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

D. Richard Brown III

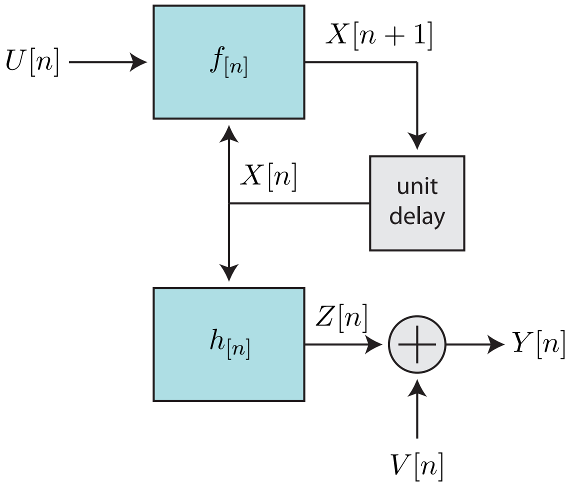
Worcester Polytechnic Institute

13-Apr-2011

Introduction

- So far, we have only considered estimation problems with fixed parameters, e.g.
  - $Y_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$.
  - $Y_k = a\cos(\omega k + \phi) + W_k$ for $\theta = [a, \phi, \omega]^\top$.
  - $Y_k \stackrel{\text{i.i.d.}}{\sim} \theta e^{-\theta y_k}$.
- Many problems require us to estimate **dynamic** or time-varying parameters, e.g.
  - Radar (position and velocity of target changing over time)
  - Communications (amplitude and phase of signal changing over time)
  - Stock market (price of shares next week changing over time)
- Step 1: We need to develop a general **dynamic model** (ECE504) for
  - How time-varying parameters evolve over time and
  - How observations are generated from the parameter state.
- Step 2: We need to develop good techniques for estimating dynamic parameters. These techniques should leverage some knowledge of the **dynamic model**.

Discrete Time Model for Dynamic Parameters



The time index is denoted as $n = 0, 1, \ldots$. All vectors are considered to be random unless otherwise specified.

$$
\begin{aligned}
U[n] &\in \mathbb{R}^s \\
X[n] &\in \mathbb{R}^m \\
Z[n] &\in \mathbb{R}^k \\
V[n] &\in \mathbb{R}^k \\
Y[n] &\in \mathbb{R}^k \\
f_{[n]} &: \mathbb{R}^s \times \mathbb{R}^m \mapsto \mathbb{R}^m \\
h_{[n]} &: \mathbb{R}^m \mapsto \mathbb{R}^k
\end{aligned}
$$

Notation and Terminology

- $U[n]$ is the dynamical system "input", $n = 0, 1, \ldots$. This input is usually random and is also called the "process noise".
- $X[n]$ is the "state" of the dynamical system, $n = 0, 1, \ldots$. This is what we want to estimate.
- $Z[n]$ is the dynamical system "output", $n = 0, 1, \ldots$.
- $f_{[n]}$ is a time varying function that updates the state based on the current state and the current input, i.e.

$$X[n+1] \quad = \quad f_{[n]}(X[n], U[n])$$

- $h_{[n]}$ is a time varying function that generates the current output based on the current state, i.e.

$$Z[n] \quad = \quad h_{[n]}(X[n])$$

- $V[n]$ is the "measurement noise" $n = 0, 1, \ldots$.
- $Y[n]$ is the "observation" $n = 0, 1, \ldots$.

Example: One-Dimensional Motion

Suppose we have a particle moving on a line with position x and velocity v updated according to

$$\begin{aligned} x[n+1] &= x[n] + Tv[n] \\ v[n+1] &= v[n] + Ta[n] \end{aligned}$$

where $a[n]$ represents the piecewise constant acceleration and T is the time between samples. The implicit assumption here is that T is small enough such that the piecewise constant approximation holds.

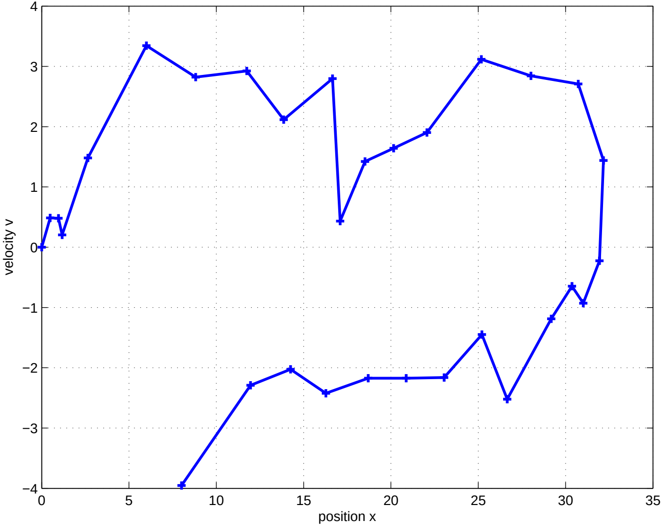
- The state $X = [x, v]^\top$.
- The input $U = a$.
- The state update equation can be written as

$$X[n+1] = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} X[n] + \begin{bmatrix} 0 \\ T \end{bmatrix} U[n]$$

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

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

Example: One-Dimensional Motion (White Gaussian Input)



Remarks

- In these types of dynamical systems, $X[k]$ is completely determined by the earlier state $X[\ell]$, $l < k$, and the inputs $\{U[\ell], \dots, U[k-1]\}$.
- The state at time ℓ completely summarizes the system in the sense that you don't need to know the details of what happened prior to time ℓ if you know $X[\ell]$.
- We are going to study problems in which we wish to estimate the dynamic state $X[\ell]$ given a sequence of observations $Y[0], \dots, Y[k]$. These problems can be categorized into three types:
  1. **Filtering**: $\ell = k$ (estimate the current state)
  2. **Prediction**: $\ell > k$ (estimate a future state)
  3. **Smoothing**: $\ell < k$ (estimate a previous state)
- Note that the notation in dynamic parameter estimation problems is different than static parameter estimation problems: We will use $X[\ell]$ to represent the quantity we wish to estimate (rather than $\theta[\ell]$).

Restriction 1: Squared Error Cost Assignment

We will only consider the squared error cost assignment, i.e.

$$\text{MSE} = \mathrm{E}\left\{\|\hat{X}[\ell] - X[\ell]\|_2^2\right\}$$

where $\hat{X}[\ell]$ is the estimate of the state $X[\ell]$.

We also assume that we know:

- the joint distribution of the inputs $U[0], U[1], \dots, U[\ell-1]$ and
- the distribution of the initial state $X[0]$.

Given the observation $Y[0], \dots, Y[k]$, what estimator $\hat{X}[\ell]$ minimizes the MSE?

Hint: Is this Bayesian estimation or non-random parameter estimation?

Restriction 2: Linear Dynamical Model

We are going to restrict our attention to systems with state update equations and output equations of the form

$$\begin{aligned} X[n+1] &= F[n]X[n] + G[n]U[n] \qquad n = 0, 1, \dots \\ Y[n] &= H[n]X[n] + V[n] \qquad n = 0, 1, \dots \end{aligned}$$

where, for each n, $F[n] \in \mathbb{R}^{m \times m}$, $G[n] \in \mathbb{R}^{m \times s}$, and $H[n] \in \mathbb{R}^{k \times m}$.

- We've already seen that one-dimensional motion fits within this linear model.
- The same is true for two- and three-dimensional motion and lots of other real-world dynamic systems.
- Many nonlinear systems can approximately fit in this model by linearizing f and h around a nominal state (Taylor series expansion).

Linear Dynamical Model Review

State update:

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

Repeating this process leads to an expression for the state at time $n+1$ in terms of the initital state $X[0]$ and the inputs $U[0], U[1], \dots$:

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

where

$$\prod_{k=0}^{t} F[n-k] := F[n]F[n-1]F[n-2]\cdots F[n-t]$$

is called the **state transition matrix** from time $n-t$ to time $n+1$.

Linear Dynamical Model Review: Time Invariant Case

When

$$
\begin{aligned}
F[n] &\equiv F \\
G[n] &\equiv G
\end{aligned}
$$

then the discrete time dynamical system is time invariant and the state at time $n+1$ is simply

$$
X[n+1] = F^{n+1}X[0] + \sum_{j=0}^{n} F^{n-j}G[j]U[j]
$$

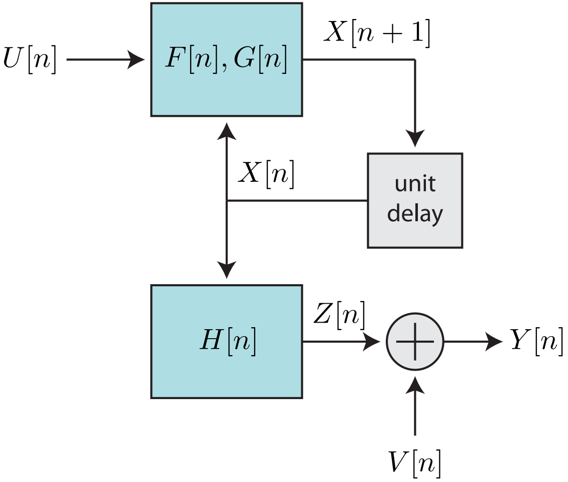
Restriction 3: Gaussian Input and Measurement Noise

To develop the main result, we will only consider systems in which the input (process noise) sequence $U[0], U[1], \dots$ and the measurement noise sequence $V[0], V[1], \dots$ are independent sequences of independent zero mean Gaussian random vectors, i.e.

$$
\begin{aligned}
\mathrm{E}\{U[k]\} &= 0 \\
\mathrm{E}\{U[k]U^\top[j]\} &= \begin{cases} Q[k] & k = j \\ 0 & \text{otherwise} \end{cases} \\
\mathrm{E}\{V[k]\} &= 0 \\
\mathrm{E}\{V[k]V^\top[j]\} &= \begin{cases} R[k] & k = j \\ 0 & \text{otherwise} \end{cases} \\
\mathrm{E}\{U[k]V^\top[j]\} &= 0 \qquad \text{for all } j \text{ and } k
\end{aligned}
$$

We also assume that the initial state $X[0] \sim \mathcal{N}(m[0], \Sigma[0])$ is a Gaussian random vector independent of $U[0], U[1], \dots$ and $V[0], V[1], \dots$.

Dynamic MMSE State Estimation (Filtering)



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 the conditional mean

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

where

$$\mathcal{Y}_0^\ell := \left[Y^\top[0], \dots, Y^\top[\ell]\right]^\top$$

is the "super vector" of all observations up to and including time ℓ.

The Brute Force Approach: Batch MMSE Estimation

Suppose we've received observations $\mathcal{Y}_0^\ell := \left[Y^\top[0], \dots, Y^\top[\ell]\right]^\top$ 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 $X[\ell]$?

Recall the conditional mean of jointly Gaussian random vectors X and Y

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

We can use this result to write

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

Recall that each $Y[n] \in \mathbb{R}^k$.

- What are the dimensions of the matrix inverse that we have to compute here?
- What happens as we get more observations?

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

Kalman's big idea was to find a computationally efficient recursion for MMSE state estimation. To see how this works, we need to approach the MMSE estimation problem from the beginning.

THE BEGINNING: We want to estimate the initial state $X[0]$ before we have any observations. What is the best estimator of the initial state?

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. 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

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

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.

We use our prior result for jointly Gaussian random vectors with

$$\mathrm{E}[X] = m[0] \qquad \mathrm{E}[Y] = H[0]m[0]$$
$$\mathrm{cov}(X,Y) = \Sigma[0]H^\top[0] \qquad \mathrm{cov}(Y,Y) = H[0]\Sigma[0]H^\top[0] + R[0]$$

to write an expression for the MMSE estimate of $X[0]$ given $Y[0]$ as

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

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

If we define the **Kalman Gain** as

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

we can write

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

Remarks:

- The term $\tilde{Y}[0 \mid -1] := Y[0] - H[0]\hat{X}[0 \mid -1]$ 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

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

Kalman Gain: Intuition

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?

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

MMSE State Estimate $\hat{X}[0]$ given $Y[0]$ (part 3)

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

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

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 Prediction

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

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

One-step MMSE state predictor:

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

ECM of MMSE one-step state predictor:

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

We have shown the following results only for the case $\ell = 0$:
MMSE state estimator:

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

ECM of MMSE state estimator:

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

We still need to show that these expressions hold for general $\ell = 0, 1, \dots$.

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 $Y[0], \ldots, Y[\ell]$.

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

$$
\begin{aligned}
\hat{X}[\ell+1\,|\,\ell+1] &:= \mathrm{E}\left\{X[\ell+1]\,|\,\mathcal{Y}_0^{\ell+1}\right\} \\
&= \mathrm{E}\left\{X[\ell+1]\,|\,\mathcal{Y}_0^{\ell}, Y[\ell+1]\right\} \\
&= \mathrm{E}\left\{X[\ell+1]\,|\,\mathcal{Y}_0^{\ell}, \tilde{Y}[\ell+1\,|\,\ell]\right\} \\
&\stackrel{\text{ur0}}{=} \mathrm{E}\left\{X[\ell+1]\,|\,\mathcal{Y}_0^{\ell}\right\} + \mathrm{cov}\left(X[\ell+1], \tilde{Y}[\ell+1\,|\,\ell]\right) \times \\
&\quad \left[\mathrm{cov}\left(\tilde{Y}[\ell+1\,|\,\ell], \tilde{Y}[\ell+1\,|\,\ell]\right)\right]^{-1} \tilde{Y}[\ell+1\,|\,\ell] \\
&= \hat{X}[\ell+1\,|\,\ell] + \mathrm{cov}\left(X[\ell+1], \tilde{Y}[\ell+1\,|\,\ell]\right) \times \\
&\quad \left[\mathrm{cov}\left(\tilde{Y}[\ell+1\,|\,\ell], \tilde{Y}[\ell+1\,|\,\ell]\right)\right]^{-1} \tilde{Y}[\ell+1\,|\,\ell]
\end{aligned}
$$

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

Induction: Conditional Mean

So we have

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

The Kalman Gain at time $\ell+1$ is

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

Also, 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

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

Useful Result Number 2

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

The four expectations that need to be computed in $\Sigma[\ell+1\,|\,\ell+1]$:

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

- 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 previously, the optimal estimates for the current state (filtering) and the next state (prediction) are given recursively as

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

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 $\Sigma[\ell\,|\,\ell-1] := \mathrm{cov}\left\{X[\ell]\,|\,Y[0],\ldots,Y[\ell-1]\right\}$ *and* $R[\ell] := \mathrm{cov}\left\{V[\ell]\right\}$.

Kalman Filter: Summary of General Recursion

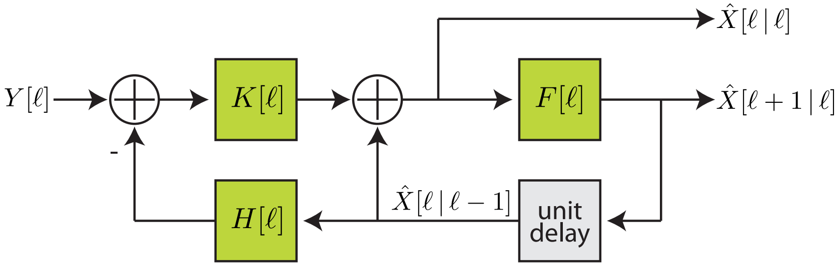
Initialization (predictions):

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

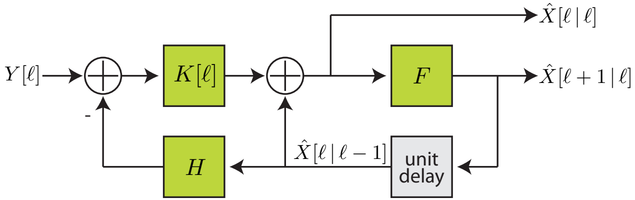
Recursion, beginning with $\ell = 0$:

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

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 $U[\ell] = a[\ell] \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2)$ (acceleration).
- The state update equation is

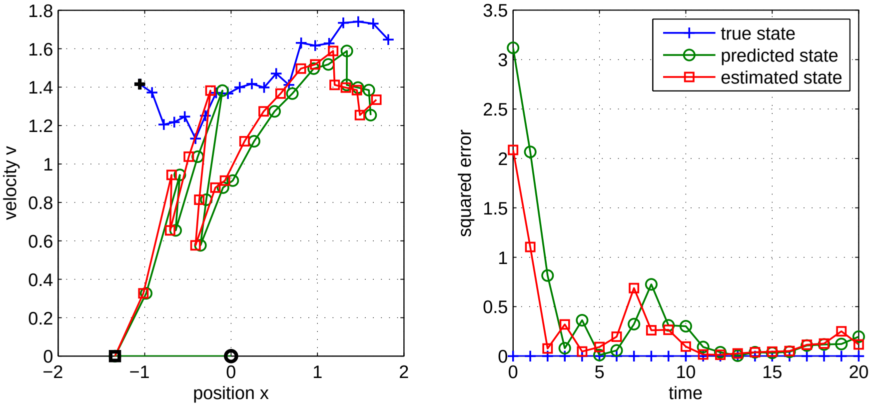
$$X[\ell+1] = \underbrace{\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}}_{F} X[\ell] + \underbrace{\begin{bmatrix} 0 \\ T \end{bmatrix}}_{G} U[\ell]$$

  where T is the sampling time.
- Suppose we observe the position of the particle in noise. Then

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

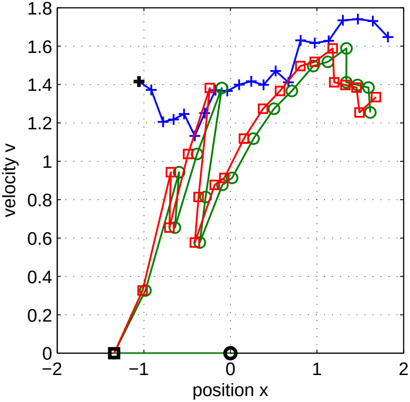
  where $V[\ell] \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_v^2)$ and is independent of $U[\ell]$.

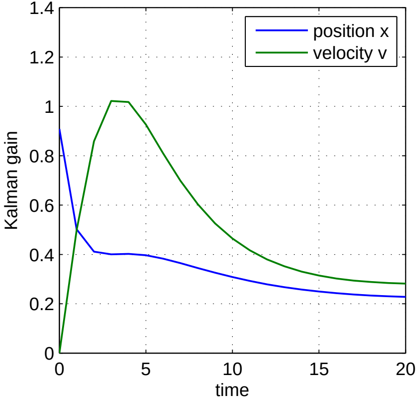
Example: One-Dimensional Motion



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

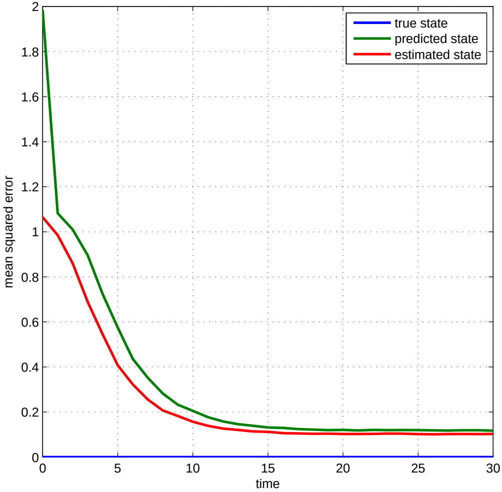
Example: One-Dimensional Motion Kalman Gain





$T = 0.1$, $\sigma_a^2 = 1$, 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: 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? What happens as more observations arrive?

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

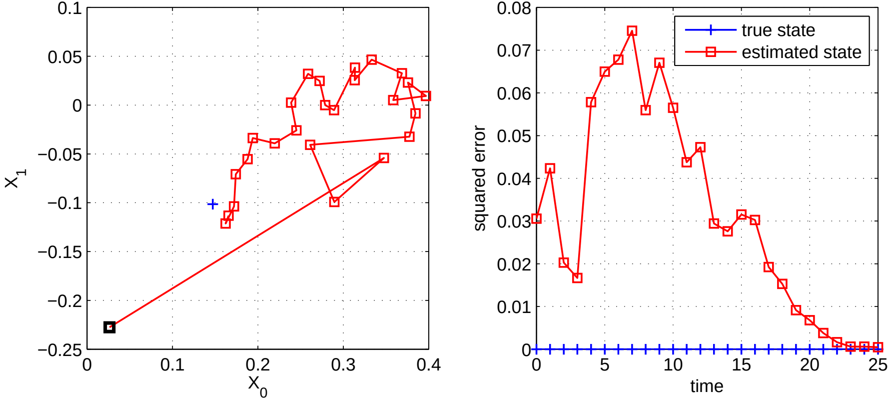
Since the predictions aren't necessary anymore, the notation and the recursion can be simplified to

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

What is the Kalman filter doing here?

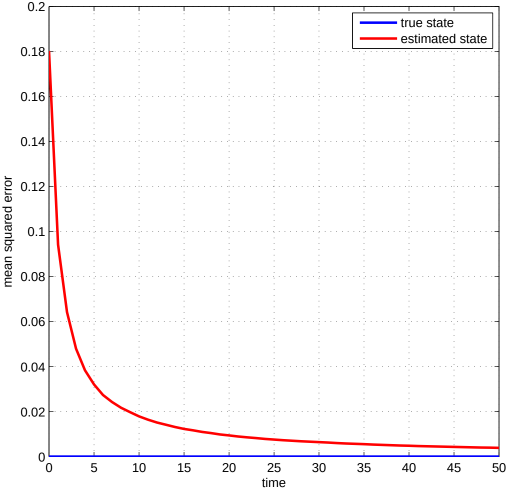
It is performing **sequential MMSE estimation** of the static parameter X. That is, given observations $Y[0], Y[1], \dots$, the Kalman filter is a computationally efficient way to sequentially refine the MMSE estimate of an unknown static parameter (state) with each new observation as they arrive (rather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

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.

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.
- 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 not covered here:
  - Prediction and smoothing.
  - 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.