star springer tracts in advanced robotics 73

Peter Corke

Robotics, Vision and Control

FUNDAMENTAL ALGORITHMS IN MATLAB®

Springer

MATLAB® and Simulink® examples

H Kalman Filter

Consider the discrete-time linear time-invariant system

$$\begin{aligned} \boldsymbol{x}\langle k+1\rangle &= \boldsymbol{F}\boldsymbol{x}\langle k\rangle + \boldsymbol{G}\boldsymbol{u}\langle k\rangle + \boldsymbol{v}\langle k\rangle \\ \boldsymbol{z}\langle k+1\rangle &= \boldsymbol{H}\boldsymbol{x}\langle k\rangle + \boldsymbol{w}\langle k\rangle \end{aligned}$$

with state vector $\boldsymbol{x} \in \mathbb{R}^n$. The vector $\boldsymbol{u} \in \mathbb{R}^m$ is the input to the system at time k, for example a velocity command, or applied forces and torques. The vector $\boldsymbol{z} \in \mathbb{R}^p$ represents the outputs of the system as measured by sensors. The matrix $\boldsymbol{F} \in \mathbb{R}^{n \times n}$ describes the dynamics of the system, that is, how the states evolve with time. The matrix $\boldsymbol{G} \in \mathbb{R}^{n \times m}$ describes how the inputs are coupled to the system states. The matrix $\boldsymbol{H} \in \mathbb{R}^{p \times n}$ describes how the system states are mapped to the observed outputs.

To account for errors in the model (represented by $\boldsymbol{F}$ and $\boldsymbol{G}$) and also unmodeled disturbances we introduce a Gaussian random variable $\boldsymbol{v} \in \mathbb{R}^n$ termed the process noise. $\boldsymbol{v}\langle k\rangle \sim N(0, \boldsymbol{V})$, that is, it has zero-mean and covariance $\boldsymbol{V}$. The sensor measurement model $\boldsymbol{H}$ is not perfect either and this is modelled by measurement noise, another Gaussian random variable $\boldsymbol{w} \in \mathbb{R}^p$ and $\boldsymbol{w}\langle k\rangle \sim N(0, \boldsymbol{W})$. The covariance matrices $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{W} \in \mathbb{R}^{p \times p}$ are symmetric and positive definite.

The general problem that we confront is:

> *given a model of the system, the known inputs $\boldsymbol{u}$ and some noisy sensor measurements $\boldsymbol{z}$, estimate the state of the system $\boldsymbol{x}$.*

In a robotic localization context $\boldsymbol{x}$ is the unknown pose of the robot, $\boldsymbol{u}$ is the commands sent to the motors and $\boldsymbol{z}$ is the output of various sensors on the robot. For a flying robot $\boldsymbol{x}$ could be the attitude, $\boldsymbol{u}$ the known forces applied to the airframe and $\boldsymbol{z}$ are the measured accelerations and angular velocities.

The Kalman filter is an optimal estimator for the case where the process and measurement noise are zero-mean Gaussian noise. The filter has two steps. The first is a prediction of the state based on the previous state and the inputs that were applied.

$$\hat{\boldsymbol{x}}\langle k+1|k\rangle = \boldsymbol{F}\hat{\boldsymbol{x}}\langle k\rangle + \boldsymbol{G}\boldsymbol{u}\langle k\rangle \tag{H.1}$$

$$\hat{\boldsymbol{P}}\langle k+1|k\rangle = \boldsymbol{F}\hat{\boldsymbol{P}}\langle k|k\rangle\boldsymbol{F}^T + \hat{\boldsymbol{V}} \tag{H.2}$$

where $\hat{\boldsymbol{x}}$ is the estimate of the state and $\hat{\boldsymbol{P}} \in \mathbb{R}^{n \times n}$ is the estimated covariance, or uncertainty, in $\hat{\boldsymbol{x}}$. This is an *open-loop* step and its accuracy depends completely on the quality of the model $\boldsymbol{F}$ and $\boldsymbol{G}$ and the ability to measure the inputs $\boldsymbol{u}$. The notation $k+1|k$ makes explicit that the left-hand side is an estimate at time $k+1$ based on information from time k.

The prediction of $\boldsymbol{P}$ involves the addition of two positive-definite matrices so the uncertainty, given no new information and the uncertainty in the process, has increased. To improve things we have to introduce new information and that comes from mea-

surements obtained using sensors. The new information that is added is known as the innovation

$$\boldsymbol{\nu}\langle k+1\rangle = \boldsymbol{z}\langle k+1\rangle - \boldsymbol{H}\hat{\boldsymbol{x}}\langle k+1|k\rangle$$

which is the difference between what the sensors measure and what the sensors are predicted to measure. Some of the difference will be due to the noise in the sensor, the measurement noise, but the remaining discrepancy indicates that the predicted state was in error and does not properly explain the sensor observations.

The second step of the Kalman filter, the *update* step, uses the Kalman gain

$$\boldsymbol{K}\langle k+1\rangle = \hat{\boldsymbol{P}}\langle k+1|k\rangle \boldsymbol{H}^T \underbrace{\left(\boldsymbol{H}\hat{\boldsymbol{P}}\langle k+1|k\rangle \boldsymbol{H}^T + \hat{\boldsymbol{W}}\right)}_{\boldsymbol{S}}{}^{-1} \tag{H.3}$$

to map the innovation into a correction for the predicted state, optimally tweaking the estimate based on what the sensors observed

$$\begin{aligned}\hat{\boldsymbol{x}}\langle k+1|k+1\rangle &= \hat{\boldsymbol{x}}\langle k+1|k\rangle + \boldsymbol{K}\langle k+1\rangle \boldsymbol{\nu}\langle k+1\rangle \\ \hat{\boldsymbol{P}}\langle k+1|k+1\rangle &= \hat{\boldsymbol{P}}\langle k+1|k\rangle - \boldsymbol{K}\langle k+1\rangle \boldsymbol{H}\hat{\boldsymbol{P}}\langle k+1|k\rangle\end{aligned}$$

Importantly we note that the uncertainty is now decreased or *deflated*, since the second term is subtracted from the predicted covariance. The term indicated by $\boldsymbol{S}$ is the estimated covariance of the innovation and comes from the uncertainty in the state and the measurement noise covariance. If the innovation has high uncertainty in relation to some states this will be reflected in the Kalman gain which will make correspondingly small adjustment to those states.

The covariance update can also be written in the Joseph form

$$\hat{\boldsymbol{P}}\langle k+1|k+1\rangle = \left(\boldsymbol{I}_{n\times n} - \boldsymbol{K}\langle k+1\rangle \boldsymbol{H}\right)\hat{\boldsymbol{P}}\langle k+1|k\rangle\left(\boldsymbol{I}_{n\times n} - \boldsymbol{K}\langle k+1\rangle \boldsymbol{H}\right)^T + \boldsymbol{K}\langle k+1\rangle \hat{\boldsymbol{V}}\boldsymbol{K}\langle k+1\rangle^T$$

which has improved numerical properties and keeps the covariance estimate symmetric, but it is computationally more costly.

The equations above constitute the classical Kalman filter which is widely used in applications from aerospace to econometrics. The filter has a number of important characteristics. Firstly it is recursive, the output of one iteration is the input to the next. Secondly, it is asynchronous. At a particular iteration if no sensor information is available we perform just the prediction step and not the update. In the case that there are different sensors, each with their own $\boldsymbol{H}$, and different sample rates, we just apply the update with the appropriate $\boldsymbol{z}$ and $\boldsymbol{H}$. The Kalman-Bucy filter is a continuous-time version of this filter.

The filter must be initialized with some reasonable value of $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{P}}$. The filter also requires our best estimates of the covariance of the process and measurement noise. In general we do not know $\boldsymbol{V}$ and $\boldsymbol{W}$ but we have some estimate $\hat{\boldsymbol{V}}$ and $\hat{\boldsymbol{W}}$ that we use in the filter. From Eq. H.2 we see that if we overestimate $\hat{\boldsymbol{V}}$ our estimate of $\boldsymbol{P}$ will be larger than it really is giving a pessimistic estimate of our certainty in the state. Conversely if we overestimate $\hat{\boldsymbol{V}}$ the filter will be *overconfident* of its estimate.

The covariance matrix $\hat{\boldsymbol{P}}$ is rich in information. The diagonal elements $\hat{\boldsymbol{P}}_{ii}$ are the variance, or uncertainty, in the state $\boldsymbol{x}_i$. The off-diagonal elements $\hat{\boldsymbol{P}}_{ij}$ are the correlations between states $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. The correlations are critical in allowing any piece of new information to *flow through* to adjust multiple states that affect a particular process output.

The term $\boldsymbol{F}\boldsymbol{P}\langle k|k\rangle \boldsymbol{F}^T\langle k\rangle$ in Eq. H.2 is interesting. Consider a one dimensional example where F is a scalar and the state estimate $\hat{x}\langle k\rangle$ has a PDF that is a Gaussian with a mean $\overline{x}\langle k\rangle$ and a variance $\sigma^2\langle k\rangle$. The prediction equation maps the state and its Gaussian dis-

tribution to a new Gaussian distribution with a mean $F\bar{x}\langle k\rangle$ and a variance $F^2\sigma^2\langle k\rangle$. The term $\boldsymbol{FP}\langle k|k\rangle\boldsymbol{F}^T\langle k\rangle$ is the matrix form of this since

$$\text{cov}(\boldsymbol{F}\boldsymbol{x}) = \boldsymbol{F}\,\text{cov}(\boldsymbol{x})\boldsymbol{F}^T$$

and appropriately scales the covariance. The term $\boldsymbol{HP}\langle k+1|k\rangle\boldsymbol{H}^T$ in Eq. H.3 *projects* the covariance of the state estimate into the observed values.

Now consider the case where the system is not linear

$$\begin{aligned}\boldsymbol{x}\langle k+1\rangle &= \boldsymbol{f}\big(\boldsymbol{x}\langle k\rangle, \boldsymbol{u}\langle k\rangle\big) + \boldsymbol{v}\langle k\rangle\\ \boldsymbol{z}\langle k+1\rangle &= \boldsymbol{h}\big(\boldsymbol{x}\langle k\rangle\big) + \boldsymbol{w}\langle k\rangle\end{aligned}$$

where $\boldsymbol{f}$ and $\boldsymbol{h}$ are now functions instead of constant matrices. $\boldsymbol{f}:\mathbb{R}^n, \mathbb{R}^m \to \mathbb{R}^n$ is a function that describes the new state in terms of the previous state and the input to the system. The function $\boldsymbol{h}:\mathbb{R}^n \to \mathbb{R}^p$ maps the state vector to the sensor measurements.

To use the linear Kalman filter with a non-linear system we first make a local linear approximation

$$\begin{aligned}\boldsymbol{x}\langle k+1\rangle &= \boldsymbol{f}\big(\hat{\boldsymbol{x}}\langle k\rangle, \boldsymbol{u}\langle k\rangle\big) + \boldsymbol{F}_x\big(\boldsymbol{x}\langle k\rangle - \hat{\boldsymbol{x}}\langle k|k\rangle\big) + \boldsymbol{F}_u\boldsymbol{u}\langle k\rangle + \boldsymbol{F}_v\boldsymbol{v}\langle k\rangle\\ \boldsymbol{z}\langle k+1\rangle &= \boldsymbol{h}\big(\hat{\boldsymbol{x}}\langle k\rangle\big) + \boldsymbol{H}_x\big(\hat{\boldsymbol{x}}\langle k+1|k\rangle - \boldsymbol{x}\langle k\rangle\big) + \boldsymbol{H}_w\boldsymbol{w}\langle k\rangle\end{aligned}$$

where $\boldsymbol{F}_x \in \mathbb{R}^{n\times n}$, $\boldsymbol{F}_u \in \mathbb{R}^{n\times m}$, $\boldsymbol{F}_v \in \mathbb{R}^{n\times n}$, $\boldsymbol{H}_x \in \mathbb{R}^{p\times n}$ and $\boldsymbol{H}_w \in \mathbb{R}^{p\times p}$ are Jacobians of the functions $\boldsymbol{f}(\cdot)$ and $\boldsymbol{h}(\cdot)$ and are evaluated at each time step.

We define a prediction error

$$\begin{aligned}\tilde{\boldsymbol{x}}\langle k+1|k\rangle &= \boldsymbol{x}\langle k\rangle - \hat{\boldsymbol{x}}\langle k+1|k\rangle\\ &= \boldsymbol{F}_x\tilde{\boldsymbol{x}}\langle k|k\rangle + \boldsymbol{F}_u\boldsymbol{u}\langle k\rangle + \boldsymbol{F}_v\boldsymbol{v}\langle k\rangle\end{aligned}$$

and a measurement residual

$$\begin{aligned}\tilde{\boldsymbol{z}}\langle k+1|k\rangle &= \boldsymbol{z}\langle k+1\rangle - \boldsymbol{h}\langle k+1|k\rangle\\ &= \boldsymbol{H}_x\tilde{\boldsymbol{x}} + \boldsymbol{H}_w\boldsymbol{w}\langle k\rangle\end{aligned}$$

which are linear and the Kalman filter equations above can be applied. The prediction step of the extended Kalman filter is

$$\begin{aligned}\hat{\boldsymbol{x}}\langle k+1|k\rangle &= \boldsymbol{f}\big(\hat{\boldsymbol{x}}\langle k\rangle, \boldsymbol{u}\langle k\rangle\big)\\ \hat{\boldsymbol{P}}\langle k+1|k\rangle &= \boldsymbol{F}_x\hat{\boldsymbol{P}}\langle k|k\rangle\boldsymbol{F}_x^T + \boldsymbol{F}_v\hat{\boldsymbol{V}}\langle k\rangle\boldsymbol{F}_v^T\end{aligned}$$

and the update step is

$$\begin{aligned}\hat{\boldsymbol{x}}\langle k+1|k+1\rangle &= \hat{\boldsymbol{x}}\langle k+1|k\rangle + \boldsymbol{K}\langle k+1\rangle\boldsymbol{\nu}\langle k+1\rangle\\ \hat{\boldsymbol{P}}\langle k+1|k+1\rangle &= \hat{\boldsymbol{P}}\langle k+1|k\rangle - \boldsymbol{K}\langle k+1\rangle\boldsymbol{H}_x\hat{\boldsymbol{P}}\langle k+1|k\rangle\end{aligned}$$

where the innovation is

$$\boldsymbol{\nu}\langle k+1\rangle = \boldsymbol{z}\langle k+1\rangle - \boldsymbol{h}\big(\hat{\boldsymbol{x}}\langle k+1|k\rangle\big)$$

and the Kalman gain is

$$\boldsymbol{K}\langle k+1\rangle = \hat{\boldsymbol{P}}\langle k+1|k\rangle\boldsymbol{H}_x^T\Big(\boldsymbol{H}_x\hat{\boldsymbol{P}}\langle k+1|k\rangle\boldsymbol{H}_x^T + \boldsymbol{H}_w\hat{\boldsymbol{W}}\boldsymbol{H}_w^T\Big)^{-1}$$

A fundamental problem with the extended Kalman filter is that PDFs of the random variables are no longer Gaussian after being operated on by the non-linear

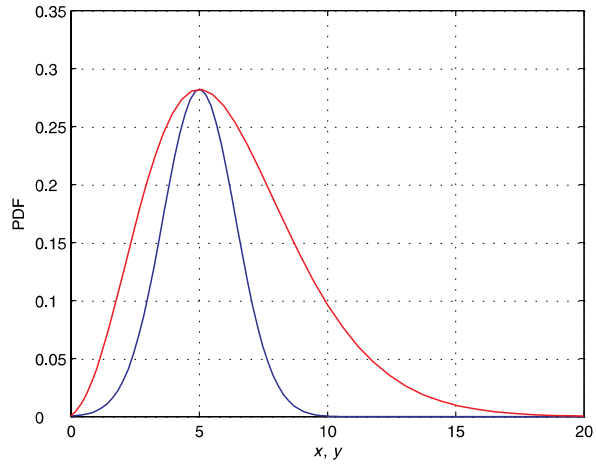

Fig. H.1.
PDF of the state x (blue) which is Gaussian $N(5, 2)$ and the PDF of the non-linear function $x^2/5$ (red)

functions $\boldsymbol{f}(\cdot)$ and $\boldsymbol{h}(\cdot)$. We can easily illustrate this by considering a scalar system with the PDF of the state estimate being the Gaussian $N(5, 2)$

```
>> x = linspace(0, 20, 100);
>> g = gaussfunc(5, 2, x);
>> plot(x, g);
```

Now consider the nonlinear function $y = x^2/5$ and we overlay the PDF of y

```
>> y = x.^2 / 5;
>> plot(y, g, 'r');
```

which is shown in Fig. H.1. We see that the PDF of y has its peak, the mode, at the same location but the distribution is no longer Gaussian. It has lost its symmetry so the mean value will actually be greater than the mode. The Jacobians that appear in the EKF equations appropriately scale the covariance but the resulting non-Gaussian distributions break the assumptions which guarantee that the Kalman filter is an optimal estimator. Alternatives include the iterated EKF described by Jazwinski (1970) or the Unscented Kalman Filter (UKF) (Julier and Uhlmann 2004) which uses discrete sample points to approximate the PDF.