 1 + h^T (zI - F)^{-1} k_p$$

This is a stable transfer function since its poles are given by the eigenvalues of $F$ and these eigenvalues lie inside the unit disc. Accordingly, $y(n)$ is a wide-sense stationary process and its $z-$spectrum is given by

$$S_y(z) = r_c L(z) \left[ L \left( \frac{1}{z^*} \right) \right]^*$$

We therefore find that the steady-state Kalman equations allow us to determine the canonical spectral factor, $L(z)$, of the process $\{y(n)\}$ (since it allows us to determine $k_p$).