3. Filtering Theory

In general terms, "filtering" means to remove something that we do not want to pass through a barrier, e.g., water purification. When the entities involved are signals, such as an electrical voltage, the barrier becomes a filter in the sense of signal processing.

To determine whether a system is performing properly, and ultimately to control the system's performance, we must know what the system is "doing" at any instant of time. In other words, we must know the state of the system; for example, in navigation, the state consists of position and velocity of the craft in question. In a bioreactor, the state may be substrate concentrations and enzyme activity. When we build measuring devices and take measurements of a system, the measurements are generally contaminated with noise caused by the electronic and mechanical components of the measuring device. The problem of determining the state of a system from noisy measurements is called estimation, or filtering.

The statistical approaches to filtering postulate that certain statistical properties are possessed by the useful part of the signal and by the unwanted noisy part of the signal. Measurements are the sum of the useful part and the noisy part of the signal, and the task is to eliminate as much of the noise as possible through processing of the measurements by a filter. The earliest statistical ideas of Wiener [6] and Kolmogorov [7] concern processes with statistical properties which do not change with time, i.e., stationary processes.

This assumption that the underlying useful signal and noise processes are stationary is crucial to the Wiener and Kolmogorov theory. However, a theory that does not require this assumption was developed by Kalman in his celebrated paper in 1960 [8]. The new theory is called the Kalman filter and is an optimal state estimation process which is applied to a dynamical system that
involves random perturbations. More precisely, the Kalman filter consists of a linear, unbiased, and minimum error variance recursive algorithm to optimally estimate the state of a dynamical system from noisy data taken at discrete real-time intervals. It has been widely used in many areas of industrial and governmental applications such as laser tracking systems, satellite navigation, ballistic missile trajectory estimation, radar, and fire control.

In this chapter, we are concerned only with Kalman filter theory. This theory will be applied to enzyme systems in the next two chapters. The unifying theme in this chapter is the probabilistic or Bayesian approach. To derive the Kalman filter of a linear dynamical system a Hilbert space approach is possible [9], but it is difficult to generalize this approach to a nonlinear dynamical system.

In the next section, we will summarize some results from probability theory that are required to understand and interpret the Kalman filter. Further information on probability theory can be found in [10], [11].

3.1. Probability Preliminaries

One of the fundamental concepts in the Kalman filter is the gaussian process, so we will start with it.

1. Gaussian Random Vector

Let \( \mathbf{X} \) be a random \( n \)-vector with expectation denoted by \( \mathbb{E}[\mathbf{X}] \) and covariance matrix \( \Sigma = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top] \). If \( \Sigma \) has a nonsingular covariance matrix, i.e., \( \det(\Sigma) \neq 0 \), then we say that \( \mathbf{X} \) is gaussian or normal if and only if its probability density function is of the form

\[
p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det \Sigma|^{1/2}} \exp\left[ -\frac{1}{2} (\mathbf{x} - \mathbb{E}[\mathbf{X}])^\top \Sigma^{-1} (\mathbf{x} - \mathbb{E}[\mathbf{X}]) \right].
\]
For a gaussian random vector, $X$, with mean $\bar{X}$ and covariance matrix $\Sigma$, we write $X \sim N(\bar{X}, \Sigma)$. If we denote $\phi_X$ to be the characteristic function (the Fourier transform of the probability density function) of $X$, then $\phi_X$ is evaluated to be

$$\phi_X = \exp \left( j s^T \bar{X} - \frac{1}{2} s^T \Sigma s \right),$$

where $j$ is $\sqrt{-1}$.

2. **Theorem 1**

Let $X$ and $Y$ be random n-vectors with $Y = f(X)$. Suppose $f^{-1}$ exists and both $f$ and $f^{-1}$ are continuously differentiable. Then the probability density function for $X$ and $Y$ are related by

$$p_Y(y) = p_X(f^{-1}(y)) \cdot \left| \det \left( \frac{\partial f^{-1}(y)}{\partial y} \right) \right|,$$

where $\det$ is the determinant.

(See [12] for the proof of Theorem 1.)

This result is important in Bayesian probability theory to compute the conditional probability density function conditional on a set of measurements. It is used in the proof of Theorem 4 which is used extensively in the derivation of the Kalman filter algorithm and is used in deriving the exact filter.

3. **Jointly Gaussian Distribution**

To say that random n-vector $X$ and random m-vector $Y$ are jointly gaussian random variables is the same as saying that the random vector $Z^T = [X, Y]^T$ is a gaussian random vector. The probability density function and the characteristic function are given by

$$p_z(z) = \frac{1}{(2\pi)^{(m+n)/2} \cdot \det(R_z)^{1/2}} \cdot \exp \left[ -\frac{1}{2} (z - \bar{Z})^T \cdot P_z \cdot (z - \bar{Z}) \right],$$

(3.2)
\[ \phi_z(u) = \exp\left(j u^T \bar{Z} - \frac{1}{2} u^T R_z u\right), \]  

where \( \bar{Z} \in \mathbb{R}^{m+n}, \bar{Z}^T = E[Z^T] = \left[ E[X^T], E[Y^T] \right] = \left[ \bar{X}^T, \bar{Y}^T \right], \) and

\[
R_z = E\left[(z - \bar{Z})(z - \bar{Z})^T\right] \\
= \begin{bmatrix}
E[(x - \bar{X})(x - \bar{X})^T] & E[(x - \bar{X})(y - \bar{Y})^T] \\
E[(y - \bar{Y})(x - \bar{X})^T] & E[(y - \bar{Y})(y - \bar{Y})^T]
\end{bmatrix} \\
= \begin{bmatrix}
R_x & R_{xy} \\
R_{yx} & R_y
\end{bmatrix}.
\]

4. **Theorem 2**

If the jointly gaussian distributed random vectors, \( X \) and \( Y \), are uncorrelated, they are independent.

(See [12] for a proof of Theorem 2.)

Theorem 2 and Theorem 3 below, are used in the proof of theorem 4.

5. **Theorem 3**

Consider the p-vector \( Z \sim \mathcal{N}(\bar{Z}, R_z) \), and let \( w = CZ + a \), where \( C \) is a \( q \times p \) constant matrix, \( a \) is a constant \( q \)-vector, and \( w \) is a random \( q \)-vector. Then \( w \sim \mathcal{N}(C \cdot \bar{Z} + a, C \cdot R_z \cdot C^T) \).

(See [12] for a proof of Theorem 3.)

This theorem essentially means that linear operations on gaussian random vectors produce gaussian random vectors.

6. **Theorem 4**

Let \( X \) and \( Y \) be jointly gaussian distributed random vectors of dimension \( n \) and \( m \), respectively. Then the conditional density of \( X \) given \( Y = y \) is gaussian with mean

\[
\bar{X} + R_{xy} R_y^{-1} (y - \bar{Y})
\]

and covariance matrix

\[
R_x - R_{xy} R_y^{-1} R_{yx}.
\]
This theorem is essential in the derivation of the Kalman filter and so we present its proof below. In the next section, we will show how to obtain the Kalman filter algorithms by repeated application of this theorem.

**Proof**

This proof combines elements found in different sources [12], [13].

Consider the transformation

\[
\begin{align*}
W_1 &= X - R_{xy} R_{y}^{-1} Y, \\
W_2 &= Y.
\end{align*}
\]  

(3.7)

From Theorem 3, \( W^T = [W_1^T, W_2^T] \) is gaussian distributed. The covariance matrix of \( W_1 \) is

\[
E[(W_1 - E[W_1])(W_1 - E[W_1])^T] = R_x - R_{xy} R_{y}^{-1} R_{yx}.
\]

An easy computation shows that \( W_1 \) and \( W_2 \) are uncorrelated. Hence, they are independent by Theorem 2. As a result, their joint probability density function is the product of their individual densities

\[
\begin{align*}
W_1 &\sim N(\overline{X} - R_{xy} R_{y}^{-1} \overline{Y}, R_x - R_{xy} R_{y}^{-1} R_{yx}), \\
W_2 &\sim N(\overline{Y}, R_{y}).
\end{align*}
\]

The joint probability density function is given by

\[
p_{w_1,w_2}(w_1, w_2) = \left[ (2\pi)^{n/2} \left| \det (R_x - R_{xy} R_{y}^{-1} R_{yx}) \right|^{1/2} \right]^{-1} \\
\cdot \exp \left[ -\frac{1}{2} (w_1 - \overline{X} - R_{xy} R_{y}^{-1} \overline{Y})^T (R_x - R_{xy} R_{y}^{-1} R_{yx}) (w_1 - \overline{X} - R_{xy} R_{y}^{-1} \overline{Y}) \right] \\
\quad \cdot \left[ (2\pi)^{m/2} \left| \det (R_{y}) \right|^{1/2} \right]^{-1} \exp \left[ -\frac{1}{2} (w_2 - \overline{Y})^T R_{y}^{-1} (w_2 - \overline{Y}) \right].
\]  

(3.8)

The jacobian of the transformation (3.7) is 1, and so according to Theorem 1, we obtain the joint probability density function \( p_{x,y}(x,y) \) by eliminating \( W_1 \) and \( W_2 \) in (3.8) in favor of \( X \) and \( Y \) via (3.7). As a result we get
\[ p_{x,y}(x,y) = \left[ (2\pi)^{n/2} \det\left( R_x - R_{xy} R_y^{-1} R_{yx} \right) \right]^{-1} \]
\[ \cdot \exp \left[ -\frac{1}{2} \left( x - \bar{X} - R_{xy} R_y^{-1} (y - \bar{Y}) \right)^T \left( R_x - R_{xy} R_y^{-1} R_{yx} \right)^{-1} \left( x - \bar{X} - R_{xy} R_y^{-1} (y - \bar{Y}) \right) \right] \]
\[ \cdot \left[ (2\pi)^{m/2} \det(R_y) \right]^{-1} \exp \left[ -\frac{1}{2} (y - \bar{Y})^T R_y^{-1} (y - \bar{Y}) \right]. \] (3.9)

The second probability density function in (3.9) is the probability density function of \( y \). The conditional probability density function of \( x \) on \( y \) is defined by

\[ p_{x|y}(x|y) = \frac{p_{x,y}(x,y)}{p_y(y)}. \] (3.10)

Comparing (3.9) and (3.10) the conditional probability density of \( x \) on \( y \) is given by

\[ p_{x|y}(x|y) = \left[ (2\pi)^{n/2} \det\left( R_x - R_{xy} R_y^{-1} R_{yx} \right) \right]^{-1} \]
\[ \cdot \exp \left[ -\frac{1}{2} \left( x - \bar{X} - R_{xy} R_y^{-1} (y - \bar{Y}) \right)^T \left( R_x - R_{xy} R_y^{-1} R_{yx} \right)^{-1} \left( x - \bar{X} - R_{xy} R_y^{-1} (y - \bar{Y}) \right) \right]. \] (3.11)

### 3.2. Kalman Filter Theory

In this section, the linear Kalman filter algorithm will be derived. Before the derivation, we define the notation used in the filter equations:

- \( \bar{X} \) denote the mean of a random vector \( X \).
- \( Z^{k-1} \) denotes the set of measurement vectors \( \{Z_1, Z_2, \ldots, Z_{k-1}\} \) where \( Z_i \) for \( i = 1, 2, \ldots, k-1 \), is the q-vector representing the measurements at time \( t_i \).
- \( X_{k|i} \) denotes the mean of the conditional probability density of \( X_k \) conditioned on \( Z^i \), e.g., \( X_{k|k-1} \) predicts the state at time \( t_k \) from measurements up to time \( t_{k-1} \), whereas \( X_{k|k} \) updates the predicted state at time \( t_k \) using all measurements up to time \( t_k \).
\( R_{XY} \) denotes the covariance matrix of random vectors \( X, Y \), with
\[
R_{XY} = E[(X - \bar{X})(Y - \bar{Y})^T].
\]
\( P_{k|i} \) denotes the covariance matrix of a random vector \( X_k \) conditioned on the measurement set \( Z^i \).

3.2.1. System Description

We shall restrict our attention to discrete-time systems, or, equivalently, systems where the underlying system equations are difference equations rather than differential equations. In some cases, the difference equations correspond to discretizations of differential equations. The reason for studying a discrete-time system is that observations are made and control strategies are implemented frequently at discrete times in a real system.

The dynamics of the state vector \( X \) is described by the following system of difference equations:
\[
X_{k+1} = F_k X_k + \xi_k, \quad k=0,1,2,3,..., \tag{3.12}
\]
where \( X_k \) is a random \( p \)-vector, \( F_k \) is a \( p \times p \) matrix independent of \( X_k \), and \( \xi_k \) is a random \( p \)-vector representing the model system error. The statistical properties of \( \xi_k \) are described in the next section.

The measurements of the system at time \( t_k \) are described by the following system of algebraic equations
\[
Z_k = H_k X_k + \eta_k, \tag{3.13}
\]
where \( H_k \) is a \( q \times p \) matrix independent of \( X_k \), \( Z_k \) is the vector of quantities measured at time \( t_k \), and \( \eta_k \) is a random \( q \)-vector representing the model system error. The statistical properties of \( \eta_k \) are described in the next section.

3.2.2. Noise Description

To derive the Kalman filter (KF), we have to make assumptions about the noise processes \( \xi_k \) and \( \eta_k \).
Assumptions on the noise processes

The noise processes \( \{ \xi_k \} \) and \( \{ \eta_k \} \) are white noise, i.e., \( \xi_k \) and \( \eta_k \) are gaussian variables with zero mean and

\[
E[\xi_k \xi_i^\top] = \delta_{ki} Q_k, \quad E[\eta_k \eta_i^\top] = \delta_{ki} S_k, \quad E[\xi_k \xi_i^\top] = E[\eta_k \eta_i^\top] = 0,
\]

(3.14)

where \( S_k \) and \( Q_k \) are \( q \times q \) and \( p \times p \) matrices, respectively, and \( \delta_{kl} \) is the Kronecker delta.

In most of the literature, the covariance matrix, \( Q_k \), is called the model error covariance matrix and the covariance matrix, \( S_k \), is called the measurement error covariance matrix, respectively.

3.2.3. Derivation of Kalman Filter Algorithm

In this section, we derive the KF algorithm using Theorem 4 which is based on Bayesian probability theory. This derivation is quite long and we will break it up to into several lemmas. Because of the recursive nature of the resulting formulas, i.e., the index appears only in the subscripts, it suffices to prove the lemmas for \( k=0 \) and \( 1 \). Each lemma summarizes an important result.

In breaking up the derivation of the KF, our objective is to use the known measurements of a system to both update its present state and predict its future state. In Lemma 1, we develop formulas to update the mean and covariance of the system state at time \( t_0 \) based on measurements up to time \( t_0 \). The distinction between the formulas in Lemma 1 and the updating formulas in Lemma 5 is that the formulas in Lemma 1 require us to supply the initial value covariance matrix. Lemmas 2, 3, 4, and 6 give results predicting the state at time \( t_k \) when the measurements up to time \( t_{k-1} \) is given.

3.2.3.1. Lemma 1

The mean \( X_{0/0} \) and the covariance matrix \( P_{0/0} \) of the conditional probability density function of \( X_0 \) conditioned on \( Z_0 \) are, respectively,
\[ X_{0/0} = \overline{X}_0 + R_{X_0 Z_0} H_0^T (H_0 R_{X_0 X_0} H_0^T + S)^{-1} (Z_0 - H_0 \overline{X}_0), \]  
\[ P_{0/0} = R_{X_0 Z_0} - R_{X_0 X_0} H_0^T (H_0 R_{X_0 X_0} H_0^T + S)^{-1} H_0 R_{X_0 X_0}, \]

where \( \overline{X}_0 \) is the mean of the probability density of \( X_0 \).

**Proof**

Applying Theorem 4 from Section 3.1, the conditional mean \( X_{0/0} \) and the conditional covariance matrix are given symbolically as

\[ X_{0/0} = \overline{X}_0 + R_{X_0 Z_0} R_{Z_0 Z_0}^{-1} (Z_0 - \overline{Z}_0), \]  
\[ P_{0/0} = R_{X_0 X_0} - R_{X_0 Z_0} R_{Z_0 Z_0}^{-1} R_{Z_0 X_0}. \]

We need to evaluate \( R_{Z_0 Z_0} \) and \( R_{X_0 Z_0} \) from the dynamical and measurement model equations (3.12)and (3.13).

Firstly, let us evaluate \( R_{X_0 Z_0} \) from its definition

\[ R_{X_0 Z_0} = E[(X_0 - \overline{X}_0)(Z_0 - \overline{Z}_0)^T]. \]

Applying the measurement model equation we get

\[ R_{X_0 Z_0} = E[(X_0 - \overline{X}_0)(H_0 (X_0 - \overline{X}_0) + \eta_0)^T]. \]

Since \( X_0 \) and \( \eta_0 \) are independent, \( R_{X_0 Z_0} \) can be simplified to

\[ R_{X_0 Z_0} = R_{X_0 X_0} H_0^T. \]  

(3.19)

Similarly, if we apply the definition of \( R_{X_0 \eta_0} \), measurement model equation and \( E[\eta_0 \eta_0^T] = S_0 \), we obtain

\[ R_{Z_0 Z_0} = H_0 R_{X_0 X_0} H_0^T + S_0. \]  

(3.20)

Finally, \( R_{Z_0 Z_0} \) is the transpose of \( R_{X_0 Z_0} \)

\[ R_{Z_0 Z_0} = H_0 R_{X_0 X_0}. \]  

(3.21)

Substitution of (3.19), (3.20), (3.21) into (3.17), and (3.18) will complete the proof.

### 3.2.3.2. Lemma 2
The mean of the probability density function of $X_1$ conditioned on $Z_0$ is given by

$$X_{1/0} = F_0 X_{0/0}.$$  \hfill (3.22)

**Proof**

Applying Theorem 4, the mean of the probability density function of $X_1$ conditioned on $Z_0$ is given by

$$X_{1/0} = \overline{X}_1 + R_{x_{1/0}}^{-1} R_{z_{0/0}}^{-1} (Z_0 - \overline{Z}_0).$$  \hfill (3.23)

Firstly, $\overline{X}_1$ is the mean of $X_1$ and can be shown to be equal to $F_0 \overline{X}_0$ from the dynamical model equation. Secondly, applying definition of $R_{x_{1/0}}$ to the dynamical and measurement model equations. Calculations similar to those in the proof of Lemma 1 show that

$$R_{x_{1/0}} = F_0 R_{x_{0/0}} H_0^T.$$  \hfill (3.24)

Substituting (3.24) and $\overline{X}_1 = F_0 \overline{X}_0$ into (3.23) and using Lemma 1 gives the desired result.

**3.2.3.3. Lemma 3**

The covariance matrix of the probability density function of $X_1$ conditioned on $Z_0$ is given by

$$P_{1/0} = F_0 P_{0/0} F_0^T + Q_0.$$  \hfill (3.25)

**Proof**

Applying Theorem 4 the covariance matrix of the probability density function of $X_1$ conditioned on $Z_0$ is given by

$$P_{1/0} = R_{x_{1/0}} - R_{x_{1/0}} R_{z_{0/0}}^{-1} R_{z_{0/0}}.$$  \hfill (3.26)

To get an expression for $P_{1/0}$, we have to calculate $R_{x_{1/0}}$ since the other terms in (3.26) have been calculated in Lemma 2.

From the definition of $R_{x_{1/0}}$ and the dynamical model equation,

$$R_{x_{1/0}} = E\left[(F_0 X_0 + \xi_0 - F_0 \overline{X}_0)(F_0 X_0 + \xi_0 - F_0 \overline{X}_0)^T\right].$$
Rearrangement of terms and the independence of $X_1$ and $\xi_0$ implies
\[ R_{x_1x_1} = F_0 R_{x_0x_0} F_0^T + Q_0, \]
where
\[ Q_0 = E[\xi_0^T \xi_0]. \]
Substitution of $R_{x_1x_1}$, $R_{x_1z_0}$, $R_{z_0z_0}^{-1}$ into (3.26) and application of Lemma 1 give
the desire result.

**3.2.3.4. Lemma 4**

Let $X_1$, $Z_1$, $Z_0$ be jointly gaussian, then the mean of the probability density function of the random vector
\[ \begin{pmatrix} X_1 \\ Z_1 \end{pmatrix} \]
conditioned on $Z_0$ has mean $\begin{pmatrix} X_{1/0} \\ Z_{1/0} \end{pmatrix}$ and
covariance matrix
\[
\begin{bmatrix}
P_{1/0} & P_{1/0} H_1^T \\
H_1 P_{1/0} & H_1 P_{1/0} H_1^T + S_1
\end{bmatrix}
\]  \hspace{1cm} (3.27)

where $S_1 = E[\eta_1 \eta_1^T]$.

**Proof**

From Theorem 4, the mean of the probability density function of $Z_1$
conditioned on $Z_0$ is given by
\[ Z_{1/0} = \bar{Z}_1 + R_{z_1z_0} R_{z_0z_0}^{-1} R_{z_0} X_1. \]

Also, consider $\begin{pmatrix} X_1 \\ Z_1 \end{pmatrix}$ as a random vector, then from application of Theorem 4,
the mean of the probability density function of $\begin{pmatrix} X_1 \\ Z_1 \end{pmatrix}$ conditioned on $Z_0$ is
given by
\[
E\left( \begin{pmatrix} X_1 \\ Z_1 \end{pmatrix} \bigg| Z_0 \right) = \begin{pmatrix} \bar{X}_1 \\ \bar{Z}_1 \end{pmatrix} + R_{x_1x_0} R_{z_0z_0}^{-1} (Z_0 - \bar{Z}_0).
\]  \hspace{1cm} (3.28)
However, from our notation $R_{\{x_1\}_{z_1|z_0}}$ is equal to $E\left[\left\{\left(\begin{array}{c} x_1 \\ Z_1 \end{array}\right) - \left(\begin{array}{c} \bar{x}_1 \\ \bar{Z}_1 \end{array}\right)\right\}(Z_0 - \bar{Z}_0)^T\right]$.

Hence,

$$R_{\{x_1\}_{z_1|z_0}} = E\left[\left(\begin{array}{c} x_1 - \bar{x}_1 \\ Z_1 - \bar{Z}_1 \end{array}\right)(Z_0 - \bar{Z}_0)^T\right] = \left(\begin{array}{c} R_{x_1|z_0} \\ R_{z_1|z_0} \end{array}\right).$$ (3.29)

Substitution of (3.29) into (3.28), we obtain

$$E\left[\left(\begin{array}{c} x_1 \\ Z_1 \end{array}\right)\right|_{Z_0} = \left(\begin{array}{c} \bar{x}_1 \\ \bar{Z}_1 \end{array}\right) + \left(\begin{array}{c} R_{x_1|z_0} \\ R_{z_1|z_0} \end{array}\right) \left(\begin{array}{c} R_{z_0|z_0} \end{array}\right)^{-1} \left(\begin{array}{c} Z_0 - \bar{Z}_0 \end{array}\right)$$

$$= \left(\begin{array}{c} \bar{x}_1 \\ \bar{Z}_1 \end{array}\right) + \left(\begin{array}{c} R_{x_1|z_0} R_{z_0|z_0}^{-1} \left(\begin{array}{c} Z_0 - \bar{Z}_0 \end{array}\right) \\ R_{z_1|z_0} R_{z_0|z_0}^{-1} \left(\begin{array}{c} Z_0 - \bar{Z}_0 \end{array}\right) \end{array}\right).$$

Comparing the above expression to $X_{1|0}$ and $Z_{1|0}$, we obtain

$$E\left[\left(\begin{array}{c} x_1 \\ Z_1 \end{array}\right)\right|_{Z_0} = \left(\begin{array}{c} X_{1|0} \\ Z_{1|0} \end{array}\right).$$ (3.30)

By assumption for the KF, $\left(\begin{array}{c} x_1 \\ Z_1 \end{array}\right)$, $Z_0$ are gaussian. Hence, the probability density function of $\left(\begin{array}{c} x_1 \\ Z_1 \end{array}\right)$ conditioned on $Z_0$ is also gaussian. The mean of this conditional probability density function is given by (3.30).

Now we calculate the covariance matrix of the probability density of $\left(\begin{array}{c} x_1 \\ Z_1 \end{array}\right)$ conditioned on $Z_0$. Applying Theorem 4 again, and denoting the covariance matrix of the probability density function of $\left(\begin{array}{c} x_1 \\ Z_1 \end{array}\right)$ conditioned on $Z_0$ by $\sigma$,

$$\sigma = R_{\{x_1\}_{z_1|z_0}} - R_{\{x_1\}_{z_0|z_1}} R_{z_0|z_0}^{-1} R_{z_0|z_1}. $$
Similar to the argument for deriving $R_{\left(x_1\right)z_0} = \begin{pmatrix} R_{x_1z_0} \\ R_{z_0z_0} \end{pmatrix}$,

$$R_{\left(x_1\right)\left(x_1\right)\left(z_1\right)} = \begin{pmatrix} R_{x_1x_1} & R_{x_1z_1} \\ R_{z_1x_1} & R_{z_1z_1} \end{pmatrix}.$$  

Substitution of the above expression into $\sigma$ and after simplification, $\sigma$ is shown to be

$$\sigma = \begin{pmatrix} R_{x_1x_1} - R_{x_1z_0} R_{z_0z_0}^{-1} R_{z_0x_1} & R_{x_1z_1} - R_{x_1z_0} R_{z_0z_0}^{-1} R_{z_0z_1} \\ R_{z_1x_1} - R_{z_1z_0} R_{z_0z_0}^{-1} R_{z_0x_1} & R_{z_1z_1} - R_{z_1z_0} R_{z_0z_0}^{-1} R_{z_0z_1} \end{pmatrix}.$$  

From Lemma 3, the first diagonal element of $\sigma$, is

$$R_{x_1x_1} - R_{x_1z_0} R_{z_0z_0}^{-1} R_{z_0x_1} = P_{1/0}.$$  

From Theorem 4 the covariance matrix of the probability density function of $Z_1$ conditioned on $Z_0$, $\theta_{1/0}$, is given by

$$\theta_{1/0} = R_{z_1z_1} - R_{z_1z_0} R_{z_0z_0}^{-1} R_{z_0z_1}.$$  

Consequently $\sigma$ can be rewritten as

$$\sigma = \begin{pmatrix} P_{1/0} & R_{x_1z_1} - R_{x_1z_0} R_{z_0z_0}^{-1} R_{z_0z_1} \\ R_{z_1x_1} - R_{z_1z_0} R_{z_0z_0}^{-1} R_{z_0x_1} & \theta_{1/0} \end{pmatrix}.$$  

Application of the measurement model equation and from the definition of the covariance matrix, $\theta_{1/0}$ can be shown to equal to

$$\theta_{1/0} = H_1 P_{1/0} H_1^T + S_1.$$  

Calculation similar to the proof in Lemma 3, the off-diagonal element

$$R_{x_1z_1} - R_{x_1z_0} R_{z_0z_0}^{-1} R_{z_0z_1}$$  

is equal to $P_{1/0} H_1^T$.

Finally the other off-diagonal element

$$R_{x_1z_1} - R_{x_1z_0} R_{z_0z_0}^{-1} R_{z_0z_1}$$  

is just the transpose of $R_{x_1z_1} - R_{x_1z_0} R_{z_0z_0}^{-1} R_{z_0z_1}$ and hence equal to

$$R_{z_1x_1} - R_{z_1z_0} R_{z_0z_0}^{-1} R_{z_0x_1} = H_1 P_{1/0}.$$  

Combining all these results $\sigma$ can be rewritten to be
\[
\sigma = \begin{bmatrix}
P_{i/0} & P_{i/0} \mathbf{H}_1^T \\
\mathbf{H}_1 & P_{i/0} \mathbf{H}_1^T + \mathbf{S}_1
\end{bmatrix}.
\]

And this is the desired result.

### 3.2.3.5. Lemma 5

Let \(X_1, Z_1, Z_0\) be jointly gaussian, then the probability density function of \(X_1\) conditioned on \((Z_1, Z_0)\) is gaussian. The mean and the covariance matrix of this conditional probability density function are, respectively,

\[
X_{1|1} = X_{1|0} + P_{i/0} \mathbf{H}_1^T (\mathbf{H}_1 P_{i/0} \mathbf{H}_1^T + \mathbf{S}_1)^{-1} (Z_1 - \mathbf{H}_1 X_{1|0}),
\]

\[
P_{1|1} = P_{i/0} - P_{i/0} \mathbf{H}_1^T (\mathbf{H}_1 P_{i/0} \mathbf{H}_1^T + \mathbf{S}_1)^{-1} \mathbf{H}_1 P_{i/0}.
\]

**Proof**

The probability density function of \(X_1\) conditioned on \((Z_1, Z_0)\) according to Bayes' Law is given by

\[
f(X_1 | Z_1, Z_0) = \frac{f(X_1, Z_1, Z_0)}{f(Z_1 | Z_0)}.
\]

The probability density function \(f(X_1, Z_1, Z_0)\) is gaussian and the mean and covariance matrix are given by Lemma 4. Also from probability theory, we can associate to this probability density function \(f(X_1, Z_1 | Z_0)\) a random vector. Notationally, we write the random vector as \((X_i | Z_0, Z_i | Z_0)\). Then in terms of \((X_1 | Z_0, Z_1 | Z_0)\) the probability density function \(f(X_1, Z_1 | Z_0)\) can be rewritten as \(f(X_1 | Z_0, Z_1 | Z_0)\). Thus,

\[
f(X_1, Z_1 | Z_0) = \frac{f(X_1 | Z_0, Z_1 | Z_0)}{f(Z_1 | Z_0)}.
\]

Hence, we can apply Theorem 4 to get the mean and covariance of \(f(X_1 | Z_1, Z_0)\) from the mean and covariance matrix of \(f(X_1 \setminus Z_0, Z_1 \setminus Z_0)\).
Firstly, from Theorem 4 the mean of the conditional probability density function of $f(X_i|Z_i, Z_0)$ is given as

$$X_{i/0} = X_{i/0} + P_{i/0} H_i^T(H_i P_{i/0} H_i^T + S_i)^{-1}(Z_i \backslash Z_0 - Z_{i/0}),$$

where $Z_i \backslash Z_0$ is the measured value of the random vector $Z_i$, so we will write $Z_i \backslash Z_0$ as $\tilde{z}_i$. On the other hand, $Z_{i/0}$ is the mean of the probability density function of $Z_i$ conditioned on $Z_0$. By Theorem 4,

$$Z_{i/0} = \overline{Z}_i + R_{z_i|z_0}^{-1} R_{z_0|z_0}^{-1}(Z_0 - \overline{Z}_0).$$

If we apply dynamic and measurement model to evaluate $R_{z_i|z_0}$, $Z_{i/0}$ can be rewritten as

$$Z_{i/0} = H_i \left[ \overline{X}_i + (F_0 R_{x_0|x_0} H_0^T) R_{z_0|z_0}^{-1}(Z_0 - \overline{Z}_0) \right].$$

From Lemmas 2 and 3, the expression inside the curly braces is just $X_{i/0}$. As a result,

$$Z_{i/0} = H_i X_{i/0}.$$

Consequently, this result implies

$$X_{i/0} = X_{i/0} + P_{i/0} H_i^T(H_i P_{i/0} H_i^T + S_i)^{-1}(\tilde{z}_i - H_i X_{i/0}).$$

Secondly, the conditional probability density function $f(X_i \backslash Z_0, Z_i \backslash Z_0)$ has covariance matrix given by Lemma 4 as

$$
\begin{bmatrix}
P_{i/0} & P_{i/0} H_i^T \\
H_i P_{i/0} & H_i P_{i/0} H_i^T + S_i
\end{bmatrix}
= 
\begin{bmatrix}
R_{x_i|z_0, x_i|z_0} & R_{x_i|z_0, z_i|z_0} \\
R_{z_i|z_0, x_i|z_0} & R_{z_i|z_0, z_i|z_0}
\end{bmatrix}.

(3.33)
$$

On the other hand the covariance matrix $P_{i/0}$ of the conditional probability density function $f(X_i|Z_0, Z_i) = f((X_i|Z_0) | (Z_i|Z_0))$ is given by Theorem 4 as

$$P_{i/0} = R_{x_i|z_0, x_i|z_0} - R_{x_i|z_0, z_i|z_0} R_{z_i|z_0, z_i|z_0}^{-1} R_{z_i|z_0, x_i|z_0}.$$

Hence, substitution of (3.33) into (3.34) gives the desire result.

3.2.3.6. Lemma 6

The probability density function of $X_2$ conditioned on $(Z_1, Z_0)$ has mean
\[ X_{2/l} = F_{l} X_{l/1}, \quad (3.35) \]

and covariance matrix
\[ P_{2/l} = F_{l} P_{l/1} F_{l}^{T} + Q_{l}. \quad (3.36) \]

**Proof**

Firstly, let us consider \((X_{2}, Z_{l})\) conditioned on \(Z_{0}\). From Theorem 4, the probability density function of \((X_{2}, Z_{l})\) conditioned on \(Z_{0}\) has mean \((X_{2|0}, Z_{1|0})\)

and covariance matrix \(\sigma\), where

\[ \sigma = R_{x_{2}x_{2}}(x_{2}) - R_{x_{2}z_{0}}(z_{0}) R_{z_{0}z_{0}}^{-1} R_{z_{0}x_{2}}(x_{2}) \]

From arguments similar to those in Lemma 4, \(\sigma\) can be rewritten as

\[ \sigma = \begin{pmatrix} R_{x_{2}x_{2}} - R_{x_{2}z_{0}} R_{z_{0}x_{2}}^{-1} R_{z_{0}z_{0}} & R_{x_{2}z_{1}} - R_{x_{2}z_{0}} R_{z_{0}z_{1}}^{-1} R_{z_{0}z_{0}} \\ R_{z_{1}x_{2}} - R_{z_{1}z_{0}} R_{z_{0}x_{2}}^{-1} R_{z_{0}x_{2}} & R_{z_{1}z_{1}} - R_{z_{1}z_{0}} R_{z_{0}z_{1}}^{-1} R_{z_{0}z_{0}} \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} \quad (3.37) \]

From the argument in the proof of Lemma 5, we obtain immediately a simplified expression of \(\theta_{22}\),

\[ \theta_{22} = H_{1} P_{l/0} H_{1}^{T} + S_{1}. \]

Applying Theorem 4 again with (3.37), we obtain the mean, \(X_{2/l}\), and covariance matrix, \(P_{2/l}\), of \(X_{2}\) on \((Z_{1}, Z_{0})\). The mean and the covariance matrix can be written symbolically as

\[ X_{2/l} = X_{2/0} + \theta_{12} \sigma_{22}^{-1} (Z_{1} - H_{1} E[Z_{1} \mid Z_{0}]), \]

\[ P_{2/l} = \theta_{11} - \theta_{12} \sigma_{22}^{-1} \theta_{21}, \]

where \(E[Z_{1} \mid Z_{0}] = H_{1} X_{1/0}\). So the proof is reduced to calculating \(\theta_{11}, \theta_{21}, \theta_{12}(= \theta_{21}^{T})\), which in turn is further reduced to calculating \(R_{x_{2}z_{1}}, R_{x_{2}z_{0}}, R_{z_{0}z_{1}}, R_{z_{0}x_{2}}\).

Application of the dynamical and measurement model equations together with Lemmas 1 and 3, \(\theta_{11}, \theta_{21}, \theta_{12}(= \theta_{21}^{T})\) can be shown to be

\[ \theta_{21} = F_{l} P_{l/0} H_{1}^{T}, \]
\[ \theta_{11} = F_1 P_{i0} F_1^T + Q_1 \text{ where } Q_i = E[\xi_i \xi_i^T]. \]

Substitution of \( \theta_{11}, \theta_{21}, \theta_{12} (= \theta_{21}^T) \) into \( P_{2i1} = \theta_{11} - \theta_{12} \theta_{22}^{-1} \theta_{21} \) with Lemma 5 results in
\[ P_{2i1} = \theta_{11} - \theta_{12} \theta_{22}^{-1} \theta_{21} \]

Finally, substitution of \( \theta_{11}, \theta_{21}, \theta_{12} (= \theta_{21}^T) \) into
\[ X_{2i1} = X_{2i0} + \theta_{12} \theta_{22}^{-1} (Z_i - H_i E[Z_i \setminus Z_0]) \]

and applying Lemma 5 results in
\[ X_{2i1} = F_1 X_{i1i}. \]

Since \( X_{2i0} \) can be shown to be (from Lemma 2)
\[ X_{2i0} = F_1 X_{i0i}. \]

This completes the proof.

We can see that (3.35) in Lemma 6 are the same as (3.22) in Lemma 2 except each index is increased by 1. However, this difference is the same as observed for Lemmas 1 and 5.

### 3.2.4. Kalman Filter Algorithm

Generalizing the indices on the formulas in Lemmas 1-6, the following KF is proposed for the dynamical and measurement model equations given by,

**Dynamical model**
\[ X_{k+1} = F_k \cdot X_k + \xi_k, \text{ where } E[\xi_k \xi_k^T] = \delta_{kk} \cdot Q_k \quad (3.38) \]

**Measurement model**
\[ Z_k = H_k \cdot X_k + \eta_k, \text{ where } E[\eta_k \eta_k^T] = \delta_{kk} \cdot S_k \quad (3.39) \]

The KF algorithm consists of several different pieces. To predict the state and covariance matrix of the system at time \( t_k \) using measurements up to time \( t_{k-1} \), we have the results

1. **state prediction**
\[ X_{k|k-1} = F_k \cdot X_{k-1|k-1}, \quad (3.40) \]
(2) propagation of covariance matrix
\[ P_{k|k-1} = F_k P_{k-1|k-1} F^T + Q_k. \quad (3.41) \]

To update the state and covariance matrix of the system at time \( t_k \) using measurements up to time \( t_k \), we compute the gain matrix

(3) Gain Matrix
\[ K_k = P_{k|k-1} H_k \left[ H_k P_{k|k-1} H_k^T + S_k \right]^{-1}, \quad (3.42) \]

and insert it into the state update and covariance matrix update given by

(4) state update
\[ X_{k|k} = X_{k|k-1} + K_k \left( Z_k - H_k X_{k|k-1} \right), \quad (3.43) \]

(5) estimated error covariance matrix update
\[ P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1}. \quad (3.44) \]

The KF algorithm is applied in Chapters 4 and 5 to simulated enzyme deactivation and separation of enantiomers.

3.2.5. Alternative Method of Derivation of Kalman Filter[14]

The Bayesian approach used to derive the KF algorithm in Section 3.2.4 has the advantage that it can be easily generalized to nonlinear filtering problems. However, the KF has the property that the covariance matrix of the error is minimized. This property is not apparent using the Bayesian approach. Therefore, we give an alternative proof in which this property is more obvious.

The proof of the KF algorithm given by (3.38)-(3.44) requires the concept of innovation. We start with the definition of an innovation process.

3.2.5.1. Innovation Process

The innovation \( v_n \) is defined by
\[ v_n = Z_n - H_n X_{n/n-1}, \]

where \( Z_n, H_n, X_{n/n} \) are defined in Section 3.2.4.
3.2.5.2. Lemma 7

The process $\nu_n$ is a gaussian process such that

\[ E[\nu_n] = 0, \]
\[ E[\nu_n \nu_n^T] = \delta_{nn} (H_n P_{n/n} H_n^T + S_n) \]

where $S_n, H_n, P_{n/n}$ are defined in Section 3.2.4.

Proof

From the measurement model equation

\[ \nu_n = H_n (X_n - X_{n/n-1}) + \eta_n. \]

Therefore,

\[ E[\nu_n | Z^{n-1}] = E[H_n (X_n - X_{n/n-1}) | Z^{n-1}] + E[\eta_n | Z^{n-1}] = 0. \] (3.45)

Also from Theorem 3, $\nu_n$ is gaussian since it is a linear combination of gaussian variables $X_n$. From the conditional expectation in probability theory, we have

\[ E[A] = E[E[A|B]]. \] (3.46)

Combining (3.45) and (3.46) results in

\[ E[\nu_n] = 0. \]

Moreover,

\[ E[\nu_n \nu_n^T] = E\left[ \left\{ H_n (X_n - X_{n/n-1}) + \eta_n \right\} \left\{ H_n (X_n - X_{n/n-1}) + \eta_n \right\}^T \right]. \]

Since $\eta_n$ is independent of $X_n$, we deduce immediately

\[ E[\nu_n \nu_n^T] = H_n P_{n/n-1} H_n^T + S_n. \]

Furthermore, for $k=1,2,\ldots,n-1$,

\[ E[\nu_n Z_k^T | Z^{n-1}] = E[\nu_n | Z^{n-1}] Z_k^T = 0. \]

From (3.46), we have,

\[ 0 = E[\nu_n Z_k^T] = E[\nu_n \nu_k^T] + E[\nu_n X_{k/k-1}^T] H_k^T. \]
But,
\[ E[\mathbf{v}_n X_{k/k-1}] = E[\mathbf{v}_n] X_{k/k-1} = 0. \]

Therefore, we conclude
\[ E[\mathbf{v}_n \mathbf{u}_k^\top] = 0, \text{ for } k=1,\ldots,n-1. \]

which completes the proof.

3.2.5.3. Theorem 5

For the dynamical and measurement model equations (3.38) and (3.39), the formulas (3.40)-(3.44) hold.

Proof

The results (3.40) and (3.41) follow from the application of the definition of \( X_{k/k-1} \) and the dynamical equation, i.e., for (3.40) we have

\[ X_{k+1/k} = E[X_{k+1} \mid Z^k] = E[F_k X_k + \xi_k \mid Z^k] \]
\[ = F_k X_{k/k}. \]

A similar argument yields (3.41).

Moreover, using the innovation, we can assert that
\[ X_{k/k} = E[x_k \mid z_1, z_2, \ldots, z_{k-1}, \mathbf{v}_k]. \]

Using the fact that \( X_{k/k} \) is also the best linear estimate of \( X_k \), given \( Z_1, Z_2, \ldots, Z_{k-1}, \mathbf{v}_k \), we have the formula
\[ x_{k/k} = x_{k/k-1} + K_k \mathbf{v}_k, \]

which corresponds to (3.43), where \( K_k \) is a gain matrix to be determined. It remains to fix \( K_k \), knowing that it minimizes the covariance of the error.

The error of the estimated state can be written as
\[ \mathbf{e}_k^+ = X - X_{k/k} \]
\[ = (X - x_{k/k-1}) - K_k \mathbf{v}_k. \]

Hence, the covariance matrix of \( \mathbf{e}_k^+ \), \( E[\mathbf{e}_k^+ \mathbf{e}_k^{+\top}] \), is given by
\[
P_{k/k} = P_{k/k-1} + K_k (H_k P_{k/k-1} H_k^T + S_k) K_k^T - K_k E[(x_k - x_{k/k-1})^T] - E[(x_k - x_{k/k-1})u_k^T] K_k^T.
\]

However,
\[
E[u_k(X_k - X_{k/k-1})^T] = E[Z_k(X_k - X_{k/k-1})^T]
= E[H_k X_k (X_k - X_{k/k-1})^T]
= H_k P_{k/k-1}.
\]

Therefore, we obtain by completing the square
\[
P_{k/k} = P_{k/k-1} + K_k (H_k P_{k/k-1} H_k^T + S_k) K_k^T - K_k H_k P_{k/k-1} - P_{k/k-1} H_k^T K_k^T
= P_{k/k-1} + \left[K_k - P_k H_k^T (H_k P_k H_k^T + S_k)^{-1} \right] (H_k P_k H_k^T + S_k)
\times \left[K_k^T - (H_k P_k H_k^T + S_k)^{-1} H_k P_k \right] - P_{k/k-1} H_k^T (H_k P_k H_k^T + S_k)^{-1} H_k P_k.
\]

It follows immediately that the best value of \(K_k\) is
\[
K_k = P_{k/k-1} H_k^T \left[H_k P_{k/k-1} H_k^T + S_k\right]^{-1},
\]
which is (3.42) and (4.47) reduces to (4.44).

Thus the proof is complete.

3.3. Extended Kalman Filter Theory

The extended Kalman filter (EKF) is a generalized version of the Kalman filter when the dynamical or measurement model equations are nonlinear. The dynamical and measurement model equations are
\[
X_{k+1} = f_k(X_k) + \xi_k, \quad (3.48)
\]
\[
Z_k = h_k(X_k) + \eta_k, \quad (3.49)
\]
where \(X_k\) is a random \(n\)-vector, \(Z_k\) is a random \(m\)-vector, \(\xi_k\) and \(\eta_k\) are sequences of white noise processes of dimension \(n\) and \(m\), respectively, \(f_k(X_k)\) is a nonlinear \(n\)-vector function, and \(h_k(X_k)\) is a nonlinear \(m\)-vector function.
It is possible to derive the EKF in a way similar to the derivation of Lemmas 1 to 6 by Taylor expansion of \( f_k(X_k) \), and \( h_k(X_k) \) about \( x_{k/k-1} \) and neglecting the moments of order higher than two. However, the simplest way to derive the EKF is to linearize the dynamical model equations about \( X_{k/k} \) and measurement model equations about \( X_{k/k-1} \), and apply Kalman Filter algorithms to the linearized equations.

The nonlinear functions \( f_k(X_k) \) and \( h_k(X_k) \) can be expanded in Taylor series about the conditional means \( X_{k/k} \) and \( X_{k/k-1} \), respectively, as

\[
\begin{align*}
    f(X_k) &= f(X_{k/k}) + F_k (X_k - X_{k/k}) + \ldots, \\
    h(X_k) &= h(X_{k/k-1}) + H_k (X_k - X_{k/k-1}) + \ldots,
\end{align*}
\]

where \( F_k = \frac{\partial f_k}{\partial X_k}(X_{k/k}) \) and \( H_k = \frac{\partial h_k}{\partial X_k}(X_{k/k-1}) \) are \( n \times n \) and \( m \times n \) matrices, respectively.

Neglecting higher-order terms and assuming we know \( X_{k/k} \) and \( X_{k/k-1} \) enables us to approximate the dynamical and measurement models as

\[
\begin{align*}
    X_{k+1} &= F_k X_k + u_k + \xi_k, \\
    Z_k &= H_k X_k + v_k + \eta_k,
\end{align*}
\]

where

\[
\begin{align*}
    u_k &= f_k(X_{k/k}) - f_k(X_{k/k}), \\
    v_k &= h_k(X_{k/k-1}) - h_k(X_{k/k-1})
\end{align*}
\]

Now if we apply the KF to the linearized dynamical and measurement model equations (3.52) and (3.53), the EKF will be obtained. The extended Kalman Filter algorithms applied to (3.48) and (3.49) is identical to (3.40)-(3.44) except that \( F_k = \frac{\partial f_k(X_{k/k})}{\partial X_k} \) and \( H_k = \frac{\partial h_k(X_{k/k-1})}{\partial X_k} \).

### 3.4. More About Error Covariance Matrices

To apply the EKF or the KF algorithms, derived in previous sections, we have to specify the model error covariance matrix, \( Q_k \), the initial state error
covariance matrix, \( P_0 \), and the measurement error covariance matrix, \( S_k \). There are no systematic methods to determine these covariance matrices. In most of the literature, the values of these covariance matrices are determined by trial-and-error based on the experience of the user. However, there is a rule to help us to decide the range of these covariance matrices.

Let us start with the model error covariance matrix, \( Q_k \). Since the diagonal elements of the model error covariance matrix are the squares of the standard deviations of noises, then, the square roots of the diagonal elements of the model error covariance matrix should be large enough to account for the error between the exact model equations and the system model equations. Similarly, the square roots of the diagonal elements of the initial state error covariance matrix should be large enough to cover the difference between the exact initial values of the state and the initial values of the state specified by the EKF or the KF. Finally, the values of the measurement error covariance matrix are dependent on the instruments that we use to perform the measurements and good values of the measurement error covariance matrix can be found in the manuals for the instruments.

3.5. Exact Filter

Since a nonlinear transformation can transform a gaussian variable into a nongaussian variable, the fundamental assumption of the KF theory can be violated. Some other techniques have been proposed to improve the performance of the EKF, for example, the gaussian sum technique [15] [16]. However, if a dynamical system is simple enough, then the conditional probability density function can be computed. In this section, a recursion procedure will be developed to compute the conditional probability.

The algorithm is composed of two parts:

1. Evolution of the conditional probability density function.
2. Update of the conditional probability density function from the measured data.

Consider the dynamical and measurement model equations of a system is given by (3.17) and (3.18).

**Lemma 8 (Evolution of conditional probability density function)**

Let \( p_k(x_k|Z^k) \) denote the probability density function of \( x_k \) conditioned on measurements \( Z^k \). If \( f_k^{-1} \) exists then the probability density function of \( x_{k+1} \) conditioned on measurements \( Z^k \) is given by

\[
p_{k+1}(x_{k+1}|Z^k) = p_k(f_k^{-1}(x_{k+1})|Z^k) \left| \frac{\partial f_k^{-1}(x_{k+1})}{\partial x_{k+1}} \right|.
\]

**Proof**

This result is a consequence of Theorem 1.

**Lemma 9 (Update of conditional probability density function from measured data)**

Let \( p_k(x_k|Z^{k-1}) \) denote the probability density function of \( x_k \) conditioned on measurements \( Z^{k-1} \), then

\[
p_k(x_k|Z^k) = \frac{p_{x_k}(z_k|x_k)p_k(x_k|Z^{k-1})}{\int p_{x_k}(z_k|\varphi)p_k(\varphi|Z^{k-1})d\varphi}.
\]

**Proof**

From the definition of \( Z^k = \{z_1, z_2, \ldots, z_k\} \)

\( p_k(x_k|Z^k) = p_k(x_k|z_k, Z^{k-1}). \)

From Bayes' rule

\[
p_k(x_k|Z^k) = \frac{p_{x_k}(z_k|x_k, Z^{k-1})p_k(x_k|Z^{k-1})}{p_{x_k}(z_k|Z^{k-1})}.
\]

Since the measurement noise is white noise, i.e., \( E[\eta_i \eta_j] = 0 \) for \( i \neq j \),
\[ p_{z_k}(z_k|x_k, Z^{k-1}) = p_{z_k}(z_k|x_k). \]

Furthermore,
\[ p_{z_k}(z_k|Z^{k-1}) = \int p_{z_k}(z_k|\varphi)p_k(\varphi|Z^{k-1})d\varphi, \]

Combining the above results, we get
\[
\frac{p_{z_k}(z_k|x_k) p_k(x_k|Z^{k-1})}{\int p_{z_k}(z_k|\varphi)p_k(\varphi|Z^{k-1})d\varphi} \quad \text{(QED)}
\]

In the following, we show how to calculate \( p_{z_k}(z_k|x_k) \) from Theorem 1 and the probability density function of \( \eta_k \). Since \( z_k = h_k(x_k) + \eta_k \), then by Theorem 1
\[ p_{z_k}(z_k|x_k) = p_{\eta_k}(z_k - h_k(x_k)). \]

Since \( \eta_k \) is gaussian (i.e., \( \eta_k \sim \mathcal{N}(0, S_k) \))
\[ p_{z_k}(z_k|x_k) = \left(\frac{2\pi}{\sqrt{\det(S_k)}}\right)^{-1/2} \exp\left(-\frac{1}{2} \left[ z_k - h_k(x_k) \right]^T S_k^{-1} \left[ z_k - h_k(x_k) \right] \right). \]

In particular in a one-dimensional problem, \( \eta_k \sim \mathcal{N}(0, \sigma^2) \), where \( \sigma \) is the standard deviation,
\[ p_{z_k}(z_k|x_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(z_k - h_k(x_k))^2}{\sigma^2} \right). \]

The exact filter algorithm will be applied to the enantiomer separation problem in Chapter 5.

We conclude this section with the remark that it is not feasible to compute the exact mean and variance except for one- and two-dimensional problems.
4. Enzyme Deactivation

In a realistic experiment or bioreactor, the activity of the enzyme may decrease as time progresses. However, the concentration of substrate decreases at a slower rate than that predicted by the enzyme-substrate reaction rate equation. It is often difficult or impossible to determine exactly the model equations for the enzyme deactivation process because there are many factors involved. An example of such a process is the enzymatic hydrolysis of cellulose, cellobiose, and glucose [17], [18]

\[ \text{cellulose} \rightarrow \text{cellobiose} \rightarrow \text{glucose} \]

Acquisition of data for the key variables and their control for the above reactions are often impeded by the lack of reliable on-line measurements of the substrate and product concentrations and the difficulties in properly modelling biological activity in processes that are time-varying. One approach to solving this problem involves the development of state estimation techniques that use available measurements to reconstruct the evolution of the state variables (i.e., the variables of the rate equation) or to estimate the parameters of the rate equation.

Enzymatic deactivation in the enzymatic hydrolysis of pretreated cellulose was studied by G. Caminal et al. [15] using the extended Kalman filter (EKF) on a model by G. Caminal et al. [16]. Traditionally, the methods used most often to study thermal deactivation of enzymes require an evaluation of activity losses by incubating the enzyme without a substrate for different periods of time and then testing its activity [19]. This methodology requires that a specific set of experiments be carried out; it does not account for the possible effects of the substrate on the enzyme's stability. The power of
the EKF is that even if we do not know the exact equation modelling the enzyme deactivation process, we can still apply the rate equation for the substrate obtained from a quasi-steady-state approximation. All other parameters are treated as constants. If we can measure the substrate concentration, then the values of all the other parameters can be predicted by the EKF during the same time period that the measurements were taken.

Application of the EKF to identify and track the slowly varying parameters of a multiple enzyme system for pretreated cellulose has the difficulty of being able to choose the system model error covariance matrix appropriately. To study the application of the EKF to an enzyme deactivation problem, a model problem is set up which simulates a single enzyme deactivation in which the parameter that represents the activity of the enzyme is slowly varying.

The EKF algorithms is developed in Section 3.3. In the following, the models consist of coupled nonlinear differential equations. The EKF algorithm is applied to Euler discretizations of these models.

4.1. Reactions and Kinetic Equations

To simplify the mathematics, we employ a single enzyme reaction in which a substrate A is turned into product P.

\[ A + E \xrightleftharpoons[k_{-1}]{k_1} C \xrightarrow[k_2]{k_1} P + E \]

where A is the substrate, E is the enzyme, P is the product, C is the intermediate enzyme-substrate complex, \(k_{-1}\), \(k_1\), \(k_2\) are the rate constants. We use the same notation, i.e., A, E, P, C, for the concentrations of these quantities. The differential equations for the rate of change of concentration of
each element in the above chemical reaction can be derived by applying the mass-action law to obtain

\[
\begin{align*}
\frac{dA}{dt} &= -k_1 A E + k_{-1} C, \\
\frac{dE}{dt} &= -k_1 A E + k_{-1} C + k_2 C, \\
\frac{dC}{dt} &= -k_2 C + k_1 A E - k_{-1} C, \\
\frac{dP}{dt} &= k_2 C.
\end{align*}
\] (4.1)

If we apply the quasi-steady-state approximation to the exact equation (equivalent to setting \(dC/dt\) to zero), the exact equations can be simplified to

\[
\begin{align*}
\frac{dA}{dt} &= -\frac{r_m A}{K_m + A}, \\
\frac{dP}{dt} &= -\frac{dA}{dt},
\end{align*}
\] (4.2)

where \(r_m = k_2 E_0\), \(K_m = (k_2 + k_{-1})/k_1\), and \(E_0 = \text{total initial enzyme concentration}\).

We simulate the enzyme deactivation by allowing \(r_m\) to decrease slowly through an exponential law, i.e., \(r_m\) is modelled by

\[
\frac{dr_m}{dt} = -\varepsilon r_m,
\] (4.3)

where \(\varepsilon\) is an artificial small parameter that controls the rate of deactivation.

In conclusion, we simulate the enzyme deactivation process by the following system of differential equations

\[
\begin{align*}
\frac{dA}{dt} &= -\frac{r_m A}{K_m + A}, \\
\frac{dr_m}{dt} &= -\varepsilon r_m.
\end{align*}
\] (4.4)

We will refer to (4.4) as the \textit{exact model}. 
4.2. On-Line Estimation of Enzyme Activity Using the Extended Kalman Filter

To apply the extended Kalman filter to monitor the activity of the enzyme which is governed by (4.4), we have to specify a system model which is a good approximation of (4.4), and then generate measurements which represent the observations of our artificial enzymes system. The measured quantity in this simulation is the substrate concentration, A.

4.2.1. System model

In a realistic problem, it is often not possible to obtain an exact mathematical model that describes the physical phenomenon. Therefore, we have to employ a system model which is an approximation to the (real) system that generates the observations. Sometimes, such an approximation is intentional. For example, it may be desirable to employ a system model of lower dimension than the real system in order to gain computational speed and simplicity. In this study, the system model is given by

\[
\frac{dA}{dt} = - \frac{r_m A}{K_m + A},
\]

\[
\frac{dr_m}{dt} = 0, \quad (4.5)
\]

\[
\frac{dK_m}{dt} = 0.
\]

4.2.2. Noiseless Measurements

4.2.2.1. Generation of Measured Data

The measured values of substrate concentration, A, are obtained by solving (4.4) using a 4th-order Runge-Kutta method with initial conditions \(A(0)=3\), \(r_m(0)=10\), \(K_m(0)=100\), and parameter \(\epsilon=0.02\). The result is shown in Figure 1.
Fig. 1 Noiseless measurements of concentration of substrate A.

4.2.2.2. Numerical Results

To apply the EKF we also have to specify the error covariance matrix of the system model, \( Q \), the error covariance matrix of the measurement model, \( R \), and the initial state error covariance matrix, \( P_0 \). Using the guidelines discussed in Chapter 3, one choice of \( Q, R, P_0 \) obtained by trial-and-error is

\[
Q = \begin{bmatrix}
10^{-8} & 0 & 0 \\
0 & 8 \times 10^{-8} & 0 \\
0 & 0 & 10^{-8}
\end{bmatrix},
R = \begin{bmatrix}
10^{-8} \\
0 \\
0
\end{bmatrix},
P_0 = \begin{bmatrix}
10^{-8} & 0 & 0 \\
0 & 10^{-8} & 0 \\
0 & 0 & 10^{-8}
\end{bmatrix}.
\]

The results of running the EKF on the above choices of error covariance matrices are shown in Figures 2 and 3. Since the estimated values of \( A \) and \( r_m \) are so close to the exact values, it is difficult to observe them in a graph. Therefore, the errors in the estimated values of \( A \) and \( r_m \) are plotted from the exact values.
Fig. 2 Error in estimated values of A from exact measurement of A. A_err = estimated value of A - exact value of A.

Fig. 3 Error in estimated values of rm. rm_err = estimated value of rm - exact value of rm.

In conclusion, the behavior of rm is being tracked correctly using the EKF with suitable choices of Q, R, and P_0. This tracking occurs in spite of the fact that the rm predicted by the system model is not varying with time and that the real rm is exponentially decreasing.

4.2.3. Small Noise Measurements

4.2.3.1. Generation of Measured Data
As in the previous case of noiseless measurements, the state that can be measured is the concentration of substrate A. However, exact measurements cannot be obtained in a realistic bioreactor. Thus, the measurements are simulated by adding white noise of zero mean and standard deviation 0.005 to the measurements in Section 4.2.2.1. (see Figure 4).

Fig. 4 Small noise measurements of concentration substrate A.

4.2.3.2. Numerical Result

Using the guidelines discussed in Chapter 3, choices of model error covariance matrix $Q$, initial state error covariance matrix $P_0$, and measurement error covariance matrix $R$ are

$$Q = \begin{bmatrix} 10^{-7} & 0 & 0 \\ 0 & 4 \times 10^{-2} & 0 \\ 0 & 0 & 10^{-8} \end{bmatrix}, \quad R = [0.005^2], \quad P_0 = \begin{bmatrix} 10^{-8} & 0 & 0 \\ 0 & 10^{-8} & 0 \\ 0 & 0 & 10^{-8} \end{bmatrix}.$$ 

The results of running the EKF on the small noise measurements using the above choices of error covariance matrices are shown in Figures 5 and 6. Again the estimated $A$ is so close to the exact values that it is difficult to observe these differences. Hence, the errors in the estimated values of $A$ from the exact values are plotted.
Fig. 5  Error in estimated $A$ using small noise measurements. $A_{\text{err}}$ = estimated value of $A$ - exact value of $A$.

Fig. 6  Estimated $r_m$ and exact $r_m$ in small noise measurement.

As expected, the error in the estimated concentration of substrate $A$ is quite small. The average absolute percentage error of the estimated concentration of substrate $A$ is 0.083%. On the other hand, the average absolute percentage error of the estimated $r_m$ is 1.92%.

4.2.4. Large Noise Measurements
4.2.4.1. Generation of Measured Data

As in Section 4.2.3.1, the measurements are simulated by adding white noise of zero mean and standard deviation 0.05 to the measurements obtained by exact solution of the exact model (see Figure 7).

![Graph showing concentration over time](image)

Fig. 7 Large noise measurements of concentration substrate A.

4.2.4.2. Numerical Result

By trial-and-error, choices of model error covariance matrix $Q$, initial state error covariance matrix $P_0$, and measurement error covariance matrix $R$ are found to be

$$Q = \begin{bmatrix} 10^{-7} & 0 & 0 \\ 0 & 8 \times 10^{-2} & 0 \\ 0 & 0 & 10^{-9} \end{bmatrix}, \quad R = [0.05^2], \quad P_0 = \begin{bmatrix} 10^{-8} & 0 & 0 \\ 0 & 10^{-8} & 0 \\ 0 & 0 & 10^{-8} \end{bmatrix}.$$  

The result of running the EKF for the above choices of error covariance matrices are shown in Figures 8 and 9. For this case, the errors in the estimated concentration of substrate A are large enough to make the difference between the estimated and the exact concentration of substrate A observable.
Fig. 8 Estimated A and exact A in large noise measurement.

Fig. 9 Estimated rm and exact rm in large noise measurement.

The average absolute percentage error of the estimated concentration of substrate A is 0.56%. On the other hand, the average absolute percentage error of the estimated rm is 4.69%.

4.2.5. The Choice of Model Error Covariance Matrix
So far there is no systematic method to determine the model covariance matrix. However, experience shows that the components of the model error covariance matrix corresponding to the state variable which is governed by a model equation with larger error have to be relatively larger. In the example of enzyme deactivation, the uncertainty in the model equation arises in the state variable \( r_m \). In this section, we are going to examine the change in the estimated \( r_m \) when the model error covariance matrix with a different \((r_m,r_m)\) component, i.e., the second diagonal element is employed.

Before doing the numerical computation, we need to introduce two quantities which indicate the accuracy of the estimated \( r_m \).

**Definition:**

The absolute percentage error, Abs\%, is defined by

\[
\text{Abs}\% = 100 \times \frac{1}{N} \sum_{n=1}^{N} \left| \frac{n^{\text{estimated value}} - n^{\text{exact value}}}{n^{\text{exact value}}} \right|
\]

where \( N \) is the total number of measurements. The relative percentage error, Rel\%, is defined by

\[
\text{Rel}\% = 100 \times \frac{1}{N} \sum_{n=1}^{N} \left( \frac{n^{\text{estimated value}} - n^{\text{exact value}}}{n^{\text{exact value}}} \right)
\]

Roughly speaking, absolute percentage error measures the amplitude of the fluctuation in the estimated values of a state variable. For example, large absolute percentage error means that the amplitude of the fluctuation in the estimation is large. Relativity percentage error can be positive or negative, and roughly speaking, it determines the best fitted curve of the estimated values.

The absolute percentage error and relative percentage error of the estimated \( r_m \) are calculated with the following sequence of model error covariance matrices, \( Q_n \), \( n=1,2,...,50 \).
\[
Q_n = \begin{bmatrix}
4 \times 10^3 & 0 & 0 \\
0 & \text{covariance} & 0 \\
0 & 0 & 10^{-9}
\end{bmatrix}
\]

where \text{covariance} = 20 \times n \times 10^{-2}.

With this choice of the model error covariance matrix, we again will examine the cases of noiseless measurement, small noise measurement, and large noise measurement.

4.2.5.1. Noiseless Measurement

The sequence of model error covariance matrices for the EKF is specified above and the measurement error covariance matrix, \( R \), and initial state error covariance, \( P_0 \), are given by

\[
R = [10^{-7}], \quad P_0 = \begin{bmatrix}
10^{-8} & 0 & 0 \\
0 & 10^{-8} & 0 \\
0 & 0 & 10^{-8}
\end{bmatrix}
\]

Figures 10 and 11 show the dependence of the absolute percentage error and relative percentage error on covariance defined in Section 4.2.5.
Fig. 10 Absolute percentage error of estimated $r_m$ as function of covariance.

Fig. 11 Relative percentage error of estimated $r_m$ as function of covariance.

4.2.5.2. Small noise measurement

For the small noise measurement case, the measurement error covariance matrix, $R$, and initial state error covariance, $P_0$, are given by

$$R = [0.005^2], P_0 = \begin{bmatrix} 10^{-8} & 0 & 0 \\ 0 & 10^{-8} & 0 \\ 0 & 0 & 10^{-8} \end{bmatrix}.$$
Figures 12 and 13 show the dependence of relative percentage error and absolute percentage error on covariance.

![Graph showing absolute percentage error as a function of covariance.]

**Fig. 12** Absolute percentage error of estimated $r_m$ as function of covariance.

![Graph showing relative percentage error as a function of covariance.]

**Fig. 13** Relative percentage error of estimated $r_m$ as function of covariance.

### 4.2.5.3. Large Noise Measurement

For the large noise measurement case, the measure error covariance matrix, $R$, and initial state error covariance, $P_0$, are chosen to be
\[ R = [0.05^2], \quad P_0 = \begin{bmatrix} 10^{-8} & 0 & 0 \\ 0 & 10^{-8} & 0 \\ 0 & 0 & 10^{-8} \end{bmatrix}. \]

Figures 14 and 15 show the dependence of relative percentage error and absolute percentage error on covariance.

**Fig. 14** Absolute percentage error of estimated \( r_m \) as function of covariance.

**Fig. 15** Relative percentage error of estimated \( r_m \) as function of covariance.
4.2.5.4. Conclusion

From the above figures, we discover that as covariance increases the absolute percentage error always increases. This is indicative of the fact that a larger uncertainty in the system model leads to larger uncertainty in our estimation of $r_m$. However, the relative percentage error decreases from positive values to negative values as covariance increases. This result shows that the value of the covariance at which the relative percentage error is zero would be a good choice. For example, for the large noise measurement case the relative percentage error of the estimated $r_m$ is zero when the covariance is about 2.25 as shown in Figure 15, and Figure 14 indicates that the absolute percentage error at this value is 17%. To reduce the absolute percentage error of the estimated $r_m$, we can adopt a more sophisticated filter algorithm such as the J-estimator invented by Jazwinski in his 1974 paper (see [20]). However, a simpler way is to sacrifice the relative percentage error. From Figure 14, the covariance decreases to about 1 when the relative percentage error of the estimated $r_m$ increases to 1%, and the absolute percentage error of the estimated $r_m$ drops to about 11%.

Unfortunately, this method cannot be applied to a realistic enzyme system because the exact values of certain state variables are not available. However, this calculation shows that a good choice of the model error covariance matrix exists.

4.3. Iteration of the Extended Kalman Filter

In an off-line analysis, we can apply the EKF again to the estimated state obtained by the application of the EKF to a set of measurements. Unfortunately, if the model error covariance matrix for the second application of the EKF is chosen to be the same as the first EKF, then the relative percentage error and the absolute percentage error of $r_m$ will become larger.
than that obtained from the first EKF. On the other hand, if the model error covariance matrix of the second EKF is chosen to smaller than the model error covariance matrix of the first EKF, then the relatively percentage error of $r_m$ will become larger. As a result, it is not practical to iterate applications of the EKF to the output obtained from the EKF.

However, a regression analysis can be done on the set of estimated values of $r_m$ which is obtained from the first EKF. For example, if $r_m$ is given in terms of a quadratic function of time, e.g., $r_m = \alpha + \beta t + \gamma t^2$, where $\alpha$, $\beta$, $\gamma$ are constants determined by the estimated values of $r_m$, then the modified model equation for $r_m$ is

$$\frac{dr_m}{dt} = \beta + \gamma t.$$

Hence, the modified system model is

$$\frac{dA}{dt} = -\frac{r_m A}{K_m + A},$$

$$\frac{dr_m}{dt} = \beta + \gamma t,$$

$$\frac{dK_m}{dt} = 0.$$

Now we can apply the EKF to this modified system model. The following schematic diagram summarizes the iteration procedure.
4.3.1. Numerical example

We apply the above iteration procedure to the case of the large noise measurements since the estimated \( r_m \) has both larger absolute percentage error and relative percentage error. Firstly, an estimation by the EKF based on the system model (4.5) is performed. Having obtained a set of estimated values of \( r_m \), a linear regression analysis is performed and the best straight line fit is calculated to be

\[
    r_m = 10.1388 - 0.1799 \, t
\]

Hence, the model equation of \( r_m \) is changed to

\[
    \frac{dr_m}{dt} = -0.1799.
\]

Consequently, the system model (4.5) is changed to:
\[
\frac{dA}{dt} = -\frac{r_m A}{K_m + A},
\]
\[
\frac{dr_m}{dt} = -0.1799,
\]
\[
\frac{dK_m}{dt} = 0.
\]

Since the system model has been changed, we have to modify the model error covariance matrix. The modified model error covariance matrix, \( Q \), is

\[
Q = \begin{bmatrix}
4 \times 10^{-3} & 0 & 0 \\
0 & 80 \times 10^{-4} & 0 \\
0 & 0 & 10^{-9}
\end{bmatrix}
\]

The absolute percentage error and the relative percentage error of the estimated values of \( r_m \) based on system model (4.6) are improved tremendously. Figure 16 is a plot of the estimated values of \( r_m \) for the original system model (4.5) and the modified system model (4.6). The following table summarizes the performance of the first and second iterations of the EKF in terms of the absolute percentage error and the relative percentage error.
Fig. 16 The estimated values of \( r_m \) for the original system model (4.5) and the modified system model (4.6).

Table 1 Absolute and relative percentage errors of estimated A and \( r_m \).

<table>
<thead>
<tr>
<th></th>
<th>1st Iteration</th>
<th>2nd Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absolute Error</td>
<td>Relative Error</td>
</tr>
<tr>
<td>A</td>
<td>0.6954%</td>
<td>0.3162%</td>
</tr>
<tr>
<td>( r_m )</td>
<td>5.8456%</td>
<td>1.9369%</td>
</tr>
</tbody>
</table>