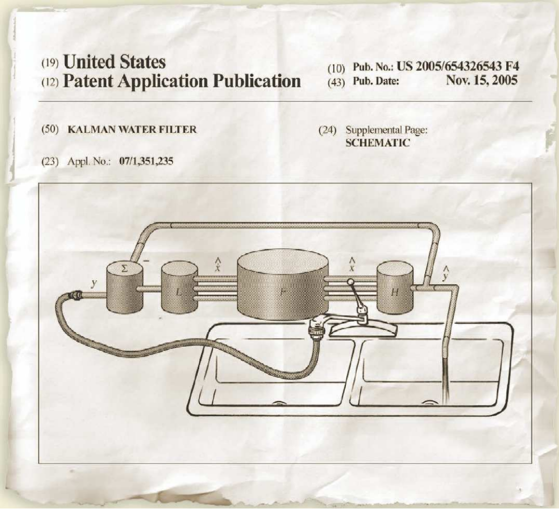
ECE531 Lecture 11: Dynamic Parameter Estimation: Kalman-Bucy Filter

D. Richard Brown III

Worcester Polytechnic Institute

09-Apr-2009

The Kalman Water Filter



(19) **United States**
(12) **Patent Application Publication**

(10) **Pub. No.: US 2005/654326543 F4**
(43) **Pub. Date: Nov. 15, 2005**

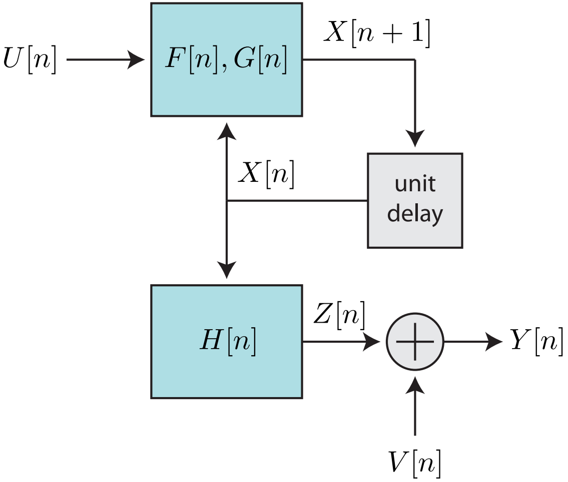
(50) **KALMAN WATER FILTER**

(23) Appl. No.: **07/1,351,235**

(24) Supplemental Page:
SCHEMATIC

(from IEEE Control Systems Magazine Feb 2006: http://ieeexplore.ieee.org/iel5/37/33354/01580172.pdf)

Dynamic MMSE State Estimation (Filtering)



For now, we will focus on the **filtering** problem, i.e. estimating $X[\ell]$ given the observations $Y[0], \dots, Y[\ell]$. We know that the MMSE state estimator is

$$\hat{X}[\ell \,|\, \ell] := \mathrm{E}\left\{X[\ell] \,|\, \mathcal{Y}_0^\ell\right\}$$

where $\mathcal{Y}_0^\ell := \{Y[0], \dots, Y[\ell]\}$. The goal here is to leverage the **restrictions** we placed on the dynamical model in Lecture 10 to derive a **computationally efficient** MMSE estimator.

MMSE State Prediction $\hat{X}[0]$ given no observations

Recall that we have assumed a Gaussian distributed initial state $X[0] \sim \mathcal{N}(m[0], \Sigma[0])$ where $m[0]$ and $\Sigma[0]$ are known. If you have no observations yet, what is the best estimator of the initial state?

Prior to receiving the first observation, the MMSE estimate (prediction) of the initial state $X[0]$ is simply

$$\hat{X}[0 \,|\, -1] \quad := \quad \mathrm{E}\left\{X[0] \,|\, \text{no observations}\right\} = m[0].$$

The error covariance matrix of this MMSE estimator (predictor) is then

$$\begin{aligned} \Sigma[0 \,|\, -1] \quad &:= \quad \mathrm{E}\left\{\left(\hat{X}[0 \,|\, -1] - X[0]\right)\left(\hat{X}[0 \,|\, -1] - X[0]\right)^\top\right\} \\ &= \quad \mathrm{E}\left\{(m[0] - X[0])\,(m[0] - X[0])^\top\right\} \\ &= \quad \Sigma[0] \end{aligned}$$

Review: Conditional Mean of Jointly Gaussian Vectors

Given

$$Y = HX + V$$

with $X \sim \mathcal{N}(m, \Sigma)$, $V \sim \mathcal{N}(0, R)$, and X and V independent.

We would like to compute the conditional mean $\mathrm{E}[X \mid Y]$. We can use the conditional pdf of jointly Gaussian random vectors to write

$$E[X \mid Y = y] = \mathrm{E}[X] + \mathrm{cov}(X, Y) \left[\mathrm{cov}(Y, Y)\right]^{-1} (y - \mathrm{E}[Y])$$

It is pretty easy to show that $\mathrm{E}[X] = m$, $\mathrm{E}[Y] = Hm$, $\mathrm{cov}(X, Y) = \Sigma H^\top$, and $\mathrm{cov}(Y, Y) = H \Sigma H^\top + R$. Hence,

$$\boxed{E[X \mid Y = y] = m + \Sigma H^\top \left[H \Sigma H^\top + R\right]^{-1} (y - Hm)}$$

MMSE State Estimate $\hat{X}[0]$ given $Y[0]$ (part 1)

At time $n = 0$, we receive the observation

$$Y[0] = H[0]X[0] + V[0]$$

where $X[0] \sim \mathcal{N}(m[0], \Sigma[0])$, $V[0] \sim \mathcal{N}(0, R[0])$, and $H[0]$ is known. Note that, under the restrictions in Lecture 10, this is just the standard linear Gaussian model.

We use our prior result to find the MMSE estimate of $X[0]$ given $Y[0]$...

$$\begin{aligned}\hat{X}[0 \,|\, 0] &:= \mathrm{E}\{X[0] \,|\, Y[0]\} \\ &= m[0] + \Sigma[0]H^\top[0]\left(H[0]\Sigma[0]H^\top[0] + R[0]\right)^{-1}(Y[0] - H[0]m[0]) \\ &= \hat{X}[0 \,|\, -1] + \\ &\quad \Sigma[0 \,|\, -1]H^\top[0]\left(H[0]\Sigma[0 \,|\, -1]H^\top[0] + R[0]\right)^{-1}(Y[0] - H[0]\hat{X}[0 \,|\, -1])\end{aligned}$$

MMSE State Estimate $\hat{X}[0]$ given $Y[0]$ (part 2)

If we define

$$\begin{aligned} K[0] &:= \mathrm{cov}(X[0], Y[0]) \left[\mathrm{cov}(Y[0], Y[0])\right]^{-1} \\ &= \Sigma[0 \,|\, -1] H^\top[0] \left(H[0]\Sigma[0 \,|\, -1]H^\top[0] + R[0]\right)^{-1} \end{aligned}$$

we can write

$$\boxed{\hat{X}[0 \,|\, 0] = \hat{X}[0 \,|\, -1] + K[0]\left(Y[0] - H[0]\hat{X}[0 \,|\, -1]\right)}$$

Remarks:

- The term $\tilde{Y}[0 \,|\, -1] := Y[0] - H[0]\hat{X}[0 \,|\, -1]$ is sometimes called the "innovation". It is the error between the MMSE **prediction** of $Y[0]$ and the actual observation $Y[0]$.
- Substituting for $Y[0]$, note that the innovation can be written as

$$\begin{aligned} \tilde{Y}[0 \,|\, -1] &= \left(H[0]X[0] + V[0] - H[0]\hat{X}[0 \,|\, -1]\right) \\ &= H[0]\left(X[0] - \hat{X}[0 \,|\, -1]\right) + V[0] \end{aligned}$$

Kalman Gain

The matrix

$$K[0] \quad := \quad \mathrm{cov}(X[0], Y[0]) \left[\mathrm{cov}(Y[0], Y[0])\right]^{-1}$$

is sometimes called the "Kalman Gain". Both covariances are unconditional here.

- Rewriting first covariance term in the Kalman Gain as

$$\mathrm{cov}(X[0], Y[0]) \quad = \quad \mathrm{E}\left\{\left(X[0] - \hat{X}[0 \,|\, -1]\right)\left(Y[0] - H[0]\hat{X}[0 \,|\, -1]\right)^\top\right\}$$

  and the second term as

$$\mathrm{cov}(Y[0], Y[0]) \quad = \quad \mathrm{E}\left\{\left(Y[0] - H[0]\hat{X}[0 \,|\, -1]\right)\left(Y[0] - H[0]\hat{X}[0 \,|\, -1]\right)^\top\right\}$$

  we can see that the Kalman Gain (in a simplistic scalar sense) is
  - "proportional" to the covariance between the state prediction error and the innovation
  - "inversely proportional" to the innovation variance
- Does this make sense in the context of our state estimation equation?

$$\hat{X}[0 \,|\, 0] = \hat{X}[0 \,|\, -1] + K[0]\left(Y[0] - H[0]\hat{X}[0 \,|\, -1]\right)$$

Additional Interpretation

- Note that $K[0]$ is only a function of the error covariance matrix of the initial state prediction $\Sigma[0\,|\,-1]$, the known state update matrix $H[0]$, and the known measurement noise covariance $R[0]$.
- Given $K[0]$ and $H[0]$, the state estimate $\hat{X}[0\,|\,0]$ is only a function of the previous state prediction $\hat{X}[0\,|\,-1]$ and the current observation $Y[0]$.
- What can we say about $\Sigma[0\,|\,0]$, the error covariance matrix of the state estimate $\hat{X}[0\,|\,0]$ (conditioned on the observation $Y[0]$)?

MMSE State Estimate $\hat{X}[0]$ given $Y[0]$

The error covariance matrix of the MMSE estimator $\hat{X}[0\,|\,0]$ can be computed as

$$\begin{aligned}\Sigma[0\,|\,0] &:= \mathrm{E}\left\{\left(\hat{X}[0\,|\,0]-X[0]\right)\left(\hat{X}[0\,|\,0]-X[0]\right)^{\top}\,|\,Y[0]\right\}\\ &= \mathrm{cov}\left\{X[0]\,|\,Y[0]\right\}\end{aligned}$$

We can use the standard result for jointly Gaussian random vectors to write

$$\begin{aligned}\Sigma[0\,|\,0] &= \Sigma[0]-\Sigma[0]H^{\top}[0]\left(H[0]\Sigma[0]H^{\top}[0]+R[0]\right)^{-1}H[0]\Sigma[0]\\ &= \Sigma[0\,|-1]-K[0]H[0]\Sigma[0\,|-1]\end{aligned}$$

where we used our definitions for $\Sigma[0\,|-1]$ and $K[0]$ in the last equality. Note that, given $K[0]$ and $H[0]$, the error covariance matrix of the MMSE state estimator after the observation $Y[0]$ is only a function of the error covariance matrix of the prediction $\Sigma[0\,|-1]$.

Remarks

What we have done so far:

1. Predicted the first state with no observations: $\hat{X}[0\,|\,-1]$.
2. Computed the error covariance matrix of this prediction: $\Sigma[0\,|\,-1]$.
3. Received the observation $Y[0]$.
4. Estimated the first state given the observation: $\hat{X}[0\,|\,0]$.
5. Computed the error covariance matrix of this estimate: $\Sigma[0\,|\,0]$.

Interesting observations:

- The **estimate** $\hat{X}[0\,|\,0]$ is expressed in terms of the **prediction** $\hat{X}[0\,|\,-1]$ and the observation $Y[0]$.
- The **estimate error covariance matrix** $\Sigma[0\,|\,0]$ is expressed in terms of the **prediction error covariance matrix** $\Sigma[0\,|\,-1]$.

The goal here is to develop a general recursion. Our first step will be to see if the prediction for the next state (and its ECM) can be expressed in terms of the estimate (and its ECM) of the previous state.

State Update from $X[\ell]$ to $X[n+1]$

From our linear dynamical model, we have

$$\begin{aligned} X[n+1] &= F[n]X[n] + G[n]U[n] \\ &= F[n]\left(F[n-1]X[n-1] + G[n-1]U[n-1]\right) + G[n]U[n] \\ &= \text{etc.} \end{aligned}$$

For $n+1 > \ell$, we can repeat this process to write

$$X[n+1] = \underbrace{\left\{\prod_{k=0}^{n-\ell} F[n-k]\right\}}_{\mathcal{F}_n^{\ell}} X[\ell] + \sum_{j=\ell}^{n} \underbrace{\left\{\prod_{k=0}^{n-j-1} F[n-k]\right\}}_{\mathcal{F}_n^{j+1}} G[j]U[j]$$

where

$$\mathcal{F}_n^t := \begin{cases} F[n]F[n-1]\cdots F[t] & t \le n \\ I & t > n \end{cases}$$

General Expression for MMSE State Prediction

For $n+1 > \ell$, we can use our prior result to write

$$
\begin{aligned}
\hat{X}[n+1\,|\,\ell] &:= \mathrm{E}\left\{X[n+1]\,|\,\mathcal{Y}_0^\ell\right\} \\
&\overset{\text{model}}{=} \mathrm{E}\left\{\mathcal{F}_n^\ell X[\ell] + \sum_{j=\ell}^{n} \mathcal{F}_n^{j+1} G[j] U[j]\,|\,\mathcal{Y}_0^\ell\right\} \\
&\overset{\text{linearity}}{=} \mathcal{F}_n^\ell \mathrm{E}\left\{X[\ell]\,|\,\mathcal{Y}_0^\ell\right\} + \sum_{j=\ell}^{n} \mathcal{F}_n^{j+1} G[j] \mathrm{E}\left\{U[j]\,|\,\mathcal{Y}_0^\ell\right\} \\
&= \mathcal{F}_n^\ell \hat{X}[\ell\,|\,\ell] + \sum_{j=\ell}^{n} \mathcal{F}_n^{j+1} G[j] \mathrm{E}\left\{U[j]\,|\,\mathcal{Y}_0^\ell\right\} \\
&\overset{\text{irrelevance}}{=} \mathcal{F}_n^\ell \hat{X}[\ell\,|\,\ell] + \sum_{j=\ell}^{n} \mathcal{F}_n^{j+1} G[j] \mathrm{E}\left\{U[j]\right\} \\
&= \mathcal{F}_n^\ell \hat{X}[\ell\,|\,\ell]
\end{aligned}
$$

Note that the one-step predictor can be expressed as $\hat{X}[\ell+1\,|\,\ell] = F[\ell]\hat{X}[\ell\,|\,\ell]$.

Conditional ECM of MMSE State Predictor

$$
\begin{aligned}
\Sigma[n+1\,|\,\ell] &:= \operatorname{cov}\left\{X[n+1]\,|\,\mathcal{Y}_0^\ell\right\} \\
&\overset{\text{model}}{=} \operatorname{cov}\left\{\mathcal{F}_n^\ell X[\ell] + \sum_{j=\ell}^{n} \mathcal{F}_n^{j+1} G[j]U[j]\,|\,\mathcal{Y}_0^\ell\right\} \\
&\overset{\text{indep}}{=} \operatorname{cov}\left\{\mathcal{F}_n^\ell X[\ell]\,|\,\mathcal{Y}_0^\ell\right\} + \operatorname{cov}\left\{\sum_{j=\ell}^{n} \mathcal{F}_n^{j+1} G[j]U[j]\,|\,\mathcal{Y}_0^\ell\right\} \\
&\overset{\text{irrelevance}}{=} \operatorname{cov}\left\{\mathcal{F}_n^\ell X[\ell]\,|\,\mathcal{Y}_0^\ell\right\} + \operatorname{cov}\left\{\sum_{j=\ell}^{n} \mathcal{F}_n^{j+1} G[j]U[j]\right\} \\
&= \mathcal{F}_n^\ell \operatorname{cov}\left\{X[\ell]\,|\,\mathcal{Y}_0^\ell\right\}(\mathcal{F}_n^\ell)^\top + \sum_{j=\ell}^{n} \mathcal{F}_n^{j+1} G[j] \operatorname{cov}\{U[j]\}\left(\mathcal{F}_n^{j+1} G[j]\right)^\top \\
&= \mathcal{F}_n^\ell \Sigma[\ell\,|\,\ell](\mathcal{F}_n^\ell)^\top + \sum_{j=\ell}^{n} \mathcal{F}_n^{j+1} G[j] Q[j]\left(\mathcal{F}_n^{j+1} G[j]\right)^\top
\end{aligned}
$$

For one-step prediction: $\Sigma[\ell+1\,|\,\ell] = F[\ell]\Sigma[\ell\,|\,\ell]F^\top[\ell] + G[\ell]Q[\ell]G^\top[\ell]$.

Summary of Main Results

MMSE state estimator (only shown so far for the case $\ell = 0$):

$$\boxed{\hat{X}[\ell \,|\, \ell] = \hat{X}[\ell \,|\, \ell - 1] + K[\ell]\left(Y[\ell] - H[\ell]\hat{X}[\ell \,|\, \ell - 1]\right)}$$

One-step MMSE state predictor:

$$\boxed{\hat{X}[\ell + 1 \,|\, \ell] = F[\ell]\hat{X}[\ell \,|\, \ell]}$$

ECM of MMSE state estimator (only shown so far for the case $\ell = 0$):

$$\boxed{\Sigma[\ell \,|\, \ell] = \Sigma[\ell \,|\, \ell - 1] - K[\ell]H[\ell]\Sigma[\ell \,|\, \ell - 1]}$$

ECM of MMSE one-step state predictor:

$$\boxed{\Sigma[\ell + 1 \,|\, \ell] = F[\ell]\Sigma[\ell \,|\, \ell]F^\top[\ell] + G[\ell]Q[\ell]G^\top[\ell]}$$

Induction: MMSE State Estimator for Arbitrary ℓ

Assume that

$$\hat{X}[\ell \,|\, \ell] = \hat{X}[\ell \,|\, \ell-1] + K[\ell] \left(Y[\ell] - H[\ell]\hat{X}[\ell \,|\, \ell-1] \right)$$

and

$$\Sigma[\ell \,|\, \ell] = \Sigma[\ell \,|\, \ell-1] - K[\ell]H[\ell]\Sigma[\ell \,|\, \ell-1]$$

are true for some value of ℓ. We want to show that these expressions are also true for $\ell+1$ using the fact that the prediction equations have already been shown to be true for any ℓ.

Fact (not too difficult to prove): The innovation

$$\tilde{Y}[\ell+1 \,|\, \ell] := Y[\ell+1] - H[\ell+1]\hat{X}[\ell+1 \,|\, \ell]$$

is a zero-mean Gaussian random vector uncorrelated with $Y[0], \dots, Y[\ell]$.

Useful Result 0

Assume that X, Y, and Z are jointly Gaussian and that Y and Z are uncorrelated with each other. Denote C as the covariance matrix between the subscripted random variables. We can use the conditional mean result for jointly Gaussian random vectors to write

$$\mathrm{E}\{X \,|\, Y, Z\} = \mathrm{E}\{X\} + \begin{bmatrix} C_{XY} & C_{XZ} \end{bmatrix} \begin{bmatrix} C_{YY} & C_{YZ} \\ C_{ZY} & C_{ZZ} \end{bmatrix}^{-1} \begin{bmatrix} Y - \mathrm{E}\{Y\} \\ Z - \mathrm{E}\{Z\} \end{bmatrix}$$

Under our assumption that Y and Z are uncorrelated, both C_{YZ} and C_{ZY} are equal to zero. The matrix inverse is then easy to compute and we have the useful result

$$\begin{aligned} \mathrm{E}\{X \,|\, Y, Z\} &= \mathrm{E}\{X\} + C_{XY} C_{YY}^{-1} (Y - \mathrm{E}\{Y\}) + C_{XZ} C_{ZZ}^{-1} (Z - \mathrm{E}\{Z\}) \\ &= \mathrm{E}\{X \,|\, Y\} + C_{XZ} C_{ZZ}^{-1} (Z - \mathrm{E}\{Z\}) . \end{aligned}$$

Induction: Conditional Mean

$$
\begin{aligned}
\hat{X}[\ell+1\,|\,\ell+1] &:= \mathrm{E}\left\{X[\ell+1]\,|\,\mathcal{Y}_0^{\ell+1}\right\} \\
&= \mathrm{E}\left\{X[\ell+1]\,|\,\mathcal{Y}_0^{\ell}, Y[\ell+1]\right\} \\
&= \mathrm{E}\left\{X[\ell+1]\,|\,\mathcal{Y}_0^{\ell}, \tilde{Y}[\ell+1\,|\,\ell]\right\} \\
&= \mathrm{E}\left\{X[\ell+1]\,|\,\mathcal{Y}_0^{\ell}\right\} + \mathrm{cov}\left(X[\ell+1], \tilde{Y}[\ell+1\,|\,\ell]\right) \times \\
&\quad \left[\mathrm{cov}\left(\tilde{Y}[\ell+1\,|\,\ell], \tilde{Y}[\ell+1\,|\,\ell]\right)\right]^{-1} \tilde{Y}[\ell+1\,|\,\ell] \\
&= \hat{X}[\ell+1\,|\,\ell] + \mathrm{cov}\left(X[\ell+1], \tilde{Y}[\ell+1\,|\,\ell]\right) \times \\
&\quad \left[\mathrm{cov}\left(\tilde{Y}[\ell+1\,|\,\ell], \tilde{Y}[\ell+1\,|\,\ell]\right)\right]^{-1} \tilde{Y}[\ell+1\,|\,\ell]
\end{aligned}
$$

where we have used "useful result 0" and the fact that $\mathrm{E}\left\{\tilde{Y}[\ell+1\,|\,\ell]\right\} = 0$ to write the second to last expression.

Induction: Conditional Mean

Defining

$$K[\ell+1] := \mathrm{cov}\left(X[\ell+1], \tilde{Y}[\ell+1\,|\,\ell]\right)\left[\mathrm{cov}\left(\tilde{Y}[\ell+1\,|\,\ell], \tilde{Y}[\ell+1\,|\,\ell]\right)\right]^{-1}$$

and substituting for the innovation

$$\tilde{Y}[\ell+1\,|\,\ell] := Y[\ell+1] - H[\ell+1]\hat{X}[\ell+1\,|\,\ell],$$

we get the result we wanted:

$$\hat{X}[\ell+1\,|\,\ell+1] \;=\; \hat{X}[\ell+1\,|\,\ell] + K[\ell+1]\left(Y[\ell+1] - H[\ell+1]\hat{X}[\ell+1\,|\,\ell]\right)$$

Induction: Error Covariance Matrix

The final step is to show

$$\Sigma[\ell+1\,|\,\ell+1] = \Sigma[\ell+1\,|\,\ell] - K[\ell+1]H[\ell+1]\Sigma[\ell+1\,|\,\ell]$$

To see this, start from the definition

$$\Sigma[\ell+1\,|\,\ell+1] \;=\; \mathrm{E}\left\{\left(\hat{X}[\ell+1\,|\,\ell+1] - X[\ell+1]\right)\left(\hat{X}[\ell+1\,|\,\ell+1] - X[\ell+1]\right)^\top\right\}$$

and substitute our result

$$\hat{X}[\ell+1\,|\,\ell+1] \;=\; \hat{X}[\ell+1\,|\,\ell] + K[\ell+1]\left(Y[\ell+1] - H[\ell+1]\hat{X}[\ell+1\,|\,\ell]\right)$$

After expanding the outer product, there are four expectations that need to be computed...

Useful Result Number 1

$$
\begin{aligned}
\operatorname{cov}\left(X[\ell+1], \tilde{Y}[\ell+1 \mid \ell]\right) &= \mathrm{E}\Big\{\left(X[\ell+1]-\hat{X}[\ell+1 \mid \ell]\right) \\
&\quad \times \left(Y[\ell+1]-H[\ell+1]\hat{X}[\ell+1 \mid \ell]\right)^\top\Big\} \\
&= \mathrm{E}\Big\{\left(X[\ell+1]-\hat{X}[\ell+1 \mid \ell]\right) \\
&\quad \times \left(H[\ell+1](X[\ell+1]-\hat{X}[\ell+1 \mid \ell])+V[\ell+1]\right)^\top\Big\} \\
&= \mathrm{E}\Big\{\left(X[\ell+1]-\hat{X}[\ell+1 \mid \ell]\right) \\
&\quad \times \left(X[\ell+1]-\hat{X}[\ell+1 \mid \ell]\right)^\top\Big\} H^\top[\ell+1] \\
&= \Sigma[\ell+1 \mid \ell] H^\top[\ell+1]
\end{aligned}
$$

Useful Result Number 2

$$
\begin{aligned}
\operatorname{cov}\left(\tilde{Y}[\ell+1\,|\,\ell], \tilde{Y}[\ell+1\,|\,\ell]\right) &= \mathrm{E}\Big\{\left(Y[\ell+1] - H[\ell+1]\hat{X}[\ell+1\,|\,\ell]\right) \\
&\quad \times \left(Y[\ell+1] - H[\ell+1]\hat{X}[\ell+1\,|\,\ell]\right)^\top\Big\} \\
&= \mathrm{E}\Big\{\left(H[\ell+1](X[\ell+1] - \hat{X}[\ell+1\,|\,\ell]) + V[\ell+1]\right) \\
&\quad \times \left(H[\ell+1](X[\ell+1] - \hat{X}[\ell+1\,|\,\ell]) + V[\ell+1]\right)^\top\Big\} \\
&= H[\ell+1]\mathrm{E}\Big\{\left(X[\ell+1] - \hat{X}[\ell+1\,|\,\ell]\right) \\
&\quad \times \left(X[\ell+1] - \hat{X}[\ell+1\,|\,\ell]\right)^\top\Big\}H^\top[\ell+1] \\
&\quad + R[\ell+1] \\
&= H[\ell+1]\Sigma[\ell+1\,|\,\ell]H^\top[\ell+1] + R[\ell+1]
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{E}\left\{\left(X[\ell+1]-\hat{X}[\ell+1\,|\,\ell]\right)\left(X[\ell+1]-\hat{X}[\ell+1\,|\,\ell]\right)^\top\right\}\\
&-K[\ell+1]\mathrm{E}\left\{\left(Y[\ell+1]-H[\ell+1]\hat{X}[\ell+1\,|\,\ell]\right)\left(X[\ell+1]-\hat{X}[\ell+1\,|\,\ell]\right)^\top\right\}\\
&-\mathrm{E}\left\{\left(X[\ell+1]-\hat{X}[\ell+1\,|\,\ell]\right)\left(Y[\ell+1]-H[\ell+1]\hat{X}[\ell+1\,|\,\ell]\right)^\top\right\}K^\top[\ell+1]+\\
&K[\ell+1]\mathrm{E}\left\{\left(Y[\ell+1]-H[\ell+1]\hat{X}[\ell+1\,|\,\ell]\right)\left(Y[\ell+1]-H[\ell+1]\hat{X}[\ell+1\,|\,\ell]\right)^\top\right\}K^\top[\ell+1]
\end{aligned}
$$

- The first line is simply $\Sigma[\ell+1\,|\,\ell]$.
- The third line was solved in "useful result number 1" and is

$$\Sigma[\ell+1\,|\,\ell]H^\top[\ell+1]K^\top[\ell+1]$$

- Inspection of $K^\top[\ell+1]$ reveals that the third line is a symmetric matrix. Hence, the second line is equal to the third line.
- The fourth line was solved in "useful result number 2"...

From useful result number 2, we can write the fourth line as

$$K[\ell+1](H[\ell+1]\Sigma[\ell+1\,|\,\ell]H^\top[\ell+1]+R[\ell+1])K^\top[\ell+1]$$

This can be simplified a bit since

$$K^\top[\ell+1] \;=\; (H[\ell+1]\Sigma[\ell+1\,|\,\ell]H^\top[\ell+1]+R[\ell+1])^{-1}H[\ell+1]\Sigma[\ell+1\,|\,\ell]$$

Substituting for the $K^\top[\ell+1]$ at the end of the equation, we can write the fourth line as

$$K[\ell+1]H[\ell+1]\Sigma[\ell+1\,|\,\ell]$$

which is the same as the second and third lines, except for a sign change. Putting it all together, we have

$$\Sigma[\ell+1\,|\,\ell+1] \;=\; \Sigma[\ell+1\,|\,\ell]-K[\ell+1]H[\ell+1]\Sigma[\ell+1\,|\,\ell]$$

which is the result we wanted to show.

The Discrete-Time Kalman-Bucy Filter (1961)

Theorem

Under the squared error cost assignment, the linear system model, and the white Gaussian input, noise, and initial state assumptions discussed in Lecture 10, the optimal estimates for the current state (filtering) and the next state (prediction) are given recursively as

$$\begin{aligned}\hat{X}[\ell \,|\, \ell] &= \hat{X}[\ell \,|\, \ell-1] + K[\ell]\left(Y[\ell] - H[\ell]\hat{X}[\ell \,|\, \ell-1]\right) \text{ for } \ell = 0,1,\dots \\ \hat{X}[\ell+1 \,|\, \ell] &= F[\ell]\hat{X}[\ell \,|\, \ell] \text{ for } \ell = 0,1,\dots\end{aligned}$$

with the initialization $\hat{X}[0 \,|\, -1] = m[0]$ *and where the matrix*

$$K[\ell] = \Sigma[\ell \,|\, \ell-1]H^\top[\ell]\left(H[\ell]\Sigma[\ell \,|\, \ell-1]H^\top[\ell] + R[\ell]\right)^{-1}$$

with $\Sigma[\ell \,|\, \ell-1] := \operatorname{cov}\{X[\ell] \,|\, Y[0],\dots,Y[\ell-1]\}$ *and* $R[\ell] := \operatorname{cov}\{V[\ell]\}$.

Kalman Filter: Summary of General Recursion

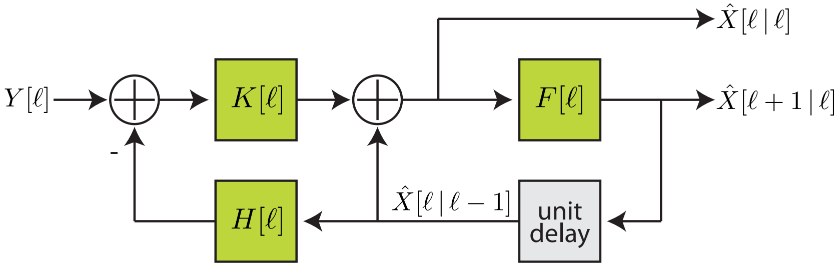
Initialization (predictions):

$$\begin{aligned}\hat{X}[0\,|\,-1] &= m[0] \\ \Sigma[0\,|\,-1] &= \Sigma[0]\end{aligned}$$

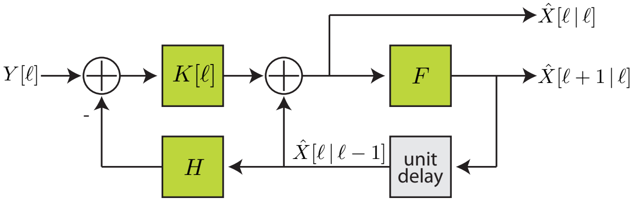
Recursion, beginning with $\ell = 0$:

$$\begin{aligned}K[\ell] &= \Sigma[\ell\,|\,\ell-1]H^\top[\ell]\left(H[\ell]\Sigma[\ell\,|\,\ell-1]H^\top[\ell] + R[\ell]\right)^{-1} \\ \hat{X}[\ell\,|\,\ell] &= \hat{X}[\ell\,|\,\ell-1] + K[\ell]\left(Y[\ell] - H[\ell]\hat{X}[\ell\,|\,\ell-1]\right) \\ \Sigma[\ell\,|\,\ell] &= \Sigma[\ell\,|\,\ell-1] - K[\ell]H[\ell]\Sigma[\ell\,|\,\ell-1] \\ \hat{X}[\ell+1\,|\,\ell] &= F[\ell]\hat{X}[\ell\,|\,\ell] \\ \Sigma[\ell+1\,|\,\ell] &= F[\ell]\Sigma[\ell\,|\,\ell]F[\ell]^\top + G[\ell]Q[\ell]G^\top[\ell]\end{aligned}$$

The Kalman Filter



The Kalman Filter: Time-Invariant Case



$$
\begin{aligned}
K[\ell] &= \Sigma[\ell \,|\, \ell-1]H^\top \left(H\Sigma[\ell \,|\, \ell-1]H^\top + R\right)^{-1} \\
\hat{X}[\ell \,|\, \ell] &= \hat{X}[\ell \,|\, \ell-1] + K[\ell]\left(Y[\ell] - H\hat{X}[\ell \,|\, \ell-1]\right) \\
\Sigma[\ell \,|\, \ell] &= \Sigma[\ell \,|\, \ell-1] - K[\ell]H\Sigma[\ell \,|\, \ell-1] \\
\hat{X}[\ell+1 \,|\, \ell] &= F\hat{X}[\ell \,|\, \ell] \\
\Sigma[\ell+1 \,|\, \ell] &= F\Sigma[\ell \,|\, \ell]F^\top + GQG^\top
\end{aligned}
$$

Even though F, G, H, Q, and R are time-invariant, $K[\ell]$ is still time-varying.

Example: One-Dimensional Motion

Our one-dimensional motion system:

- The state $X[\ell] = \left[x[\ell], v[\ell]\right]^\top$ (position, velocity).
- The input $U[\ell] = a[\ell] \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2)$ (acceleration).
- The state update equation is

$$X[\ell+1] \quad = \quad \underbrace{\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}}_{F} X[\ell] + \underbrace{\begin{bmatrix} 0 \\ T \end{bmatrix}}_{G} U[\ell]$$

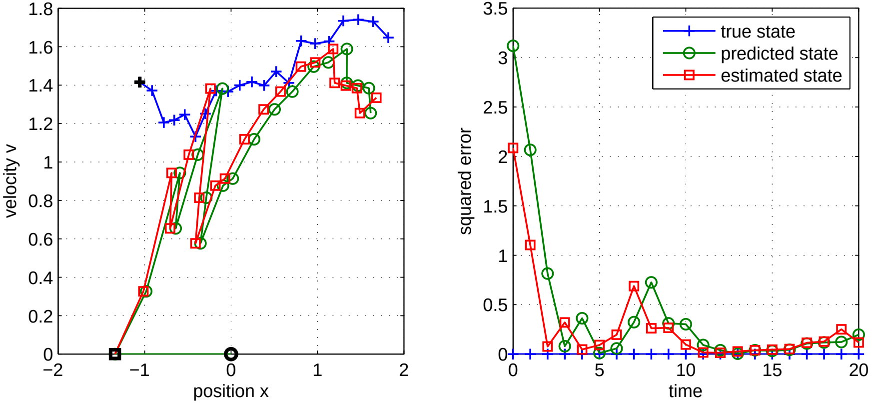
  where T is the sampling time.

- Suppose we observe the position of the particle in noise. Then

$$Y[n] \quad = \quad \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{H} X[n] + V[n]$$

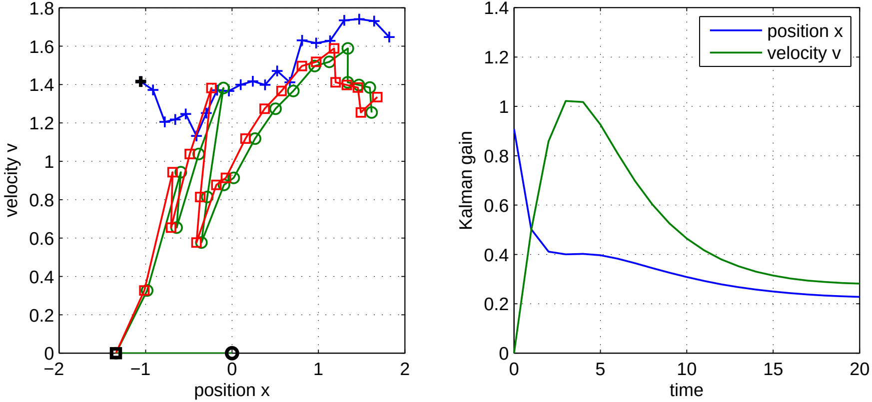
  where $V[\ell] \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_v^2)$ and is independent of $U[\ell]$.

Example: One-Dimensional Motion



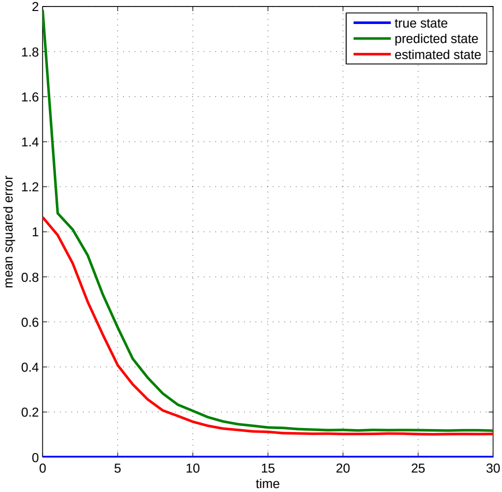
$T = 0.1$, $\sigma_a^2 = 1$, and $\sigma_v^2 = 0.1$ in this example.

Example: One-Dimensional Motion Kalman Gain



$T = 0.1$, $\sigma_a^2 = 1$, and $\sigma_v^2 = 0.1$ in this example.

Example: One-Dimensional Motion MSE



$T = 0.1$, $\sigma_a^2 = 1$, $\sigma_v^2 = 0.1$; averaged over 5000 runs.

Computational Requirements: Batch MMSE Estimation

Suppose we've received observations $Y[0],\dots,Y[\ell]$ and we wish to estimate the state $X[\ell]$ from these observations. If we just did this as a batch operation, what is the MMSE estimate of $X[\ell]$?

$$\begin{aligned}\hat{X}[\ell\,|\,\ell] &= \mathrm{E}\left\{X[\ell]\,|\,Y[0],\dots,Y[\ell]\right\}\\ &= \mathrm{E}\{X[\ell]\} + \mathrm{cov}\{X[\ell], Y_0^\ell\}\left[\mathrm{cov}\{Y_0^\ell, Y_0^\ell\}\right]^{-1}\left(Y_0^\ell - \mathrm{E}\{Y_0^L\}\right)\end{aligned}$$

where $Y_0^\ell := \left[Y^\top[0],\dots,Y^\top[\ell]\right]^\top$.

Recall that, for each ℓ, $Y[\ell] \in \mathbb{R}^k$. What are the dimensions of the matrix inverse that we have to compute here?

Computational Requirements: Kalman Filter

Note that the Kalman filter recursion also has a matrix inverse.

$$K[\ell] = \Sigma[\ell \,|\, \ell-1] H^\top[\ell] \left(H[\ell]\Sigma[\ell \,|\, \ell-1] H^\top[\ell] + R[\ell] \right)^{-1}$$

What are the dimensions of the matrix inverse that we have to compute here?

Computational Requirements: Kalman Filter

Note that the error covariance matrix and Kalman gain updates

$$\begin{aligned} K[\ell] &= \Sigma[\ell\,|\,\ell-1]H^\top[\ell]\left(H[\ell]\Sigma[\ell\,|\,\ell-1]H^\top[\ell]+R[\ell]\right)^{-1} \\ \Sigma[\ell\,|\,\ell] &= \Sigma[\ell\,|\,\ell-1]-K[\ell]H[\ell]\Sigma[\ell\,|\,\ell-1] \\ \Sigma[\ell+1\,|\,\ell] &= F[\ell]\Sigma[\ell\,|\,\ell]F[\ell]^\top+G[\ell]Q[\ell]G^\top[\ell] \end{aligned}$$

do not depend on the observation. This means that, if $F[\ell]$, $G[\ell]$, $H[\ell]$, $Q[\ell]$, and $R[\ell]$ are known in advance, e.g. they are time invariant, the error covariance matrices (prediction and estimation) as well as the Kalman gain can be **computed in advance**.

Pre-computation of $K[\ell]$, $\Sigma[\ell\,|\,\ell]$, and $\Sigma[\ell+1\,|\,\ell]$ makes the **real-time** computational requirements of the Kalman filter very modest:

- State estimate: One matrix-vector product and one vector addition.
- State prediction: One matrix-vector product.

Static State Estimation with the Kalman Filter

Consider the special case $F[\ell] \equiv I$ and $G[\ell] \equiv 0$. In this case, we have a static parameter estimation problem since the state does not change over time. It should be clear that $X[\ell] \equiv X[0] = X$.

What does the Kalman filter do in this scenario? Let's look at the recursion replacing $F[\ell] \equiv I$ and $G[\ell] \equiv 0$ (for simplicity, we also assume that H and R are time-invariant).

$$
\begin{aligned}
K[\ell] &= \Sigma[\ell \,|\, \ell-1]H^\top \left(H\Sigma[\ell \,|\, \ell-1]H^\top + R\right)^{-1} \\
\hat{X}[\ell \,|\, \ell] &= \hat{X}[\ell \,|\, \ell-1] + K[\ell]\left(Y[\ell] - H\hat{X}[\ell \,|\, \ell-1]\right) \\
\Sigma[\ell \,|\, \ell] &= \Sigma[\ell \,|\, \ell-1] - K[\ell]H\Sigma[\ell \,|\, \ell-1] \\
\hat{X}[\ell+1 \,|\, \ell] &= \hat{X}[\ell \,|\, \ell] \\
\Sigma[\ell+1 \,|\, \ell] &= \Sigma[\ell \,|\, \ell]
\end{aligned}
$$

Both predictions are equal to the last estimates (this should make sense).

Static State Estimation with the Kalman Filter

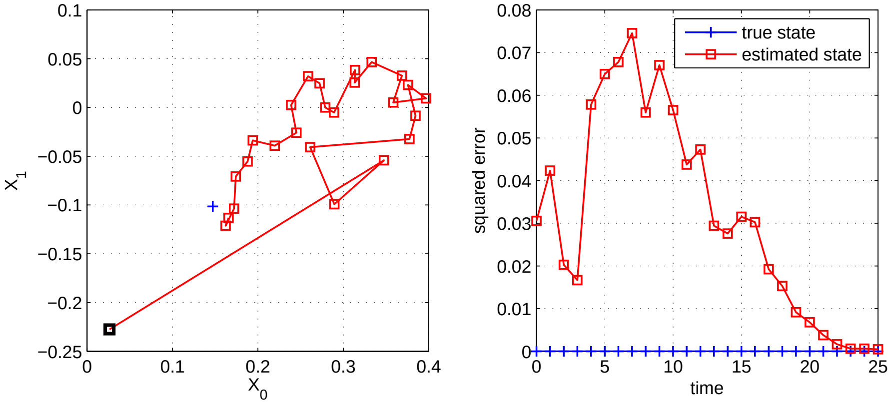
Since the predictions aren't necessary anymore, the notation and the recursion can be simplified to

$$\begin{aligned} K[\ell] &= \Sigma[\ell-1]H^\top \left(H\Sigma[\ell-1]H^\top + R\right)^{-1} \\ \hat{X}[\ell] &= \hat{X}[\ell-1] + K[\ell]\left(Y[\ell] - H\hat{X}[\ell-1]\right) \\ \Sigma[\ell] &= \Sigma[\ell-1] - K[\ell]H\Sigma[\ell-1] \end{aligned}$$

What is the Kalman filter doing here?

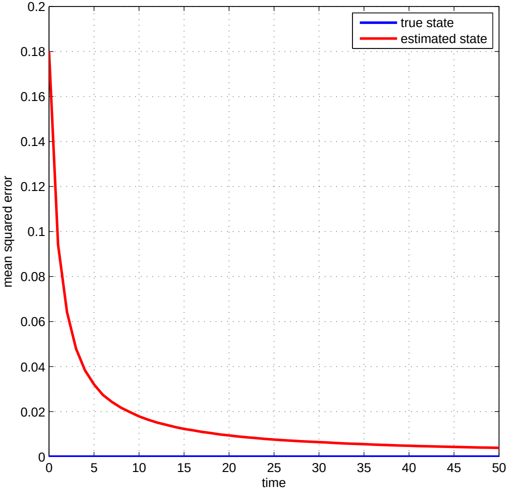
It is performing **sequential MMSE estimation** of the static parameter X. That is, given observations $Y[0], Y[1], \dots$, the Kalman filter is a computationally efficient way to sequentially refine the MMSE estimate of an unknown static parameter (state) with each new observation.

Example: Sequential MMSE Estimation of Static Param.



$\sigma_v^2 = 0.1$ and $H = I$ in this example.

Example: Sequential MMSE Estimation of Static Param.



$\sigma_v^2 = 0.1$ and $H = I$; averaged over 5000 runs.

Extension 1: Non-White Input Sequence

Suppose the input sequence to our discrete-time dynamical system model was zero-mean and wide-sense stationary, but not white, i.e.:

$$\mathrm{cov}\left\{U[k], U[\ell]\right\} = Q[k-\ell]$$

This causes the state of the system to no longer be a Markov sequence. A lot of our earlier analysis is not going to be valid anymore. What can we do?

The basic idea is to write an equivalent discrete-time dynamical system model with a white input signal and an augmented state. Then the usual Kalman filter can be applied.

Non-White Input Sequence Example

Example: Suppose we have our one-dimensional motion model again but this time the Gussian acceleration input has autocorrelation

$$Q[k-\ell] = \text{cov}\{a[k], a[\ell]\} = \sigma_a^2 \alpha^{|k-\ell|}$$

for $\alpha \in (0,1)$. Note that this covariance can be achieved by the difference equation

$$a[\ell+1] = \alpha a[\ell] + W[\ell]$$

where $W[\ell]$ is a zero-mean stationary white Gaussian sequence with $\text{var}\{W[\ell]\} = \sigma_w^2 = (1-\alpha^2)\sigma_a^2$.

The new system has input $U[\ell] = W[\ell]$ and state $X[\ell] = [x[\ell], v[\ell], a[\ell]]^\top$. What do the matrices F and G look like?

Extension 2: Deterministic Input Component

Sometimes the state update equation for the discrete-time dynamical model is specified as

$$X[\ell+1] \;=\; F[\ell]X[\ell] + G[\ell]U[\ell] + s[\ell]$$

where $s[\ell]$ is a known sequence. How does this affect the Kalman Filter recursion?

- Intuitively, the presence of a known sequence in the state update equation shouldn't degrade the performance of the MMSE estimator.
- It is not too difficult to show that the only change to the Kalman filter recursion is in the prediction step:

$$\hat{X}[\ell+1\,|\,\ell] \;=\; F[\ell]\hat{X}[\ell\,|\,\ell] + s[\ell]$$

- The error covariance matrix of the prediction is unchanged. Why?
- Hence the expressions for the Kalman gain, the state estimate, and the error covariance matrix of the estimate remain the same.
- The details are left as an exercise (Poor Chapter V.2).

Extension 3: Correlated Input and Measurement Noise

Suppose now that the current input to the dynamical system and the current measurement noise are correlated, i.e.:

$$\mathrm{cov}\{U[k], V[\ell]\} = \begin{cases} C & k = \ell \\ 0 & \text{otherwise} \end{cases}$$

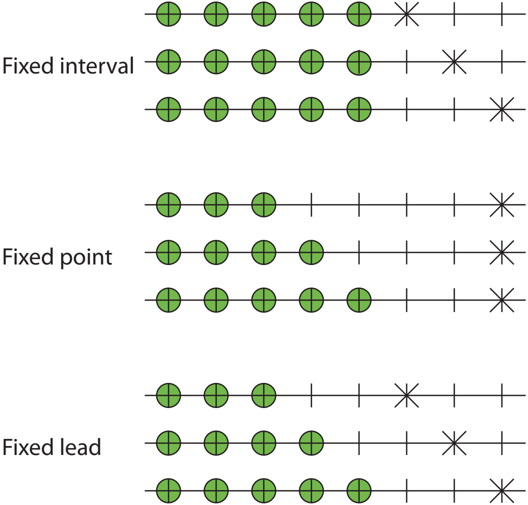
Again, a lot of our earlier analysis is not going to be valid anymore. What can we do?

The basic idea here is transform the dynamical system to an equivalent dynamical system with uncorrelated input and measurement noise sequences. After deriving the new white input sequence ($\tilde{U}[k]$, uncorrelated with $V[k]$ and with covariance $Q - CR^{-1}C^\top$), you will end up with a new state update equation

$$X[\ell+1] = (F - CR^{-1}H)X[\ell] + G\tilde{U}[k] + CR^{-1}s[k]$$

where $CR^{-1}s[k]$ is a known input (see Extension 2). The details are left as an exercise (Poor Chapter V.4).

Types of State Prediction



Fixed Interval Prediction

Given observations $Y[0], \dots, Y[L-1]$, we want to predict the states $X[L+j]$ for $j = 0, 1, \dots$ in a computationally efficient manner. Earlier, we derived a general expression for MMSE prediction of the state $X[n+1]$ given observations $Y[0], \dots, Y[\ell]$:

$$\hat{X}[n+1 \,|\, \ell] \quad = \quad \mathcal{F}_n^{\ell} \hat{X}[\ell \,|\, \ell] \tag{1}$$

where

$$\mathcal{F}_n^t \quad := \quad \begin{cases} F[n]F[n-1]\cdots F[t] & t \leq n \\ I & t > n. \end{cases}$$

This result can be written recursively for our problem as

$$\begin{aligned} \hat{X}[L+j \,|\, L-1] \quad &= \quad \mathcal{F}_{L+j-1}^{L-1} \hat{X}[L-1 \,|\, L-1] \\ &= \quad F[L+j-1]\hat{X}[L+j-1 \,|\, L-1] \end{aligned}$$

The associated error covariance matrix recursion can be computed similarly using our earlier results.

Fixed Point Prediction

Given observations $Y[0],\dots,Y[\ell]$, we want to predict the state $X[L]$ in a computationally efficient manner for $\ell = 0,\dots,L-1$.

We can use the standard Kalman recursion to rewrite (1) as

$$\begin{aligned}
\hat{X}[L\,|\,\ell] &= \mathcal{F}^{\ell}_{L-1}\hat{X}[\ell\,|\,\ell] \\
&= \mathcal{F}^{\ell}_{L-1}\left\{\hat{X}[\ell\,|\,\ell-1] + K[\ell]\tilde{Y}[\ell\,|\,\ell-1]\right\} \\
&= \mathcal{F}^{\ell}_{L-1}\left\{F[\ell-1]\hat{X}[\ell-1\,|\,\ell-1] + K[\ell]\tilde{Y}[\ell\,|\,\ell-1]\right\} \\
&= \mathcal{F}^{\ell-1}_{L-1}\hat{X}[\ell-1\,|\,\ell-1] + \mathcal{F}^{\ell}_{L-1}K[\ell]\tilde{Y}[\ell\,|\,\ell-1] \\
&= \hat{X}[L\,|\,\ell-1] + \mathcal{F}^{\ell}_{L-1}K[\ell]\tilde{Y}[\ell\,|\,\ell-1]
\end{aligned}$$

which gives a recursion for $\hat{X}[n+1\,|\,\ell]$. Note that it is assumed that $K[\ell]$ and the innovations $\tilde{Y}[\ell\,|\,\ell-1]$ were computed as part of the standard Kalman filter recursion.

A similar procedure can be followed to derive the recursion for the ECM.

Fixed Lead Prediction

Given observations $Y[0], \dots, Y[\ell]$, we want to estimate the state $X[\ell + L]$ in a computationally efficient manner for $\ell = 0, 1, \dots$. For a fixed lead of L samples, the recursion can be derived as

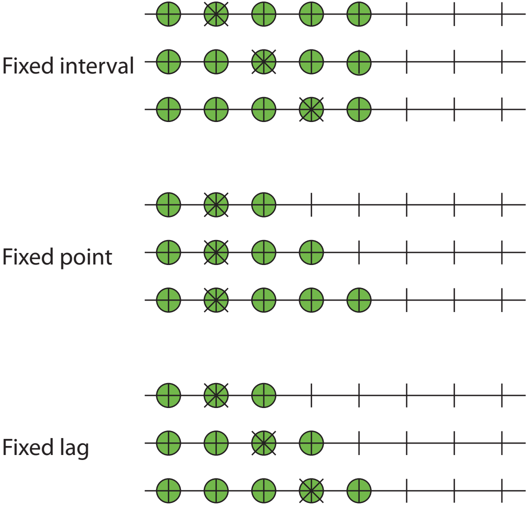
$$\hat{X}[\ell + L \,|\, \ell] \;\; = \;\; F[\ell + L - 1]\hat{X}[\ell + L - 1 \,|\, \ell - 1] + \mathcal{F}^{\ell}_{\ell+L-1} K[\ell]\tilde{Y}[\ell \,|\, \ell - 1]$$

The error covariance matrix recursion can be computed by propagating the ECM from time $\ell + 1$ to time $\ell + L$ using the standard Kalman filter ECM recursion

$$\begin{aligned}\Sigma[\ell + k \,|\, \ell] \;\; &= \;\; F[\ell + k - 1]\Sigma[\ell + k - 1 \,|\, \ell]F^{\top}[\ell + k - 1] \\ &\quad + G[\ell + k - 1]Q[\ell + k - 1]G^{\top}[\ell + k - 1]\end{aligned}$$

starting at $k = 1$ and recursing up to $k = L$.

Types of State Smoothing



Fixed Interval Smoothing

- Given observations $Y[0],\dots,Y[L-1]$, we want to estimate (smooth) the state $X[\ell]$ for $\ell < L-1$ in a computationally efficient manner.
- We assume that the standard Kalman estimates $\hat{X}[\ell\,|\,\ell]$ and predictions $\hat{X}[\ell\,|\,\ell-1]$ and associated error covariance matrices for $\ell = 0,1,\dots,L-1$ have already been computed, but now we want to go back and smooth our earlier estimates in light of the full set of observations $Y[0],\dots,Y[L-1]$.
- The (backward) recursion then is to specify $\hat{X}[\ell\,|\,L-1]$ in terms of $\hat{X}[\ell+1\,|\,L-1]$ for $\ell = L-2, L-3, \dots, 0$.
- We will also need a backward recursion for the error covariance matrix.

The main idea behind the derivation:

1. Set up the joint pdf of the states $X[\ell]$ and $X[\ell+1]$, conditioned on the observations $Y[0],\dots,Y[L-1]$. This conditional PDF will, of course, be Gaussian.
2. Write an expression for the conditional mean at time ℓ in terms of the conditional mean at time $\ell+1$ (backward recursion).

Fixed Interval Smoothing: The Main Results

The backward state recursion can be derived as

$$\hat{X}[\ell \,|\, L-1] \;=\; \hat{X}[\ell \,|\, \ell] + C[\ell] \underbrace{\left\{\hat{X}[\ell+1 \,|\, L-1] - \hat{X}[\ell+1 \,|\, \ell]\right\}}_{\text{smoothing residual}}$$

where

$$C[\ell] \;=\; \Sigma[\ell \,|\, \ell] F^\top[\ell] \Sigma^{-1}[\ell+1 \,|\, \ell]$$

is the "smoother gain". The backward recursion for the smoothed state error covariance matrix can also be derived as

$$\Sigma[\ell \,|\, L-1] \;=\; \Sigma[\ell \,|\ell] + C[\ell] \left\{\Sigma[\ell+1 \,|\, L-1] - \Sigma[\ell+1 \,|\, \ell]\right\} C^\top[\ell].$$

The backward recursions for other types of smoothing can also be derived similarly.

Conclusions

- The discrete-time Kalman-Bucy filter is often called the "Workhorse of Estimation".
- Computationally efficient and no loss of optimality in the linear Gaussian case.
- We will see that the discrete-time Kalman-Bucy filter is also optimum in dynamic (or static) MMSE parameter estimation problems among the class of linear MMSE estimators even if the noise is non-Gaussian.
- Even more extensions:
  - Continuous-time Kalman-Bucy filter.
  - Extended Kalman filter (nonlinear but differentiable state update, input, and output functions).
  - Unscented Kalman filter (also allows for nonlinear functions, better convergence properties than the EKF in some cases).
  - etc.