An introduction to Kalman filters

Jeremy Fix

March 19, 2012

Contents

1 Preliminaries

1.1 Notation

In the following, we denote x a scalar, $\boldsymbol{x}$ a vector and $\mathbf{X}$ a matrix. Vectors are supposed to be column vectors, i.e. $\boldsymbol{x}^T\boldsymbol{y}$ is the scalar product between $\boldsymbol{x}$ and $\boldsymbol{y}$. We denote $A_{i,j}$ the component at row i, column j of a matrix $\mathbf{A}$. The trace of a matrix $A \in M(n,m)$ equals : $\mathrm{Tr}(A) = \sum_{i=1}^{min(n,m)} A_{ii}$.

1.2 Matrix derivative

The derivation of a matrix A according to a matrix X (it equals zero if A is not an expression depending on X) reads :

$$\frac{\partial \mathbf{A}}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial \mathbf{A}}{\partial X_{1,1}} & \cdots & \frac{\partial \mathbf{A}}{\partial X_{1,m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{A}}{\partial X_{n,1}} & \cdots & \frac{\partial \mathbf{A}}{\partial X_{n,m}} \end{bmatrix} \tag{1}$$

We can then demonstrate the following identities :

$$\frac{d\,\mathrm{Tr}(\mathbf{AC})}{d\mathbf{A}} = \begin{bmatrix} \frac{\partial \sum_i \sum_k \mathbf{A}_{i,k}\mathbf{C}_{k,i}}{\partial A_{1,1}} & \cdots & \frac{\partial \sum_i \sum_k A_{i,k}C_{k,i}}{\partial A_{1,m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sum_i \sum_k A_{i,k}C_{k,i}}{\partial A_{n,1}} & \cdots & \frac{\partial \sum_i \sum_k A_{i,k}C_{k,i}}{\partial A_{n,m}} \end{bmatrix} = \begin{bmatrix} C_{1,1} & \cdots & C_{m,1} \\ \vdots & \ddots & \vdots \\ C_{1,n} & \cdots & C_{m,n} \end{bmatrix} = \mathbf{C}^T \tag{2}$$

$$\frac{d\,\mathrm{Tr}(\mathbf{ABA}^T)}{d\mathbf{A}} = \mathbf{AB} + \mathbf{AB}^T \tag{3}$$

$$= 2\mathbf{AB}, \text{ if } \mathbf{B} \text{ is symmetric} \tag{4}$$

1.3 Variance/covariance of random vectors

The covariance matrix of a random vector $\boldsymbol{x}$ is defined as :

$$\mathrm{cov}(\boldsymbol{x}) = E[(\boldsymbol{x} - E[\boldsymbol{x}]).(\boldsymbol{x} - E[\boldsymbol{x}])^T]$$

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are random vectors, their covariance is defined as :

$$\begin{aligned} \mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) &= E[(\boldsymbol{x} - E[\boldsymbol{x}]).(\boldsymbol{y} - E[\boldsymbol{y}])^T] \\ &= \mathrm{cov}(\boldsymbol{y}, \boldsymbol{x})^T \end{aligned}$$

The covariance of the sum of two random vectors $\boldsymbol{x}, \boldsymbol{y}$ reads :

$$\mathrm{cov}(\boldsymbol{x} + \boldsymbol{y}) = \mathrm{cov}(\boldsymbol{x}) + \mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) + \mathrm{cov}(\boldsymbol{y}, \boldsymbol{x}) + \mathrm{cov}(\boldsymbol{y})$$

By definition, if the two random vectors are independent, then $\mathrm{cov}(\boldsymbol{x}, \boldsymbol{y}) = 0$. Therefore, the covariance matrix of the sum $\boldsymbol{x} + \boldsymbol{y}$ of two independent random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ equals the sum of the covariance matrix of the two vectors :

$$\text{If } \boldsymbol{x} \text{ and } \boldsymbol{y} \text{ are independent}, \mathrm{cov}(\boldsymbol{x} + \boldsymbol{y}) = \mathrm{cov}(\boldsymbol{x}) + \mathrm{cov}(\boldsymbol{y}) \tag{5}$$

The covariance of a linear transformation of random vector $\boldsymbol{x}$ reads :

$$\mathrm{cov}(\mathbf{A}\boldsymbol{x}) = \mathbf{A}\,\mathrm{cov}(\boldsymbol{x})\mathbf{A}^T$$

Finally, for $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}$ three random vectors, we have the identity :

$$\mathrm{cov}(\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}) = \mathrm{cov}(\boldsymbol{x}_1, \boldsymbol{y}) + \mathrm{cov}(\boldsymbol{x}_2, \boldsymbol{y})$$

1.4 Assumptions behind kalman filtering

Todo
- The whiteness of noise
- only the first and second order moments are propagated

2 Discrete Kalman filter with linear state-space formulation

The presentation below follows Welch and Bishop (2006). Consider a state $x_k \in \mathbb{R}^n$ for which we get an observation $y_k \in \mathbb{R}^m$, the state evolving according to a linear recursive equation under the influence of an external input u_k and the observations being linearly linked to the state :

$$
\begin{aligned}
\boldsymbol{x}_k &= \mathbf{A}_k \boldsymbol{x}_{k-1} + \mathbf{B}_k \boldsymbol{u}_k + \boldsymbol{w}_k & (6) \\
\boldsymbol{y}_k &= \mathbf{H}_k \boldsymbol{x}_k + \boldsymbol{v}_k & (7)
\end{aligned}
$$

the evolution and observation being subject to some white and independent noises w_k, v_k. In addition to be zero-mean, these noises are supposed to be normally distributed with a variance/covariance matrix denoted respectively Q_k and R_k : $w_k \sim N(0, Q_k), v_k \sim N(0, R_k)$

Kalman filtering seeks to get the best estimate of the state x_k given the observations y_k, starting from an initial estimate $\hat{x}_0$, the uncertainty on this prior estimate P_0, and using a linear update of the form : $\hat{x}_k = \hat{x}_{k-1} + K(y_k - H\hat{x}_{k-1})$.

Let's denote the a priori estimate of the state $\hat{x}_k^-$ and the a posteriori estimate of the state $\hat{x}_k$. Let's denote e_k^- and e_k respectively the a priori and the a posteriori errors, and P_k^-, P_k their respective covariances :

$$
\begin{aligned}
e_k^- &= x_k - \hat{x}_k^- \\
P_k^- &= E[e_k^- e_k^{-T}] \\
e_k &= x_k - \hat{x}_k \\
P_k &= E[e_k e_k^T] & (8)
\end{aligned}
$$

The problem is to derive the expression of the Kalman gain that allows to get a linear update of an a priori estimate with the difference between the observation and a predicted observation and minimizing the a posteriori error on all the components of the state i.e. the trace of the covariance $e_k^T e_k = \text{Tr}(P_k)$.

2.1 Prediction

The first step of the Kalman filter is to predict the state, its covariance as well as the observations. It will be followed by a correction step which seeks to correct our estimate of the state given the discrepancy between the predicted observation and the real observation. Given an estimation of the state $\hat{\boldsymbol{x}}_{k-1}$ and its associated covariance $\mathbf{P}_{k-1}$, a prediction of the state at time k is defined as :

$$
\hat{\boldsymbol{x}}_k^- = \mathbf{A}_k \hat{\boldsymbol{x}}_{k-1} + \mathbf{B}_k \boldsymbol{u}_k \qquad (9)
$$

Given this prediction, we can compute its associated covariance using the evolution equation :

$$
\begin{aligned}
\mathbf{P}_{k-1}^- &= E[\boldsymbol{e}_k^- \boldsymbol{e}_k^{-T}] = E[(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-).(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-)^T] & (10) \\
&= E[(\boldsymbol{x}_k - \mathbf{A}_k \hat{\boldsymbol{x}}_{k-1} - \mathbf{B}_k \boldsymbol{u}_k).(\boldsymbol{x}_k - \mathbf{A}_k \hat{\boldsymbol{x}}_{k-1} - \mathbf{B}_k \boldsymbol{u}_k)^T] & (11) \\
&= E[(\mathbf{A}_k \boldsymbol{x}_{k-1} + \mathbf{B}_k \boldsymbol{u}_k + \boldsymbol{w}_k - \mathbf{A}_k \hat{\boldsymbol{x}}_{k-1} - \mathbf{B}_k \boldsymbol{u}_k).(\mathbf{A}_k \boldsymbol{x}_{k-1} + \mathbf{B}_k \boldsymbol{u}_k + \boldsymbol{w}_k - \mathbf{A}\hat{\boldsymbol{x}}_{k-1} - \mathbf{B}_k \boldsymbol{u}_k)^T] & (12) \\
&= E[\mathbf{A}_k(\boldsymbol{x}_{k-1} + \hat{\boldsymbol{x}}_{k-1}).(\boldsymbol{x}_{k-1} + \hat{\boldsymbol{x}}_{k-1})^T \mathbf{A}_k^T] + E[\boldsymbol{w}_k \boldsymbol{w}_k^T] & (13) \\
&= \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{Q}_k & (14)
\end{aligned}
$$

From equation (12) to equation (13), we make use of the identity (5) since $\mathbf{A}(\boldsymbol{x}_{k-1} + \hat{\boldsymbol{x}}_{k-1})$ and $\boldsymbol{w}_k$ are independent.

2.2 Correction

Given our prediction of the state, its covariance and the associated observations, we now want to linearly correct these estimates. Therefore we want to derive the expression of the Kalman gain $\mathbf{K}_k$ such that the correction $\hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k^- + \mathbf{K}_k(\boldsymbol{y}_k - \mathbf{H}\hat{\boldsymbol{x}}_k^-)$ minimizes the a posterior error made on each component of the state. The expression (8) can be expanded using the equations (7):

$$
\begin{aligned}
\mathbf{P}_k &= E[\boldsymbol{e}_k \boldsymbol{e}_k^T] = E[(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k)^T] \\
&= E[(\boldsymbol{x}_k - (\hat{\boldsymbol{x}}_k^- + \mathbf{K}_k(\boldsymbol{y}_k - \mathbf{H}_k\hat{\boldsymbol{x}}_k^-)))(\boldsymbol{x}_k - (\hat{\boldsymbol{x}}_k^- + \mathbf{K}_k(\boldsymbol{y}_k - \mathbf{H}_k\hat{\boldsymbol{x}}_k^-)))^T] \\
&= E[(\boldsymbol{x}_k - (\hat{\boldsymbol{x}}_k^- + \mathbf{K}_k(\mathbf{H}_k\boldsymbol{x}_k + \boldsymbol{v}_k - \mathbf{H}_k\hat{\boldsymbol{x}}_k^-)))(\boldsymbol{x}_k - (\hat{\boldsymbol{x}}_k^- + \mathbf{K}_k(\mathbf{H}_k\boldsymbol{x}_k + \boldsymbol{v}_k - \mathbf{H}_k\hat{\boldsymbol{x}}_k^-)))^T] \\
&= E[((\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-) - \mathbf{K}_k\boldsymbol{v}_k)((\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-) - \mathbf{K}_k\boldsymbol{v}_k)^T]
\end{aligned}
\tag{15}
$$

Given the noise w is white and independent, we can simplify the previous equation :

$$
\begin{aligned}
\mathbf{P}_k &= E[((\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-))((\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-))^T] + E[(K_k v_k)(K_k v_k)^T] \\
&= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k).E[(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-))(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-))^T].(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k E[\boldsymbol{v}_k \boldsymbol{v}_k^T]\mathbf{K}_k^T \\
&= (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_k^-(\mathbf{I} - \mathbf{K}_k\mathbf{H})^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\
&= \mathbf{P}_k^- + (\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T\mathbf{K}_k^T) - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^- - \mathbf{P}_k^-\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T
\end{aligned}
$$
$$\tag{16}$$
$$\tag{17}$$

The equation 16 is known as the Joseph form of the covariance update equation. As we said before, we seek to minimize the norm of the a posteriori error which is equivalent to minimizing the trace of the covariance matrix $\mathbf{P}_k$. Given the matrix $\mathbf{P}_k^-$ is symmetric ($\mathbf{P}_k^{-T} = \mathbf{P}_k^-$), we get the identity :

$$
\begin{aligned}
\mathrm{Tr}(\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-) &= \mathrm{Tr}((\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-)^T) = \mathrm{Tr}(\mathbf{P}_k^{-T}\mathbf{H}_k^T\mathbf{K}_k^T) \\
&= \mathrm{Tr}(\mathbf{P}_k^-\mathbf{H}_k^T\mathbf{K}_k^T)
\end{aligned}
$$

The trace of $\mathbf{P}_k$ therefore reads :

$$
\mathrm{Tr}(\mathbf{P}_k) = \mathrm{Tr}(\mathbf{P}_k^-) + \mathrm{Tr}(\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T\mathbf{K}_k^T) - 2\,\mathrm{Tr}(\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-) + \mathrm{Tr}(\mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T)
\tag{18}
$$

In order to minimize the previous trace, we compute its derivative relative to the gain K_k. Given the identities provided in section 1, we get :

$$
\frac{d\,\mathrm{Tr}(\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T\mathbf{K}_k^T)}{d\mathbf{K}_k} = 2\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}^T \text{ since } \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T \text{ is symmetric}
\tag{19}
$$

$$
\frac{d\,\mathrm{Tr}(\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-)}{d\mathbf{K}_k} = \mathbf{P}_k^-\mathbf{H}_k^T
\tag{20}
$$

$$
\frac{d\,\mathrm{Tr}(\mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T)}{d\mathbf{K}_k} = 2\mathbf{K}_k\mathbf{R}_k
\tag{21}
$$

$$
\frac{d\,\mathrm{Tr}(\mathbf{P}_k^-)}{d\mathbf{K}_k} = 0
\tag{22}
$$

Making use of the previous identities leads to:

$$
\frac{d\,\mathrm{Tr}(\mathbf{P}_k)}{d\mathbf{K}_k} = 2\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T - 2\mathbf{P}_k^-\mathbf{H}_k^T + 2\mathbf{K}_k\mathbf{R}_k
\tag{23}
$$

To get the optimal kalman gain, it is sufficient to set the previous derivative to zero (see Gelb (1986), p109) :

$$
\begin{aligned}
\mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T - \mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{K}_k\mathbf{R}_k &= 0 \tag{24} \\
\mathbf{K}_k(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k) &= \mathbf{P}_k^-\mathbf{H}_k^T \tag{25} \\
\mathbf{K}_k &= \mathbf{P}_k^-\mathbf{H}_k^T.(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{26}
\end{aligned}
$$

The previous equation gives the expression of the Kalman gain. We can now update the estimate of the state $\hat{\boldsymbol{x}}_k$ from the previous estimate $\hat{\boldsymbol{x}}_{k-1}$:

$$
\begin{aligned}
\hat{\boldsymbol{x}}_k^- &= \mathbf{A}_k\hat{\boldsymbol{x}}_{k-1} \tag{27} \\
\hat{\boldsymbol{x}}_k &= \hat{\boldsymbol{x}}_k^- + \mathbf{K}_k(\boldsymbol{y}_k - \mathbf{H}_k\hat{\boldsymbol{x}}_k^-) \tag{28}
\end{aligned}
$$

We finally need to know how to update the covariance matrix $\mathbf{P}_k$ from the previous estimate of the covariance of the state $\mathbf{P}_{k-1}$, which is simplified by injecting 26 into 17 :

$$
\begin{align}
\mathbf{P}_k^- &= \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}^T + \mathbf{Q}_k \tag{29} \\
\mathbf{P}_k &= \mathbf{P}_k^- + (\mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T \mathbf{K}_k^T) - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^- - \mathbf{P}_k^- \mathbf{H}_k^T \mathbf{K}_k^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T \tag{30} \\
&= \mathbf{P}_k^- - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^- + \mathbf{K}_k (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k) \mathbf{K}_k^T - \mathbf{P}_k^- \mathbf{H}_k^T \mathbf{K}_k^T \tag{31} \\
&= \mathbf{P}_k^- - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^- + \mathbf{P}_k^- \mathbf{H}_k^T \mathbf{K}_k^T - \mathbf{P}_k^- \mathbf{H}_k^T \mathbf{K}_k^T \tag{32} \\
&= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^- \tag{33}
\end{align}
$$

2.3 Summary

In summary, given an a priori estimate $\hat{\boldsymbol{x}}_0$ of the state and its associated covariance (uncertainty) $\mathbf{P}_0$, the equations of Kalman filtering are the following :

State-space formulation :

$$
\begin{align}
\boldsymbol{x}_k &= \mathbf{A}_k \boldsymbol{x}_{k-1} + \mathbf{B}_k \boldsymbol{u}_k + \boldsymbol{w}_k \tag{34} \\
\boldsymbol{y}_k &= \mathbf{H}_k \boldsymbol{x}_k + \boldsymbol{v}_k \tag{35}
\end{align}
$$

with $w_k \sim N(0, Q_k), v_k \sim N(0, R_k)$.

Prediction step:

$$
\begin{align}
\hat{\boldsymbol{x}}_k^- &= \mathbf{A}_k \hat{\boldsymbol{x}}_{k-1} + \mathbf{B}_k \boldsymbol{u}_k \tag{36} \\
\mathbf{P}_k^- &= \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{Q}_k \tag{37}
\end{align}
$$

Correction step:

$$
\begin{align}
\mathbf{K}_k &= \mathbf{P}_k^- \mathbf{H}_k^T . (\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{38} \\
\mathbf{P}_k &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^- \tag{39} \\
\hat{\boldsymbol{x}}_k &= \hat{\boldsymbol{x}}_k^- + \mathbf{K}_k (\boldsymbol{y}_k - \mathbf{H}_k \hat{\boldsymbol{x}}_k^-) \tag{40}
\end{align}
$$

2.4 Comments and extensions

2.4.1 Intuition about the shape of the equations

We may bring few commentaries on the previous equations. Especially the expression of the Kalman gain. The Kalman gain is proportionnal to the uncertainty on the state estimate $\mathbf{P}_k^-$ meaning that the mesurements will have more impact if we are less certain on the state. The kalman gain is also inversely proportionnal to the measurement uncertainty $(\mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k)$; i.e. if a measurement is very uncertain, it will be integrated with a small kalman gain and therefore have few impact on the state estimate.

TODO : we can reason on the uncertainty of the state; we may need to link P_k à P_{k-1} and see how the uncertainty on the measurements influences. Indeed it is equation 33 ?

see http://research.microsoft.com/en-us/um/people/zhang/INRIA/Publis/Tutorial-Estim/node15.html

TODO : can we say something on the norms of the different matrices ?

2.4.2 Extensions for speeding up and handling numerical instabilities

Square root KF

2.5 Examples

Let's consider the following problem. We throw a ball of a known mass m. We have a rather good estimate of its initial position (x_0, y_0) and a rather bad estimate of its initial velocity $\dot{v}_0$. The object is moving in the gravitational field and we observe only its position through a noisy sensor at a certain frequency $f = 1/\Delta t$ (see fig. 1). We want to estimate its position and velocity. The state space formulation of the problem, with an Euler discretization of the equation of motion, is the following :
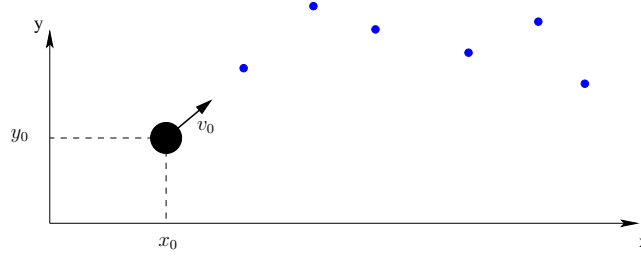

Figure 1: We seek to estimate the position and velocity of an object, of known mass, in the gravitational field. Only the position of the object is regularly sampled at frequency $1/\Delta t$ with a noisy sensor of known tolerance.

$$
\begin{bmatrix} x_k \\ y_k \\ \dot{x}_k \\ \dot{y}_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{k-1} \\ y_{k-1} \\ \dot{x}_{k-1} \\ \dot{y}_{k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ -mg\Delta t \end{bmatrix} + \boldsymbol{w}_k \tag{41}
$$

$$
\rightarrow \boldsymbol{x}_k = \mathbf{A}_k \boldsymbol{x}_{k-1} + \boldsymbol{u}_k + \boldsymbol{w}_k \tag{42}
$$

$$
\boldsymbol{z}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_k \\ y_k \\ \dot{x}_k \\ \dot{y}_k \end{bmatrix} + \boldsymbol{v}_k \tag{43}
$$

$$
\rightarrow \boldsymbol{z}_k = \mathbf{H}_k \boldsymbol{x}_k + \boldsymbol{v}_k \tag{44}
$$

Simulating the above example gives the illustration 2. We started the simulation by providing an estimate of the position at $(0,0)$ and a null velocity. The kalman filter integrates the observations every 100 ms while the simulator updates the position of the object every 10 ms. The observations are perturbated by a white noise with a standard deviation of 0.8. The ellipses in blue depict the uncertainty of the position. We used the diagonal elements of the covariance matrix $\mathbf{P}_k$ for the width and height of the ellipses, the ellipses being centered on the position estimate. The evolution noise is set to $1e - 2$ and the observation noise to 1.0. The mass of the ball is set to 0.1 kg and the gravity to 9.81. The object is initially thrown at from $(0,0)$ with a ramp at 45 with a speed amplitude of 4 m.s^{-1}. The initial estimate of the state of the Kalman filter is set $(x, y) = (0, 5), (\dot{x}, \dot{y}) = (0, 0)$ with a covariance set to identity.
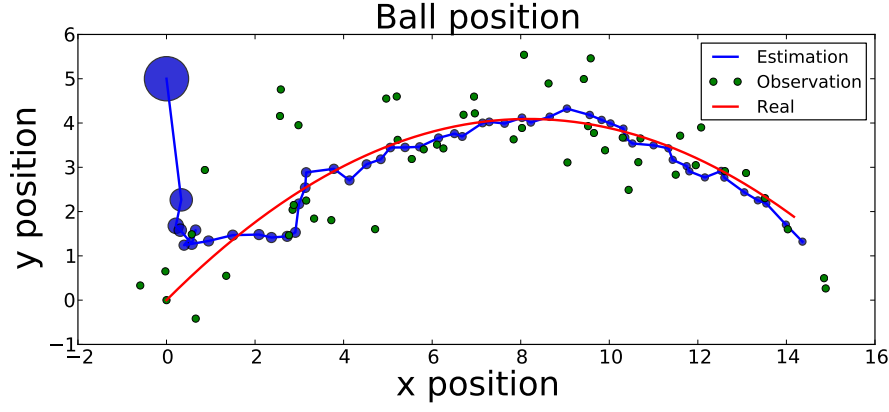
Figure 2: Simulation of a Kalman filter trying to estimate the trajectory of a ball launched with an initially unknown position and velocity.

3 EKF : using the jacobian to linearize a state-space formulation

3.1 Introduction

The Extended Kalman Filter can be applied to estimate the state of a system x_k given observations $y_k^\star$. The evolution function f characterizing the evolution of the system's state and the observation function h linking the state to the observations can be non-linear. However, EKF requires to compute the gradient of these functions relative to the state. EKF relies on a linearisation of the evolution and observation functions which are good approximations of the original functions if these functions are close to linear. The state-space formulation of EKF reads :

$$\begin{array}{rcl} \boldsymbol{x}_k & = & f(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_k \\ \boldsymbol{y}_k & = & h(\boldsymbol{x}_k) + \boldsymbol{v}_k \end{array} \tag{45}$$

where w and v are independent white noises. As always with Kalman filters, we look for the optimal gain K_k (the so-called Kalman gain) which updates an a priori estimate of the state $\hat{x}_k^-$ into an a posteriori estimate $\hat{x}_k$, minimizing the error between this a posteriori estimate and the the real state, given the observations. This can be done recursively given an a priori on the covariance of the error. A standard Kalman filter considers f and h to be linear (the evolution function is written $x_{k+1} = Fx_k + w_k$, the observation function is written $y_k = Hx_k + v_k$, F, H being matrices). Given this assumption, one can derive analytically the Kalman gain and get an update of the shape $\hat{x}_k = \hat{x}_k^- + K_k(y_k - H.\hat{x}_k^-)$. Non-linear evolution and observation functions are handled within EKF by linearising these functions around some estimates of the state; for example for the evolution function is linearized around the previous estimate of the state $\hat{x}_k$: $f(x_k) \approx f(\hat{x}_k) + \frac{\partial f}{\partial x}(\hat{x}_k)(x_k - \hat{x}_k)$. A standard Kalman filter can be applied on a new state-space formulation which is a linearisation of equations (45).

We start EKF with an estimate of the state $\hat{\boldsymbol{x}}_0$ and an associated uncertainty $\mathbf{P}_0$. The first step in applying EKF is to linearize the evolution function around the previous estimate of the state $\hat{\boldsymbol{x}}_{k-1}$:

$$\begin{array}{rcl} \boldsymbol{x}_k & \approx & f(\hat{\boldsymbol{x}}_{k-1}) + \mathbf{F}_{k-1}(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1}) + \boldsymbol{w}_k \\ & = & \mathbf{F}_{k-1}\boldsymbol{x}_{k-1} + (f(\hat{\boldsymbol{x}}_{k-1}) - \mathbf{F}_{k-1}\hat{\boldsymbol{x}}_{k-1}) + \boldsymbol{w}_k \qquad (46) \\ & = & \mathbf{F}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{u}_k + \boldsymbol{w}_k \qquad (47) \end{array}$$

where $\mathbf{F}_{k-1}$ is the Jacobian of the evolution function taken in $\hat{\boldsymbol{x}}_{k-1}$:

$$\mathbf{F}_{k-1} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\boldsymbol{x}) \end{bmatrix}_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{k-1}} \tag{48}$$

In equation (47) we recognize a linear evolution function for which we can apply the prediction step of Kalman filtering on linear state space equations. Therefore, we get the following prediction and associated covariance :

$$\hat{\boldsymbol{x}}_k^- = \mathbf{F}_{k-1}\hat{\boldsymbol{x}}_{k-1} + (f(\hat{\boldsymbol{x}}_{k-1}) - \mathbf{F}_{k-1}\hat{\boldsymbol{x}}_{k-1}) = f(\hat{\boldsymbol{x}}_{k-1}) \tag{49}$$

$$\mathbf{P}_k^- = \mathbf{F}_{k-1}\mathbf{P}_{k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}_k \tag{50}$$

For the correction step, we first linearize the observation function around the estimate of the state $\hat{\boldsymbol{x}}_k^-$:

$$\boldsymbol{y}_k \approx h(\hat{\boldsymbol{x}}_k^-) + \mathbf{H}_k.(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k^-) + \boldsymbol{v}_k \tag{51}$$

$$= \mathbf{H}_k\boldsymbol{x}_k + (h(\hat{\boldsymbol{x}}_k^-) - \mathbf{H}_k\hat{\boldsymbol{x}}_k^-) + \boldsymbol{v}_k \tag{52}$$

$$\boldsymbol{y}_k - (h(\hat{\boldsymbol{x}}_k^-) - \mathbf{H}_k\hat{\boldsymbol{x}}_k^-) \approx \mathbf{H}_k.(\boldsymbol{x}_k + \boldsymbol{v}_k \tag{53}$$

where $\mathbf{H}_k$ is the Jacobian of the observation function taken in $\hat{\boldsymbol{x}}_k^-$:

$$\mathbf{H}_k = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial h_1}{\partial x_n}(\boldsymbol{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial h_n}{\partial x_n}(\boldsymbol{x}) \end{bmatrix}_{\boldsymbol{x}=\hat{\boldsymbol{x}}_k^-} \tag{54}$$

We can now apply the correction step of kalman filters for linear state space :

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{55}$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^- \tag{56}$$

$$\hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k^- + \mathbf{K}_k(\boldsymbol{y}_k - (h(\hat{\boldsymbol{x}}_k^-) - \mathbf{H}_k\hat{\boldsymbol{x}}_k^-) - \mathbf{H}_k\hat{\boldsymbol{x}}_k^-)$$

$$= \hat{\boldsymbol{x}}_k^- + \mathbf{K}_k(\boldsymbol{y}_k - h(\hat{\boldsymbol{x}}_k^-)) \tag{57}$$

3.2 Summary

In summary, given an a priori estimate $\hat{\boldsymbol{x}}_0$ of the state and its associated covariance (uncertainty) $\mathbf{P}_0$, the equations of Extended Kalman filtering are the following :
State-space formulation :

$$\boldsymbol{x}_k = f(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_k \tag{58}$$

$$\boldsymbol{y}_k = h(\boldsymbol{x}_k) + \boldsymbol{v}_k \tag{59}$$

with $w_k \sim N(0, Q_k), v_k \sim N(0, R_k)$.
Prediction step:

$$\hat{\boldsymbol{x}}_k^- = f(\hat{\boldsymbol{x}}_{k-1}) \tag{60}$$

$$\mathbf{P}_k^- = \mathbf{F}_{k-1}\mathbf{P}_{k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}_k \tag{61}$$

Correction step:

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T.(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{62}$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^- \tag{63}$$

$$\hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k^- + \mathbf{K}_k(\boldsymbol{y}_k - h(\hat{\boldsymbol{x}}_k^-)) \tag{64}$$

with the following definitions of $\mathbf{F}_{k-1}$ and $\mathbf{H}_k$:

$$\mathbf{F}_{k-1} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\boldsymbol{x}) \end{bmatrix}_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{k-1}} \tag{65}$$

$$\mathbf{H}_k = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial h_1}{\partial x_n}(\boldsymbol{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial h_n}{\partial x_n}(\boldsymbol{x}) \end{bmatrix}_{\boldsymbol{x}=\hat{\boldsymbol{x}}_k^-} \tag{66}$$
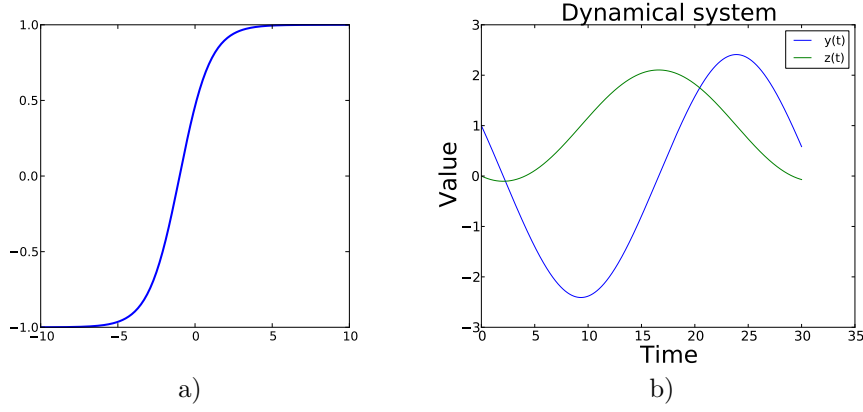
Figure 3: a) The transfer function is a sigmoid, almost linear in the xrange $[-3; 3]$ and highly non-linear outside of this range. b) The simulated system is oscillatory. Depending on the factor k, the system may have phases where $y(t)$ enters the non-linear range of the sigmoid and other phases where the evolution function can be well linearly approximated.

3.3 Examples

We can now handle non-linear evolution and observation functions. We first consider an example of this ability. Let's consider the following dynamical system :

$$\dot{x}(t) = \frac{2.0}{1 + exp(-(y(t) - 1))} - 1.0 \tag{67}$$

$$\dot{y}(t) = -k * x(t) \tag{68}$$

$$\begin{bmatrix} \dot{x}(t) \\ \dot{y}(t) \end{bmatrix} = F(\begin{bmatrix} x(t) \\ y(t) \end{bmatrix}) \tag{69}$$

$$x(0) = 1.0 \tag{70}$$

$$y(0) = 0.0 \tag{71}$$

The evolution of $x(t)$ depends non-linearly on $y(t)$, through a sigmoid which may introduce weak or strong non-linearity in the evolution function depending on the value of k which increases the range of evolution of $y(t)$ and ultimately impacts on whether or not the sigmoid is applied in its linear or non-linear range. The fig 3a shows a plot of the above sigmoid. Roughly speaking, if we are in the range $[-3; 3]$ the sigmoid is linear and it becomes non-linear out of this range. The figure 3b shows a simulation of this system. It is in the rising phase of $y(t)$ that the evolution function becomes non-linear. We will apply EKF to see how it behaves in this situation compared to situation where the evolution function is linear. The dynamical system is updated with a forward Euler method which keeps the evolution function simple but forces us to use a very small time step, here of $5e^{-4}s.$.

In order to use EKF, we need to define the state-space formulation. Using a forward Euler scheme, the state-space formulation reads :

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} = f(\begin{bmatrix} x_{k-1} \\ y_{k-1} \end{bmatrix}) + \boldsymbol{w}_k$$

$$= \begin{bmatrix} x_{k-1} \\ y_{k-1} \end{bmatrix} + \delta t F(\begin{bmatrix} x_{k-1} \\ y_{k-1} \end{bmatrix})$$

$$\boldsymbol{z}_k = h(\begin{bmatrix} x_k \\ y_k \end{bmatrix}) + \boldsymbol{v}_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix} + \boldsymbol{v}_k \tag{72}$$

For the simulations, we used $\delta t = 5.10^{-4}$. The results of applying EKF to estimate the state $(y(t), z(t))$ from noisy observations of the real system are shown on fig 4.
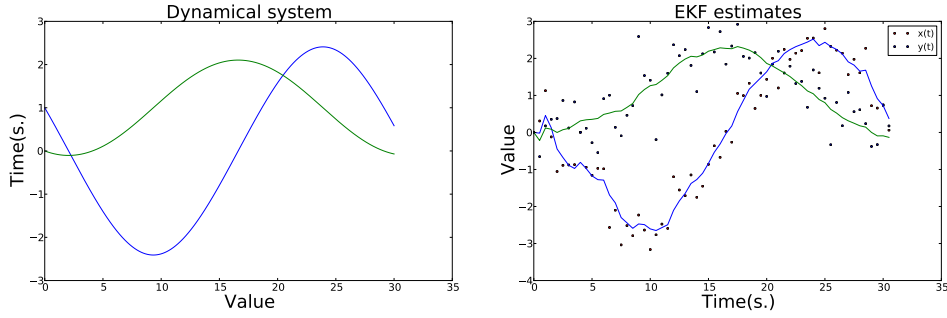
Figure 4: Application of EKF to estimate the state of a non-linear dynamical system.

4 UKF : estimating an arbitrary non-linear transformation of a random variable

TODO ....

As we saw in the previous section, the Extended Kalman Filter deals with non-linear evolution and observation functions by approximating them at the first order. It also requires to be able to compute the jacobian of these functions. This leads to two flaws of EKF : approximation errors when the evolution and observation functions are not locally linear and the inability of EKF to handle non-derivable functions. The Unscented Kalman Filter is another approach for handling non-linear evolution and observation functions. It relies on the idea that it is easier to approximate any non-linear transformation of a random variable than to approximate arbitrary non-linear functions (as done with the Jacobian in EKF) Julier and Uhlmann (2004). A thorough presentation of the unscented kalman filter algorithms can be found in van der Merwe (2004).

4.1 Introduction

The Unscented Kalman Filter relies on the Unscented Transform which tries to solve the following problem : *Let x be a random variable with mean $\bar{x}$ and covariance $\mathbf{P}_{xx}$. A second random variable y related to x as $y = f(x)$ is introduced. We wish to calculte the first two moments of y : its mean $\bar{y}$ and covariance $\mathbf{P}_{yy}$.*

For this problem, three desired properties are introduced :

- consistency : $\mathbf{P}_{yy} - E[(y - \bar{y}).(y - \bar{y})^T] \geq 0$

- efficiency : $\mathbf{P}_{yy} - E[(y - \bar{y}).(y - \bar{y})^T]$ must be minimized

- unbiased : $\bar{y} \approx E[y]$

Consistency means that the covariance of y is never under estimated. This is an important feature of Kalman filters. If the covariance is underestimated, the resulting kalman filter may be diverging. A consistent, efficient and unbiased transformation procedure can be examined by considering the Taylor series expansion of the expression of y around $\bar{x}$ Julier and Uhlmann (1997).

4.2 The Unscented Transform

La UT pourrait être utiliser pour estimer comment l'incertitude sur les valeurs des paramètres d'un modèle se propagent sur l'incertitude des réponses
It relies on the Unscented Transform schematically depicted on figure 5.

The problem that the unscented transform seeks to solve is to estimate the mean and variance of a non-linear transformation of a random variable. Starting with a random variable X of mean $\bar{X}$ and covariance Σ_X, we start
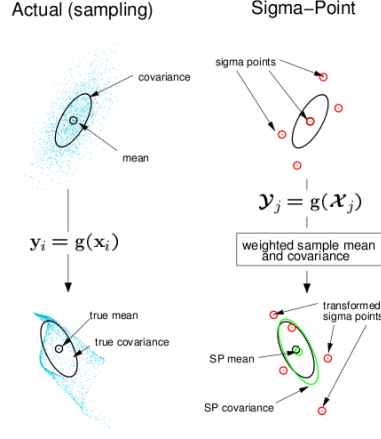
10

Figure 5: Illustration of the Unscented Transform

by defining a set of $2n + 1$ sigma points X_j (where n is the number of components of X). These sigma-points are defined such that their mean and covariance are the same than the random variable X, namely :

$$\overline{X} \;=\; \frac{1}{2n+1} \sum_j X_j \tag{73}$$

$$\Sigma_X \;=\; \frac{1}{2n+1} \sum_j (X_j - \overline{X})(X_j - \overline{X})^T \tag{74}$$

Then we compute the image, denoted $Y_j = g(X_j)$, of these $2n + 1$ sigma-points through the non-linear transformation. We now introduce a set of $2n + 1$ weights W_j such that the weighted mean and covariance of the images is as close as possible to the true mean and covariance of the transformed random variable :

$$\overline{g(X)} \;\approx\; \overline{Y} = \frac{1}{2n+1} \sum_j W_j Y_j \tag{75}$$

$$\Sigma_{g(X)} \;\approx\; \Sigma_Y = \frac{1}{2n+1} W_j (Y_j - \overline{Y})(Y_j - \overline{Y})^T \tag{76}$$

Now the question remains on how to define the sigma-points X_j and which values should we use for the weights W_j.

4.3 UKF for parameter estimation

4.4 UKF for state estimation

4.5 Dual estimation

References

Gelb, A. (1986). *Applied Optimal Estimation*. MIT Press.

Julier, S. and Uhlmann, J. (1997). A new extension of the kalman filter to nonlinear systems. In *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, volume 3, pages 21–24. Citeseer.

Julier, S. and Uhlmann, J. (2004). Unscented filtering and nonlinear estimation. In *Proc IEEE*, pages 401–422.

van der Merwe, R. (2004). *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space models*. PhD thesis, OGI School of Science & Engeneering at Oregon Health & Science University, Portland, OR, USA.

Welch, G. and Bishop, G. (2006). An introduction to the kalman filter. Technical Report Technical Report TR95-041, Department of Computer science - University of North California.