10. The Extended Kalman Filter

We discuss the *Extended Kalman Filter* (EKF) as an extension of the KF to nonlinear systems. The EKF is derived by linearizing the nonlinear system equations about the latest state estimate.

10.1 Discrete-Time EKF

We consider the nonlinear discrete-time system

$$x(k) = q_{k-1}\big(x(k-1), u(k-1), v(k-1)\big) \qquad \mathrm{E}\left[x(0)\right] = x_0, \mathrm{Var}\left[x(0)\right] = P_0 \tag{10.1}$$

$$\mathrm{E}\left[v(k-1)\right] = 0, \mathrm{Var}\left[v(k-1)\right] = Q(k-1)$$

$$z(k) = h_k\big(x(k), w(k)\big) \qquad \mathrm{E}\left[w(k)\right] = 0, \mathrm{Var}\left[w(k)\right] = R(k) \tag{10.2}$$

for $k = 1, 2, \ldots$, and where $x(0)$, $\{v(\cdot)\}$, and $\{w(\cdot)\}$ are mutually independent. We assume that q_{k-1} is continuously differentiable with respect to $x(k-1)$ and $v(k-1)$, and that h_k is continuously differentiable with respect to $x(k)$ and $w(k)$.

The key idea in the derivation of the EKF is simple: in order to obtain a state estimate for the nonlinear system above, we linearize the system equations about the current state estimate, and we then apply the (standard) KF prior and measurement update equations to the linearized equations.

10.1.1 Process update

Assume we have computed $\hat{x}_m(k-1)$ and $P_m(k-1)$ as (approximations of) the conditional mean and variance of the state $x(k-1)$ given the measurements $z(1{:}k-1)$. Linearizing (10.1)

Last update: 2018-02-20 at 14:01:13

about $x(k{-}1) = \hat{x}_m(k{-}1)$ and $v(k{-}1) = \mathrm{E}\left[v(k{-}1)\right] = 0$ yields

$$
\begin{aligned}
x(k) &\approx q_{k-1}(\hat{x}_m(k{-}1), u(k{-}1), 0) \\
&\qquad + \underbrace{\frac{\partial q_{k-1}(\hat{x}_m(k{-}1), u(k{-}1), 0)}{\partial x}}_{=:A(k-1)} \cdot (x(k{-}1) - \hat{x}_m(k{-}1)) + \underbrace{\frac{\partial q_{k-1}(\hat{x}_m(k{-}1), u(k{-}1), 0)}{\partial v}}_{=:L(k-1)} \cdot v(k{-}1) \\
&= A(k{-}1)x(k{-}1) + \underbrace{L(k{-}1)v(k{-}1)}_{=:\widetilde{v}(k-1)} + \underbrace{q_{k-1}(\hat{x}_m(k{-}1), u(k{-}1), 0) - A(k{-}1)\hat{x}_m(k{-}1)}_{=:\xi(k-1)} \\
&= A(k{-}1)x(k{-}1) + \widetilde{v}(k{-}1) + \xi(k{-}1),
\end{aligned}
$$

where $\xi(k{-}1)$ is treated as a known input, and the process noise $\widetilde{v}(k{-}1)$ has zero-mean and variance $\mathrm{Var}\left[\widetilde{v}(k{-}1)\right] = L(k{-}1)Q(k{-}1)L^T(k{-}1)$. We can now apply the KF prior update equations to the linearized process equation:

$$
\begin{aligned}
\hat{x}_p(k) &= A(k{-}1)\hat{x}_m(k{-}1) + \xi(k{-}1) \\
&= q_{k-1}(\hat{x}_m(k{-}1), u(k{-}1), 0) \qquad \text{(by substituting } \xi(k{-}1)\text{)} \\
P_p(k) &= A(k{-}1)P_m(k{-}1)A^T(k{-}1) + L(k{-}1)Q(k{-}1)L^T(k{-}1).
\end{aligned}
$$

Intuition: predict the mean state estimate forward using the nonlinear process model and update the variance according to the linearized equations.

10.1.2 Measurement update

We linearize (10.2) about $x(k) = \hat{x}_p(k)$ and $w(k) = \mathrm{E}\left[w(k)\right] = 0$:

$$
\begin{aligned}
z(k) &\approx h_k(\hat{x}_p(k), 0) + \underbrace{\frac{\partial h_k(\hat{x}_p(k), 0)}{\partial x}}_{=:H(k)} \cdot (x(k) - \hat{x}_p(k)) + \underbrace{\frac{\partial h_k(\hat{x}_p(k), 0)}{\partial w}}_{=:M(k)} \cdot w(k) \\
&= H(k)x(k) + \underbrace{M(k)w(k)}_{=:\widetilde{w}(k)} + \underbrace{h_k(\hat{x}_p(k), 0) - H(k)\hat{x}_p(k)}_{=:\zeta(k)} \\
&= H(k)x(k) + \widetilde{w}(k) + \zeta(k),
\end{aligned}
$$

where $\widetilde{w}(k)$ has zero mean and variance $\mathrm{Var}\left[\widetilde{w}(k)\right] = M(k)R(k)M^T(k)$. Compared to the measurement equation that we used in the derivation of the KF, there is the additional term $\zeta(k)$, which is known. It is straightforward to extend the KF measurement update to this case (for example, by introducing the auxiliary measurement $\widetilde{z}(k) := z(k) - \zeta(k)$).

Applying the KF measurement update to the linearized measurement equation yields:

$$\begin{aligned}
K(k) &= P_p(k)H^T(k)\left(H(k)P_p(k)H^T(k) + M(k)R(k)M^T(k)\right)^{-1}\\
\hat{x}_m(k) &= \hat{x}_p(k) + K(k)\left(z(k) - H(k)\hat{x}_p(k) - \zeta(k)\right)\\
&= \hat{x}_p(k) + K(k)\left(z(k) - h_k(\hat{x}_p(k), 0)\right) \qquad \text{(by substituting } \zeta(k)\text{)}\\
P_m(k) &= \big(I - K(k)H(k)\big)P_p(k).
\end{aligned}$$

Intuition: correct for the mismatch between the actual measurement $z(k)$ and its nonlinear prediction $h_k(\hat{x}_p(k), 0)$.

10.1.3 Summary

The discrete-time EKF equations are given by:

Initialization: $\hat{x}_m(0) = x_0$, $P_m(0) = P_0$.

Step 1 (S1): Prior update/Prediction step

$$\begin{aligned}
\hat{x}_p(k) &= q_{k-1}(\hat{x}_m(k{-}1), u(k{-}1), 0)\\
P_p(k) &= A(k{-}1)P_m(k{-}1)A^T(k{-}1) + L(k{-}1)Q(k{-}1)L^T(k{-}1)
\end{aligned}$$

where

$$A(k{-}1) := \frac{\partial q_{k-1}(\hat{x}_m(k{-}1), u(k{-}1), 0)}{\partial x} \quad \text{and} \quad L(k{-}1) := \frac{\partial q_{k-1}(\hat{x}_m(k{-}1), u(k{-}1), 0)}{\partial v}.$$

Step 2 (S2): A posteriori update/Measurement update step

$$\begin{aligned}
K(k) &= P_p(k)H^T(k)\left(H(k)P_p(k)H^T(k) + M(k)R(k)M^T(k)\right)^{-1}\\
\hat{x}_m(k) &= \hat{x}_p(k) + K(k)\left(z(k) - h_k(\hat{x}_p(k), 0)\right)\\
P_m(k) &= \big(I - K(k)H(k)\big)P_p(k)
\end{aligned}$$

where

$$H(k) := \frac{\partial h_k(\hat{x}_p(k), 0)}{\partial x} \quad \text{and} \quad M(k) := \frac{\partial h_k(\hat{x}_p(k), 0)}{\partial w}.$$

10.1.4 Remarks

- The matrices $A(k{-}1)$, $L(k{-}1)$, $H(k)$, and $M(k)$ are obtained from linearizing the system equations about the current state estimate (which depends on real-time measurements).

Hence, the EKF gains cannot be computed off-line, even if the model and noise distributions are known for all k.

- If the actual state and noise values are close to the values that we linearize about (i.e. if $x(k\text{-}1) - \hat{x}_m(k\text{-}1)$, $v(k\text{-}1)$, $x(k) - \hat{x}_p(k)$, and $w(k)$ are all close to zero), then the linearization is a good approximation of the actual nonlinear dynamics. This assumption may, however, be bad. In the case of Gaussian noise, for example, the above quantities are not guaranteed to be small since the noise is actually unbounded.

- The EKF variables $\hat{x}_p(k)$, $\hat{x}_m(k)$, $P_p(k)$, and $P_m(k)$ do not capture the true conditional mean and variance of $x(k)$ (let alone the full conditional PDF). They are only approximations of mean and variance.

  For example, in the prior update, $\hat{x}_p(k)$ would only accurately capture the mean update if the expected value operator $\mathrm{E}[\cdot]$ and q_{k-1} commuted; that is, if

  $$\mathrm{E}\left[q_{k-1}\big(x(k\text{-}1), u(k\text{-}1), v(k\text{-}1)\big)\right] = q_{k-1}\big(\mathrm{E}[x(k\text{-}1)], \mathrm{E}[u(k\text{-}1)], \mathrm{E}[v(k\text{-}1)]\big).$$

  This is not true for a general nonlinear function q_{k-1}, and may be a really bad assumption in the case of strong nonlinearities (it holds, however, for linear q_{k-1}).

  Even though the EKF variables do not capture the true conditional mean and variance, they are often still referred to as the prior/posterior mean and variance.

- Despite the fact that the EKF is a (possibly crude) approximation of the Bayesian state estimator and general convergence guarantees cannot be given, the EKF has proven to work well in many practical applications. As a rule of thumb, an EKF often works well for (mildly) nonlinear systems with unimodal distributions.

  Solving the Bayesian state estimation problem for a general nonlinear system is often computationally intractable. Hence, the EKF may be seen as a computationally tractable approximation (trade-off: tractability vs. accuracy).

10.2 Example of a "nice" system with poor performance

Consider the following scalar system, with the functions

$$\begin{aligned} x(k) &= f\left(x(k\text{-}1), v(k\text{-}1)\right) = 2\arctan\left(x(k\text{-}1) + v(k\text{-}1)\right) \\ z(k) &= h(x(k), w(k)) = x(k) + w(k) \end{aligned}$$

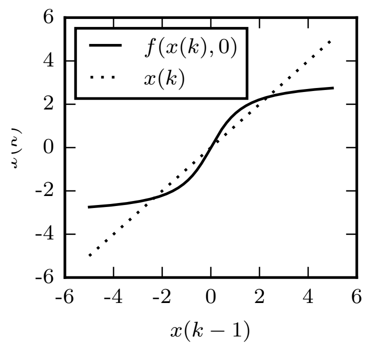
and $x(0) \sim \mathcal{N}(x_0, 1)$, $v(k\text{-}1) \sim \mathcal{N}(0, 0.1)$, and $w(k) \sim \mathcal{N}(0, 10)$.

We notice that the measurement is linear, and that the dynamics are smooth (infinitely differentiable):

$$\frac{\partial f\left(x,v\right)}{\partial x}=\frac{2}{(x+v)^2+1}=\frac{\partial f\left(x,v\right)}{\partial v} \tag{10.3}$$

The process uncertainty is small, and the measurement uncertainty large.

The dynamics are illustrated below:



The deterministic system (when we set $v(k) = 0$) has three equilibria, at $x(k) = 0$ and $x(k) \approx \pm 2.33$. The equilibrium at zero is unstable, the other two are stable.

We can run the EKF algorithm on generated data many times, to get a feeling for the system's performance. We can compare the EKF performance for $x_0 = 4$, and $x_0 = 0$.

We note, for $x_0 = 4$: the EKF does a good job of tracking the state (note: the 1σ line is true only for one run of the EKF (the last one we did), not for all runs). The true state clumps around the stable equilibrium at $x \approx 2.33$.

For $x_0 = 0$: the EKF does *not* do a good job of tracking the state. Usually, it's OK, but sometimes the true state goes to one equilibrium, and the estimator to the other. Note

- this is true even though we can measure the state,
- it doesn't matter how long we run the filter, it doesn't correct itself,
- there is no model mismatch.

If we look at a plot of the true state and the estimate for such a case, we see the problem: the EKF estimate initially moved in the opposite direction from the true state (perhaps due to process noise). At the same time, the EKF is overly confident of its estimate (as can be seen by the small covariance). In this case, we would say that the EKF has *diverged*.