

# Kalman and Extended Kalman Filters: Concept, Derivation and Properties

Maria Isabel Ribeiro  
Institute for Systems and Robotics  
Instituto Superior Técnico  
Av. Rovisco Pais, 1  
1049-001 Lisboa PORTUGAL  
{mir@isr.ist.utl.pt}

©M. Isabel Ribeiro, 2004

February 2004

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>The Filtering Problem</b>	<b>3</b>
<b>3</b>	<b>Estimation of Random Parameters. General Results</b>	<b>8</b>
3.1	Problem Formulation . . . . .	8
3.2	Problem Reformulation . . . . .	10
3.3	Particularization for Gaussian Random Vectors . . . . .	12
<b>4</b>	<b>The Kalman Filter</b>	<b>14</b>
4.1	Kalman Filter dynamics . . . . .	15
4.2	One-step ahead prediction dynamics . . . . .	22
4.3	Kalman filter dynamics for a linear time-invariant system . . . . .	23
4.4	Steady-state Kalman filter . . . . .	24
4.5	Initial conditions . . . . .	25
4.6	Innovation Process . . . . .	27
4.7	The Kalman filter dynamics and the error ellipsoids . . . . .	29
<b>5</b>	<b>The Extended Kalman Filter</b>	<b>31</b>
5.1	Derivation of Extended Kalman Filter dynamics . . . . .	34

# Chapter 1

## Introduction

This report presents and derives the Kalman filter and the Extended Kalman filter dynamics. The general filtering problem is formulated and it is shown that, under linearity and Gaussian conditions on the systems dynamics, the general filter particularizes to the Kalman filter. It is shown that the Kalman filter is a linear, discrete time, finite dimensional time-varying system that evaluates the state estimate that minimizes the mean-square error.

The Kalman filter dynamics results from the consecutive cycles of prediction and filtering. The dynamics of these cycles is derived and interpreted in the framework of Gaussian probability density functions. Under additional conditions on the system dynamics, the Kalman filter dynamics converges to a steady-state filter and the steady-state gain is derived. The innovation process associated with the filter, that represents the novel information conveyed to the state estimate by the last system measurement, is introduced. The filter dynamics is interpreted in terms of the error ellipsoids associated with the Gaussian pdf involved in the filter dynamics.

When either the system state dynamics or the observation dynamics is non-linear, the conditional probability density functions that provide the minimum mean-square estimate are no longer Gaussian. The optimal non-linear filter propagates these non-Gaussian functions and evaluate their mean, which represents an high computational burden. A non optimal approach to solve the problem, in the frame of linear filters, is the Extended Kalman filter (EKF). The EKF implements a Kalman filter for a system dynamics that results from the linearization of the original non-linear filter dynamics around the previous state estimates.

# Chapter 2

## The Filtering Problem

This section formulates the general filtering problem and explains the conditions under which the general filter simplifies to a Kalman filter (KF).

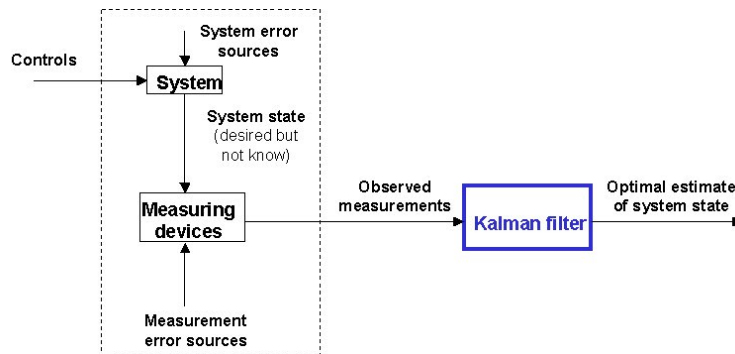


Figure 2.1: Typical application of the Kalman Filter

Figure 2.1, reproduced from [4], illustrates the application context in which the Kalman Filter is used. A physical system, (e.g., a mobile robot, a chemical process, a satellite) is driven by a set of external inputs or controls and its outputs are evaluated by measuring devices or sensors, such that the knowledge on the system’s behavior is solely given by the inputs and the observed outputs. The observations convey the errors and uncertainties in the process, namely the sensor noise and the system errors.

Based on the available information (control inputs and observations) it is required to obtain an estimate of the system’s state that optimizes a given criteria. This is the role played by a filter. In particular situations, explained in the following sections, this filter is a Kalman Filter.

The general filtering problem may be formulated along the following lines. Let

$$x(k+1) = f(x(k), u(k), w(k)) \quad (2.1)$$

$$y(k) = h(x(k), v(k)) \quad (2.2)$$

be the state dynamics of a general non-linear time-varying system, where

- $x \in \mathcal{R}^n$  is the system state vector,
- $f(\cdot, \cdot, \cdot)$  defines the system's dynamics,
- $u \in \mathcal{R}^m$  is the control vector,
- $w$  is the vector that conveys the system error sources,
- $y \in \mathcal{R}^r$  is the observation vector,
- $h(\cdot, \cdot, \cdot)$  is the measurement function,
- $v$  is the vector that represents the measurement error sources.

### Given

- $f, h$ , the noise characterization, the initial conditions,
- the set of controls,  $u(0), u(1), \dots, u(k-1)$ ,
- the set of measurements,  $y(1), y(1), y(2), \dots, y(k)$ ,

### obtain

- the **best estimate** of  $x(k)$ .

Any type of filter tries to obtain an optimal estimate of the desired quantities (the system's state) from data provided by a noisy environment. The concept of optimality expressed by the words **best estimate** corresponds to the minimization of the state estimation error in some respect.

Taking a Bayesian viewpoint, the filter propagates the conditional probability density function of the desired quantities, conditioned on the knowledge of the actual data coming from the measuring devices, i.e., the filter evaluates and propagates the conditional pdf

$$p(x(k)|y(1), \dots, y(k), u(0), \dots, u(k-1)) \quad (2.3)$$

for increasing values of  $k$ . This pdf conveys the amount of certainty on the knowledge of the value of  $x(k)$ .

Consider that, for a given time instant  $k$ , the sequence of past inputs and the sequence of past measurements are denoted by<sup>1</sup>

$$U_0^{k-1} = \{u_0, u_1, \dots, u_{k-1}\} \quad (2.4)$$

$$Y_1^k = \{y_1, y_2, \dots, y_k\}. \quad (2.5)$$

The entire system evolution and filtering process, may be stated in the following steps, [1], that considers the systems dynamics (2.1)-(2.2):

- **Given**  $x_0$ 
  - Nature apply  $w_0$ ,
  - We apply  $u_0$ ,
  - The system moves to state  $x_1$ ,
  - We make a measurement  $y_1$ .
- **Question:** which is the best estimate of  $x_1$ ?  
**Answer:** is obtained from  $p(x_1|Y_1^1, U_0^0)$ 
  - Nature apply  $w_1$ ,
  - We apply  $u_1$ ,
  - The system moves to state  $x_2$ ,
  - We make a measurement  $y_2$ .
- **Question:** which is the best estimate of  $x_2$ ?  
**Answer:** is obtained from  $p(x_2|Y_1^2, U_0^1)$ 
  - ...
  - ...
  - ...
  - ...
- **Question:** which is the best estimate of  $x_{k-1}$ ?  
**Answer:** is obtained from  $p(x_{k-1}|Y_1^{k-1}, U_0^{k-2})$ 
  - Nature apply  $w_{k-1}$ ,
  - We apply  $u_{k-1}$ ,
  - The system moves to state  $x_k$ ,

---

<sup>1</sup>Along this text  $u(i) = u_i$ ,  $y(i) = y_i$  and  $x(i) = x_i$ .

- We make a measurement  $y_k$ .
- **Question:** which is the best estimate of  $x_k$ ?  
**Answer:** is obtained from  $p(x_k|Y_1^k, U_0^{k-1})$
- ...
- ...
- ...
- ...

Therefore, aiming at obtaining the best state estimate, the filter propagates the conditional pdf for increasing values of  $k$ , and for each  $k$ , it obtains the estimate  $\hat{x}_k$  that optimizes a chosen criteria, as represented in the following diagram.

$$\begin{array}{lll}
 p(x_0) & & \\
 p(x_1|Y_1^1, U_0^0) & \longrightarrow & \hat{x}_1 \\
 p(x_2|Y_1^2, U_0^1) & \longrightarrow & \hat{x}_2 \\
 \vdots & \longrightarrow & \vdots \\
 p(x_{k-1}|Y_1^{k-1}, U_0^{k-2}) & \longrightarrow & \hat{x}_{k-1} \\
 p(x_k|Y_1^k, U_0^{k-1}) & \longrightarrow & \hat{x}_k
 \end{array}$$

Different optimization criteria may be chosen, leading to different estimates of the system's state vector. The estimate can be

- the **mean**, i.e., the center of the probability mass, corresponding to the minimum mean-square error criteria,
- the **mode** that corresponds to the value of  $x$  that has the highest probability, corresponding to the Maximum a Posterior (MAP) criteria,
- the **median**, where the estimate is the value of  $x$  such that half the probability weight lies to the left and half to the right of it.

For the conditional pdf represented in Figure 2.2 these criteria leads to different state estimates. So far, we formulated the general filtering problem. Under a set of particular conditions related with the linearity of the system (state and observation) dynamics and the normality of the random vectors involved (e.g., initial condition, state and measurement noise), the conditional probability density functions propagated by the filter are Gaussian for every  $k$ . The involved pdf are thus completely characterize by the mean vector and the covariance matrix. Rather

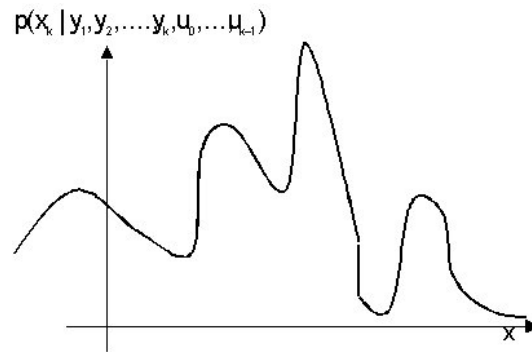


Figure 2.2: General conditional pdf

than propagating the entire pdf, the filter only propagates (recursively) the first and second moments of the conditional pdf. The general filter simplifies to what is known as the Kalman filter, whose dynamics is be derived in Section 4.

The Kalman filter dynamics will be derived as a general random parameter vector estimation. The KF filter evaluates the minimum mean-square error estimate of the random vector that is the system's state.

Results on the estimation of a general random parameter vector are presented in Section 3.

# Chapter 3

## Estimation of Random Parameters. General Results

This section presents basic results on the estimation of a random parameter vector based on a set of observations. This is the framework in which the Kalman filter will be derived, given that the state vector of a given dynamic system is interpreted as a random vector whose estimation is required. Deeper presentations of the issues of parameter estimation may be found, for example, in [3], [5], [10].

Let

$$\theta \in R^n \quad (3.1)$$

be a random vector, from which the available information is given by a finite set of observations

$$Y_1^k = [y(1), y(2), \dots, y(k-1), y(k)] \quad (3.2)$$

with no assumption on the dependency between  $y(i)$  and  $\theta$ .

Denote by

$$p(\theta, Y_1^k), \quad p(\theta|Y_1^k) \text{ e } p(Y_1^k)$$

the joint probability density function (pdf) of  $\theta$  and  $Y_1^k$ , the conditional pdf of  $\theta$  given  $Y_1^k$ , and the pdf of  $Y_1^k$ , respectively.

### 3.1 Problem Formulation

The estimation problem of the random vector  $\theta$  is stated, in general terms, as follows: given the observations  $y(1), y(2), \dots, y(k)$ , evaluate an estimate of  $\theta$ , i.e.,

$$\hat{\theta}(k) = f[y(i), \quad i = 1, \dots, k] \quad (3.3)$$

that optimizes a given criteria. Common optimization criteria are:

- the mean square error,
- the maximum a posterior.

In the sequel we will consider the mean-square error estimator, and therefore, the estimated value of the random vector is such that the cost function

$$J[\tilde{\theta}(k)] = E[\tilde{\theta}(k)^T \tilde{\theta}(k)] \quad (3.4)$$

is minimized, where  $\tilde{\theta}(k)$  stands for the estimation error given by

$$\tilde{\theta}(k) \triangleq \theta - \hat{\theta}(k). \quad (3.5)$$

According to the above formulated problem, the estimate  $\hat{\theta}(k)$  is given by

$$\hat{\theta}(k) = \operatorname{argmin} E[(\theta - \hat{\theta}(k))^T (\theta - \hat{\theta}(k))]. \quad (3.6)$$

We now show that minimizing  $E[\tilde{\theta}(k)^T \tilde{\theta}(k)]$  relative to  $\hat{\theta}(k)$  is equivalent to minimize the condition mean  $E[\tilde{\theta}(k)^T \tilde{\theta}(k) | Y_1^k]$  relative to  $\hat{\theta}(k)$ . In fact, from the definition of the mean operator, we have

$$E[\tilde{\theta}(k)^T \tilde{\theta}(k)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{\theta}(k)^T \tilde{\theta}(k) p(\theta, Y_1^k) d\theta dY_1^k \quad (3.7)$$

where  $d\theta = d\theta_1 d\theta_2 \dots d\theta_n$  and  $dY_1^k = dy_1 dy_2 \dots dy_k$ . Using the result obtained from Bayes law, (see e.g., [8])

$$p(\theta, Y_1^k) = p(\theta | Y_1^k) p(Y_1^k) \quad (3.8)$$

in (3.7) yields:

$$E[\tilde{\theta}(k)^T \tilde{\theta}(k)] = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \tilde{\theta}(k)^T \tilde{\theta}(k) p(\theta | Y_1^k) d\theta \right] p(Y_1^k) dY_1^k.$$

Moreover, reasoning about the meaning of the integral inside the square brackets, results

$$E[\tilde{\theta}(k)^T \tilde{\theta}(k)] = \int_{-\infty}^{\infty} E[\tilde{\theta}(k)^T \tilde{\theta}(k) | Y_1^k] p(Y_1^k) dY_1^k.$$

Therefore, minimizing the mean value of the left hand side of the previous equality relative to  $\hat{\theta}(k)$  is equivalent to minimize, relative to the same vector, the mean value  $E[\tilde{\theta}(k)^T \tilde{\theta}(k) | Y_1^k]$  on the integral on the right hand side. Consequently, the estimation of the random parameter vector can be formulated in a different way, as stated in the following subsection.

## 3.2 Problem Reformulation

Given the set of observations  $y(1), y(2), \dots, y(k)$ , the addressed problem is the derivation of an estimator of  $\theta$  that minimizes the conditional mean-square error, i.e.,

$$\hat{\theta}(k) = \operatorname{argmin} E[\tilde{\theta}(k)^T \tilde{\theta}(k) | Y_1^k]. \quad (3.9)$$

**Result 3.2.1** : *The estimator that minimizes the conditional mean-square error is the conditional mean, [5], [10],*

$$\hat{\theta}(k) = E[\theta | Y_1^k]. \quad (3.10)$$

**Proof:** From the definition of the estimation error in (3.5), the cost function in (3.9) can be rewritten as

$$J = E[(\theta - \hat{\theta}(k))^T (\theta - \hat{\theta}(k)) | Y_1^k] \quad (3.11)$$

or else,

$$J = E[\theta^T \theta - \theta^T \hat{\theta}(k) - \hat{\theta}(k)^T \theta + \hat{\theta}(k)^T \hat{\theta}(k) | Y_1^k] \quad (3.12)$$

$$= E[\theta^T \theta | Y_1^k] - E[\theta^T | Y_1^k] \hat{\theta}(k) - \hat{\theta}(k)^T E[\theta | Y_1^k] + E[\hat{\theta}(k)^T \hat{\theta}(k) | Y_1^k]. \quad (3.13)$$

The last equality results from the fact that, by definition (see (3.3)),  $\theta(k)$  is a function of  $Y_1^k$  and thus

$$E[\hat{\theta}(k) | Y_1^k] = \hat{\theta}(k).$$

If we add and subtract  $E[\theta^T | Y_1^k] E[\theta | Y_1^k]$  to (3.13) yields

$$J = E[\theta^T \theta | Y_1^k] - E[\theta^T | Y_1^k] E[\theta | Y_1^k] + [\hat{\theta}(k) - E[\theta | Y_1^k]]^T [\hat{\theta}(k) - E[\theta | Y_1^k]]$$

where the first two terms in the right hand side do not depend on  $\hat{\theta}(k)$ . The dependency of  $\hat{\theta}(k)$  on  $J$  results from a quadratic term, and therefore it is immediate that  $J$  achieves a minimum when the quadratic term is zero, and thus

$$\hat{\theta}(k) = E[\theta | Y_1^k],$$

which concludes the proof. □

**Corollary 3.2.1** : *Consider that  $f(Y_1^k)$  is a given function of the observations  $Y_1^k$ . Then, the estimation error is orthogonal to  $f(Y_1^k)$ ,  $\theta - \hat{\theta}(k) \perp f(Y_1^k)$ , this meaning that*

$$E[(\theta - \hat{\theta}(k)) f^T(Y_1^k)] = 0. \quad (3.14)$$

**Proof:** For the proof we use the following result on jointly distributed random variables. Let  $x$  and  $y$  be jointly distributed random variables and  $g(y)$  a function of  $y$ . It is known that, [8]

$$E[xg(y)] = E[E(x|y)g(y)] \quad (3.15)$$

where the outer mean-value operator in the right hand side is defined relative to the random variable  $y$ . Using (3.15) in the left hand side of (3.14) results

$$E[\tilde{\theta}(k)f^T(Y_1^k)] = E[E(\tilde{\theta}(k)|Y_1^k)f^T(Y_1^k)]. \quad (3.16)$$

Evaluating the mean value of the variable inside the square brackets in (3.16) leads to

$$E[\tilde{\theta}(k)|Y_1^k] = E[\theta|Y_1^k] - \hat{\theta}(k) \quad (3.17)$$

because  $\hat{\theta}(k)$  is known when  $Y_1^k$  is given. Therefore, (3.17) is zero, from where (3.14) holds, this concluding the proof.  $\square$

The particularization of the corollary for the case where  $f(Y_1^k) = \hat{\theta}(k)$  yields,

$$E[\tilde{\theta}(k)\hat{\theta}^T(k)] = 0. \quad (3.18)$$

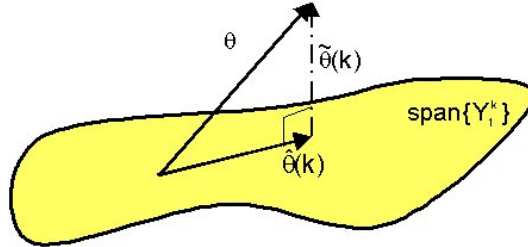


Figure 3.1: Minimum mean-square error orthogonal to  $\hat{\theta}(k)$

Figure 3.1 presents a graphical interpretation of the meaning of Corollary 3.2.1 and of its particularization. The space spanned by  $Y_1^k$  is represented by the yellow area. The parameter vector exact value,  $\theta$ , is not known. The information we have to estimate  $\theta$  is exclusively given by the set of observations, and therefore  $\hat{\theta}(k)$  lies in the space generated by  $Y - 1^k$ . The minimum-mean square error estimator is such that the estimation error is minimized. From the figure it is evident that the error is minimum when  $\hat{\theta}(k)$  is the orthogonal projection of  $\theta$  in the space

spanned by  $Y_1^k$ . Therefore, the estimation error  $\tilde{\theta}(k)$  is orthogonal to the space of the observations, as expressed in (3.18).

The results derived so far, made no assumptions on the type of the probability density functions involved. In the next subsection the previous results are particularized for the Gaussian case.

### 3.3 Particularization for Gaussian Random Vectors

The Result 3.2.1 is valid for any joint distribution of  $\theta$  and  $Y_1^k$ , i.e., it does not particularize the joint pdf of these variables.

It is well known from the research community dealing with estimation and filtering theory that many results simplify when assuming that the involved variables are Gaussian. This subsection discusses the simplifications resulting from considering that  $\theta$  and  $Y_1^k$  in Result 3.2.1 are jointly Gaussian.

**Result 3.3.1** *If  $\theta$  e  $Y_1^k$  are jointly Gaussian random vectors, then,*

$$E[\theta|Y_1^k] = E[\theta] + R_{\theta Y_1^k} R_{Y_1^k}^{-1} [Y_1^k - E[Y_1^k]], \quad (3.19)$$

where

$$R_{\theta Y_1^k} = E[(\theta - E(\theta))(Y_1^k - E(Y_1^k))^T], \quad (3.20)$$

$$R_{Y_1^k Y_1^k} = E[(Y_1^k - E(Y_1^k))(Y_1^k - E(Y_1^k))^T]. \quad (3.21)$$

The previous result is very important. It states that, when  $\theta$  e  $Y_1^k$  are jointly Gaussian, the estimator of  $\theta$  that minimizes the conditional mean-square error is a **linear** combination of the observations. In fact, note that (3.19) may be rewritten as

$$E[\theta|Y_1^k] = f(E(\theta), E(Y_1^k)) + \sum_{i=1}^k W_i Y_i, \quad (3.22)$$

making evident the linear combination of the observations involved.

When  $\theta$  and  $Y_1^k$ , are not jointly Gaussian then, in general terms,  $E[\theta|Y_1^k]$  is a **non linear function** of the observations.

**Result 3.3.2** *In the situation considered in Result 3.3.1,  $\hat{\theta}(k)$  is an unbiased estimate of  $\theta$ , i.e.,*

$$E[\hat{\theta}(k)] = E[\theta]. \quad (3.23)$$

**Result 3.3.3** *In the situation considered in Result 3.3.1,  $\hat{\theta}(k)$  is a minimum variance estimator.*

**Result 3.3.4** *In the situation considered in Result 3.3.1,  $\hat{\theta}(k)$  and  $\tilde{\theta}(k)$  are jointly distributed Gaussian random vectors.*

For the proofs of the three previous results see [5]. A result, related with Result 3.3.1, is now presented.

**Result 3.3.5** *Consider that  $\theta$  e  $Y_1^k$  are **not** jointly Gaussian, but  $E[\theta]$ ,  $E[Y_1^k]$ ,  $R_{Y_1^k Y_1^k}$  and  $R_{\theta Y_1^k}$  are known. Then, the **linear estimator** that minimizes the mean square error is (still) given by*

$$\hat{\theta}(k) = E[\theta] + R_{\theta Y_1^k} R_{Y_1^k}^{-1} (Y_1^k - E[Y_1^k]). \quad (3.24)$$

Note that the minimization in Result 3.3.5 is subject to the constraint of having a linear estimator while in Result 3.3.1 no constraint is considered. If the linear estimator constraint in Result 3.3.5 was not considered, the minimum mean square error estimator will generally yield an estimate  $\hat{\theta}(k)$  as a **non-linear** function of the observations.

# Chapter 4

## The Kalman Filter

Section 2 presented the filtering problem for a general nonlinear system dynamics. Consider now that the system represented in Figure 2.1 has a linear time-varying dynamics, i.e., that (2.1)-(2.2) particularizes to,

$$x_{k+1} = A_k x_k + B_k u_k + G w_k \quad k \geq 0 \quad (4.1)$$

$$y_k = C_k x_k + v_k \quad (4.2)$$

where  $x(k) \in \mathcal{R}^n$ ,  $u(k) \in \mathcal{R}^m$ ,  $w(k) \in \mathcal{R}^n$ ,  $v(k) \in \mathcal{R}^r$ ,  $y(k) \in \mathcal{R}^r$ ,  $\{w_k\}$  and  $\{v_k\}$  are sequences of white, zero mean, Gaussian noise with zero mean

$$E[w_k] = E[v_k] = 0, \quad (4.3)$$

and joint covariance matrix

$$E \left[ \begin{pmatrix} w_k \\ v_k \end{pmatrix} (w_k^T \ v_k^T) \right] = \begin{bmatrix} Q_k & 0 \\ 0 & R_k \end{bmatrix}. \quad (4.4)$$

The initial state,  $x_0$ , is a Gaussian random vector with mean

$$E[x_0] = \bar{x}_0 \quad (4.5)$$

and covariance matrix

$$E[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T] = \Sigma_0. \quad (4.6)$$

The sequence  $\{u_k\}$  is deterministic.

The problem of state estimation was formulated in Section 2. It can also be considered as the estimation of a random parameter vector, and therefore the results in Section 3 hold.

For the system (4.1)-(4.2), the Kalman filter is the filter that obtains the minimum mean-square state error estimate. In fact, when  $x(0)$  is a Gaussian vector, the state and observations noises  $w(k)$  and  $v(k)$  are white and Gaussian and the state and observation dynamics are linear,

1. the conditional probability density functions  $p(x_k|Y_1^k, U_0^{k-1})$  are Gaussian for any  $k$ ,
2. the mean, the mode, and the median of this conditional pdf coincide,
3. the Kalman filter, i.e., the filter that propagates the conditional pdf  $p(x_k|Y_1^k, U_0^{k-1})$  and obtains the state estimate by optimizing a given criteria, is the best filter among all the possible filter types and it optimizes any criteria that might be considered.

Let

$$p(x_k|Y_1^k, U_0^{k-1}) \sim \mathcal{N}(\hat{x}(k|k), P(k|k)) \quad (4.7)$$

represent the conditional pdf as a Gaussian pdf. The state estimate  $\hat{x}(k|k)$  is the conditional mean of the pdf and the covariance matrix  $P(k|k)$  quantifies the uncertainty of the estimate,

$$\begin{aligned} \hat{x}(k|k) &= E[x(k)|Y_1^k, U_0^{k-1}] \\ P(k|k) &= E[(x(k) - \hat{x}(k|k))(x(k) - \hat{x}(k|k))^T | Y_1^k, U_0^{k-1}]. \end{aligned}$$

Therefore, rather than propagating the entire conditional pdf, the Kalman filter only propagates the first and second moments. This is illustrated in Figure 4.1.

Subsection 4.1 derives the filter dynamics in terms of the mean and covariance matrix of the conditional pdf, i.e., it shows how the filter propagates the mean and the covariance matrix. This dynamics is recursive in the sense that to evaluate  $\hat{x}(k+1|k+1)$ , the Kalman filter only requires the previous estimate,  $\hat{x}(k|k)$  and the new observation,  $y(k+1)$ .

## 4.1 Kalman Filter dynamics

When  $v_k$ ,  $w_k$  and  $x_0$  are Gaussian vectors, the random vectors  $x_k$ ,  $x_{k+1}$ ,  $Y_1^k$  are jointly Gaussian. As discussed before, the Kalman filter propagates the Gaussian

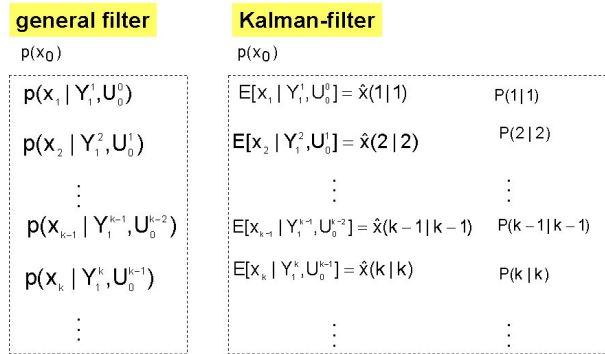


Figure 4.1: Propagation of the conditional pdf: General filter and Kalman filter

pdf  $p(x_k | Y_1^k, U_0^{k-1})$  and therefore the filter dynamics defines the general transition from  $p(x_k | Y_1^k, U_0^{k-1})$  to  $p(x_{k+1} | Y_1^{k+1}, U_0^k)$

$$\begin{array}{ll}
 Y_1^k = \{y_1, y_2, \dots, y_k\} & Y_1^{k+1} = \{Y_1^k, y_{k+1}\} \\
 U_0^{k-1} = \{u_0, u_1, \dots, u_{k-1}\} & U_0^k = \{U_0^{k-1}, u_k\}
 \end{array}$$

$$p(x_k | Y_1^k, U_0^{k-1}) \longrightarrow p(x_{k+1} | Y_1^{k+1}, U_0^k)$$

where both pdf are Gaussian and the input and observation information available at time instant  $k$  and  $k+1$  are displayed. Rather than being done directly, this transition is implemented as a two step-procedure, a prediction cycle and a filtering or update cycle, as represented in the diagram of Figure 4.2, where

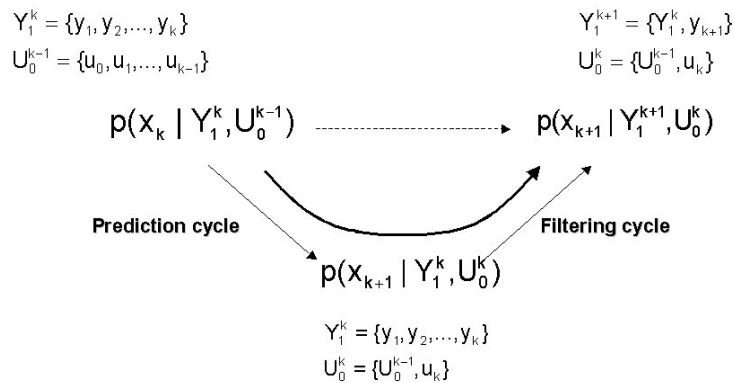


Figure 4.2: Prediction and Filtering cycles in the Kalman Filter dynamics

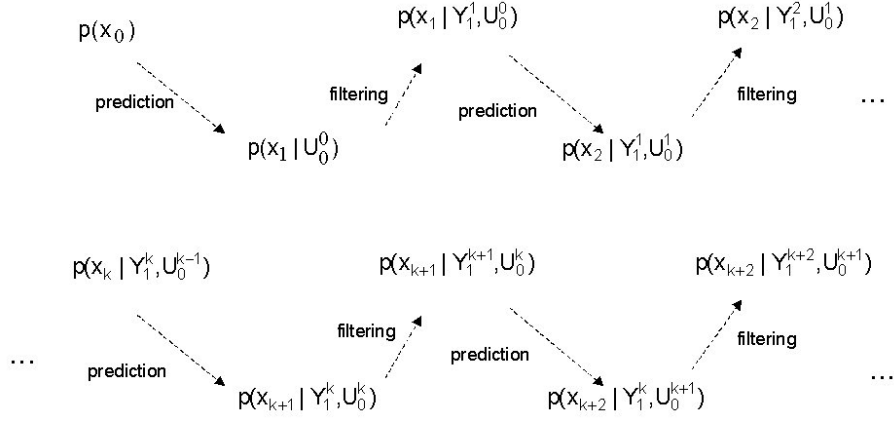


Figure 4.3: Consecutive prediction and filtering cycles on Kalman Filter dynamics

- $p(x_{k+1}|Y_1^k, U_0^k)$ , defined for time instant  $k + 1$ , represents what can be said about  $x(k + 1)$  **before** making the observation  $y(k + 1)$ .
- The filtering cycle states how to improve the information on  $x(k + 1)$  **after** making the observation  $y(k + 1)$ .

In summary, the Kalman filter dynamics results from a recursive application of prediction and filtering cycles, as represented in Figure 4.3.

Let

$$p(x_k|Y_1^k, U_0^{k-1}) \sim \mathcal{N}(\hat{x}(k|k), P(k|k)) \quad (4.8)$$

$$p(x_{k+1}|Y_1^k, U_0^k) \sim \mathcal{N}(\hat{x}(k+1|k), P(k+1|k)) \quad (4.9)$$

where  $\hat{x}(k|k)$  e  $\hat{x}(k+1|k)$  are given by

$$\hat{x}(k|k) = E[x(k)|Y_1^k, U_0^{k-1}] \quad (4.10)$$

$$\hat{x}(k+1|k) = E[x(k+1)|Y_1^k, U_0^k] \quad (4.11)$$

and

$$P(k|k) = E[(x_k - \hat{x}(k|k))(x_k - \hat{x}(k|k))^T | Y_1^k, U_0^{k-1}] \quad (4.12)$$

$$P(k+1|k) = E[(x_{k+1} - \hat{x}(k+1|k))(x_{k+1} - \hat{x}(k+1|k))^T | Y_1^k, U_0^k] \quad (4.13)$$

For the derivation of the filter's dynamics, assume, at this stage, that  $p(x_k|Y_1^k, U_0^{k-1})$ , is known, i.e.,  $\hat{x}(k|k)$  and  $P(k|k)$  are given.

**Step 1:** Evaluation of  $p(x_{k+1}|Y_1^k, U_0^k)$

State PREDICTION

This Gaussian pdf is completely characterized by the mean and covariance matrix. Applying the mean value operator on both sides of (4.1), yields

$$E[x_{k+1}|Y_1^k, U_0^k] = A_k E[x_k|Y_1^k, U_0^k] + B_k E[u_k|Y_1^k, U_0^k] + G E[w_k|Y_1^k, U_0^k]. \quad (4.14)$$

Taking (4.8) and (4.9) into account, considering that  $w_k$  e  $Y_1^k$  are independent random vectors and that  $w_k$  has zero mean, we obtain:

$$\hat{x}(k+1|k) = A_k \hat{x}(k|k) + B_k u_k. \quad (4.15)$$

Defining the **prediction error** as

$$\tilde{x}(k+1|k) \triangleq x(k+1) - \hat{x}(k+1|k) \quad (4.16)$$

and replacing in this expression the values of  $x(k+1)$  and  $\hat{x}(k+1|k)$  yields:

$$\tilde{x}(k+1|k) = A_k x_k + B_k u_k + G_k w_k - A_k \hat{x}(k|k) - B_k u_k = A_k \tilde{x}(k|k) + G_k w_k \quad (4.17)$$

where the **filtering error** was defined similarly to (4.16)

$$\tilde{x}(k|k) \triangleq x(k) - \hat{x}(k|k). \quad (4.18)$$

Given that  $\tilde{x}(k|k)$  and  $w_k$  are independent, from (4.17) we have

$$E[\tilde{x}(k+1|k)\tilde{x}(k+1|k)^T|Y_1^k, U_0^k] = A_k E[\tilde{x}(k|k)\tilde{x}(k|k)^T|Y_1^k, U_0^k] A_k^T + G_k Q G_k^T. \quad (4.19)$$

Including in (4.19) the notations (4.12) and (4.13) results:

$$P(k+1|k) = A_k P(k|k) A_k^T + G_k Q G_k^T. \quad (4.20)$$

The predicted estimate of the system's state and the associated covariance matrix in (4.15) and (4.20), correspond to the best knowledge of the system's state at time instant  $k+1$  before making the observation at this time instant. Notice that the prediction dynamics in (4.15) follows exactly the system's dynamics in (4.1), which is the expected result given that the system noise has zero mean.

**Step 2:** Evaluation of  $p(y_{k+1}|Y_1^k, U_0^k)$

Measurement PREDICTION

From (4.2), it is clear that

$$p(y_{k+1}|Y_1^k, U_0^k) = p(C_{k+1}x_{k+1} + v_{k+1}|Y_1^k, U_0^k) \quad (4.21)$$

and thus, as this is a Gaussian pdf, the **predicted measurement** is given by

$$\hat{y}(k+1|k) = E[y_{k+1}|Y_1^k, U_0^k] = C_{k+1}\hat{x}_{k+1|k}. \quad (4.22)$$

Defining the measurement prediction error as

$$\tilde{y}(k+1|k) \triangleq y_{k+1} - \hat{y}(k+1|k), \quad (4.23)$$

and replacing the values of  $y(k+1)$  and  $\hat{y}(k+1|k)$  results:

$$\tilde{y}(k+1|k) = C_{k+1}\hat{x}(k+1|k) + v_{k+1}. \quad (4.24)$$

Therefore, the covariance matrix associated to (4.24) is given by

$$P_{\tilde{y}}(k+1|k) = C_{k+1}P(k+1|k)C_{k+1}^T + R_k. \quad (4.25)$$

Multiplying  $x_{k+1}$  on the right by  $\tilde{y}(k+1|k)^T$  and using (4.24) we obtain:

$$x_{k+1}\tilde{y}^T(k+1|k) = x_{k+1}\hat{x}(k+1|k)^T C_{k+1}^T + x_{k+1}v_{k+1}^T$$

from where

$$E[x_{k+1}\tilde{y}^T(k+1|k)] = P(k+1|k)C_{k+1}^T. \quad (4.26)$$

Given the predicted estimate of the state at time instant  $k+1$  knowing all the observations until  $k$ ,  $\hat{x}(k+1|k)$  in (4.15), and taking into account that, in the linear observation dynamics (4.2) the noise has zero mean, it is clear that the predicted measurement (4.22) follows the same observation dynamics of the real system.

**Step 3:** Evaluation of  $p(x_{k+1}|Y_1^{k+1}, U_0^k)$

FILTERING

To evaluate the conditional mean of  $x_{k+1}$  note that

$$Y_1^{k+1} \text{ e } \{Y_1^k, \tilde{y}(k+1|k)\}$$

are equivalent from the view point of the contained information. Therefore,

$$E[x_{k+1}|Y_1^{k+1}, U_0^k] = E[x_{k+1}|Y_1^k, \tilde{y}(k+1|k), U_0^k]. \quad (4.27)$$

On the other hand,  $Y_1^k$  and  $\tilde{y}(k+1|k)$  are independent (see Corollary 3.2.1 in Section 3) and therefore

$$\hat{x}(k+1|k+1) = E[x(k+1)|Y_1^k] + E[x_{k+1}, \tilde{y}^T(k+1|k)P_{\tilde{y}(k+1|k)}^{-1}]\tilde{y}(k+1|k)$$

which is equivalent to,

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + P(k+1|k)C_{k+1}^T [C_{k+1}P(k+1|k)C_{k+1}^T + R]^{-1} [y(k+1) - C_{k+1}\hat{x}(k+1|k)] \quad (4.28)$$

Defining the Kalman gain as

$$K(k+1) = P(k+1|k)C_{k+1}^T [C_{k+1}P(k+1|k)C_{k+1}^T + R]^{-1} \quad (4.29)$$

equation (4.28) may be rewritten as

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + \underbrace{P(k+1|k)C_{k+1}^T [C_{k+1}P(k+1|k)C_{k+1}^T + R]^{-1}}_{K(k+1)} \underbrace{[y(k+1) - C_{k+1}\hat{x}(k+1|k)]}_{\hat{y}(k+1|k)} \quad (4.30)$$

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + K(k+1)[y(k+1) - C_{k+1}\hat{x}(k+1|k)] \quad (4.31)$$

from where we can conclude that, the filtered state estimate is obtain from the predicted estimate as,

$$\text{filtered state estimate} = \text{predicted state estimate} + \text{Gain} * \text{Error}$$

The Gain is the Kalman gain defined in (4.29). The gain multiplies the error. The error is given by  $[y(k+1) - C_{k+1}\hat{x}(k+1|k)]$ , i.e., is the difference between the real measurement obtained at time instant  $k+1$  and measurement prediction obtained from the predicted value of the state. It states the novelty or the new information that the new observation  $y(k+1)$  brought to the filter relative to the state  $x(k+1)$ .

Defining the filtering error as,

$$\tilde{x}(k+1|k+1) \triangleq x(k+1) - \hat{x}(k+1|k+1)$$

and replacing in (4.28) yields:

$$\tilde{x}(k+1|k+1) = \tilde{x}(k+1|k) - P(k+1|k)C_{k+1}^T [C_{k+1}P_{k+1|k}C_{k+1}^T + R]^{-1} [C_{k+1}\tilde{x}(k+1|k) + v_{k+1}]$$

from where

$$P(k+1|k+1) = P(k+1|k) - P(k+1|k)C_{k+1}^T [C_{k+1}P_{k+1|k}C_{k+1}^T + R]^{-1} C_{k+1}P(k+1|k). \quad (4.32)$$

**Summary:**

Prediction:

$$\hat{x}(k+1|k) = A_k \hat{x}(k|k) + B_k u_k \quad (4.33)$$

$$P(k+1|k) = A_k P(k|k) A_k^T + G_k Q G_k^T \quad (4.34)$$

### Filtering

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)[y(k) - C_k \hat{x}(k|k-1)] \quad (4.35)$$

$$K(k) = P(k|k-1) C_k^T [C_k P(k|k-1) C_k^T + R]^{-1} \quad (4.36)$$

$$P(k|k) = [I - K(k) C_k] P(k|k-1) \quad (4.37)$$

### Initial conditions

$$\hat{x}(0|-1) = \bar{x}_0 \quad (4.38)$$

$$P(0|-1) = \Sigma_0 \quad (4.39)$$

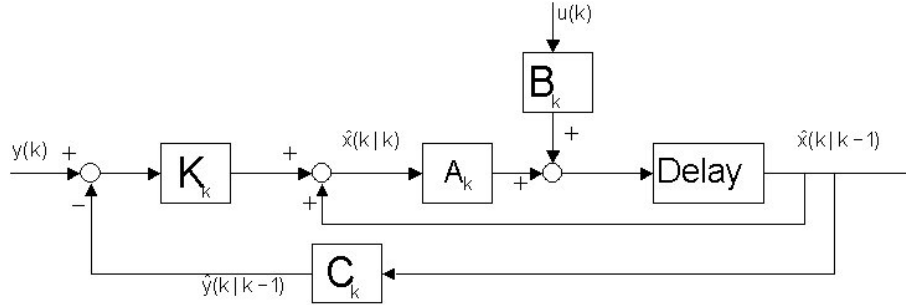


Figure 4.4: Block diagram of the Kalman filter

Figure 4.4 presents the block diagram of the Kalman filter. We list a number of important properties of the Kalman Filter:

- the Kalman Filter is a linear, discrete time, finite dimensional time-varying system, whose inputs are the system inputs,  $u_0, u_1, \dots, u_{k-1}$ , and the system measurements,  $y_0, y_1, \dots, y_k$ . The output is the process  $\hat{x}(k|k-1)$  or  $\hat{x}(k|k)$ ,
- The conditional error covariance matrix

$$P(k|k-1) = E[(x(k) - \hat{x}(k|k-1) (x(k) - \hat{x}(k|k-1))^T | Y_1^{k-1}, U_0^{k-1}$$

is actually independent of  $Y_1^{k-1}$ , which means that no one set of measurements helps any more than other to eliminate some uncertainty about  $x(k)$ . The filter gain,  $K(k)$  is also independent of  $Y_1^{k-1}$ . Because of this, the error covariance  $P(k|k-1)$  and the filter gain  $K(k)$  can be computed before the filter is actually run. This is not generally the case in nonlinear filters.

Some other useful properties will be discussed in the following sections.

## 4.2 One-step ahead prediction dynamics

Using simultaneously (4.33) and (4.35) the filter dynamics is written in terms of the state predicted estimate,

$$\hat{x}(k+1|k) = A_k[I - K(k)C_k]\hat{x}(k|k-1) + B_k u_k + A_k K(k)y_k \quad (4.40)$$

with initial condition

$$\hat{x}(0|-1) = \bar{x}_0 \quad (4.41)$$

where,

$$K(k) = P(k|k-1)C_k^T [C_k P(k|k-1)C_k^T + R]^{-1} \quad (4.42)$$

$$P(k+1|k) = A_k P(k|k-1)A_k^T - A_k K(k)C_k P(k|k-1)A(k)^T + G_k Q G_k^T \quad (4.43)$$

$$P(0|-1) = \Sigma_o \quad (4.44)$$

Equation (4.44) may be rewritten differently by replacing the gain  $K(k)$  by its value given by (4.42),

$$P(k+1|k) = A_k P(k|k-1)A_k^T + G_k Q G_k^T - A_k P(k|k-1)C_k^T [C_k P(k|k-1)C_k^T + R]^{-1} C_k P(k|k-1)A_k^T \quad (4.45)$$

or else,

$$P(k+1|k) = A_k P(k|k-1)A_k^T + G_k Q G_k^T - A_k K(k)[C_k P(k|k-1)C_k^T + R]K^T(k)A_k^T. \quad (4.46)$$

which is a Riccati equation.

From the definition of the predicted and filtered errors in (4.16) and (4.18), and the above recursions, it is immediate that

$$\tilde{x}(k+1|k) = A_k \tilde{x}(k|k) + G_k w_k \quad (4.47)$$

$$\tilde{x}(k|k) = [I - K(k)C_k]\tilde{x}(k|k-1) - K(k)v_k \quad (4.48)$$

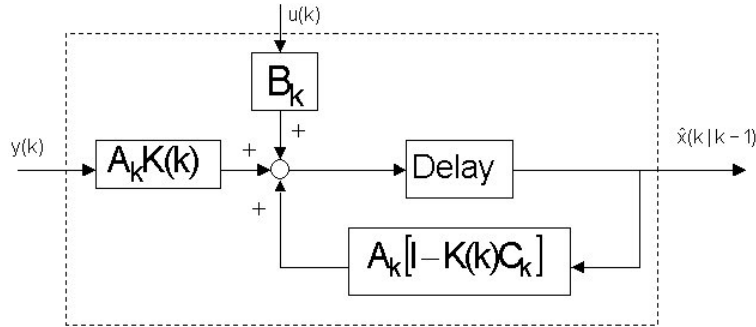


Figure 4.5: Block diagram of the Kalman filter prediction dynamics

from where it is possible to write the dynamics of the one-step prediction error,

$$\tilde{x}(k+1|k) = A_k[I - K(k)C_k]\tilde{x}(k|k-1) - A_k K(k)v_k + G_k w_k. \quad (4.49)$$

Evaluating the mean of the above equation, and taking into account that  $w_k$  and  $v_k$  are zero mean sequences, yields,

$$E[\tilde{x}(k+1|k)] = A_k[I - K(k)C_k]E[\tilde{x}(k|k-1)]. \quad (4.50)$$

showing that it satisfies an homogeneous dynamics.

The one step-ahead prediction dynamics of the Kalman filter is represented in Figure 4.5.

### 4.3 Kalman filter dynamics for a linear time-invariant system

Consider now that the linear system (4.1)-(4.2) is time-invariant, i.e.  $A_k = A$ ,  $B_k = B$ ,  $G_k = G$ ,  $C_k = C$ ,  $\forall k \geq 0$ , and that  $Q_k = Q$ ,  $R_k = R$ , i.e.,

$$x_{k+1} = Ax_k + Bu_k + Gw_k \quad k \geq 0 \quad (4.51)$$

$$y_k = Cx_k + v_k \quad (4.52)$$

with  $w_k$  and  $v_k$  mutually independent sequences of zero mean white Gaussian noise with joint covariance matrix

$$E \left[ \begin{pmatrix} w_k \\ v_k \end{pmatrix} (w_k^T v_k^T) \right] = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}. \quad (4.53)$$

The initial condition,  $x(0)$  is Gaussian with mean  $\bar{x}_0$  and covariance  $\Sigma_0$ .

The Kalman filter dynamics is obtained by the particularization of the general time-varying dynamics for the time-invariant situation, i.e.,

$$\hat{x}(k+1|k) = A\hat{x}(k|k-1) + Bu_k + K(k)[y_k - C\hat{x}_{k|k-1}] \quad (4.54)$$

$$K(k) = P(k|k-1)C^T[CP(k|k-1)C^T + R]^{-1} \quad (4.55)$$

$$P(k+1|k) = AP(k|k-1)A^T + GQG^T - AK(k)[CP(k|k-1)C^T + R]K^T(k)A \quad (4.56)$$

Note that, even though the original system is time-invariant, the Kalman Filter is a **time-varying** linear system, given that in (4.54) the Kalman gain is a function of  $k$ .

Equation (4.56) is known as a discrete Riccati equation. In the sequel, we discuss the conditions under which the Riccati equation converges.

Under certain conditions, detailed in the following subsection, the Kalman gain converges to a steady-state value. The corresponding filter is known as the steady-state Kalman filter.

## 4.4 Steady-state Kalman filter

Consider the system dynamics (4.51)-(4.52) and assume the following additional assumptions:

1. The matrix  $Q = Q^T > 0$ , i.e., is a positive definite matrix,
2. The matrix  $R = R^T > 0$ , i.e., is a positive definite matrix,
3. The pair  $(A, G)$  is controllable, i.e.,

$$\text{rank}[G \mid AG \mid A^2G \mid \dots \mid A^{n-1}G] = n,$$

4. The pair  $(A, C)$  is observable, i.e.,

$$\text{rank}[C^T \mid A^T C^T \mid A^{T^2} C^T \mid \dots \mid A^{T^{n-1}} C^T] = n.$$

**Result 4.4.1** *Under the above conditions,*

1. The prediction covariance matrix  $P(k|k-1)$  converges to a constant matrix,

$$\lim_{k \rightarrow \infty} P(k|k-1) = \bar{P}$$

where  $\bar{P}$  is a symmetric positive definite matrix,  $P = P^T > 0$ .

2.  $\bar{P}$  is the unique positive definite solution of the discrete algebraic Riccati equation

$$\bar{P} = A\bar{P}A^T - A\bar{P}C^T[C\bar{P}C^T + R]^{-1}C\bar{P}A^T \quad (4.57)$$

3.  $\bar{P}$  is independent of  $\Sigma_0$  provided that  $\Sigma_0 \geq 0$ .

**Proof:** see [2].

As a consequence of Result 4.4.1, the filter gain in (4.55) converges to

$$\bar{K} = \lim_{k \rightarrow \infty} K(k) = \bar{P}C^T[C\bar{P}C^T + R]^{-1} \quad (4.58)$$

i.e., in steady-state the Kalman gain is constant and the filter dynamics is time-invariant.

## 4.5 Initial conditions

In this subsection we discuss the initial conditions considered both for the system and for the Kalman filter. With no loss of generality, we will particularize the discussion for null control inputs,  $u_k = 0$ .

### System

Let

$$\begin{cases} x_{k+1} &= Ax_k + Gw_k, k \geq 0 \\ y_k &= Cx_k + v_k \end{cases} \quad (4.59)$$

where

$$E[x_0] = \bar{x}_0 \quad (4.60)$$

$$\Sigma_0 = E[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T] \quad (4.61)$$

and the sequences  $\{v_k\}$  and  $\{w_k\}$  have the statistical characterization presented in Section 2.

Applying the mean value operator to both sides of (4.59) yields

$$E[x_{k+1}] = AE[x_k]$$

whose solution is

$$E[x_k] = A^k \bar{x}_0. \quad (4.62)$$

Thus, if  $\bar{x}_0 \neq 0$ ,  $\{x_k\}$  is not a stationary process. Assume that the following hypothesis hold:

**Hypothesis:**  $x_0 = 0$

The constant variation formula applied to (4.59) yields

$$x(l) = A^{l-k}x(k) + \sum_{j=k}^{l-1} A^{l-1-j}Gw_j. \quad (4.63)$$

Multiplying (4.63) on the right by  $x^T(k)$  and evaluating the mean value, results:

$$E[x(l)x^T(k)] = A^{l-k}E[x(k)x(k)^T], l \geq k.$$

Consequently, for  $x(k)$  to be stationary,  $E[x(l)x^T(k)]$  should not depend on  $k$ . Evaluating  $E[x(k)x(k)^T]$  for increasing values of  $k$  we obtain:

$$\begin{aligned} E[x(0)x(0)^T] &= \Sigma_0 & (4.64) \\ E[x(1)x(1)^T] &= E[(Ax(0) + Gw(0))(x^T(0)A^T + w^T(0)G^T)] = A\Sigma_0A^T + GQG^T & (4.65) \\ E[x(2)x(2)^T] &= AE[(x(1)x(1)^T)A^T + GQG^T] = A^2\Sigma_0A^{2T} + AGQG^TA^T + GQG^T & (4.66) \end{aligned}$$

from where

$$E[x(k)x(k)^T] = AE[(x(k-1)x(k-1)^T)A^T + GQG^T]. \quad (4.67)$$

Therefore, the process  $\{x_k\}$  is stationary if and only if

$$\Sigma_0 = A\Sigma_0A^T + GQG^T.$$

Remark, however, that this stationarity condition is not required for the application of the Kalman filter nor it degrades the filter performance.

**Kalman filter**

The filter initial conditions, given, for example, in terms of the one-step prediction are:

$$\hat{x}(0 | -1) = \bar{x}_0 \quad (4.68)$$

$$P(0 | -1) = \Sigma_0, \quad (4.69)$$

which means that the first state prediction has the same statistics as the initial condition of the system. The above conditions have an intuitive explanation. In the absence of system measurements (i.e., formally at time instant  $k = -1$ ), the **best** that can be said in terms of the state prediction at time instant 0 is that this prediction coincides with the mean value of the random vector that is the system initial state.

As will be proved in the sequel, the choice of (4.68) and (4.69) leads to unbiased state estimates for all  $k$ . When the values of  $\bar{x}_0$  and  $\Sigma_0$  are not a priori known, the filter initialization cannot reflect the system initial conditions. A possible choice is

$$\hat{x}(0 | -1) = 0 \quad (4.70)$$

$$P(0 | -1) = P_0 = \alpha I. \quad (4.71)$$

## 4.6 Innovation Process

The process

$$e(k) = y(k) - \hat{y}(k|k-1) \quad (4.72)$$

is known as the innovation process. It represents the component of  $y(k)$  that cannot be predicted at time instant  $k-1$ . In other words, it represents the innovation, the novelty that  $y(k)$  brings to the system at time instant  $k$ . This process has some important characteristics, that we herein list.

**Property 4.6.1** *The innovation process has zero mean.*

**Prof:**

$$\begin{aligned} E[e(k)] &= E[y(k) - \hat{y}(k|k-1)] \\ &= E[Cx(k) + v(k) - C\hat{x}(k|k-1)] \\ &= CE[\tilde{x}(k|k-1)] \end{aligned}$$

given that  $\{v_k\}$  is zero mean. For a time-invariant system, the prediction error dynamics is given by (4.50) that is a homogeneous dynamics. The same conclusion holds for a time-invariant system. For  $k = 0$ ,  $E[\tilde{x}(0| - 1)] = 0$  given that

$$E[\tilde{x}(0| - 1)] = E[x_0] - \hat{x}(0| - 1)$$

and we choose  $\hat{x}(0| - 1) = \bar{x}_0$  (see 4.38). Therefore, the mean value of the prediction error is zero, and in consequence, the innovation process has zero mean.  $\square$

The above proof raises a comment relative to the initial conditions chosen for the Kalman filter. According to (4.50) the prediction error has a homogeneous dynamics, and therefore an initial null error leads to a null error for every  $k$ . If  $\hat{x}_{0|-1} \neq \bar{x}_0$  the initial prediction error is not zero. However, under the conditions for which there exists a steady solution for the discrete Riccati equation, the error asymptotically converges to zero.

**Property 4.6.2** *The innovation process is white.*

**Proof:** In this proof we will consider that  $\hat{x}_{0|-1} = \bar{x}_0$ , and thus  $E[e(k)] = 0$ , i.e., the innovation process is zero mean. We want to prove that

$$E[e(k)e^T(j)] = 0$$

for  $k \neq j$ . For simplicity we will consider the situation in which  $j = k + 1$ ; this is not the entire proof, but rather a first step towards it. From the definition of the innovation process, we have :

$$e(k) = C\tilde{x}(k|k - 1) + v_k$$

and thus

$$\begin{aligned} E[e(k)e^T(k + 1)] &= CE[\tilde{x}(k|k - 1)\tilde{x}(k + 1|k)]C^T + CE[\tilde{x}(k|k - 1)v_{k+1}^T] \\ &\quad + E[v_kv_k^T(k + 1|k)]C^T + E[v_kv_{k+1}^T] \end{aligned} \quad (4.73)$$

As  $\{v_k\}$  has zero mean and is a white process, the second and fourth terms in (4.73) are zero. We invite the reader to replace (4.33) and (4.35) in the above equality and to conclude the demonstration.

**Property 4.6.3**

$$E[e(k)e^T(k)] = CP(k|k - 1)C^T + R$$

**Property 4.6.4**

$$\lim_{k \rightarrow \infty} E[e(k)e^T(k)] = C\bar{P}C^T + R$$

## 4.7 The Kalman filter dynamics and the error ellipsoids

In previous subsections we demonstrate that,

$$\begin{aligned} p(x(k+1)|Y_1^k, U_0^k) &\sim \mathcal{N}(\hat{x}(k+1|k), P(k+1|k)) \\ p(x(k+1)|Y_1^{k+1}, U_0^k) &\sim \mathcal{N}(\hat{x}(k+1|k+1), P(k+1|k+1)). \end{aligned}$$

Moreover, according to known results on Gaussian random vectors, [9] it is known that

$$[x(k+1) - \hat{x}(k+1|k)]^T P(k+1|k)^{-1} [x(k+1) - \hat{x}(k+1|k)] \leq K$$

$$[x(k+1) - \hat{x}(k+1|k+1)]^T P(k+1|k+1)^{-1} [x(k+1) - \hat{x}(k+1|k+1)] \leq K$$

represent contours of equal probability of the random variable around its mean value.

All the pdf involved in the Kalman filter are Gaussian, and therefore, associated with the filter dynamics, we may consider the corresponding locus of equal probability around the predicted and estimated values of the state, that constitute the mean of the conditional pdf that the filter propagates.

Figures 4.6 and 4.7 illustrate that interpretation. In these figures the ellipses represent the contour of equal probability (in the particular case of second order Gaussian random variables) around the mean, the dashes lines corresponds to the actual filter dynamics that involves the mean values and the solid lines represent the exact values of the random variables (the ones that the filter estimates).

Prediction cycle

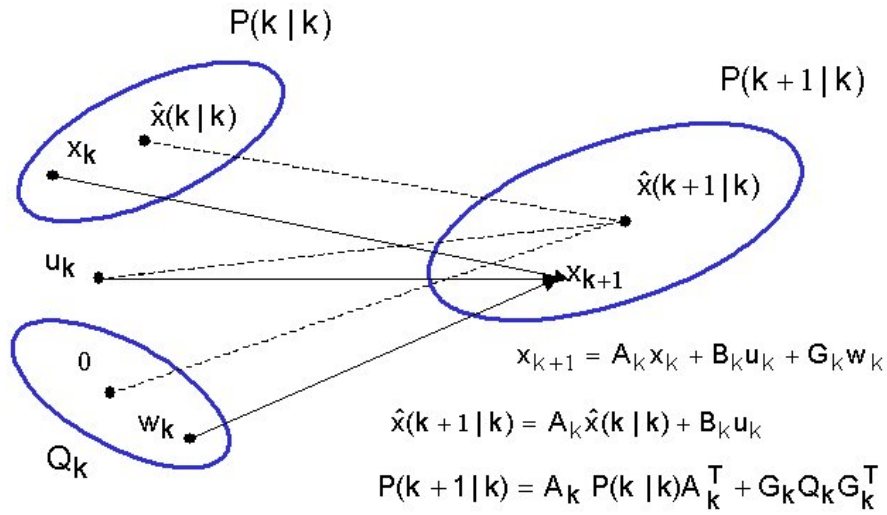


Figure 4.6: Error ellipsoid propagation in the Kalman filter prediction cycle

Filtering cycle

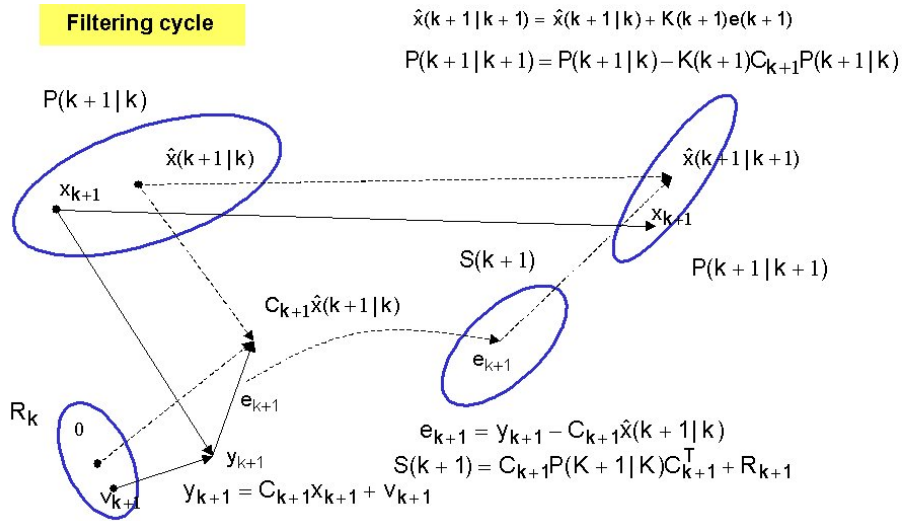


Figure 4.7: Error ellipsoid propagation in the Kalman filter filtering cycle

# Chapter 5

## The Extended Kalman Filter

In this section we address the filtering problem in case the system dynamics (state and observations) is nonlinear. With no loss of generality we will consider that the system has no external inputs. Consider the non-linear dynamics

$$x_{k+1} = f_k(x_k) + w_k \quad (5.1)$$

$$y_k = h_k(x_k) + v_k \quad (5.2)$$

where,

$$\begin{aligned} x_k &\in \mathcal{R}^n, & f_k(x_k) : \mathcal{R}^n &\longrightarrow \mathcal{R}^n \\ y_k &\in \mathcal{R}^r & h_k(x_k) : \mathcal{R}^n &\longrightarrow \mathcal{R}^r \\ v_k &\in \mathcal{R}^r \\ w_k &\in \mathcal{R}^n \end{aligned} \quad (5.3)$$

and  $\{v_k\}, \{w_k\}$  are white Gaussian, independent random processes with zero mean and covariance matrix

$$E[v_k v_k^T] = R_k, \quad E[w_k w_k^T] = Q_k \quad (5.4)$$

and  $x_0$  is the system initial condition considered as a Gaussian random vector,

$$x_0 \sim N(\bar{x}_0, \Sigma_0).$$

Let  $Y_1^k = \{y_1, y_2, \dots, y_k\}$  be a set of system measurements. The filter's goal is to obtain an estimate of the system's state based on these measurements.

As presented in Section 2, the estimator that minimizes the mean-square error evaluates the condition mean of the pdf of  $x_k$  given  $Y_1^k$ . Except in very particular cases, the computation of the conditional mean requires the knowledge of

the entire conditional pdf. One of these particular cases, referred in Section 4, is the one in which the system dynamics is linear, the initial conditional is a Gaussian random vector and system and measurement noises are mutually independent white Gaussian processes with zero mean. As a consequence, the conditional pdf  $p(x(k) | Y_1^k)$ ,  $p(x(k+1) | Y_1^k)$  and  $p(x(k+1) | Y_1^{k+1})$  are Gaussian.

With the non linear dynamics (5.1)-(5.2), these pdf are non Gaussian. To evaluate its first and second moments, the optimal nonlinear filter has to propagate the entire pdf which, in the general case, represents a heavy computational burden.

The Extended Kalman filter (EKF) gives an approximation of the optimal estimate. The non-linearities of the systems's dynamics are approximated by a linearized version of the non-linear system model around the last state estimate. For this approximation to be valid, this linearization should be a good approximation of the non-linear model in all the uncertainty domain associated with the state estimate.

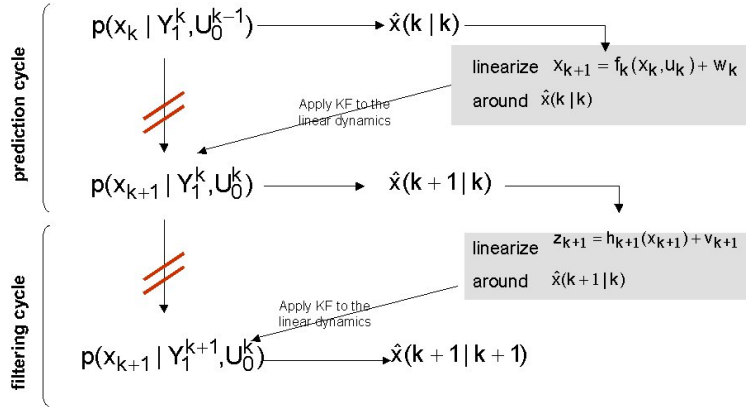


Figure 5.1: Extended Kalman filter dynamic concept

Figure 5.1 represents one cycle of consecutive prediction and filtering updates with the consecutive pdf transitions,

$$p(x_k | Y_1^k, U_0^{k-1}) \longrightarrow p(x_{k+1} | Y_1^k, U_0^k) \longrightarrow p(x_{k+1} | Y_1^{k+1}, U_0^k)$$

Rather than propagating the non Gaussian pdf, the Extended Kalman filter considers, at each cycle, a linearization of the non-linear dynamics (5.1)-(5.2) around the last consecutive predicted and filtered estimates of the state, and for the linearized dynamics, it applies the Kalman Filter.

One iteration of the EKF is composed by the following consecutive steps:

1. Consider the last filtered state estimate  $\hat{x}(k|k)$ ,
2. Linearize the system dynamics,  $x_{k+1} = f(x_k) + w_k$  around  $\hat{x}(k|k)$ ,
3. Apply the prediction step of the Kalman filter to the linearized system dynamics just obtained, yielding  $\hat{x}(k+1|k)$  and  $P(k+1|k)$ ,
4. Linearize the observation dynamics,  $y_k = h(x_k) + v_k$  around  $\hat{x}(k+1|k)$ ,
5. Apply the filtering or update cycle of the Kalman filter to the linearized observation dynamics, yielding  $\hat{x}(k+1|k+1)$  and  $P(k+1|k+1)$ .

Let  $F(k)$  and  $H(k)$  be the Jacobian matrices of  $f(\cdot)$  and  $h(\cdot)$ , denoted by

$$\begin{aligned} F(k) &= \nabla f_k |_{\hat{x}(k|k)} \\ H(k+1) &= \nabla h |_{\hat{x}(k+1|k)} \end{aligned}$$

The Extended Kalman filter algorithm is stated below:

### Predict Cycle

$$\begin{aligned} \hat{x}(k+1|k) &= f_k(\hat{x}(k|k)) \\ P(k+1|k) &= F(k)P(k|k)F^T(k) + Q(k) \end{aligned}$$

### Filtered Cycle

$$\begin{aligned} \hat{x}(k+1|k+1) &= \hat{x}(k+1|k) + K(k+1)[y_{k+1} - h_{k+1}(\hat{x}(k+1|k))] \\ K(k+1) &= P(k+1|k)H^T(k+1)[H(k+1)P(k+1|k)H^T(k+1) + R(k+1)]^{-1} \\ P(k+1|k+1) &= [I - K(k+1)H(k+1)]P(k+1|k) \end{aligned}$$

It is important to state that the EKF is not an optimal filter, but rather it is implemented based on a set of approximations. Thus, the matrices  $P(k|k)$  and  $P(k+1|k)$  do not represent the true covariance of the state estimates.

Moreover, as the matrices  $F(k)$  and  $H(k)$  depend on previous state estimates and therefore on measurements, the filter gain  $K(k)$  and the matrices  $P(k|k)$  and  $P(k+1|k)$  cannot be computed off-line as occurs in the Kalman filter.

Contrary to the Kalman filter, the EKF may diverge, if the consecutive linearizations are not a good approximation of the linear model in all the associated uncertainty domain.

## 5.1 Derivation of Extended Kalman Filter dynamics

This subsection presents the formal derivation of the EKF dynamics.

### Prediction

Assume that  $p(x_k | Y_1^k)$  is a Gaussian pdf with mean  $\eta_F^{n-1}$  and covariance matrix  $V_F^n$ , i.e.,

$$p(x_k | Y_1^k) \sim \mathcal{N}(x_k - \eta_F^k, V_F^k) = \mathcal{N}(x_k - \hat{x}(k|k), P(k|k)). \quad (5.5)$$

From the non-linear system dynamics,

$$x_{k+1} = f_k(x_k) + w_k, \quad (5.6)$$

and the Bayes law, the conditional pdf of  $x_{k+1}$  given  $Y_1^k$  is given by

$$p(x_{k+1} | Y_1^k) = \int_{-\infty}^{\infty} p(x_{k+1} | x_k) p(x_k | Y_1^k) dx_k,$$

or also,

$$p(x_{k+1} | Y_1^k) = \int_{-\infty}^{\infty} p_{w_k}(x_{k+1} - f_k(x_k)) p(x_k | Y_1^k) dx_k \quad (5.7)$$

where

$$p_{w_k}(x_{k+1} - f_k(x_k)) = \frac{1}{(2\pi)^{n/2} [\det Q_k]^{1/2}} \exp\left[-\frac{1}{2}(x_{k+1} - f_k(x_k))^T Q_k^{-1} (x_{k+1} - f_k(x_k))\right]. \quad (5.8)$$

The previous expression is **not** a Gaussian pdf given the nonlinearity in  $x_k$ . We will linearize  $f_k(x_k)$  in (5.6) around  $\eta_F^k = \hat{x}(k|k)$  neglecting higher order terms, this yielding

$$\begin{aligned} f_k(x_k) &\cong f_k(\eta_F^k) + \nabla f_k |_{\eta_F^k} \cdot [x_k - \eta_F^k] \\ &= \overbrace{f_k(\eta_F^k) - \nabla f_k |_{\eta_F^k} \cdot \eta_F^k}^{s_k} + \nabla f_k |_{\eta_F^k} \cdot x_k. \end{aligned} \quad (5.9)$$

where  $\nabla f_k$  is the Jacobian matrix of  $f(\cdot)$ ,

$$\nabla f_k = \frac{\partial f(x(k))}{\partial x(k)} \Big|_{\eta_F^k}$$

---

<sup>1</sup>F - refers filtering

With this linearization, the system dynamics may be written as:

$$x_{k+1} = \nabla f_k \big|_{\eta_F^k} \cdot x_k + w_k + \underbrace{[f_k(\eta_F^k) - \nabla f_k \big|_{\eta_F^k} \cdot \eta_F^k]}_{s_k} \quad (5.10)$$

or, in a condensed format,

$$\boxed{x_{k+1} = \nabla f_k \big|_{\eta_F^k} \cdot x_k + w_k + s_k} \quad (5.11)$$

Note that (5.11) represents a linear dynamics, in which  $s_k$  is known, has a null conditional expected value and depends on previous values of the state estimate. According to (5.9) the pdf in (5.7) can be written as:

$$\begin{aligned} p(x_{k+1} | Y_1^k) &= \int_{-\infty}^{\infty} p_{w_k}(x_{k+1} - \nabla f_k \big|_{\eta_F^k} \cdot x_k - s_k) \cdot p(x_k | Y_1^k) dx_k \\ &= \int_{-\infty}^{\infty} \mathcal{N}(x_{k+1} - \nabla f_k \big|_{\eta_F^k} \cdot x_k - s_k, Q_k) \cdot \mathcal{N}(x_k - \eta_F^k, V_F^k) dx_k \\ &= \int_{-\infty}^{\infty} \mathcal{N}(x_{k+1} - s_k - \nabla f_k \big|_{\eta_F^k} \cdot x_k, Q_k) \mathcal{N}(x_k - \eta_F^k, V_F^k) dx_k \end{aligned} \quad (5.12)$$

To simplify the computation of the previous pdf, consider the following variable transformation

$$z_k = \nabla f_k \cdot x_k. \quad (5.13)$$

where we considered, for the sake of simplicity, the simplified notation  $\nabla f_k$  to represent  $\nabla f_k \big|_{\eta_F^k}$ .

Evaluating the mean and the covariance matrix of the random vector (5.13) results:

$$E[y_k] = \nabla f_k \cdot E[x_k] = \nabla f_k \cdot \eta_F^k \quad (5.14)$$

$$E[y_k y_k^T] = \nabla f_k \cdot V_F^k \cdot \nabla f_k^T. \quad (5.15)$$

From the previous result, the pdf of  $x_k$  in (5.5) may be written as:

$$\begin{aligned} \mathcal{N}(x_k - \eta_F^k, V_F^k) &= \\ &= \frac{1}{(2\pi)^{n/2} (\det V_F^k)^{1/2}} \exp\left[-\frac{1}{2} (x_k - \eta_F^k)^T (V_F^k)^{-1} (x_k - \eta_F^k)\right] = \\ &= \frac{1}{(2\pi)^{n/2} (\det V_F^k)^{1/2}} \exp\left[-\frac{1}{2} (\nabla f_k x_k - \nabla f_k \cdot \eta_F^k)^T (\nabla f_k)^{-T} (V_F^k)^{-1} (\nabla f_k)^{-1} (\nabla f_k x_k - \nabla f_k \eta_F^k)\right] = \\ &= \frac{1}{(2\pi)^{n/2} (\det V_F^k)^{1/2}} \exp\left[-\frac{1}{2} (\nabla f_k x_k - \nabla f_k \eta_F^k)^T (\nabla f_k \cdot V_F^k \nabla f_k^T)^{-1} (\nabla f_k x_k - \nabla f_k \eta_F^k)\right] = \\ &= \det \nabla f_k \cdot \frac{1}{(2\pi)^{n/2} (\det \nabla f_k V_F^k \nabla f_k^T)^{n/2}} \\ &= \exp\left[-\frac{1}{2} (\nabla f_k x_k - \nabla f_k \eta_F^k)^T (\nabla f_k V_F^k \nabla f_k^T)^{-1} (\nabla f_k x_k - \nabla f_k \eta_F^k)\right]. \end{aligned}$$

We thus conclude that

$$\mathcal{N}(x_k - \eta_F^k, V_F^k) = \det \nabla f_k \cdot \mathcal{N}(\nabla f_k x_k - \nabla f_k \eta_F^k, \nabla f_k V_F^k \nabla f_k^T). \quad (5.16)$$

Replacing (5.16) in (5.12) yields:

$$\begin{aligned} p(x_{k+1} | Y_1^k) &= \\ &= \int_{-\infty}^{\infty} \mathcal{N}(x_{k+1} - s_k - \nabla f_k x_k, Q_k) \mathcal{N}(\nabla f_k x_k - \nabla f_k \eta_F^k, \nabla f_k V_F^k \nabla f_k^T) d(\nabla f_k \cdot x_k) \\ &= \mathcal{N}(x_{k+1} - s_k, Q_k) \star \mathcal{N}(x_{k+1} - \nabla f_k \cdot \eta_F^k, \nabla f_k V_F^k \nabla f_k^T) \end{aligned}$$

where  $\star$  represents the convolution of the two functions. We finally conclude that,

$$p(x_{k+1} | Y_1^k) = \mathcal{N}(x_{k+1} - \nabla f_k |_{\eta_F^k} \cdot \eta_F^k - s_k, Q_k + \nabla f_k |_{\eta_F^k} V_F^k \nabla f_k |_{\eta_F^k}^T) \quad (5.17)$$

We just conclude that,

if  $p(x_k | Z_1^k)$  is a Gaussian pdf with

1. mean  $\eta_F^n$ ,
2. covariance matrix  $V_F^n$

then, the linearization of the dynamics around  $\eta_F^n$  yields  $p(x_{k+1} | Z_1^k)$ , which is a Gaussian pdf with

1. mean  $\eta_P^{k+1}$
2. covariance matrix  $V_P^{k+1}$

where

$$\eta_P^{k+1} = \nabla f_k |_{\eta_F^k} \cdot \eta_F^k + f_k(\eta_F^k) - \nabla f_k |_{\eta_F^k} \cdot \eta_F^k \quad (5.18)$$

or else, given the value of  $s_k$  given in (5.10), can be simplified to

$$\eta_P^{k+1} = f_k(\eta_F^k) \quad (5.19)$$

$$V_P^{k+1} = Q_k + \nabla f_k |_{\eta_F^k} \cdot V_F^k \cdot \nabla f_k^T |_{\eta_F^k}. \quad (5.20)$$

These values are taken as the predicted state estimate and the associated covariance obtained by the EKF, i.e.,

$$\hat{x}(k+1|k) = \eta_P^{k+1} \quad (5.21)$$

$$P(k+1|k) = V_P^{k+1}, \quad (5.22)$$

representing the predicted dynamics,

$$\begin{aligned}\hat{x}(k+1|k) &= f_k(\hat{x}(k|k)) \\ P(k+1|k) &= \nabla f_k |_{\eta_F^k} \cdot P(k|k) \cdot \nabla f_k^T |_{\eta_F^k}\end{aligned}$$

## Filtering

In the filtering cycle, we use the system measurement at time instant  $k+1$ ,  $y_{k+1}$  to update the pdf  $p(x_{k+1} | Y_1^k)$  as represented

$$p(x_{k+1} | Y_1^k) \xrightarrow{y_{k+1}} p(x_{k+1} | Y_1^{k+1})$$

According to Bayes law,

$$p(x_{k+1} | Y_1^{k+1}) = \frac{p(Y_1^k)}{p(Y_1^{k+1})} \cdot [p(y_{k+1} | x_{k+1}) \cdot p(x_{k+1} | Y_1^k)]. \quad (5.23)$$

Given that

$$y_{k+1} = h_{k+1}(x_{k+1}) + v_{k+1}, \quad (5.24)$$

the pdf of  $y_{k+1}$  conditioned on the state  $x_{k+1}$  is given by

$$p(y_{k+1} | x_{k+1}) = \frac{1}{(2\pi)^{r/2} (\det R_{k+1})^{1/2}} \exp\left[-\frac{1}{2} (y_{k+1} - h_{k+1}(x_{k+1}))^T R_{k+1}^{-1} (y_{k+1} - h_{k+1}(x_{k+1}))\right]. \quad (5.25)$$

With a similar argument as the one used on the prediction cycle, the previous pdf may be simplified through the linearization of the observation dynamics.

Linearizing  $h_{k+1}(x_{k+1})$  around  $\eta_P^{k+1}$  and neglecting higher order terms results

$$h_{k+1}(x_{k+1}) \simeq h_{k+1}(\eta_P^{k+1}) + \nabla h |_{\eta_P^{k+1}} (x_{k+1} - \eta_P^{k+1}), \quad (5.26)$$

and so the system observation equation may be approximated by,

$$\boxed{y_{k+1} \simeq \nabla h |_{\eta_P^{k+1}} \cdot x_{k+1} + v_{k+1} + r_{k+1}} \quad (5.27)$$

with

$$r_{k+1} = h_{k+1}(\eta_P^{k+1}) - \nabla h |_{\eta_P^{k+1}} \cdot \eta_P^{k+1}. \quad (5.28)$$

being a known term in the linearized observation dynamics, (5.27). After the linearization around the predicted state estimate - that corresponds to  $\eta_P^{k+1} = \hat{x}_{k+1|k}$  (see (5.21), - the observation dynamics may be considered linear, and the computation of  $p(y_{k+1} | x_{k+1})$  in (5.25) is immediate. We have,

$$p(y_{k+1} | x_{k+1}) = \mathcal{N}(y_{k+1} - r_{k+1} - \nabla h_{k+1} |_{\eta_P^{k+1}} \cdot x_{k+1}, R_{k+1}). \quad (5.29)$$

Expression (5.29) may be rewritten as:

$$p(y_{k+1} | x_{k+1}) = \mathcal{N}(\nabla h_{k+1} |_{\eta_P^{k+1}} \cdot x_{k+1} + r_{k+1} - y_{k+1}, R_{k+1}). \quad (5.30)$$

Using a variable transformation similar to the one used in the prediction cycle, the previous pdf may be expressed as

$$p(x_{k+1} | Y_1^k) = \det \nabla h |_{\eta_P^{k+1}} \mathcal{N}(\nabla h_{k+1} |_{\eta_P^{k+1}} \cdot x_{k+1} - \nabla h_{k+1} |_{\eta_P^{k+1}} \cdot \eta_P^{k+1}, \nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T) \quad (5.31)$$

Multiplying expressions (5.30) and (5.31) as represented in the last product in (5.23) yields:

$$p(y_{k+1} | x_{k+1}) \cdot p(x_{k+1} | Y_1^k) \sim \mathcal{N}(\nabla h_{k+1} |_{\eta_P^{k+1}} \cdot x_{k+1} - \mu, V) \quad (5.32)$$

where the mean and covariance matrix are given by:

$$\begin{aligned} \mu &= \nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T (\nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} [-r_{k+1} + y_{k+1}] \\ &\quad + R_{k+1} (\nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} \nabla h_{k+1} \cdot \eta_P^{k+1}, \end{aligned} \quad (5.33)$$

$$V = \nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T (\nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} R_{k+1}. \quad (5.34)$$

Replacing in (5.33) the expression (5.28) we obtain:

$$\begin{aligned} \mu &= \nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T (\nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} \\ &\quad [-h_{k+1}(\eta_P^{k+1}) + \nabla h_{k+1} \cdot \eta_P^{k+1} + z_{k+1}] \\ &\quad + R_{k+1} (\nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} \nabla h_{k+1} \eta_P^{k+1} \\ &= \nabla h_{k+1} \eta_P^{k+1} + \nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T (\nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} [y_{k+1} - h_{k+1}(\eta_P^{k+1})] \end{aligned} \quad (5.35)$$

$$V = \nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T (\nabla h_{k+1} V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} R_{k+1} \quad (5.37)$$

where we use the short notation

$$\nabla h_{k+1} = \nabla h_{k+1} |_{\eta_P^{k+1}}. \quad (5.38)$$

Note that (5.32) expresses the pdf of  $\nabla h_{k+1} |_{\eta_P^{k+1}} \cdot x_{k+1}$  and not that of  $x_{k+1}$  as desired. In fact, the goal is to evaluate the mean and covariance matrix in

$$\mathcal{N}(x_{k+1} - \mu_1, V_1). \quad (5.39)$$

Note that (5.32) can be obtained from (5.39). We know that:

$$\begin{aligned} \mathcal{N}(x_{k+1} - \mu_1, V_1) &= \det \nabla h_{k+1} \cdot \mathcal{N}(\nabla h_{k+1} x_{k+1} - \nabla h_{k+1} \mu_1, \nabla h_{k+1} V_1 \nabla h_{k+1}^T) \\ &= \det \nabla h_{k+1} \mathcal{N}(\nabla h_{k+1} x_{k+1} - \mu, V), \end{aligned} \quad (5.40)$$

where  $\mu$  and  $V$  are given by (5.33) and (5.34).

Comparing (5.40) with (5.40) yields:

$$\begin{aligned}\nabla h_{k+1}\mu_1 &= \nabla h_{k+1}\eta_P^{k+1} + \nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T (\nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} [y_{k+1} - h_{k+1}(\eta_P^{k+1})] \\ \mu_1 &= \eta_P^{k+1} + V_P^{k+1} \nabla h_{k+1}^T (\nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} [y_{k+1} - h_{k+1}(\eta_P^{k+1})].\end{aligned}$$

We thus conclude that:

$$\boxed{\eta_F^{k+1} = \eta_P^{k+1} + V_P^{k+1} \nabla h_{k+1}^T (\nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} [y_{k+1} - h_{k+1}(\eta_P^{k+1})]} \quad (5.41)$$

Comparing (5.40) and (5.40) in terms of the covariance matrices, yields:

$$V = \nabla h_{k+1}V_1 \nabla h_{k+1}^T. \quad (5.42)$$

Replacing in this expression  $V$  by its value given by (5.37) result,

$$\begin{aligned}V &= \nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T (\nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T + R_{k+1})^{-1} R_{k+1} \\ &= \nabla h_{k+1}V_F^{k+1} \nabla h_{k+1}^T\end{aligned}$$

that has to be solved relative to  $V_F^{k+1}$ . From the above equalities, we successively obtain:

$$\begin{aligned}\nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T &= \nabla h_{k+1}V_F^{k+1} \nabla h_{k+1}^T R_{k+1}^{-1} (\nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T + R_{k+1}) \\ &= \nabla h_{k+1}V_F^{k+1} \nabla h_{k+1}^T R_{k+1}^{-1} \nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T + \nabla h_{k+1}V_F^{k+1} \nabla h_{k+1}^T\end{aligned}$$

or else,

$$\begin{aligned}V_P^{k+1} &= V_F^{k+1} \nabla h_{k+1}^T R_{k+1}^{-1} \nabla h_{k+1}V_P^{k+1} + V_F^{k+1} \\ V_P^{k+1} &= V_F^{k+1} [I + \nabla h_{k+1}^T R_{k+1}^{-1} \nabla h_{k+1}V_P^{k+1}] \\ V_F^{k+1} &= V_P^{k+1} [I + \nabla h_{k+1}^T R_{k+1}^{-1} \nabla h_{k+1}V_P^{k+1}]^{-1}.\end{aligned}$$

Using the lemma of the inversion of matrices,

$$\begin{aligned}V_F^{k+1} &= V_P^{k+1} [I - \nabla h_{k+1}^T R_{k+1}^{-1} (I + \nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T R_{k+1}^{-1})^{-1} \nabla h_{k+1}V_P^{k+1}] \\ &= V_P^{k+1} [I - \nabla h_{k+1}^T [R_{k+1} + \nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T]^{-1} \nabla h_{k+1}V_P^{k+1}]\end{aligned}$$

$$\boxed{V_F^{k+1} = V_P^{k+1} - V_P^{k+1} \nabla h_{k+1}^T [R_{k+1} + \nabla h_{k+1}V_P^{k+1} \nabla h_{k+1}^T]^{-1} \nabla h_{k+1}V_P^{k+1}} \quad (5.43)$$

Therefore, if we consider that  $p(x_{k+1}|Y_1^k)$  is a Gaussian pdf, have access to the measurement  $y_{k+1}$  and linearize the system observation dynamics around

$\eta_P^{k+1} = \hat{x}(k+1|k)$  we obtain a Gaussian pdf  $p(x_{k+1} | Y_1^{k+1})$  with mean  $\eta_F^{k+1}$  and covariance matrix  $V_F^{k+1}$  given by (5.41) and (5.43), respectively.

Finally, we summarize the previous results and interpret the Extended Kalman filter as a Kalman filter applied to a linear time-varying dynamics.

Let:

$$\begin{aligned}\eta_P^{k+1} &= \hat{x}(k+1|k) \\ V_P^{k+1} &= P(k+1|k) \\ \eta_F^{k+1} &= \hat{x}(k+1|k+1) \\ V_F^{k+1} &= P(k+1|k+1)\end{aligned}$$

and consider

$$\begin{aligned}\nabla f_k |_{\eta_F^k} &= \nabla f_k |_{\hat{x}(k|k)} = F(k) \\ \nabla h_{k+1} |_{\eta_P^{k+1}} &= \nabla h |_{\hat{x}(k+1|k)} = H(k+1) \\ s(k) &= f_k(\hat{x}(k|k)) - F(k) \cdot \hat{x}(k|k) \\ r(k+1) &= h_{k+1}(\hat{x}(k+1|k)) - H(k+1) \cdot \hat{x}(k+1|k).\end{aligned}$$

Assume the linear system in whose dynamics the just evaluated quantities are included.

$$x(k+1) = F(k)x(k) + w_k + s(k) \quad (5.44)$$

$$y(k+1) = H(k+1)x(k+1) + v_{k+1} + r(k+1) \quad (5.45)$$

where  $w_k$  and  $v_{k+1}$  are white Gaussian noises,  $s(k)$  and  $r(k)$  are known quantities with null expected value.

The EKF applies the Kalman filter dynamics to (5.44)-(5.45), where the matrices  $F(k)$  and  $H(k)$  depend on the previous state estimates, yielding

$$\begin{aligned}\hat{x}(k+1|k) &= f_k(\hat{x}(k|k)) \\ \hat{x}(k+1|k+1) &= \hat{x}(k+1|k) + K(k+1)[y_{k+1} - h_{k+1}(\hat{x}(k+1|k))]\end{aligned}$$

where  $K(k+1)$  is the filter gain and

$$\begin{aligned}K(k+1) &= P(k+1|k)H^T(k+1)[H(k+1)P(k+1|k)H^T(k+1) + R(k+1)]^{-1} \\ P(k+1|k) &= F(k)P(k|k)F^T(k) + Q(k) \\ P(k+1|k+1) &= P(k+1|k) - P(k+1|k)H^T(k+1) \\ &\quad [H(k+1)P(k+1|k)H^T(k+1) + R(k+1)]^{-1}H(k+1)P(k+1|k)\end{aligned}$$

Expression (5.46) may be rewritten as:

$$\begin{aligned}
 P(k+1|k+1) &= P(k+1|n) \times & (5.47) \\
 &\times [I - P(k+1|n)H^T(k+1)[H(k+1) + P(k+1|k)H^T(k+1) + R_{k+1}]^{-1}H(k+1)] & (5.48)
 \end{aligned}$$

$$\boxed{P(k+1|k+1) = [I - K(k+1)H(k+1)]P(k+1|k)} \quad (5.49)$$

# Bibliography

- [1] Michael Athans, "Dynamic Stochastic Estimation, Prediction and Smoothing," Series of Lectures, Spring 1999.
- [2] T. Kailath, "Lectures Notes on Wiener and Kalman Filtering," Springer-Verlag, 1981.
- [3] Thomas Kailath, Ali H. Sayed, Babak Hassibi, "Linear Estimation," Prentice Hall, 2000.
- [4] Peter S. Maybeck, "The Kalman Filter: An Introduction to Concepts," in Autonomous Robot Vehciles, I.J. Cox, G. T. Wilfong (eds), Springer-Verlag, 1990.
- [5] J. Mendel, "Lessons in Digital Estimation Theory", Prentice-Hall, 1987.
- [6] K. S. Miller, "Multidimensional Gaussian Distributions," John Wiley & Sons, 1963.
- [7] J. M. F. Moura, "Linear and Nonlinear Stochastic Filtering," NATO Advanced Study Institute, Les Houches, September 1985.
- [8] A. Papoulis, "Probability, Random Variables and Stochastic Processes," McGraw-Hill, 1965.
- [9] M. Isabel Ribeiro, "Gaussian Probability Density Functions: Properties and Error Characterization," Institute for Systems and Robotics, Technical Report, February 2004.
- [10] H. L. Van Trees, "Detection, Estimation and Modulation Theory," John Wiley & Sons, 1968.

## ERRATA

The equation (4.37) has an error. The correct version is

$$P(k | k) = [I - K(k)C]P(k | k-1)$$

Thanks to Sergio Trimboli that pointed out the error in a preliminary version

23.March.2008