

# Discrete Kalman Filter Tutorial

Gabriel A. Terejanu  
Department of Computer Science and Engineering  
University at Buffalo, Buffalo, NY 14260  
terejanu@buffalo.edu

## 1 Introduction

Consider the following stochastic dynamic model and the sequence of noisy observations  $\mathbf{z}_k$ :

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}, k) \quad (1)$$

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k, k) \quad (2)$$

Also let  $\mathbf{x}_0$  be the random initial condition of the system and

$$\mathbf{Z}_k = \{\mathbf{z}_i | 1 \leq i \leq k\} \quad (3)$$

be the set of  $k$  observations. Finding  $\mathbf{x}_k^a$ , the estimate or analysis of the state space  $\mathbf{x}_k$ , given  $\mathbf{Z}_k$  and the initial conditions is called the **filtering** problem. When the dynamic model for the process,  $\mathbf{f}(\cdot)$ , and for the measurements,  $\mathbf{h}(\cdot)$ , are linear, and the random components  $\mathbf{x}_0$ ,  $\mathbf{w}_k$ ,  $\mathbf{v}_k$  are uncorrelated Gaussian random vectors, then the solution is given by the classical **Kalman filter** equations [7].

The Kalman filter is named after Rudolph E. Kalman, who in 1960 published his famous paper describing a recursive solution to the discrete-data linear filtering problem (Kalman 1960) [11]. It is the **optimal estimator** for a large class of problems, finding the most probable state as an unbiased linear minimum variance estimate of a system based on discrete observations of the system and a model which describes the evolution of the system [5].

## 2 Dynamic process

A **stochastic time-variant linear system** is described by the difference equation and the observation model:

$$\mathbf{x}_k = \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{B}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1} \quad (4)$$

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \quad (5)$$

Where the **control input**  $\mathbf{u}_k$  is a known nonrandom vector. The **initial state**  $\mathbf{x}_0$  is a random vector with known mean  $\mu_0 = E[\mathbf{x}_0]$  and covariance  $\mathbf{P}_0 = E[(\mathbf{x}_0 - \mu_0)(\mathbf{x}_0 - \mu_0)^T]$ .

In the following we assume that the random vector  $\mathbf{w}_k$  captures uncertainties in the model and  $\mathbf{v}_k$

denotes the measurement noise. Both are temporally uncorrelated (white noise), zero-mean random sequences with known covariances and both of them are uncorrelated with the initial state  $\mathbf{x}_0$ .

$$E[\mathbf{w}_k] = 0 \quad E[\mathbf{w}_k \mathbf{w}_k^T] = \mathbf{Q}_k \quad E[\mathbf{w}_k \mathbf{w}_j^T] = 0 \text{ for } k \neq j \quad E[\mathbf{w}_k \mathbf{x}_0^T] = 0 \text{ for all } k \quad (6)$$

$$E[\mathbf{v}_k] = 0 \quad E[\mathbf{v}_k \mathbf{v}_k^T] = \mathbf{R}_k \quad E[\mathbf{v}_k \mathbf{v}_j^T] = 0 \text{ for } k \neq j \quad E[\mathbf{v}_k \mathbf{x}_0^T] = 0 \text{ for all } k \quad (7)$$

Also the two random vectors  $\mathbf{w}_k$  and  $\mathbf{v}_k$  are uncorrelated:

$$E[\mathbf{w}_k \mathbf{v}_j^T] = 0 \text{ for all } k \text{ and } j \quad (8)$$

The assumptions about the unbiasedness and uncorrelation are not that critical, extensions of Kalman Filter can be derived in these situations.

Dimension and description of variables:

$\mathbf{x}_k$	$n \times 1$	– State vector
$\mathbf{u}_k$	$l \times 1$	– Input/control vector
$\mathbf{w}_k$	$n \times 1$	– Process noise vector
$\mathbf{z}_k$	$m \times 1$	– Observation vector
$\mathbf{v}_k$	$m \times 1$	– Measurement noise vector
$\mathbf{A}_k$	$n \times n$	– State transition matrix
$\mathbf{B}_k$	$n \times l$	– Input/control matrix
$\mathbf{H}_k$	$m \times n$	– Observation matrix
$\mathbf{Q}_k$	$n \times n$	– Process noise covariance matrix
$\mathbf{R}_k$	$m \times m$	– Measurement noise covariance matrix

### 3 KF derivation

The optimal (minimum variance unbiased) estimate is the conditional mean and is computed in two steps: the **forecast step** using the model difference equations and the **data assimilation** step. Hence the Kalman Filter has a "predictor-corrector" structure.

#### Model Forecast Step

Initially, since the only available information is the mean,  $\mu_0$ , and the covariance,  $\mathbf{P}_0$ , of the initial state then the initial optimal estimate  $\mathbf{x}_0^a$  and error covariance is:

$$\mathbf{x}_0^a = \mu_0 = E[\mathbf{x}_0] \quad (9)$$

$$\mathbf{P}_0 = E[(\mathbf{x}_0 - \mathbf{x}_0^a)(\mathbf{x}_0 - \mathbf{x}_0^a)^T] \quad (10)$$

Assume now that we have an **optimal estimate**  $\mathbf{x}_{k-1}^a \equiv E[\mathbf{x}_{k-1} | \mathbf{Z}_{k-1}]$  with  $\mathbf{P}_{k-1} \equiv E[(\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^a)(\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^a)^T]$  covariance at time  $k-1$ . The predictable part of  $\mathbf{x}_k$  is given by:

$$\begin{aligned} \mathbf{x}_k^f &\equiv E[\mathbf{x}_k | \mathbf{Z}_{k-1}] \\ &= E[\mathbf{A}_{k-1} \mathbf{x}_{k-1} + \mathbf{B}_{k-1} \mathbf{u}_{k-1} + \mathbf{w}_{k-1}] \\ &= \mathbf{A}_{k-1} \mathbf{x}_{k-1}^a + \mathbf{B}_{k-1} \mathbf{u}_{k-1} \end{aligned} \quad (11)$$

The forecast error is:

$$\begin{aligned}
\mathbf{e}_k^f &\equiv \mathbf{x}_k - \mathbf{x}_k^f & (12) \\
&= \mathbf{A}_{k-1}(\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^a) + \mathbf{w}_{k-1} \\
&= \mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1}
\end{aligned}$$

The forecast error covariance is given by:

$$\begin{aligned}
\mathbf{P}_k^f &\equiv E[\mathbf{e}_k^f(\mathbf{e}_k^f)^T] & (13) \\
&= E[(\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1})(\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1})^T] \\
&= \mathbf{A}_{k-1}E[\mathbf{e}_{k-1}(\mathbf{e}_{k-1})^T]\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1} \\
&= \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1} & (14)
\end{aligned}$$

### Data Assimilation Step

At time  $k$  we have two pieces of information: the forecast value  $\mathbf{x}_k^f$  with the covariance  $\mathbf{P}_k^f$  and the measurement  $\mathbf{z}_k$  with the covariance  $\mathbf{R}_k$ . We know that:

$$\begin{aligned}
\mathbf{x}_k^a &\equiv E[\mathbf{x}_k|Z_k] & (15) \\
&= E[\mathbf{x}_k|Z_{k-1}] + E[\mathbf{x}_k|\mathbf{z}_k]
\end{aligned}$$

Assume that the last term is a linear operation on the **innovation**  $\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^f$  [10](see also [9]: Projection theorem p.408 and Kalman innovations p.443). The innovation represents the new information contained in the observation  $\mathbf{z}_k$ .

$$E[\mathbf{x}_k|\mathbf{z}_k] = \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^f) \quad (16)$$

Therefore:

$$\begin{aligned}
\mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^f) \\
&= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{x}_k^f + \mathbf{K}_k\mathbf{z}_k
\end{aligned}$$

So, the easiest way to combine the two bits of information is to assume that the unbiased estimate  $\mathbf{x}_k^a$  is a linear combination of both the forecast and the measurement.

$$\mathbf{x}_k^a = \mathbf{L}_k\mathbf{x}_k^f + \mathbf{K}_k\mathbf{z}_k \quad \text{where} \quad \mathbf{L}_k = \mathbf{I} - \mathbf{K}_k\mathbf{H}_k \quad (17)$$

Or, the **optimal estimate** at time  $k$  is equal to the best prediction plus a correction term of an **optimal weighting value**,  $\mathbf{K}_k$ , times the innovation as in (17) [8].

Substitute (4), (5) and (11) into (16):

$$\begin{aligned}
\mathbf{x}_k^a &= \mathbf{A}_{k-1}\mathbf{x}_{k-1}^a + \mathbf{B}_{k-1}\mathbf{u}_{k-1} + \mathbf{K}_k(\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k(\mathbf{A}_{k-1}\mathbf{x}_{k-1}^a + \mathbf{B}_{k-1}\mathbf{u}_{k-1})) & (18) \\
&= \mathbf{A}_{k-1}\mathbf{x}_{k-1}^a + \mathbf{B}_{k-1}\mathbf{u}_{k-1} + \mathbf{K}_k(\mathbf{H}_k\mathbf{A}_{k-1}(\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^a) + \mathbf{H}_k\mathbf{w}_{k-1} + \mathbf{v}_k)
\end{aligned}$$

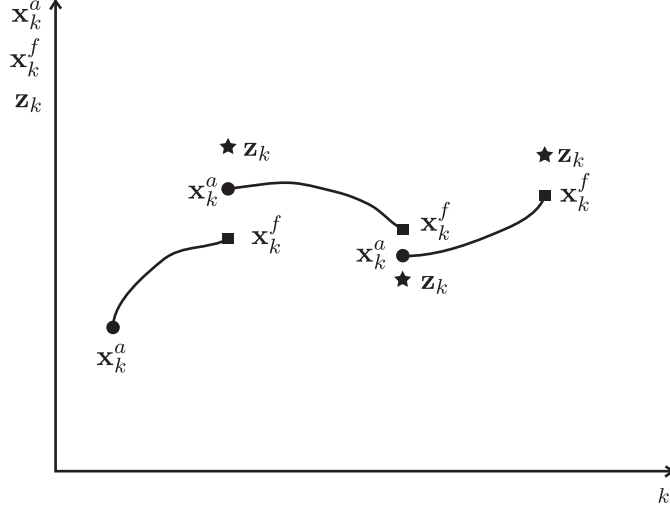


Figure 1: Sequential assimilation

The error in the estimate  $\mathbf{x}_k^a$  is:

$$\begin{aligned}
\mathbf{e}_k &\equiv \mathbf{x}_k - \mathbf{x}_k^a \\
&= \mathbf{A}_{k-1}\mathbf{e}_{k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{A}_{k-1}\mathbf{e}_{k-1} + (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{w}_{k-1} - \mathbf{K}_k\mathbf{v}_k \\
&= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1}) - \mathbf{K}_k\mathbf{v}_k
\end{aligned} \tag{19}$$

Then, the posterior covariance of the new estimate is:

$$\begin{aligned}
\mathbf{P}_k &\equiv E[\mathbf{e}_k\mathbf{e}_k^T] \\
&= E[(\mathbf{L}_k(\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1}) - \mathbf{K}_k\mathbf{v}_k)(\mathbf{L}_k(\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1}) - \mathbf{K}_k\mathbf{v}_k)^T] \\
&= \mathbf{L}_k E[(\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1})(\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1})^T] \mathbf{L}_k^T + \mathbf{K}_k E[\mathbf{v}_k\mathbf{v}_k^T] \mathbf{K}_k^T \\
&= \mathbf{L}_k(\mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1})\mathbf{L}_k^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\
&= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\
&= \mathbf{P}_k^f - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^f - \mathbf{P}_k^f\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\mathbf{D}_k\mathbf{K}_k^T
\end{aligned} \tag{21}$$

where

$$\mathbf{D}_k = \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k \tag{22}$$

The posterior covariance formula holds for any  $\mathbf{K}_k$ . The cross terms canceled because  $\mathbf{x}_{k-1}$ ,  $\mathbf{w}_{k-1}$  and  $\mathbf{v}_k$  are uncorrelated and  $\mathbf{e}_{k-1}$  is a function of  $\mathbf{x}_{k-1}$ .

Our goal is to minimize the error in the estimate  $\mathbf{e}_{ki}$  for any state  $i = 1, n$ . The problem is constructed as a mean squared error minimiser. The cost functional to be minimized is given by [1]:

$$J = E \left[ \sum_{i=1}^n \mathbf{e}_{ki}^2 \right] \tag{23}$$

This is the sum of error variances for each state variable. Therefore the cost functional can be expressed as the trace of the error covariance:

$$J = \text{tr}(\mathbf{P}_k) \quad (24)$$

Since  $\text{tr}(\mathbf{P}_k)$  is a function of  $\mathbf{K}_k$  and  $\mathbf{K}_k$  is the only unknown, we request to minimize the  $\text{tr}(\mathbf{P}_k)$  w.r.t.  $\mathbf{K}_k$ .

$$\frac{\partial \text{tr}(\mathbf{P}_k)}{\partial \mathbf{K}_k} = 0 \quad (25)$$

The partial derivative of the trace is easily given using matrix calculus rules [4].

$$\begin{aligned} \frac{\partial \text{tr}(\mathbf{P}_k^f - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^f - \mathbf{P}_k^f \mathbf{H}_k^T \mathbf{K}_k^T + \mathbf{K}_k \mathbf{D}_k \mathbf{K}_k^T)}{\partial \mathbf{K}_k} &= 0 \\ -(\mathbf{H}_k \mathbf{P}_k^f)^T - \mathbf{P}_k^f \mathbf{H}_k^T + 2\mathbf{K}_k \mathbf{D}_k &= 0 \end{aligned} \quad (26)$$

Thus, the **Kalman gain** is given by:

$$\begin{aligned} \mathbf{K}_k &= \mathbf{P}_k^f \mathbf{H}_k^T \mathbf{D}_k^{-1} \\ &= \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \end{aligned} \quad (27)$$

Substituting this back into (20):

$$\begin{aligned} \mathbf{P}_k &= \mathbf{P}_k^f - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^f - \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{D}_k^{-1})^T \mathbf{H}_k (\mathbf{P}_k^f)^T + \mathbf{P}_k^f \mathbf{H}_k^T \mathbf{D}_k^{-1} \mathbf{D}_k (\mathbf{D}_k^{-1})^T \mathbf{H}_k (\mathbf{P}_k^f)^T \\ &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f \end{aligned} \quad (28)$$

Note that the covariance  $\mathbf{P}_k$  does not directly depend on observations,  $\mathbf{z}_k$ , or the input vector. This property makes it possible to compute and to analyse the covariance matrix in absence of any observation.

## 4 Summary of Kalman Filter

Model and Observation:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}_{k-1} \mathbf{x}_{k-1} + \mathbf{B}_{k-1} \mathbf{u}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{z}_k &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \end{aligned}$$

Initialization:

$$\mathbf{x}_0^a = \mu_0 \text{ with error covariance } \mathbf{P}_0$$

Model Forecast Step/Predictor:

$$\begin{aligned} \mathbf{x}_k^f &= \mathbf{A}_{k-1} \mathbf{x}_{k-1}^a + \mathbf{B}_{k-1} \mathbf{u}_{k-1} \\ \mathbf{P}_k^f &= \mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1} \end{aligned}$$

Data Assimilation Step/Corrector:

$$\begin{aligned} \mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^f) \\ \mathbf{K}_k &= \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\ \mathbf{P}_k &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f \end{aligned}$$

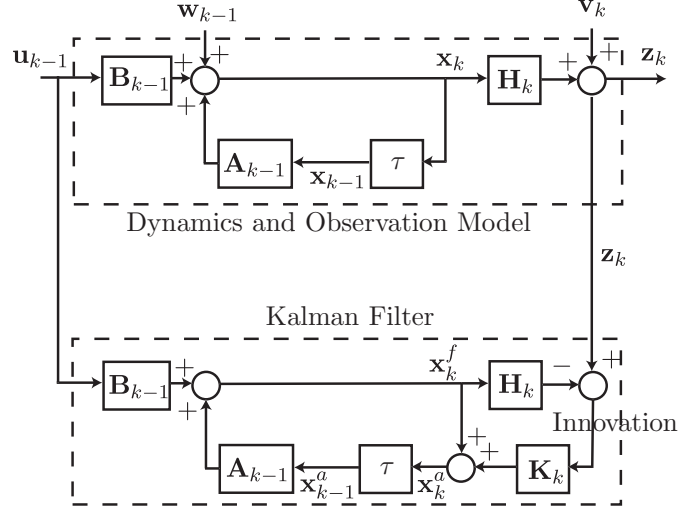


Figure 2: The block diagram for Kalman Filter

## 5 KF original derivation

The following derivation respects Kalman original concept of derivation [10]. The notation that has been changed for the consistency of the tutorial. The optimal estimate for the system (4)-(5) is derived using orthogonal projections on the vector space of random variables.

### Orthogonal Projection

Let the vector space  $\mathcal{Z}_k$  be the set of all linear combinations of the random variables (observations)  $\mathbf{z}_k$ .  $\mathcal{Z}_k$  is a finite-dimensional subspace of the space of all possible observations.

$$\mathcal{Z}_k \equiv \left\{ z_k \mid z_k = \sum_{i=1}^k \alpha_i \mathbf{z}_i \right\} \quad (29)$$

Two vectors  $\mathbf{u}, \mathbf{v} \in \mathcal{Z}_k$  are orthogonal if their correlation is zero. Any vector  $\mathbf{x}$  can be uniquely decomposed in two parts:  $\bar{\mathbf{x}} \in \mathcal{Z}_k$  and  $\tilde{\mathbf{x}} \perp \mathcal{Z}_k$ .

$$\mathbf{x} = \bar{\mathbf{x}} + \tilde{\mathbf{x}} \quad (30)$$

*Theorem [10]:* Let  $\{\mathbf{x}_k\}, \{\mathbf{z}_k\}$  be random processes with zero mean. If either (1) the random processes are Gaussian or (2) the optimal estimate is restricted to be a linear function of the observed random variables, and the loss function  $L(\mathbf{e}_k)$  is quadratic in  $\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$ , where  $L(\cdot)$  is a positive non-decreasing function of error

$$\begin{aligned} \hat{\mathbf{x}}_k &= \text{optimal estimate of } \mathbf{x}_k \text{ given } \{\mathbf{z}_k\} \\ &= \text{orthogonal projection } \bar{\mathbf{x}}_k \text{ of } \mathbf{x}_k \text{ on } \mathcal{Z}_k \end{aligned} \quad (31)$$

### Derivation

Assume  $\mathcal{Z}_{k-1}$  is known and  $\mathbf{z}_k$  is measured. Let  $\tilde{\mathbf{z}}_k$  be the component of  $\mathbf{z}_k$  orthogonal on  $\mathcal{Z}_{k-1}$ . The component  $\tilde{\mathbf{z}}_k$  generates a linear manifold  $\tilde{\mathcal{Z}}_k$ .

$$\mathcal{Z}_k = \mathcal{Z}_{k-1} \cup \tilde{\mathcal{Z}}_k \quad (32)$$

Every vector in  $\tilde{\mathcal{Z}}_k$  is orthogonal to every vector in  $\mathcal{Z}(k-1)$ .

Assume that  $\hat{\mathbf{x}}_{k-1}$  is know, then:

$$\begin{aligned}\mathbf{x}_k^a &\equiv E[\mathbf{x}_k|\mathcal{Z}_k] \\ &= E[\mathbf{x}_k|\mathcal{Z}_{k-1}] + E[\mathbf{x}_k|\tilde{\mathcal{Z}}_k] \\ &= \mathbf{x}_k^f + E[\mathbf{x}_k|\tilde{\mathcal{Z}}_k]\end{aligned}\tag{33}$$

Where the forecast value  $\mathbf{x}_k^f$  of  $\mathbf{x}_k$  can be obtained as in (11) and the forecast covariance matrix is given by (13):

$$\mathbf{x}_k^f = \mathbf{A}_{k-1}\mathbf{x}_{k-1}^a + \mathbf{B}_{k-1}\mathbf{u}_{k-1}\tag{34}$$

$$\mathbf{P}_k^f = \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}\tag{35}$$

Assume that the last term in (33) is a linear operation on the random variable  $\tilde{\mathbf{z}}_k$  (called **innovation**):

$$E[\mathbf{x}_k|\tilde{\mathcal{Z}}_k] = \mathbf{K}_k\tilde{\mathbf{z}}_k\tag{36}$$

where

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - \bar{\mathbf{z}}_k\tag{37}$$

but  $\bar{\mathbf{z}}_k$  is the orthogonal projection of  $\mathbf{z}_k$  on  $\mathcal{Z}_{k-1}$ . So:

$$\begin{aligned}\bar{\mathbf{z}}_k &= E[\mathbf{z}_k|\mathcal{Z}_{k-1}] \\ &= E[\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k|\mathcal{Z}_{k-1}] \\ &= \mathbf{H}_k\mathbf{x}_k^f\end{aligned}\tag{38}$$

Hence, (33) becomes:

$$\begin{aligned}\mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^f) \\ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{x}_k^f + \mathbf{K}_k\mathbf{z}_k\end{aligned}\tag{39}$$

The estimate error:

$$\begin{aligned}\mathbf{e}_k &\equiv \mathbf{x}_k - \mathbf{x}_k^a \\ &= \mathbf{A}_{k-1}\mathbf{e}_{k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{A}_{k-1}\mathbf{e}_{k-1} + (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{w}_{k-1} - \mathbf{K}_k\mathbf{v}_k \\ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1}) - \mathbf{K}_k\mathbf{v}_k\end{aligned}\tag{40}$$

Therefore the error covariance matrix, derived as in (20), is:

$$\begin{aligned}\mathbf{P}_k &\equiv E[\mathbf{e}_k\mathbf{e}_k^T] \\ &= \mathbf{P}_k^f - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^f - \mathbf{P}_k^f\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\mathbf{D}_k\mathbf{K}_k^T\end{aligned}\tag{41}$$

where

$$\mathbf{D}_k = \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k\tag{42}$$

We have to find an explicit formula for  $\mathbf{K}_k$  by noting that the residual  $\mathbf{x}_k - E[\mathbf{x}_k|\tilde{\mathcal{Z}}_k]$  is orthogonal to  $\tilde{\mathcal{Z}}_k$ , therefore it is orthogonal to  $\tilde{\mathbf{z}}_k$ . Results:

$$\begin{aligned} 0 &= E[(\mathbf{x}_k - E[\mathbf{x}_k|\tilde{\mathcal{Z}}_k])\tilde{\mathbf{z}}_k^T] \\ &= E[(\mathbf{x}_k - \mathbf{K}_k\tilde{\mathbf{z}}_k)\tilde{\mathbf{z}}_k^T] \\ &= E[\mathbf{x}_k\tilde{\mathbf{z}}_k^T] - \mathbf{K}_k E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T] \end{aligned} \quad (43)$$

We know that  $\mathbf{x}_k = \bar{\mathbf{x}}_k + \tilde{\mathbf{x}}_k$ , where  $\bar{\mathbf{x}}_k \in \mathcal{Z}_{k-1}$ , therefore  $\bar{\mathbf{x}}_k \perp \tilde{\mathcal{Z}}_k$  so  $\bar{\mathbf{x}}_k \perp \tilde{\mathbf{z}}_k$ .

$$\begin{aligned} E[\mathbf{x}_k\tilde{\mathbf{z}}_k^T] &= E[(\bar{\mathbf{x}}_k + \tilde{\mathbf{x}}_k)\tilde{\mathbf{z}}_k^T] \\ &= E[\tilde{\mathbf{x}}_k\tilde{\mathbf{z}}_k^T] \\ &= E[(\mathbf{x}_k - \bar{\mathbf{x}}_k)\tilde{\mathbf{z}}_k^T] \\ &= E[(\mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{B}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1} - E[\mathbf{x}_k|\mathcal{Z}_{k-1}])\tilde{\mathbf{z}}_k^T] \\ &= E[(\mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{B}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1} - \mathbf{A}_{k-1}\mathbf{x}_{k-1}^a - \mathbf{B}_{k-1}\mathbf{u}_{k-1})\tilde{\mathbf{z}}_k^T] \\ &= E[(\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{w}_{k-1})\tilde{\mathbf{z}}_k^T] \\ &= \mathbf{A}_{k-1}E[\mathbf{e}_{k-1}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^f)^T] + E[\mathbf{w}_{k-1}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^f)^T] \end{aligned} \quad (44)$$

We can obtain an expression of the innovation as function of the estimate error:

$$\begin{aligned} \mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k^f &= \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k\mathbf{x}_k^f \\ &= \mathbf{H}_k(\mathbf{x}_k - \mathbf{x}_k^f) + \mathbf{v}_k \\ &= \mathbf{H}_k(\mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{B}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1} - \mathbf{A}_{k-1}\mathbf{x}_{k-1}^a - \mathbf{B}_{k-1}\mathbf{u}_{k-1}) + \mathbf{v}_k \\ &= \mathbf{H}_k\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{H}_k\mathbf{w}_{k-1} + \mathbf{v}_k \end{aligned} \quad (45)$$

Substituting this into (44):

$$\begin{aligned} E[\mathbf{x}_k\tilde{\mathbf{z}}_k^T] &= \mathbf{A}_{k-1}E[\mathbf{e}_{k-1}(\mathbf{H}_k\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{H}_k\mathbf{w}_{k-1} + \mathbf{v}_k)^T] \\ &\quad + E[\mathbf{w}_{k-1}(\mathbf{H}_k\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{H}_k\mathbf{w}_{k-1} + \mathbf{v}_k)^T] \\ &= \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T\mathbf{H}_k^T + \mathbf{Q}_{k-1}\mathbf{H}_k^T \\ &= \mathbf{P}_k^f\mathbf{H}_k^T \end{aligned} \quad (46)$$

The last term from (43) is:

$$\begin{aligned} E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T] &= E[(\mathbf{H}_k\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{H}_k\mathbf{w}_{k-1} + \mathbf{v}_k)(\mathbf{H}_k\mathbf{A}_{k-1}\mathbf{e}_{k-1} + \mathbf{H}_k\mathbf{w}_{k-1} + \mathbf{v}_k)^T] \\ &= \mathbf{H}_k\mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T\mathbf{H}_k^T + \mathbf{H}_k\mathbf{Q}_{k-1}\mathbf{H}_k^T + \mathbf{R}_k \\ &= \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k \end{aligned} \quad (47)$$

With (46) and (47), (43) becomes:

$$\begin{aligned} 0 &= E[\mathbf{x}_k\tilde{\mathbf{z}}_k^T] - \mathbf{K}_k E[\tilde{\mathbf{z}}_k\tilde{\mathbf{z}}_k^T] \\ &= \mathbf{P}_k^f\mathbf{H}_k^T - \mathbf{K}_k(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k) \end{aligned} \quad (48)$$

Results that the gain matrix  $\mathbf{K}_k$  is:

$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (49)$$

The Kalman Filter equations are given by (34), (35), (33), (49) and (41). Note that in the original Kalman paper the gain derived there,  $\Delta_k$ , is given by  $\Delta_k = \mathbf{A}_{k-1}\mathbf{K}_k$ .



## 6 Information form

In the information filter (inverse-covariance filter) the estimated vector and the covariance matrix has been replaced by the **information state**  $\mathbf{y}_k$  and the **information matrix**  $\mathbf{Y}_k$  respectively.

$$\mathbf{y}_k^a \equiv \mathbf{Y}_k \mathbf{x}_k^a \quad (50)$$

$$\mathbf{Y}_k \equiv \mathbf{P}_k^{-1} \quad (51)$$

The same information form has the forecast estimate and covariance matrix.

$$\mathbf{y}_k^f \equiv \mathbf{Y}_k^f \mathbf{x}_k^f \quad (52)$$

$$\mathbf{Y}_k^f \equiv (\mathbf{P}_k^f)^{-1} \quad (53)$$

With this changes we try to write Kalman filter equations in the information form [6]. So the data assimilation equations become:

$$\begin{aligned} \mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^f) & (54) \\ \mathbf{P}_k \mathbf{y}_k^a &= \mathbf{P}_k^f \mathbf{y}_k^f + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k \mathbf{P}_k^f \mathbf{x}_k^f) \\ &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f \mathbf{y}_k^f + \mathbf{K}_k \mathbf{z}_k \\ &= \mathbf{P}_k \mathbf{y}_k^f + \mathbf{K}_k \mathbf{z}_k \\ \mathbf{y}_k^a &= \mathbf{y}_k^f + \mathbf{P}_k^{-1} \mathbf{K}_k \mathbf{z}_k \\ &= \mathbf{y}_k^f + (\mathbf{P}_k^f)^{-1} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^{-1} \mathbf{K}_k \mathbf{z}_k \\ &= \mathbf{y}_k^f + [\mathbf{K}_k^{-1} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f]^{-1} \mathbf{z}_k \\ &= \mathbf{y}_k^f + (\mathbf{K}_k^{-1} \mathbf{P}_k^f - \mathbf{H}_k \mathbf{P}_k^f)^{-1} \mathbf{z}_k \\ &= \mathbf{y}_k^f + [(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k) (\mathbf{H}_k^T)^{-1} (\mathbf{P}_k^f)^{-1} \mathbf{P}_k^f - \mathbf{H}_k \mathbf{P}_k^f]^{-1} \mathbf{z}_k \\ \mathbf{y}_k^a &= \mathbf{y}_k^f + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{z}_k & (55) \end{aligned}$$

Derivation of the information matrix follows immediately from the posterior covariance matrix formula.

$$\begin{aligned} \mathbf{P}_k^{-1} &= (\mathbf{P}_k^f)^{-1} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^{-1} & (56) \\ &= (\mathbf{P}_k^f)^{-1} [\mathbf{K}_k (\mathbf{K}_k^{-1} - \mathbf{H}_k)]^{-1} \\ &= (\mathbf{P}_k^f)^{-1} \left[ \mathbf{K}_k [(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k) (\mathbf{H}_k^T)^{-1} (\mathbf{P}_k^f)^{-1} - \mathbf{H}_k] \right]^{-1} \\ &= (\mathbf{P}_k^f)^{-1} \left[ \mathbf{K}_k \mathbf{R}_k (\mathbf{H}_k^T)^{-1} (\mathbf{P}_k^f)^{-1} \right]^{-1} \\ &= \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{K}_k^{-1} \\ &= \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k) (\mathbf{H}_k^T)^{-1} (\mathbf{P}_k^f)^{-1} \\ &= \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{R}_k (\mathbf{H}_k^T)^{-1} (\mathbf{P}_k^f)^{-1} \\ \mathbf{Y}_k &= \mathbf{Y}_k^f + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k & (57) \end{aligned}$$

Provided that  $\mathbf{A}_{k-1}$  is nonsingular, equations of the model forecast become:

$$\begin{aligned}
\mathbf{x}_k^f &= \mathbf{A}_{k-1}\mathbf{x}_{k-1}^a + \mathbf{B}_{k-1}\mathbf{u}_{k-1} & (58) \\
\mathbf{y}_k^f &= (\mathbf{P}_k^f)^{-1}\mathbf{A}_{k-1}\mathbf{P}_{k-1}\hat{\mathbf{y}}_{k-1} + \mathbf{Y}_k^f\mathbf{B}_{k-1}\mathbf{u}_{k-1} \\
&= [\mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}]^{-1}\mathbf{A}_{k-1}\mathbf{P}_{k-1}\hat{\mathbf{y}}_{k-1} + \mathbf{Y}_k^f\mathbf{B}_{k-1}\mathbf{u}_{k-1} \\
&= \left[ \mathbf{P}_{k-1}^{-1}\mathbf{A}_{k-1}^{-1}[\mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}] \right]^{-1} \hat{\mathbf{y}}_{k-1} + \mathbf{Y}_k^f\mathbf{B}_{k-1}\mathbf{u}_{k-1} \\
&= (\mathbf{A}_{k-1}^T + \mathbf{Y}_{k-1}\mathbf{A}_{k-1}^{-1}\mathbf{Q}_{k-1})^{-1} + \mathbf{Y}_k^f\mathbf{B}_{k-1}\mathbf{u}_{k-1} \\
&= (\mathbf{I} + (\mathbf{A}_{k-1}^{-1})^T\mathbf{Y}_{k-1}\mathbf{A}_{k-1}^{-1}\mathbf{Q}_{k-1})^{-1}(\mathbf{A}_{k-1}^T)^{-1} + \mathbf{Y}_k^f\mathbf{B}_{k-1}\mathbf{u}_{k-1} \\
\mathbf{y}_k^f &= (\mathbf{I} + \mathbf{M}_{k-1}\mathbf{Q}_{k-1})^{-1}(\mathbf{A}_{k-1}^T)^{-1} + \mathbf{Y}_k^f\mathbf{B}_{k-1}\mathbf{u}_{k-1} & (59)
\end{aligned}$$

where  $\mathbf{M}_k = (\mathbf{A}_k^{-1})^T\mathbf{Y}_k\mathbf{A}_k^{-1}$ . Using the forecast covariance matrix recurrence formula we can derive its counterpart information matrix.

$$\begin{aligned}
\mathbf{Y}_k^f &= (\mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1})^{-1} & (60) \\
&= (\mathbf{I} + \mathbf{M}_{k-1}\mathbf{Q}_{k-1})^{-1}\mathbf{M}_{k-1}
\end{aligned}$$

The summary of the information form of the Kalman filter:

Initialization (given  $\mu_0$  and  $\mathbf{P}_0$ ):

$$\begin{aligned}
\mathbf{Y}_0 &= \mathbf{P}_0^{-1} \\
\mathbf{y}_0^a &= \mathbf{Y}_0\mu_0
\end{aligned}$$

Model Forecast Step/Predictor:

$$\begin{aligned}
\mathbf{M}_k &= (\mathbf{A}_k^{-1})^T\mathbf{Y}_k\mathbf{A}_k^{-1} \\
\mathbf{Y}_k^f &= (\mathbf{I} + \mathbf{M}_{k-1}\mathbf{Q}_{k-1})^{-1}\mathbf{M}_{k-1} \\
\mathbf{y}_k^f &= (\mathbf{I} + \mathbf{M}_{k-1}\mathbf{Q}_{k-1})^{-1}(\mathbf{A}_{k-1}^T)^{-1} + \mathbf{Y}_k^f\mathbf{B}_{k-1}\mathbf{u}_{k-1}
\end{aligned}$$

Data Assimilation Step/Corrector:

$$\begin{aligned}
\mathbf{y}_k^a &= \mathbf{y}_k^f + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{z}_k \\
\mathbf{Y}_k &= \mathbf{Y}_k^f + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k
\end{aligned}$$

## 7 Innovation approach

The method solves the estimation problem much easier using the innovation process, which is the observed process converted into a white-noise process. The innovation represents the new information measure in the observation variable  $\mathbf{z}_k$ , being given all the past observations and the old information deduced therefrom. It is defined as:

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k - E[\mathbf{z}_k|\mathbf{Z}_{k-1}] \quad (61)$$

Several properties of the innovation [3]:

1. The innovation  $\tilde{\mathbf{z}}_k$ , associated with the current observation, is uncorrelated with the past observations:  $E[\tilde{\mathbf{z}}_k \mathbf{z}_j^T] = 0$  where  $j = 1, 2, \dots, k-1$ .
2. The innovations are orthogonal to each other:  $E[\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_j^T] = 0$  where  $i \neq j$ .
3. There is a one-to-one correspondence between the innovation and the associated observation.
4. It has zero mean.

We know that  $\mathbf{x}_k^a \equiv E[\mathbf{x}_k | \mathbf{Z}_k]$ . Since the innovation sequence  $\{\tilde{\mathbf{z}}_k\}$  “contains all the information in the observation sequence”  $\{\mathbf{z}_k\}$  [2], the estimate can be assumed to be a linear combination of all the innovations up to  $k$  [9]:

$$\mathbf{x}_k^a = \sum_{i=1}^k \mathbf{I}_i \tilde{\mathbf{z}}_i \quad (62)$$

where  $\mathbf{I}_i$  is an  $n \times m$  matrix to be determined. We know by the Projection Theorem that the forecast error  $\mathbf{e}_k$  is uncorrelated with the innovation sequence. Therefore, for all  $i$  up to  $k$ :

$$\begin{aligned} 0 &= E[\mathbf{e}_k \tilde{\mathbf{z}}_i^T] \\ &= E[(\mathbf{x}_k - \mathbf{x}_k^a) \tilde{\mathbf{z}}_i^T] \\ E[\mathbf{x}_k \tilde{\mathbf{z}}_i^T] &= E[\mathbf{x}_k^a \tilde{\mathbf{z}}_i^T] \\ E[\mathbf{x}_k \tilde{\mathbf{z}}_i^T] &= \sum_{l=1}^k \mathbf{I}_l E[\tilde{\mathbf{z}}_l \tilde{\mathbf{z}}_i^T] \\ &= \mathbf{I}_i E[\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^T] \end{aligned} \quad (63)$$

All the terms up to  $k-1$  go to zero since the innovations are temporally uncorrelated.

Where  $E[\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^T]$  is the innovation covariance  $Cov(\tilde{\mathbf{z}}_i)$ . Results:

$$\mathbf{I}_i = E[\mathbf{x}_k \tilde{\mathbf{z}}_i^T] Cov^{-1}(\tilde{\mathbf{z}}_i) \quad (64)$$

Substitute this into (62):

$$\begin{aligned} \mathbf{x}_k^a &= \sum_{i=1}^k E[\mathbf{x}_k \tilde{\mathbf{z}}_i^T] Cov^{-1}(\tilde{\mathbf{z}}_i) \tilde{\mathbf{z}}_i \\ &= \sum_{i=1}^{k-1} E[\mathbf{x}_k \tilde{\mathbf{z}}_i^T] Cov^{-1}(\tilde{\mathbf{z}}_i) \tilde{\mathbf{z}}_i + E[\mathbf{x}_k \tilde{\mathbf{z}}_k^T] Cov^{-1}(\tilde{\mathbf{z}}_k) \tilde{\mathbf{z}}_k \\ &= \sum_{i=1}^{k-1} E[(\mathbf{A}_{k-1} \mathbf{x}_{k-1} + \mathbf{B}_{k-1} \mathbf{u}_{k-1} + \mathbf{w}_{k-1}) \tilde{\mathbf{z}}_i^T] Cov^{-1}(\tilde{\mathbf{z}}_i) \tilde{\mathbf{z}}_i + E[\mathbf{x}_k \tilde{\mathbf{z}}_k^T] Cov^{-1}(\tilde{\mathbf{z}}_k) \tilde{\mathbf{z}}_k \\ &= \mathbf{A}_{k-1} \sum_{i=1}^{k-1} E[\mathbf{x}_{k-1} \tilde{\mathbf{z}}_i^T] Cov^{-1}(\tilde{\mathbf{z}}_i) \tilde{\mathbf{z}}_i + E[\mathbf{x}_k \tilde{\mathbf{z}}_k^T] Cov^{-1}(\tilde{\mathbf{z}}_k) \tilde{\mathbf{z}}_k \\ &= \mathbf{A}_{k-1} \mathbf{x}_{k-1}^a + E[\mathbf{x}_k \tilde{\mathbf{z}}_k^T] Cov^{-1}(\tilde{\mathbf{z}}_k) \tilde{\mathbf{z}}_k \\ &= \mathbf{A}_{k-1} \mathbf{x}_{k-1}^a + \mathbf{K}_k \tilde{\mathbf{z}}_k \end{aligned} \quad (65)$$

where  $\mathbf{K}_k = E[\mathbf{x}_k \tilde{\mathbf{z}}_k^T] Cov^{-1}(\tilde{\mathbf{z}}_k)$ .

The error in the estimate is:

$$\begin{aligned} \mathbf{e}_k &\equiv \mathbf{x}_k - \mathbf{x}_k^a \\ &= \mathbf{A}_{k-1} \mathbf{e}_{k-1} + \mathbf{w}_{k-1} - \mathbf{K}_k \tilde{\mathbf{z}}_k \end{aligned} \quad (66)$$

Then, the posterior covariance of the new estimate is:

$$\begin{aligned} \mathbf{P}_k &\equiv E[\mathbf{e}_k \mathbf{e}_k^T] \\ &= \mathbf{A}_{k-1} E[\mathbf{e}_{k-1} \mathbf{e}_{k-1}^T] \mathbf{A}_{k-1}^T + \mathbf{A}_{k-1} E[\mathbf{e}_{k-1} \tilde{\mathbf{z}}_k^T] \mathbf{K}_k^T + E[\mathbf{w}_{k-1} \mathbf{w}_{k-1}^T] - E[\mathbf{w}_{k-1} \tilde{\mathbf{z}}_k^T] \mathbf{K}_k^T \\ &\quad - \mathbf{K}_k E[\tilde{\mathbf{z}}_k \mathbf{e}_{k-1}^T] \mathbf{A}_{k-1}^T - \mathbf{K}_k E[\tilde{\mathbf{z}}_k \mathbf{w}_{k-1}^T] + \mathbf{K}_k E[\tilde{\mathbf{z}}_k \tilde{\mathbf{z}}_k^T] \mathbf{K}_k^T \\ &= \mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^T + \mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^T \mathbf{H}_k \mathbf{K}_k^T + \mathbf{Q}_{k-1} - \mathbf{Q}_{k-1} \mathbf{H}_k^T \mathbf{K}_k^T \\ &\quad - \mathbf{K}_k \mathbf{H}_k \mathbf{A}_{k-1} \mathbf{A}_{k-1} \mathbf{P}_{k-1}^T - \mathbf{K}_k \mathbf{H}_k \mathbf{Q}_{k-1}^T + \mathbf{K}_k Cov(\tilde{\mathbf{z}}_k) \mathbf{K}_k^T \end{aligned} \quad (67)$$

Denote  $\mathbf{P}_k^f = \mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}$  and substitute this back into (67) yields:

$$\begin{aligned} \mathbf{P}_k &= \mathbf{P}_k^f - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^f - \mathbf{P}_k^f \mathbf{H}_k^T \mathbf{K}_k^T + \mathbf{K}_k Cov(\tilde{\mathbf{z}}_k) \mathbf{K}_k^T \\ &= \mathbf{P}_k^f - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k^f - \mathbf{P}_k^f \mathbf{H}_k^T \mathbf{K}_k^T + E[\mathbf{x}_k \tilde{\mathbf{z}}_k^T] \mathbf{K}_k^T \end{aligned} \quad (68)$$

But  $E[\mathbf{x}_k \tilde{\mathbf{z}}_k^T] = \mathbf{P}_k^f \mathbf{H}_k^T$  (see (46)) and  $Cov(\tilde{\mathbf{z}}_k) = \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k$  (see (47)). Then:

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f \quad (69)$$

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \quad (70)$$

The equations (61), (62), (69) and (70) define the Kalman Filter algorithm.

## 8 Properties of Kalman Filter

### Stability - Theorem Jaswinski

Asymptotic stability of the KF means that its solution will gradually become insensitive to its initial conditions, provided that the norms of the noise covariance matrices,  $\mathbf{Q}_k, \mathbf{R}_k$  are bounded.

If the system (4) and (5) with  $\mathbf{x}_0, \mathbf{w}_k, \mathbf{v}_k$ , independent, is uniformly completely observable and uniformly completely controllable and if  $\mathbf{P}_0 \geq 0$  then the discrete time KF is uniformly asymptotically stable.

### Filter Divergence

This phenomenon occurs when the filter seems to behave well, having low error variance, but the estimate is far away from the truth. This is due to errors in the system modeling: the model error is higher than expected, the system model has the wrong form, the system is unstable or has bias errors.

## 9 Remarks

1. The filter produce the error covariance matrix  $\mathbf{P}_k$  which is an important estimate for the accuracy of the **estimate**.

2. The filter is **optimal for Gaussian** sequences only.
3. While the measurement noise covariance  $\mathbf{R}_k$  is possible to be determined, the process noise covariance matrix  $\mathbf{Q}_k$  has to be computed to adjust to different dynamics. We are not able to directly observe the process we are estimating. Therefore a **tuning** on  $\mathbf{Q}_k$  has to be performed for superior filter performances.

## 10 Conclusion

While most of the filters are formulated in the frequency domain, the Kalman Filter is a purely **time-domain filter**.

The main issue remains how the uncertainties are represented.

## References

- [1] Michael Athans. *The Control Handbook*, chapter Kalman Filtering, pages 589–594. CRC Press, 1996.
- [2] A.V.Balakrishnan. *Kalman Filtering Theory*. Optimization Software, Inc., 1984.
- [3] Mourad Barkat. *Signal Detection and Estimation*. Artech House Inc, 2005.
- [4] Jon Dattorro. *Convex Optimization & Euclidean Distance Geometry*, chapter Matrix Calculus. Meboo Publishing USA, 2006.
- [5] Henk Eskes. Data Assimilation: The Kalman Filter.
- [6] Mohinder S. Grewal and Angus P. Andrews. *Kalman Filtering. Theory and Practice using Matlab 2nd*. John Wiley & Sons, 2001.
- [7] John M. Lewis and S.Lakshmivarahan. *Dynamic Data Assimilation, A Least Squares Approach*. 2006.
- [8] Peter S. Mayback. *Introduction to Random Signals and Applied Kalman Filtering*. Academic Press, 1979.
- [9] Athanasios Papoulis. *Probability, Random Variables, and Stochastic Process*. McGraw-Hill, Inc., 2nd edition, 1965.
- [10] R.E.Kalman. A New Approach to Linear Filtering and Prediction Problems. *Trans.ASME*, 1960.
- [11] Greg Welch and Gary Bishop. An Introduction to the Kalman Filter. *SIGGRAPH, ACM*, 2001.