Lecture 14: Discrete-Time Kalman Filter

This lecture will be our first exposure to a *state estimation* algorithm called the Kalman filter (which comes in many variants). The purpose of the Kalman filter is to infer the state of a dynamical system using the system's noisy outputs. That is, a Kalman filter can improve the estimate of a state that is directly measured by a noisy sensor and it can also infer unmeasured states. Much like the Luenberg observer, the Kalman filter uses the dynamics and measurement model of the system to perform this inference; however, the Kalman filter also considers the noise properties of the system (process and measurement noise) and is an optimal filter.

Historical Context. In the introduction to his book "Advanced Kalman Filtering", Bruce Gibbs describes four reasons for the growth in the field of state estimation in the decades following World War II: (1) the development of new radar, sonar, and communication systems led to an expanded interest in signal processing theory, (2) the development of digital computers allowed implementation of advanced algorithms, (3) the start of space exploration required state-space based approaches to estimation and control, and (4) the seminal papers by Kalman (1960) and Kalman and Bucy (1961) provided practical estimation algorithms that could be applied to many systems. Today, Kalman filters are found in numerous aerospace and automotive applications and consumer products such as GPS receivers.

Derivation [1, Secs. 3.3, 5.1]

We will start of by describing our dynamic system in terms of the dynamic model and measurements, then we will derive the equations that describe how to propagate the mean and covariance of the state through time.

Motion and Measurement Models. Consider the linear discrete-time system of the form

$$\boldsymbol{x}_k = \boldsymbol{F}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{w}_{k-1} \tag{1}$$

where $\boldsymbol{u}_{k-1}$ is a known input and $\boldsymbol{w}_{k-1}$ is zero mean independent and identically distributed (i.i.d.) Gaussian noise with covariance $\boldsymbol{Q}_k = E[\boldsymbol{w}_k\boldsymbol{w}_k^T]$ and $E[\cdot]$ is the expectation operator. Assume the measurement equation is linear

$$\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{v}_k \tag{2}$$

where $\boldsymbol{v}_k$ is a zero mean Gaussian measurement noise with variance $\boldsymbol{R}_k = E[\boldsymbol{v}_k\boldsymbol{v}_k^T]$. We are interested in recursively estimating the state $\hat{\boldsymbol{x}}_k$ accounting for the system dynamics and the measurement model.

Motion Update: Propagating the State Expected Value and Covariance. Our goal is to find recursive expressions for the expected value of the state $\hat{\boldsymbol{x}}_k = E[\boldsymbol{x}_k]$ and its covariance $\boldsymbol{P}_k =$

$E[\{\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k\}\{\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k\}^T]$. Begin by taking the expected value of the system dynamics:

$$\begin{aligned}
\hat{\boldsymbol{x}}_k = E[\boldsymbol{x}_k] &= E[\boldsymbol{F}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{w}_{k-1}] && (3)\\
&= E[\boldsymbol{F}_{k-1}\boldsymbol{x}_{k-1}] + \underbrace{E[\boldsymbol{G}_{k-1}\boldsymbol{u}_{k-1}]}_{\text{known exactly}} + \underbrace{E[\boldsymbol{w}_{k-1}]}_{=0 \text{ (since zero mean)}} && (4)\\
&= \boldsymbol{F}_{k-1}E[\boldsymbol{x}_{k-1}] + \boldsymbol{G}_{k-1}\boldsymbol{u}_{k-1} && (5)\\
&= \boldsymbol{F}_{k-1}\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{u}_{k-1} && (6)
\end{aligned}$$

Thus, the mean state of the system at discrete-time k can be found by simply propagating the mean state at discrete time $k-1$ through the system dynamics with the known control input at $k-1$. Next, we compute the covariance by using its definition, substituting in the system dynamics, and the result above, and simplifying:

$$\begin{aligned}
\boldsymbol{P}_k &= E[\{\boldsymbol{x}_k - \hat{\boldsymbol{x}}\}\{\boldsymbol{x}_k - \hat{\boldsymbol{x}}\}^T] && (7)\\
&= E[\{(\boldsymbol{F}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{w}_{k-1}) - (\boldsymbol{F}_{k-1}\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{u}_{k-1})\}\{\cdots\}^T] && (8)\\
&= E[\{\boldsymbol{F}_{k-1}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1}) + \boldsymbol{w}_{k-1}\}\{\cdots\}^T] && (9)\\
&= E[\{\boldsymbol{F}_{k-1}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1}) + \boldsymbol{w}_{k-1}\}\{(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})^T\boldsymbol{F}_{k-1}^T + \boldsymbol{w}_{k-1}^T\}] && (10)\\
&= E[\boldsymbol{F}_{k-1}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})^T\boldsymbol{F}_{k-1}^T + \boldsymbol{w}_{k-1}\boldsymbol{w}_{k-1}^T && (11)\\
&\quad + \boldsymbol{F}_{k-1}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})\boldsymbol{w}_{k-1}^T + \boldsymbol{w}_{k-1}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})^T\boldsymbol{F}_{k-1}^T] && (12)\\
&= \boldsymbol{F}_{k-1}E[(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})^T]\boldsymbol{F}_{k-1}^T + E[\boldsymbol{w}_{k-1}\boldsymbol{w}_{k-1}^T] && (13)\\
&\quad + \boldsymbol{F}_{k-1}E[(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})\boldsymbol{w}_{k-1}^T] + E[\boldsymbol{w}_{k-1}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})^T]\boldsymbol{F}_{k-1}^T] && (14)
\end{aligned}$$

Assume the noise $\boldsymbol{w}_{k-1}$ is uncorrelated with the state $\boldsymbol{x}_{k-1}$, then $E[(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})\boldsymbol{w}_{k-1}^T] = 0$ and $E[\boldsymbol{w}_{k-1}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})^T] = 0$. Also notice that the term $E[(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1})^T]$ is simply the definition of the covariance at the $(k-1)$th timestep, $\boldsymbol{P}_{k-1}$. Thus, the above simplifies to

$$\boldsymbol{P}_k = \boldsymbol{F}_{k-1}\boldsymbol{P}_{k-1}\boldsymbol{F}_{k-1}^T + \boldsymbol{Q}_{k-1} \tag{15}$$

which has the form of a discrete time Lyapunov Equation.

Proof that the State is Gaussian. Assume the initial state $\boldsymbol{x}_0$ is Gaussian. At $k=1$, we have $\boldsymbol{x}_1 = \boldsymbol{A}_0\boldsymbol{x}_0 + \boldsymbol{B}_0\boldsymbol{u}_0 + \mathbf{w}_0$. At $k=2$

$$\boldsymbol{x}_2 = \boldsymbol{A}_1\underbrace{(\boldsymbol{A}_0\boldsymbol{x}_0 + \boldsymbol{B}_0\boldsymbol{u}_0 + \mathbf{w}_0)}_{\boldsymbol{x}_1} + \boldsymbol{B}_1\boldsymbol{u}_1 + \mathbf{w}_1 \tag{16}$$

at $k=3$

$$\boldsymbol{x}_3 = \boldsymbol{A}_2\underbrace{\boldsymbol{A}_1(\boldsymbol{A}_0\boldsymbol{x}_0 + \boldsymbol{B}_0\boldsymbol{u}_0 + \mathbf{w}_0) + \boldsymbol{B}_1\boldsymbol{u}_1 + \mathbf{w}_1}_{\boldsymbol{x}_2} + \boldsymbol{B}_2\boldsymbol{u}_2 + \mathbf{w}_2 \tag{17}$$

at an arbitrary k

$$\boldsymbol{x}_k = \boldsymbol{A}_{k,0}\boldsymbol{x}_0 + \sum_{i=0}^{k-1}(\boldsymbol{A}_{k,i+1}\boldsymbol{B}_i\boldsymbol{u}_i + \boldsymbol{A}_{k,i+1}\mathbf{w}_i) \tag{18}$$

where

$$\boldsymbol{A}_{k,i} = \begin{cases} \boldsymbol{A}_{k-1}\boldsymbol{A}_{k-2}\cdots\boldsymbol{A}_i & k > i \\ I & k = i \\ 0 & k < i \end{cases} \tag{19}$$

It is clear that $\boldsymbol{x}_k$ is a linear combination of the Gaussian random variables $\boldsymbol{x}_0$, $\boldsymbol{w}_{1:k}$ and the known input sequence $\boldsymbol{u}_{1:k}$. It follows that $\boldsymbol{x}_k$ is a Gaussian. Since we have characterized its mean and covariance we state that

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{1:k-1}) = \mathcal{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_k, \boldsymbol{P}_k) \tag{20}$$

In this context, the subscript notation denotes a set of random values starting from the initial time i.e., $\boldsymbol{y}_{1:k-1}$ is the set of all $k-1$ measurements, $\boldsymbol{y}_{1:k-1} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{k-1}\}$.

Intermediate Step: Recursive Estimator of a Constant Vector. As an intermediate step, consider the process of estimating a constant parameter vector $\boldsymbol{\theta} \in \mathbb{R}^n$ from noisy measurements in a recursive manner. That is, we update the estimate of $\boldsymbol{\theta}$ at each instant a new measurement is obtained. The linear recursive estimator is the form

$$\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{\theta} + \boldsymbol{v}_k \tag{21}$$

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{H}_k\hat{\boldsymbol{\theta}}_{k-1}) \tag{22}$$

where $\boldsymbol{\theta}$ is the true value of the constant vector being estimated and $\boldsymbol{y}_k \in \mathbb{R}^m$ is a measurement that is related to $\boldsymbol{\theta}$ (via the matrix $\boldsymbol{H}_k \in \mathbb{R}^{m\times n}$). The measurement is corrupted with zero-mean additive Gaussian measurement noise $\boldsymbol{v}_k \in \mathbb{R}^m$ with covariance $\boldsymbol{R}_k \in \mathbb{R}^{m\times m}$. The state estimator produces an estimate $\hat{\boldsymbol{\theta}}$ of the true quantity $\boldsymbol{\theta}$, by scaling the correction term or *innovation* $(\boldsymbol{y}_k - \boldsymbol{H}_k\hat{\boldsymbol{\theta}}_{k-1})$ by a factor $\boldsymbol{K}_k \in \mathbb{R}^{n\times m}$, called the *estimator gain*. At a given time step k, the error in the estimate is the random variable $\boldsymbol{\epsilon}_{\boldsymbol{\theta},k} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_k$. Let $\boldsymbol{\Sigma}_k = E[\boldsymbol{\epsilon}_{\boldsymbol{\theta},k}\boldsymbol{\epsilon}_{\theta,k}^T]$ be the covariance of this estimation error. The expected value of the estimation error is

$$\begin{aligned}
E[\boldsymbol{\epsilon}_{\boldsymbol{\theta},k}] &= E[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_k] && (23)\\
&= E[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{k-1} - \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{H}_k\hat{\boldsymbol{\theta}}_{k-1})] && (24)\\
&= E[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{k-1} - \boldsymbol{K}_k(\boldsymbol{H}_k\boldsymbol{\theta} + \boldsymbol{v}_k - \boldsymbol{H}_k\hat{\boldsymbol{\theta}}_{k-1})] && (25)\\
&= E[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{k-1} - \boldsymbol{K}_k\boldsymbol{H}_k(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{k-1}) - \boldsymbol{K}_k\boldsymbol{v}_k] && (26)\\
&= E[\boldsymbol{\epsilon}_{\boldsymbol{\theta},k-1} - \boldsymbol{K}_k\boldsymbol{H}_k\boldsymbol{\epsilon}_{\boldsymbol{\theta},k-1} - \boldsymbol{K}_k\boldsymbol{v}_k] && (27)\\
&= (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)E[\boldsymbol{\epsilon}_{\boldsymbol{\theta},k-1}] - \boldsymbol{K}_kE[\boldsymbol{v}_k] && (28)\\
&= (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)E[\boldsymbol{\epsilon}_{\boldsymbol{\theta},k-1}] && (29)
\end{aligned}$$

Suppose measurements start at $k = 0$. If the initial estimate is equal to the true state (i.e., $\hat{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}$) then $\boldsymbol{\epsilon}_{\boldsymbol{\theta},0} = 0$. Further if the measurement noise is zero mean, then regardless what future noisy measurements are obtained and regardless of the value of the estimator gain $\boldsymbol{K}_k$, all future estimates $\boldsymbol{\theta}_k$ will be *on average* equal to the true value (i.e. $E[\epsilon_{\theta,k}] = 0$ for all future k). We call this an *unbiased* estimator.

Similarly, compute the covariance:

$$\boldsymbol{\Sigma}_k = E[\boldsymbol{\epsilon}_{\theta,k}\boldsymbol{\epsilon}_{\theta,k}^T] \tag{30}$$

$$= E[((\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\epsilon}_{\theta,k-1} - \boldsymbol{K}_k\boldsymbol{v}_k)(\cdots)^T] \tag{31}$$

$$= E[(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\epsilon}_{\theta,k-1}\boldsymbol{\epsilon}_{\theta,k-1}^T(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)^T - (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\epsilon}_{\theta,k-1}\boldsymbol{v}_k^T\boldsymbol{K}_k^T \tag{32}$$

$$- \boldsymbol{K}_k\boldsymbol{v}_k\boldsymbol{\epsilon}_{\theta,k-1}^T(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)^T + \boldsymbol{K}_k\boldsymbol{v}_k\boldsymbol{v}_k^T\boldsymbol{K}_k^T] \tag{33}$$

Note that $\boldsymbol{\epsilon}_{\theta,k-1}$ (the estimation error at $k-1$) is uncorrelated with $\boldsymbol{v}_k$. So that $E[\boldsymbol{\epsilon}_{\theta,k-1}\boldsymbol{v}_k] = E[\boldsymbol{\epsilon}_{\theta,k-1}]E[\boldsymbol{v}_k] = 0$ for zero-mean noise. Then substituting into the above equation gives an expression on how to recursively update the covariance

$$\boldsymbol{\Sigma}_k = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)E[\boldsymbol{\epsilon}_{\theta,k-1}\boldsymbol{\epsilon}_{\theta,k-1}^T](\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)^T - (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)E[\boldsymbol{\epsilon}_{\theta,k-1}\boldsymbol{v}_k^T]\boldsymbol{K}_k^T \tag{34}$$

$$- \boldsymbol{K}_kE[\boldsymbol{v}_k\boldsymbol{\epsilon}_{\theta,k-1}^T](\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)^T + \boldsymbol{K}_kE[\boldsymbol{v}_k\boldsymbol{v}_k^T]\boldsymbol{K}_k^T \tag{35}$$

$$\boldsymbol{\Sigma}_k = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\Sigma}_{k-1}(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)^T + \boldsymbol{K}_k\boldsymbol{R}_k\boldsymbol{K}_k^T \tag{36}$$

Next, we must determine an optimal value for the estimator gain $\boldsymbol{K}_k$. Suppose we wish to minimize the sum of the variances of the estimation errors at time k, this cost is expressed as

$$J_k = E[(p_1 - \hat{p}_1)^2] + \cdots + E[(p_n - \hat{p}_n)^2] \tag{37}$$

$$= E[\boldsymbol{\epsilon}_{\theta,k}^T\boldsymbol{\epsilon}_{\theta,k}] \tag{38}$$

$$= E[\text{Tr}(\boldsymbol{\epsilon}_{\theta,k}\boldsymbol{\epsilon}_{\theta,k}^T)] \tag{39}$$

$$= \text{Tr}(\boldsymbol{\Sigma}_k) \tag{40}$$

Recall that the trace of a matrix is the sum of the diagonal elements (a scalar). We wish to find the gain matrix $\boldsymbol{K}_k$ that will minimize the cost J_k.

Note: Suppose $\boldsymbol{A}$ is a $m \times n$ matrix and $\boldsymbol{G}$ is a $n \times n$ matrix. Then $\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^\text{T}$ is a $m \times m$ matrix. The partial derivative of the $\text{Tr}(\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^T)$ is also a $m \times m$ matrix with each entry being the partial with respect to a_{ij} for $i, j = 1, \ldots, m$. In general,

$$\frac{\partial \text{Tr}(\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^T)}{\partial \boldsymbol{A}} = \boldsymbol{A}\boldsymbol{G}^\text{T} + \boldsymbol{A}\boldsymbol{G} \tag{41}$$

but for the special case of a symmetric matrix $\boldsymbol{G}$ (where $\boldsymbol{G} = \boldsymbol{G}^T$) the following derivative rule holds

$$\frac{\partial \text{Tr}(\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}^T)}{\partial \boldsymbol{A}} = 2\boldsymbol{A}\boldsymbol{G} \tag{42}$$

Then

$$\frac{\partial J_k}{\partial \boldsymbol{K}_k} = \frac{\partial}{\partial \boldsymbol{K}_k}\text{Tr}(\boldsymbol{\Sigma}_k) = \frac{\partial}{\partial \boldsymbol{K}_k}\text{Tr}\left((\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\Sigma}_{k-1}(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)^T + \boldsymbol{K}_k\boldsymbol{R}_k\boldsymbol{K}_k^T\right) \tag{43}$$

$$= \frac{\partial}{\partial \boldsymbol{K}_k}\text{Tr}\left((\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\Sigma}_{k-1}(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)^T\right) + \frac{\partial}{\partial \boldsymbol{K}_k}\text{Tr}\left(\boldsymbol{K}_k\boldsymbol{R}_k\boldsymbol{K}_k^T\right) \tag{44}$$

$$= 2(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\Sigma}_{k-1}\left(\frac{\partial}{\partial \boldsymbol{K}_k}(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\right) + 2\boldsymbol{K}_k\boldsymbol{R}_k \tag{45}$$

$$= 2(\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\Sigma}_{k-1}\left(-\boldsymbol{H}_k^T\right) + 2\boldsymbol{K}_k\boldsymbol{R}_k \tag{46}$$

which is a $n \times m$ matrix of partial derivatives. To find the minimizer we set the above derivative to zero and solve for the optimal gain matrix $\boldsymbol{K}_k$:

$$
\begin{aligned}
\boldsymbol{K}_k \boldsymbol{R}_k &= (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k)\boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T && (47)\\
\boldsymbol{K}_k \boldsymbol{R}_k &= \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T - \boldsymbol{K}_k \boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T && (48)\\
\boldsymbol{K}_k(\boldsymbol{R}_k + \boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T) &= \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T && (49)\\
\boldsymbol{K}_k &= \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T(\underbrace{\boldsymbol{R}_k + \boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T}_{\text{Innovation Covariance,}\boldsymbol{S}_k})^{-1} && (50)
\end{aligned}
$$

The above form of $\boldsymbol{\Sigma}_k$ in (36) holds for any gain $\boldsymbol{K}_k$. However if we use the optimal gain we have just derived then we can simplify this further. First, postmultiply the above expression by $\boldsymbol{S}_k \boldsymbol{K}_k^{\mathrm{T}}$

$$
\begin{aligned}
\boldsymbol{K}_k &= \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T \boldsymbol{S}_k^{-1} && (51)\\
\boldsymbol{K}_k \boldsymbol{S}_k \boldsymbol{K}_k^{\mathrm{T}} &= \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T \boldsymbol{S}_k^{-1}\boldsymbol{S}_k \boldsymbol{K}_k^{\mathrm{T}} && (52)\\
\boldsymbol{K}_k \boldsymbol{S}_k \boldsymbol{K}_k^{\mathrm{T}} &= \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T \boldsymbol{K}_k^{\mathrm{T}} && (53)
\end{aligned}
$$

Then expanding terms in (36)

$$
\begin{aligned}
\boldsymbol{\Sigma}_k &= (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k)\boldsymbol{\Sigma}_{k-1}(\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k)^T + \boldsymbol{K}_k \boldsymbol{R}_k \boldsymbol{K}_k^T && (54)\\
&= (\boldsymbol{\Sigma}_{k-1} - \boldsymbol{K}_k \boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1})(\boldsymbol{I} - \boldsymbol{H}_k^T \boldsymbol{K}_k^T) + \boldsymbol{K}_k \boldsymbol{R}_k \boldsymbol{K}_k^T && (55)\\
&= \boldsymbol{\Sigma}_{k-1} - \boldsymbol{K}_k \boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1} - \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T \boldsymbol{K}_k^T + \boldsymbol{K}_k \boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T \boldsymbol{K}_k^T + \boldsymbol{K}_k \boldsymbol{R}_k \boldsymbol{K}_k^T && (56)\\
&= \boldsymbol{\Sigma}_{k-1} - \boldsymbol{K}_k \boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1} - \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T \boldsymbol{K}_k^T + \boldsymbol{K}_k(\boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T + \boldsymbol{R}_k)\boldsymbol{K}_k^T && (57)\\
&= \boldsymbol{\Sigma}_{k-1} - \boldsymbol{K}_k \boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1} - \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T \boldsymbol{K}_k^T + \boldsymbol{K}_k \boldsymbol{S}_k \boldsymbol{K}_k^T && (58)
\end{aligned}
$$

Using the identity $\boldsymbol{K}_k \boldsymbol{S}_k \boldsymbol{K}^{\mathrm{T}} = \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T \boldsymbol{K}^{\mathrm{T}}$, this simplifies to

$$
\begin{aligned}
\boldsymbol{\Sigma}_k &= \boldsymbol{\Sigma}_{k-1} - \boldsymbol{K}_k \boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1} - \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T \boldsymbol{K}_k^T + \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T \boldsymbol{K}^{-1} && (59)\\
&= (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k)\boldsymbol{\Sigma}_{k-1} && (60)
\end{aligned}
$$

In summary, we have developed a recursive estimator that optimally estimes a constant parameter $\boldsymbol{\theta}$ (in a least-square error sense) given measurements

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{\theta} + \boldsymbol{v}_k \tag{61}$$

where $E[\boldsymbol{v}_k] = \mathbf{0}$ and $E[\boldsymbol{v}_k \boldsymbol{v}_k^{\mathrm{T}}] = \boldsymbol{R}_k$ for $k = 1, 2, \ldots$ using the recursive equations

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_k &= \hat{\boldsymbol{\theta}}_{k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{H}_k \hat{\boldsymbol{\theta}}_{k-1}) && (62)\\
\boldsymbol{\Sigma}_k &= (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k)\boldsymbol{\Sigma}_{k-1} && (63)
\end{aligned}
$$

where

$$\boldsymbol{K}_k = \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T(\boldsymbol{R}_k + \boldsymbol{H}_k \boldsymbol{\Sigma}_{k-1}\boldsymbol{H}_k^T)^{-1} \tag{64}$$

and an initial guess $\boldsymbol{\theta}_0$ and initial covariance $\boldsymbol{\Sigma}_0$ are given.

Measurement Update: Incorporating New Information. The measurement update step of the Kalman filter applies the recursive least-square estimator above with the following changes:

- Let (61) represent the measurement of the dynamic system (2) where the constant parameter vector $\boldsymbol{\theta}$ is replaced with the true system state, $\boldsymbol{\theta} \to \boldsymbol{x}_k$.
- Let $\hat{\boldsymbol{\theta}}_{k-1}$ in (62) represent the motion-update from the previous time-step, $\hat{\boldsymbol{\theta}}_{k-1} \to \boldsymbol{x}_{k|k-1}$ and the updated parameter estimate represent the measurement-update (posterior mean), $\hat{\boldsymbol{\theta}}_k \to \boldsymbol{x}_{k|k}$. Also $\boldsymbol{H}_k\hat{\boldsymbol{\theta}}_{k-1}$ becomes the predicted measurement given the motion update, $\boldsymbol{H}_k\hat{\boldsymbol{\theta}}_{k-1} \to \boldsymbol{H}_k\boldsymbol{x}_{k|k-1}$. The state posterior is then

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{y}_{k|k-1}) \tag{65}$$

- Similarly, let $\boldsymbol{\Sigma}_{k-1}$ in (63)–(64) represent the covariance after motion update $\boldsymbol{\Sigma}_{k-1} \to \boldsymbol{P}_{k|k-1}$ and $\boldsymbol{\Sigma}_k$ represent the covariance after the measurement update $\boldsymbol{\Sigma}_{k-1} \to \boldsymbol{P}_{k|k}$ so that the covariance measurement update equations become:

$$\boldsymbol{P}_{k|k} = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{P}_{k|k-1} \tag{66}$$
$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^T\boldsymbol{S}_k^{-1} \tag{67}$$

where $\boldsymbol{S}_k = \boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^T$.

Summary of Discrete LTI Kalman Filter Equations. Assume the $\boldsymbol{F}_k$, $\boldsymbol{G}_k$, $\boldsymbol{R}_k$ and $\boldsymbol{Q}_k$ matrices are fixed. We combine the above results for state and covariance propogation with recursive least squares estimation:

$$\text{State Prior}: \quad \hat{\boldsymbol{x}}_{k-1|k-1} \tag{68}$$
$$\text{Covariance Prior}: \quad \boldsymbol{P}_{k-1|k-1} \tag{69}$$
$$\text{Current Measurement}: \quad \boldsymbol{y}_k \tag{70}$$
$$\text{State Prediction}: \quad \hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{F}\hat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{G}\boldsymbol{u}_{\boldsymbol{k}-1} \tag{71}$$
$$\text{Covariance Prediction}: \quad \boldsymbol{P}_{k|k-1} = \boldsymbol{F}\boldsymbol{P}_{k-1|k-1}\boldsymbol{F}^T + \boldsymbol{Q} \tag{72}$$
$$\text{Measurement Prediction}: \quad \boldsymbol{y}_{k|k-1} = \boldsymbol{H}\hat{\boldsymbol{x}}_{k|k-1} \tag{73}$$
$$\text{Inovation Covariance}: \quad \boldsymbol{S}_k = \boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^T \tag{74}$$
$$\text{Kalman Gain}: \quad \boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^T\boldsymbol{S}_k^{-1} \tag{75}$$
$$\text{State Posterior}: \quad \hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{y}_{k|k-1}) \tag{76}$$
$$\text{Covariance Posterior}: \quad \boldsymbol{P}_{k|k} = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{P}_{k|k-1} \tag{77}$$

Notation: In some cases the notation $\boldsymbol{x}_{k-1|k-1}$, $\boldsymbol{x}_{k|k-1}$, and $\boldsymbol{x}_{k|k}$, is replaced with $\boldsymbol{x}_{k-1}^+$ (called the prior), $\boldsymbol{x}_k^-$ (the motion update), and $\boldsymbol{x}_k^+$ (the measurement update/posterior).

Algorithm: Discrete-time Kalman Filter

1. Ensure your system is in the form:

$$\boldsymbol{x}_k = \boldsymbol{F}\boldsymbol{x}_{k-1} + \boldsymbol{G}\boldsymbol{u}_{k-1} + \boldsymbol{w}_{k-1} \tag{78}$$
$$\boldsymbol{y}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{v}_k \tag{79}$$

where the matrices $\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{H}$ are known and the control input $\boldsymbol{u}_k$ is known for all $k = 0, 1, 2, \ldots$. The process and measurement noise covariances $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ are also known where

$$E[\boldsymbol{w}_k \boldsymbol{w}_k^T] = \boldsymbol{Q}_k \tag{80}$$
$$E[\boldsymbol{v}_k \boldsymbol{v}_k^T] = \boldsymbol{R}_k \tag{81}$$

and it is assumed that $E[\boldsymbol{w}_k] = 0$ and $E[\boldsymbol{v}_k] = 0$ for all $k = 0, 1, 2, \ldots$ and $E[\boldsymbol{w}_k \boldsymbol{w}_q^T] = 0$ and $E[\boldsymbol{v}_k \boldsymbol{v}_q^T] = 0$ for all $q \neq k$.

2. Select an initial guess

$$\hat{\boldsymbol{x}}_{0|0} = E[\boldsymbol{x}_0] \tag{82}$$
$$\boldsymbol{P}_{0|0} = E[(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0|0})(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0|0})^{\mathrm{T}}] \tag{83}$$

for the filter and an initial covariance. In the absence of other information you may choose $\hat{\boldsymbol{x}}_0 = \boldsymbol{0}$ and $\boldsymbol{P}_0 = \boldsymbol{1}_{n \times n}\sigma_0^2$ where σ_0^2 is a large number (e.g., 10E6).

3. For $k = 1, 2, 3, \ldots$ perform the following:

   (a) Obtain the measurement $\boldsymbol{y}_k$ from the sensor/data stream.

   (b) Retrieve the posterior from the last step — this becomes the current state and covariance prior $\hat{\boldsymbol{x}}_{k-1|k-1}$ and $\boldsymbol{P}_{k-1|k-1}$.

   (c) Motion update: compute the following

$$\text{State Prediction :} \quad \hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{F}\hat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{G}\boldsymbol{u}_{\boldsymbol{k}-1} \tag{84}$$
$$\text{Covariance Prediction :} \quad \boldsymbol{P}_{k|k-1} = \boldsymbol{F}\boldsymbol{P}_{k-1|k-1}\boldsymbol{F}^T + \boldsymbol{Q} \tag{85}$$

   (d) Measurement update: compute the following

$$\text{Measurement Prediction :} \quad \boldsymbol{y}_{k|k-1} = \boldsymbol{H}\hat{\boldsymbol{x}}_{k|k-1} \tag{86}$$
$$\text{Inovation Covariance :} \quad \boldsymbol{S}_k = \boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^T + \boldsymbol{R}_k \tag{87}$$
$$\text{Kalman Gain :} \quad \boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^T\boldsymbol{S}_k^{-1} \tag{88}$$
$$\text{State Posterior :} \quad \hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{y}_{k|k-1}) \tag{89}$$
$$\text{Covariance Posterior :} \quad \boldsymbol{P}_{k|k} = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{P}_{k|k-1} \tag{90}$$

   (e) Store the posterior and covariance to be used as the prior in the next iteration.

Example

Consider the following example of the two-dimensional motion of a constant velocity object. The state of the system is $\boldsymbol{x} = [x, \dot{x}, y, \dot{y}]^{\mathrm{T}}$ where (x, y) is the object's position and $(\dot{x}, \dot{y})$ are the components of the object's velocity. We assume there is no additional input, $\boldsymbol{G}_k = 0$ and $\boldsymbol{u}_k = 0$. The system dynamics are then:

$$\boldsymbol{x}_k = \boldsymbol{F}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1} \tag{91}$$

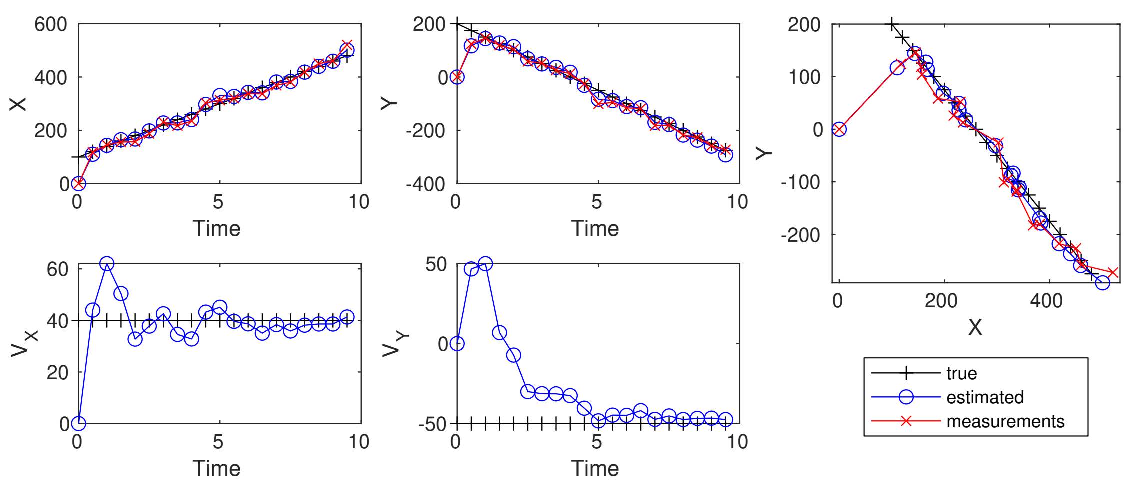
$$\begin{bmatrix} x \\ \dot{x} \\ y \\ \dot{y} \end{bmatrix}_k = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ y \\ \dot{y} \end{bmatrix}_{k-1} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}_{k-1} \tag{92}$$

where $\boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q})$ and $\boldsymbol{Q} = \mathrm{diag}([\sigma_1^2, \sigma_2^2, \sigma_1^2, \sigma_2^2]^T)$ with $\sigma_1 = 5$ and $\sigma_2 = 0.1$ and $T = 0.5$ sec is the time step. Suppose that the system measures only measures position of the object, thus the measurements are

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x}_{k-1} + \boldsymbol{v}_k \tag{93}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ v_x \\ y \\ v_y \end{bmatrix}_k + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_k \tag{94}$$

and the noise is $E[v_k^2] = \sigma_v^2$ with $\sigma_v = 20$. Suppose the true initial state of the system is $\boldsymbol{x}_0 = [100, 40, 200, -50]^T$ and simulate the system for a total of 10 seconds. Initialize the filter with the guess $\hat{\boldsymbol{x}}_0 = [0, 0, 0, 0]^T$ and initial covariance $\boldsymbol{P}_0 = (1E4)\boldsymbol{I}$. Notice that the KF velocity esimtates converge quite well to the actual values, even those these variables are not measured directly. The right-most panel shows the track of the object predicted by the KF compared to the mesurements — we see that the estimated track seems to more accurately match the actual track.



References

[1] Dan Simon. *Optimal State Estimation: Kalman, H infinity, and Nonlinear Approaches.* John Wiley & Sons, 2006.