Dynamical Systems and the Kalman Filter

Suppose you periodically make noisy observations of an object that you wish to track. We assume that the state of the object is evolving under known dymanics. This is a reasonable assumption in many scenarios, where the dynamics might be known. For example, the system might be evolving under well-understood Newtonian mechanics.

The *Kalman filter* is a probabilistic mechanism for reasoning about a sequence of state variables at discrete time steps:

$$\{\mathbf{x}_0, \mathbf{x}_1, \dots\}$$

evolving (noisily) under known mechanics that are assumed to be linear. Specifically, the assumption we make is

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{w}_t,$$

where $\mathbf{F}_t$ is a known, possibly time-dependent matrix and $\mathbf{w}_t$ is a noise term. We wish to maintain a probabilistic belief about the state variables $\{\mathbf{x}_t\}$.

To inform our beliefs, we assume that we make noisy measurements of $\mathbf{x}$ at every time step. Specifically, in the Kalman filter, we assume that we make a noisy measurement $\mathbf{z}_t$ at time t that is related to $\mathbf{x}_t$ via a linear transformation and the possible addition of more noise:

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t,$$

where $\mathbf{H}_t$ is again a known, possibly time-dependent matrix and $\mathbf{v}_t$ is a noise term.

Kalman filter: modeling details

The Kalman filter repeatedly applies Gaussian identities to reason about the evolution of a hidden state $\mathbf{x}$ given the observation sequence $\{\mathbf{z}_t\}$.

In the Kalman filter, we maintain a probabilistic belief about $\mathbf{x}_t$ that is a multivariate Gaussian distribution. Specifically, we will assume (for the sake of induction) that our belief about the state $\mathbf{x}_{t-1}$ given all observations up to time $t-1$ was a multivariate Gaussian distribution:

$$p(\mathbf{x}_{t-1} \mid \mathbf{Z}_{t-1}) = \mathcal{N}(\mathbf{x}_{t-1}; \hat{\mathbf{x}}_{t-1|t-1}, \mathbf{P}_{t-1|t-1}),$$

where we have defined $\mathbf{Z}_{t-1}$ to indicate all observations up to time $t-1$ and have adopted standard Kalman filter notation, using the subscript $t \mid t'$ to indicate our belief about $\mathbf{x}_t$ given observations $\mathbf{Z}_{t'}$.

To make inference tractable, the Kalman filter additionally assumes that the noise term $\mathbf{w}_t$ is independent of $\mathbf{x}_t$, with zero mean and known covariance $\mathbf{Q}_t$:

$$p(\mathbf{w}_t) = \mathcal{N}(\mathbf{w_t}; \mathbf{0}, \mathbf{Q}_t).$$

Recalling our assumption about the dynamics of $\mathbf{x}$:

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{w}_t,$$

we may derive our belief about $\mathbf{x}_t$ given $\mathbf{Z}_{t-1}$; we have simply linearly transformed $\mathbf{x}_{t-1}$ and added independent Gaussian noise. Applying standard results, we may derive:

$$p(\mathbf{x}_t \mid \mathbf{Z}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t-1}, \mathbf{P}_{t|t-1}),$$

where

$$
\begin{aligned}
\hat{\mathbf{x}}_{t|t-1} &= \mathbf{F}_t \hat{\mathbf{x}}_{t-1|t-1} \\
\mathbf{P}_{t|t-1} &= \mathbf{F}_t \mathbf{P}_{t-1|t-1} \mathbf{F}_t^\top + \mathbf{Q}_t.
\end{aligned}
$$

Additionally, we may again apply standard results to derive the joint distribution of $\mathbf{x}_t$ and the observation $\mathbf{z}_t$ *before* we receive observation $\mathbf{z}_t$. This will be a multivariate Gaussian distribution that we will condition on the true value of the observation, updating our belief about $\mathbf{x}_t$ given $\mathbf{Z}_t$.

We recall the linear observation assumption:

$$
\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t.
$$

Again, the Kalman filter assumes that the observation noise at time t is independent and zero-mean with known covariance $\mathbf{R}_t$:

$$
p(\mathbf{v}_t) = \mathcal{N}(\mathbf{v_t}; \mathbf{0}, \mathbf{R}_t).
$$

Now the joint distribution of $\mathbf{x}_t$ and $\mathbf{z}_t$ given the previous observations $\mathbf{Z}_{t-1}$ takes a familiar form:

$$
p(\mathbf{x}_t, \mathbf{z}_t \mid \mathbf{Z}_{t-1}) = \mathcal{N}\left(\begin{bmatrix} \mathbf{x}_t \\ \mathbf{z}_t \end{bmatrix}; \begin{bmatrix} \hat{\mathbf{x}}_{t|t-1} \\ \mathbf{H}_t \hat{\mathbf{x}}_{t|t-1} \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{t|t-1} & \mathbf{P}_{t|t-1} \mathbf{H}_t^\top \\ \mathbf{H}_t \mathbf{P}_{t|t-1} & \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^\top + \mathbf{R}_t \end{bmatrix} \right).
$$

This is called the *predict* step of the Kalman filter, and was made tractable due to the linearity of the state and observation dynamics and the multivariate Gaussian assumptions made about the previous state $\mathbf{x}_{t-1}$ and the noise terms $\mathbf{w}_t$ and $\mathbf{v}_t$.

Now we observe the value $\mathbf{z}_t$ and condition the above distribution to derive the new belief $p(\mathbf{x}_t \mid \mathbf{Z}_t)$. First, we define

$$
\begin{aligned}
\mathbf{S}_t &= \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^\top + \mathbf{R}_t \\
\mathbf{K}_t &= \mathbf{P}_{t|t-1} \mathbf{H}_t^\top \mathbf{S}_t^{-1}.
\end{aligned}
$$

The first term is simply the covariance of $\mathbf{z}_t$ given $\mathbf{Z}_{t-1}$, and the second term is traditionally called the *Kalman gain*. The Kalman gain simply falls out from the standard conditioning formula for multivariate Gaussians. Now

$$
p(\mathbf{x}_t \mid \mathbf{Z}_t) = \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}),
$$

where

$$
\begin{aligned}
\hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t (\mathbf{z}_t - \mathbf{H}_t \hat{\mathbf{x}}_{t-1|t}) \\
\mathbf{P}_{t|t} &= (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_{t|t-1}.
\end{aligned}
$$

This is called the *update* step of the Kalman filter. Notice that our new belief about $\mathbf{x}_t$ given $\mathbf{Z}_t$ is once again a multivariate Gaussian distribution, so we may continue this process recursively.