8. 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. First, we introduce the *discrete-time EKF* as an estimator for discrete-time nonlinear process and measurement equations. We then discuss the case where the process is governed by nonlinear continuous-time dynamics, which yields the *hybrid EKF*.

8.1 Discrete-Time EKF

We consider the nonlinear discrete-time system

$$x(k) = q_{k-1}\big(x(k\text{-}1), u(k\text{-}1), v(k\text{-}1)\big) \qquad \mathrm{E}[x(0)] = x_0, \mathrm{Var}[x(0)] = P_0 \tag{8.1}$$
$$\mathrm{E}[v(k\text{-}1)] = 0, \mathrm{Var}[v(k\text{-}1)] = Q(k\text{-}1)$$
$$z(k) = h_k\big(x(k), w(k)\big) \qquad \mathrm{E}[w(k)] = 0, \mathrm{Var}[w(k)] = R(k) \tag{8.2}$$

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 $x(k\text{-}1)$ and $v(k\text{-}1)$, and that h_k is continuously differentiable with respect to $x(k)$ and $w(k)$.

8.1.1 Derivation

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.

Process update

Assume we have computed $\hat{x}_m(k\text{-}1)$ and $P_m(k\text{-}1)$ as (approximations of) the conditional mean and variance of the state $x(k\text{-}1)$ given the measurements $z(1{:}k\text{-}1)$. Linearizing (8.1) about $x(k\text{-}1) = \hat{x}_m(k\text{-}1)$ and $v(k\text{-}1) = \mathrm{E}[v(k\text{-}1)] = 0$ yields

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

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

Date compiled: May 7, 2013

process equation:

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

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

Measurement update

We linearize (8.2) about $x(k) = \hat{x}_p(k)$ and $w(k) = \mathrm{E}[w(k)] = 0$:

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

where $\tilde{w}(k)$ has zero mean and variance $\mathrm{Var}[\tilde{w}(k)] = M(k)R(k)M^T(k)$. 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 $\tilde{z}(k) := z(k) - \zeta(k)$). Applying the KF measurement update to the linearized measurement equation yields:

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

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

8.1.2 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

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

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

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

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

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

8.1.3 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 $x(k-1) - \hat{x}_m(k-1)$, $v(k-1)$, $x(k) - \hat{x}_p(k)$, and $w(k)$ are all close to zero), then the linearization is a good approximation of the actual nonlinear dynamics. This approximation 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 no longer 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

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

  This is not true for a general nonlinear function q_{k-1} and, it 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).

8.2 Hybrid EKF

In practice, one often encounters problems where the process is governed by continuous-time dynamics, while the measurements are taken at discrete time instants. Hence, the system is described by

$$\begin{aligned} \dot{x}(t) &= q\big(x(t), u(t), v(t), t\big) \\ z[k] &= h_k\big(x[k], w[k]\big) \qquad\qquad \mathrm{E}[w[k]] = 0, \mathrm{Var}[w[k]] = R, \end{aligned}$$

where we consider constant measurement noise variance R for simplicity, and the process noise characteristics are to be discussed below. In this section[1], we use parentheses $(\cdot)$ to denote continuous-time signals and square brackets $[\cdot]$ to denote discrete-time samples of continuous-time signals; that is,

$$x[k] := x(kT), \qquad T = \text{constant sampling time.}$$

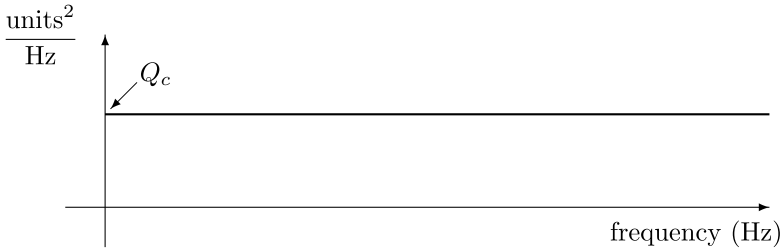
For consistency, we keep using parentheses $(\cdot)$ for the EKF variables, even though these are only updated in discrete steps.

One way to deal with this problem is to discretize the process dynamics and then implement the discrete-time EKF (discussed in the recitation). Here we present a different solution, where we work directly with the continuous-time dynamics.

[1] This is the only time in this course where we deal with continuous-time signals and dynamics.

8.2.1 Continuous-time process noise

We assume that the continuous-time process noise $v(t)$ is white noise, which is defined to have a constant power spectral density Q_c over all frequencies:



- Truly continuous-time white noise does not occur in real-world problems since it has infinite power (integration over all frequencies). Yet, the concept of continuous-time white noise is useful for theoretical analysis, and many continuous-time noise signals can be approximated by white noise.
- One can show[2] that $v(t)$ being a white noise process with constant power spectral density implies[3]

  $$\mathrm{E}[v(t)] = 0 \quad \text{and} \quad \mathrm{E}[v(t)v^T(t+\tau)] = Q_c\delta(\tau),$$

  where $\delta(\tau)$ is the continuous-time Dirac pulse, which has the following properties: it is zero everywhere except at $\tau = 0$ where it is infinite, and its integral (from $-\infty$ to ∞) is equal to one. The Dirac pulse can be defined by the property

  $$\int_{-\infty}^{\infty} \xi(\tau)\delta(\tau)\, d\tau = \xi(0),$$

  where $\xi(\tau)$ is a real-valued function.

8.2.2 Continuous-time process update

Process equation with noise characteristics:

$$\dot{x}(t) = q(x(t), u(t), v(t), t), \qquad \begin{aligned} &\mathrm{E}[x(0)] = x_0,\ \mathrm{Var}[x(0)] = P_0 \\ &\mathrm{E}[v(t)] = 0,\ E[v(t)v^T(t+\tau)] = Q_c\delta(\tau) \end{aligned}$$

and $v(\cdot)$, $x(0)$ are independent.

We consider $0 \leq t \leq T$ for now and generalize this later. Assume $\hat{x}_m(0) = \mathrm{E}[x(0)]$ and $P_m(0) = \mathrm{Var}[x(0)]$. Analogously to the discrete-time case, we predict the estimate forward using the process model. This time, we integrate the continuous-time dynamics instead of performing a discrete update step.

Mean

Let $\hat{x}(t)$ solve

$$\dot{\hat{x}}(t) = q(\hat{x}(t), u(t), 0, t), \quad \text{for } 0 \leq t \leq T \text{ and } \hat{x}(0) = \hat{x}_m(0).$$

We *assume* that $\hat{x}(t) \approx \mathrm{E}[x(t)]$, which is the case if the expected value operator and function $q(\cdot)$ commute (again, this holds if $q(\cdot)$ is linear, but can be bad for strongly nonlinear $q(\cdot)$).

Hence, we set $\hat{x}_p(1) = \hat{x}(T)$.

---

[2] A rigorous treatment of this is beyond the scope of this class.

[3] Note that the discrete-time counterpart to this is

$$\mathrm{E}[v(k)] = 0 \quad \text{and} \quad \mathrm{E}[v(k)v^T(k+\kappa)] = Q\delta_d(\kappa), \quad \text{where } \kappa \in \mathbb{Z} \text{ and } \delta_d(\kappa) := \begin{cases} 1 & \text{if } \kappa = 0 \\ 0 & \text{otherwise} \end{cases},$$

which is true for the system considered in the derivation of the KF (independence of $\{v(\cdot)\}$).

Variance

We next consider the variance. Let $\tilde{x}(t) = x(t) - \hat{x}(t)$. *Assuming* that $\tilde{x}(t)$ and $v(t)$ are small (is this a good assumption?),

$$\dot{\tilde{x}}(t) \approx A(t)\tilde{x}(t) + L(t)v(t),$$

$$A(t) := \frac{\partial q}{\partial x}(\hat{x}(t), u(t), 0, t), \quad L(t) := \frac{\partial q}{\partial v}(\hat{x}(t), u(t), 0, t).$$

For small τ, we have

$$\tilde{x}(t+\tau) \approx \tilde{x}(t) + \int_t^{t+\tau} A(\xi)\tilde{x}(\xi) + L(\xi)v(\xi)\, d\xi = \tilde{x}(t) + \tau A(t)\tilde{x}(t) + L(t)\int_t^{t+\tau} v(\xi)\, d\xi + O(\tau^2),$$

where $O(\tau^2)$ denotes second and higher-order terms.

Note that $P(t) := \mathrm{Var}[x(t)] \approx \mathrm{E}[\tilde{x}(t)\tilde{x}^T(t)]$, since we assume $\hat{x}(t) \approx \mathrm{E}[x(t)]$. We therefore have:

$$P(t+\tau) \approx P(t) + \tau A(t)P(t) + \tau P(t)A^T(t) + L(t)\underbrace{\int_t^{t+\tau}\int_t^{t+\tau} \mathrm{E}[v(\xi)v^T(s)]\, d\xi\, ds}_{\tau Q_c}\, L^T(t) + O(\tau^2)$$

Taking the limit as $\tau \to 0$:

$$\dot{P}(t) = \lim_{\tau\to 0} \frac{P(t+\tau) - P(t)}{\tau} = A(t)P(t) + P(t)A^T(t) + L(t)Q_cL^T(t).$$

We can thus solve for $P(t)$ by solving the above matrix differential equation for $0 \leq t \leq T$ with initial condition $P(0) = P_m(0)$. We then set $P_p(1) = P(T)$.

8.2.3 Summary

The hybrid EKF equations are given by:

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

Step 1 (S1): Prior update/Prediction step
Solve

$$\dot{\hat{x}}(t) = q(\hat{x}(t), u(t), 0, t), \quad \text{for } (k-1)T \leq t \leq kT \text{ and } \hat{x}((k-1)T) = \hat{x}_m(k-1).$$

Then $\hat{x}_p(k) := \hat{x}(kT)$.

Solve

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) + L(t)Q_cL^T(t), \quad \text{for } (k-1)T \leq t \leq kT \text{ and } P((k-1)T) = P_m(k-1),$$

where $A(t) = \frac{\partial q}{\partial x}(\hat{x}(t), u(t), 0, t)$ and $L(t) = \frac{\partial q}{\partial v}(\hat{x}(t), u(t), 0, t)$. Then $P_p(k) := P(kT)$.

Step 2 (S2): A posteriori update/Measurement update step
The measurement update step is identical to the one for the discrete-time EKF since the measurement model is still discrete-time.

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

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

8.2.4 Remarks

- The hybrid EKF requires the solution of a vector and a matrix ordinary differential equation (ODE), which is typically done using numerical ODE solvers (such as Runge-Kutta methods as, for example, implemented in Matlab's `ode45`). The accuracy of the numerical integration largely depends on the order of the solver. Generally, numerical accuracy is at the cost of increased computation.

- We have discussed continuous-time dynamics within the setting of the EKF (i.e. for nonlinear systems). The same discussion, of course, applies to linear systems (which are just a special case of nonlinear systems). Furthermore, one can extend this discussion to the case where also the measurement model is a continuous-time model (linear or nonlinear) and derive the continuous-time KF (for a continuous-time linear system) and the continuous-time EKF (for a continuous-time nonlinear system).