UNIVERSITY OF TARTU
FACULTY OF SCIENCE AND TECHNOLOGY
Institute of Mathematics and Statistics
Mathematics

Johanna Adele Järvsoo

Kalman Filter and Extended Kalman Filter

Bachelor's thesis (9 ECTS)

Supervisors: Jüri Lember,
Andris Slavinskis

Tartu 2017

Kalman Filter and Extended Kalman Filter

Bachelor's thesis

Johanna Adele Järvsoo

Abstract. In the Bachelor's thesis we describe the Kalman filtering algorithm for linear-Gaussian state space models and give an example of its application. We describe the extended Kalman filter for differentiable Gaussian state space models and give examples of its application. We show that for linear-Gaussian state space models the extended Kalman filter gives the same results as the Kalman filter.

CERCS research specialisation. P160 Statistics, operations research, programming, actuarial mathematics.

Keywords. Random variables, estimation, normal distribution, noise, Kalman filters.

Kalmani filter ja laiendatud Kalmani filter

Bakalaureusetöö

Johanna Adele Järvsoo

Lühikokkuvõte. Bakalaureusetöös kirjeldatakse Kalmani filtrit lineaarsete normaaljaotusega müraga mudelite jaoks ja antakse näide Kalmani filtri rakendamisest. Lisaks kirjeldatakse laiendatud Kalmani filtrit diferentseeruvate normaaljaotuega müraga mudelite jaoks. Me toome näiteid laiendatud Kalmani filteri kasutamisest ja näitame, et lineaarsete mudelite jaoks annab laiendatud Kalmani filter sama tulemuse kui Kalmani filter.

CERCS teaduseriala. P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika.

Märksõnad. Juhuslikud suurused, hindamine, normaaljaotus, müra, Kalmani filtrid.

Contents

Introduction

The Kalman filter is a recursive state estimation algorithm. It estimates the current state Z_t of the system from the previous state Z_{t-1} and measurements $\{Y_1 = y_1, \ldots, Y_t = y_t\}$ up to the time t which we denote by $Y_{1:t} = y_{1:t}$.

In 1960 R. E. Kalman published his paper [2]. Since then the Kalman filter has become a widely used tool in engineering, especially in the field of navigation. Kalman filter can be used only for the linear-Gaussian state space models. The extended Kalman filter broadens the type of models it can be used for to differentiable Gaussian models.

In this theses we give a description of the Kalman filter and the extended Kalman filter. We also give examples of the application of the Kalman filter and the extended Kalman filter. The thesis is mainly based on the book *Machine Learning: A Probabilistic Perspective* by Kevin P. Murphy [3].

This thesis consists of four sections. In the first section we give the prerequsites necessary for the following sections.

In the second section we define the state space models. In the first subsection we give examples of state space models. In the second subsection we define the different types of state space model problems. In the third subsection we describe the linear-Gaussian state space models.

In the third section we describe the Kalman filter algorithm. In the first subsection we describe the Kalman filter prediction step where it is shown that if the state Z_0 has Gaussian distribution $Z_t | Y_{1:t-1} = y_{1:t-1}$ has Gaussian distribution. In the second subsection we describe the Kalman filter measurement step where it is shown, that $Z_t | Y_{1:t} = y_{1:t}$ has Gaussian distribution. In the third subsection we give an example of the application of the Kalman filter.

In the fourth section we describe the algorithm of the extended Kalman filter. In the first subsection we describe the prediction step for the extended Kalman filter. In the second subsection we describe the measurement step for the extended Kalman filter. In the third subsection we give examples of the application of the extended Kalman filter. In the fourth subsection we show that for linear-Gaussian systems the extended Kalman filter gives the same results as the Kalman filter.

The Python codes used to generate the figures for the examples are presented in the appendices.

1 Prerequisites

Let us denote random vectors by upper-case letters, that is $X = (X_1, \ldots, X_d)^T$ is a d-dimensional random vector. Let the lower-case letters represent possible numerical values of random vectors, that is $x = (x_1, \ldots, x_d)^T$ represents the possible numerical values of the vector $X = (X_1, \ldots, X_d)^T$.

Let $\mathbf{A}$ be a matrix. We denote the transpose of $\mathbf{A}$ by $\mathbf{A}^T$.

Let us denote the density function of the random vector X by $p(x)$ and the joint density function of random vectors $X_1, \ldots, X_n$ by $p(x_1, \ldots, x_n)$.

Definition 1.1. Let us look at a d-dimensional random vector X that has normal distribution with mean μ and covariance matrix $\mathbf{\Sigma}$, that is $X \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. Let us denote the density function of X by $\phi(x|\mu, \mathbf{\Sigma})$, that is

$$\phi(x|\mu, \mathbf{\Sigma}) := \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \mathbf{\Sigma}^{-1}(x-\mu)\right]. \qquad (1.1)$$

Claim 1.1. *Let a d-dimensional random vector X have normal distribution where $X \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. Suppose that $\mathbf{A}$ is a $m \times d$ matrix and $\mathbf{b}$ is a m-element vector. Let Y be a m-dimensional random vector for which $Y = \mathbf{A}X + \mathbf{b}$. Then Y has normal distribution and moreover $Y \sim \mathcal{N}(\mathbf{A}\mu + \mathbf{b}, \mathbf{A\Sigma A}^T)$ [1, p. 181].*

Claim 1.2. *Let $X_1, \ldots, X_n$ be random vectors, where X_n is a d-dimensional random vector. The rule of total probability [3, p. 29] states, that the joint probability density function of the random vectors $X_1, \ldots, X_{n-1}$ can written as:*

$$p(x_1, \ldots, x_{n-1}) = \int_{\mathbb{R}^d} p(x_1, \ldots, x_n) \mathrm{d}x_n. \qquad (1.2)$$

Let X and Y be random vectors. Let us denote the conditional density function $p(x|Y{=}y)$ by $p(x|y)$, i.e., $p(x|y) := p(x|Y{=}y)$.

Claim 1.3. *Let X, Y be random vectors. The Bayes' rule states that*

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(y|x)p(x)}{p(y)}. \qquad (1.3)$$

Claim 1.4. *Let X be a random vector that has Gaussian distribution where*

$$X \sim \mathcal{N}(\mu_X, \mathbf{\Sigma}_X) \qquad (1.4)$$

and let Y be a random vector, for which

$$Y := \mathbf{A}X + \mathbf{b} + V \tag{1.5}$$

where

$$V \sim \mathcal{N}(0, \mathbf{\Sigma}_y). \tag{1.6}$$

Then according to the Bayes rule for linear Gaussian systems [3, p. 119] $X|Y = y$ *has normal distribution and the conditional density* $p(x|y)$ *is in the form*

$$p(x|y) = \phi(x|\mu_{x|y}, \mathbf{\Sigma}_{x|y}), \tag{1.7}$$

where

$$\mathbf{\Sigma}_{x|y}^{-1} = \mathbf{\Sigma}_x^{-1} + \mathbf{A}^T\mathbf{\Sigma}_y^{-1}\mathbf{A} \tag{1.8}$$

$$\mu_{x|y} = \mathbf{\Sigma}_{x|y}[\mathbf{A}^T\mathbf{\Sigma}_y^{-1}(y - \mathbf{b}) + \mathbf{\Sigma}_x^{-1}\mu_x]. \tag{1.9}$$

Claim 1.5. *Let* K *be a d-dimensional random vector where* $K \sim \mathcal{N}(\mu, \mathbf{\Sigma})$*. Let* L *be a random vector for which* $L = \mathbf{A}K + \mathbf{b} + V$*, where the random vector* V *has a normal distribution and* $V \sim \mathcal{N}(0, \mathbf{Q})$*. The following applies:*

$$\int_{\mathbb{R}^d} \phi(l|\mathbf{A}k + \mathbf{b}, \mathbf{Q})\phi(k|\mu, \mathbf{\Sigma})\mathrm{d}k = \phi(l|\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T + \mathbf{Q})$$

Proof. Using the rule of total probability (1.2) we get that the density function of the random vector L is:

$$p(l) = \int_{\mathbb{R}^d} p(l|k)p(k)\mathrm{d}k. \tag{1.10}$$

The random vector $K \sim \mathcal{N}(\mu, \mathbf{\Sigma})$ and so $p(k) = \phi(k|\mu, \mathbf{\Sigma})$. The random vector $L = \mathbf{A}K + \mathbf{b} + V$, where $V \sim \mathcal{N}(0, \mathbf{Q})$. According to Claim 1.1 L has Gaussian distributiona and

$$L \sim \mathcal{N}(\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T + \mathbf{Q}).$$

So the density of L is $p(l) = \phi(l|\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T + \mathbf{Q})$. Now the distribution of $L|K{=}k$ is

$$L|K{=}k \sim \mathcal{N}(\mathbf{A}k + \mathbf{b}, \mathbf{Q})$$

and the density of $L|K=k$ is $p(l|K=k) = \phi(k|\mathbf{A}l + \mathbf{b}, \mathbf{Q})$.

The rule of total probability (1.10) becomes

$$\phi(l|\mathbf{A}\mu + \mathbf{b}, \mathbf{A\Sigma A}^T + \mathbf{Q}) = \int_{\mathbb{R}^d} \phi(l|\mathbf{A}k + \mathbf{b}, \mathbf{Q})\phi(k|\mu, \mathbf{\Sigma})\mathrm{d}k. \tag{1.11}$$

■

Definition 1.2. Let $X = (X_1, \ldots, X_d)^T$ be a random vector and let $\mathbf{f}$ be a function. The first order Taylor expansion of the function $\mathbf{f}$ where $\mathbf{f}(X) = (f_1(X), \ldots, f_m(X))$ is:

$$\mathbf{f}(X) \approx \mathbf{f}(\mathbf{a}) + \mathbf{Df}(\mathbf{a})(X - \mathbf{a}), \tag{1.12}$$

where $\mathbf{a}$ is a point where $\mathbf{f}$ is differentiable and $\mathbf{Df}(\mathbf{a})$ is the matrix of partial derivatives, that is

$$\mathbf{Df}(\mathbf{a}) = \begin{bmatrix} \frac{\partial \mathbf{f}_1(\mathbf{a})}{\partial x_1} & \frac{\partial \mathbf{f}_1(\mathbf{a})}{\partial x_2} & \cdots & \frac{\partial \mathbf{f}_1(\mathbf{a})}{\partial x_d} \\ \frac{\partial \mathbf{f}_2(\mathbf{a})}{\partial x_1} & \frac{\partial \mathbf{f}_2(\mathbf{a})}{\partial x_2} & \cdots & \frac{\partial \mathbf{f}_2(\mathbf{a})}{\partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{f}_m(\mathbf{a})}{\partial x_1} & \frac{\partial \mathbf{f}_m(\mathbf{a})}{\partial x_2} & \cdots & \frac{\partial \mathbf{f}_m(\mathbf{a})}{\partial x_d} \end{bmatrix}. \tag{1.13}$$

2 State Space Models

Let $Z_0, Z_1, Z_2 \ldots$ be a sequence of random vectors called "states". Suppose that for every $t = 1, 2, \ldots$ the value of Z_t only depends on the previous state value Z_{t-1} and some independent random vector V_t called "noise", but not the state values $Z_1, \ldots, Z_{t-2}$. Then Z_t can be written as a function of the previous state Z_{t-1} and the noise V_t, more precisely we get the following state space model:

$$Z_t = \mathbf{g}_t(Z_{t-1}, V_t) \qquad t = 1, 2, \ldots. \tag{2.1}$$

Now let us look at some examples of state space models.

2.1 Examples of State Space Models

Let us look at a state space where the state Z_t is a location in a one dimensional space. Then Z_t is a random variable. Let the state model have a linear noise

V_t, that is, let the state space model be in the form

$$Z_t = g(Z_{t-1}) + V_t \qquad t = 1, 2, \dots . \tag{2.2}$$

For this state space model the expectation of state Z_t becomes

$$\mathrm{E}[Z_t] = \mathrm{E}[g(Z_{t-1}) + V_t] = \mathrm{E}[g(Z_{t-1})] + \mathrm{E}[V_t]. \tag{2.3}$$

The previous state Z_{t-1} and the noise V_t are independent random variables. So $g(Z_{t-1})$ and V_t are also independent. The variance of state Z_t becomes

$$\mathrm{Var}[Z_t] = \mathrm{Var}[g(Z_{t-1}) + V_t] = \mathrm{Var}[g(Z_{t-1})] + \mathrm{Var}[V_t]. \tag{2.4}$$

Let us assume, that the noise V_t has expectation 0 and variance σ_V^2 . Then the expectation and variance of state Z_t become

$$\mathrm{E}[Z_t] = \mathrm{E}[g(Z_{t-1})], \tag{2.5}$$

$$\mathrm{Var}[Z_t] = \mathrm{Var}[g(Z_{t-1})] + \sigma_V^2. \tag{2.6}$$

Let Z_{t-1} have expectation $\mathrm{E}[Z_{t-1}] = \mu$ and variance $\mathrm{Var}[Z_{t-1}] = \sigma^2$.

Example 2.1. Let g be linear, that is let $g(Z_t) = aZ_{t-1} + b$, where $a, b \in \mathbb{Z}$. For this model the expectation and variance of the state Z_t become

$$\mathrm{E}[Z_t] = \mathrm{E}[g(Z_{t-1})] = \mathrm{E}[aZ_{t-1} + b] = a\mathrm{E}[Z_{t-1}] + b = a\mu + b \tag{2.7}$$

and

$$\begin{aligned}\mathrm{Var}[Z_t] &= \mathrm{Var}[g(Z_{t-1})] + \sigma_V^2 = \mathrm{Var}[aZ_{t-1} + b] + \sigma_V^2 = a^2\mathrm{Var}[Z_{t-1}] + \sigma_V^2 \\ &= a^2\sigma + \sigma_V^2.\end{aligned} \tag{2.8}$$

When we assume, that the noise V_t has normal distribution and that the previous state Z_{t-1} also has Gaussian distribution, then using Claim 1.1 the state $Z_t = aZ_{t-1} + V_t$ has normal distribution.

Example 2.2. Let $g(Z_{t-1}) = Z_{t-1}^3 + Z_{t-1}$. Let us assume, that the noise V_t has normal distribution more precisely $V_t \sim \mathcal{N}(0, \sigma_V^2)$. We also assume, that the previous state Z_{t-1} has normal distribution with expectation μ and variance σ^2, that is $Z_{t-1} \sim \mathcal{N}(\mu, \sigma^2)$.

Even though the previous state Z_{t-1} and the noise V_t have normal distribution the random variable $Z_t = g(Z_{t-1}) + V_t$ does not have normal distribution.

2.2 Types of State Space Model Problems

Let us have the state space model

$$Z_t = \mathbf{g}_t(Z_{t-1}, V_t), \qquad t = 1, 2, \ldots. \tag{2.9}$$

Let

$$Y_t = \mathbf{h}_t(Z_t, W_t) \qquad t = 1, 2, \ldots \tag{2.10}$$

be the random vector of "measurements", where the noise W_t is a Gaussian random variable, which is independent from Z_t.

Suppose we have measured the values $y_0, \ldots, y_{t_1}$, which we denote by $y_{1:t_1} := \{y_0, \ldots, y_{t_1}\}$, we denote the corresponding random variables $Y_0, \ldots, Y_{t_1}$ by $Y_{1:t_1} := \{Y_0, \ldots, Y_{t_1}\}$ and when $\{Y_0 = y_0, \ldots, Y_{t-1} = y_{t_1}\}$ we write that $Y_{1:t_1} = y_{1:t_1}$.

The state estimation problem in general consists of estimating the distribution of Z_{t_2} given the observations $y_{1:t_1}$. Depending on the time t_2 the state-estimation problems can be divided as:

- **filtering**, estimating the distribution of the state at time $t_2 = t_1$.
- **smoothing**, estimating the distribution of the state at time $t_2 < t_1$.
- **prediction**, estimating the distribution of the state at time $t_2 > t_1$.

2.3 Linear-Gaussian State Space Model

Let Z_t be the state at time t that can be presented as a function of Z_{t-1} and noise vector V_t. Let Y_t be a random vector, which can be presented as a function of Z_t and the random "measurement noise" vector W_t. The state space model can be written in the form

$$Z_t = \mathbf{g}_t(Z_{t-1}, V_t) \qquad t = 1, 2, \ldots, \tag{2.11}$$

$$Y_t = \mathbf{h}_t(Z_t, W_t) \qquad t = 1, 2, \ldots. \tag{2.12}$$

The model (2.11) is called the "system model" and the model (2.12) is called the "measurement model".

Let us look at a linear-Gaussian state space model, that is a state space model for which

- the system noise V_t and the measurement noise W_t are independent and Gaussian with the expectation 0 and covariance matrices $\mathbf{Q}_t$ and $\mathbf{R}_t$ respectively, that is:

$$V_t \sim \mathcal{N}(0, \mathbf{Q}_t), \tag{2.13}$$

$$W_t \sim \mathcal{N}(0, \mathbf{R}_t). \tag{2.14}$$

- the function $\mathbf{g}_t$ is a linear function, which means the system model can be written as

$$Z_t = \mathbf{A}_t Z_{t-1} + V_t \qquad t = 1, 2, \ldots, \tag{2.15}$$

  where $\mathbf{A}_t$ is a matrix defining the linear function $\mathbf{g}_t$.

- the function $\mathbf{h}_t$ is a linear function, which means, that the measurement model can be written as:

$$Y_t = \mathbf{C}_t Z_t + W_t \qquad t = 1, 2, \ldots, \tag{2.16}$$

  where $\mathbf{C}_t$ is a matrix defining the linear function $\mathbf{h}_t$.

For any $t_1, t_2 \in \mathbb{N}$ and $y_{1:t_1}$ let us denote the expectation $\mathrm{E}[Z_{t_2}|Y_{1:t_1} = y_{1:t_1}] =: \mu_{t_2|t_1}$ and the covariance matrix $\mathbf{Cov}[Z_{t_2}|Y_{1:t_1} = y_{1:t_1}] =: \mathbf{\Sigma}_{t_2|t_1}$. Both $\mu_{t_2|t_1}$ and $\mathbf{\Sigma}_{t_2|t_1}$ depend on $y_{1:t_1}$, but $y_{1:t_1}$ is left out of the notation because it is fixed. If $t_1 = t_2 =: t$, we denote $\mu_{t|t} =: \mu_t$ and $\mathbf{\Sigma}_{t|t} =: \mathbf{\Sigma}_t$.

Claim 2.1. *If Z_0 Gaussian random vector then Z_t is a Gaussian random vector for every $t = 1, 2, \ldots$.*

Proof. Let Z_0 be a Gaussian random vector. We prove by induction that Z_t is Gaussian. Firstly for the induction base we have that Z_0 is Gaussian.

For the induction step we assume, that state Z_t is Gaussian and that $Z_t \sim \mathcal{N}(\mu_{t|0}, \mathbf{\Sigma}_{t|0})$. We show, that the state Z_{t+1} is also Gaussian. According to the system model (2.9) the state $Z_{t+1} = \mathbf{A}_{t+1} Z_t + V_{t+1}$. As Z_t and V_t are Gaussian then according to Claim 1.1 the state Z_{t+1} has Gaussian distribution and

$$Z_{t+1} \sim \mathcal{N}(\mathbf{A}_t \mu_{t|0}, \mathbf{A}_t \mathbf{\Sigma}_{t|0} \mathbf{A}_t^T + \mathbf{Q}_{t+1}) =: \mathcal{N}(\mu_{t+1|0}, \mathbf{\Sigma}_{t+1|0}). \tag{2.17}$$

■

Let us assume, that Z_0 is a Gaussian random variable, more precisely $Z_0 \sim \mathcal{N}(\mu_0, \mathbf{\Sigma}_0)$. Let us now look at an example of a linear-Gaussian state space model.

Example 2.3. Let us consider an object moving in a 2-dimensional plane, where $z_{1,t}$ and $z_{2,t}$ are the horizontal and vertical location coordinates and $\dot{z}_{1,t}$ and $\dot{z}_{2,t}$ are the corresponding velocities. We can represent the described system with a state vector:

$$Z_t = \begin{pmatrix} Z_{1,t} \\ Z_{2,t} \\ \dot{Z}_{1,t} \\ \dot{Z}_{2,t} \end{pmatrix} \tag{2.18}$$

Let us assume that the object is moving at constant velocity with random Gaussian noise, that is the velocities are in the form $\dot{z}_{1,t} = \dot{z}_{1,t-1} + V_{3,t}$ and $\dot{z}_{2,t} = \dot{z}_{2,t-1} + V_{4,t}$ where the noise $V_t = (V_{1,t}, V_{2,t}, V_{3,t}, V_{4,t})^T$ is Gaussian, and $V_t \sim \mathcal{N}(0, \mathbf{Q}_t)$.

This means we can model the system as follows:

$$Z_t = \begin{pmatrix} Z_{1,t} \\ Z_{2,t} \\ \dot{Z}_{1,t} \\ \dot{Z}_{2,t} \end{pmatrix} = \begin{pmatrix} Z_{1,t} + \dot{Z}_{1,t-1} \cdot \Delta + V_{1,t} \\ Z_{2,t} + \dot{Z}_{2,t-1} \cdot \Delta + V_{2,t} \\ \dot{Z}_{1,t-1} + V_{3,t} \\ \dot{Z}_{2,t-1} + V_{4,t} \end{pmatrix} \tag{2.19}$$

where Δ is the sampling period.

This equation can be written in the matrix form as:

$$Z_t = \mathbf{A}_t Z_{t-1} + V_t \tag{2.20}$$

where

$$A = \begin{pmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.21}$$

Now suppose we can only observe the location of the object but not its velocity. Let $y_{1,t}$ and $y_{2,t}$ be the measured location coordinates. Then $y_t = (y_{1,t}, y_{2,t})^T$ is the observed location which is subject to Gaussian noise. We can model the measurement as follows:

$$Y_t = \begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} c_1 Z_{1,t} + W_{1,t} \\ c_2 Z_{2,t} + W_{2,t} \end{pmatrix} \tag{2.22}$$

where $W_t = (W_{1,t}, W_{2,t})^T$ is the measurement noise, which has normal distribution and $W_t \sim \mathcal{N}(0, \mathbf{R}_t)$ and $a_1, a_1 \in \mathbb{R}$

We can write the measurement model in the matrix form as follows:

$$Y_t = \mathbf{C}_t Z_t + \boldsymbol{\delta}_t \tag{2.23}$$

where

$$\mathbf{C}_t = \begin{pmatrix} c_1 & 0 & 0 & 0 \\ 0 & c_2 & 0 & 0 \end{pmatrix}. \tag{2.24}$$

3 Kalman filter

The Kalman filter is a recursive filtering algorithm, which evaluates the state Z_t for linear-Gaussian state space models from Z_{t-1} and measurements $Y_{1:t}$.

Let us have a linear-Gaussian state space model as described before. Let us assume, that Z_0 is Gaussian and that $Z_0 \sim \mathcal{N}(\mu_0, \mathbf{\Sigma}_0)$. We have already shown, that then Z_t is Gaussian for every $t = 1, 2, \ldots$ and we have denoted $Z_t \sim \mathcal{N}(\mu_{t|0}, \mathbf{\Sigma}_{t|0})$.

We show, that the conditional density $p(Z_t | Y_{1:t} = y_{1:t})$ is Gaussian. We do this by induction. For the induction basis we know that Z_0 is Gaussian. We divide the induction step into two parts:

- **prediction step**, show that $Z_{t-1} | Y_{1:t-1} = y_{1:t-1}$ has Gaussian distribution,

- **measurment step**, show that $Z_{t-1}|Y_{1:t}=y_{1:t}$ has Gaussian distribution.

Let us assume, that the distribution of $Z_{t-1}|Y_{1:t-1}=y_{1:t-1}$ is available and Gaussian, that is $p(z_{t-1}|y_{1:t-1}) = \phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1})$.

3.1 Prediction step

Let the dimension of Z_t for every $t = 1, 2, \dots$ be d. Using the law of total probability (1.2) we get that the joint density function of Z_t and $Y_{1:t-1}$ can be written as:

$$p(z_t, y_{1:t-1}) = \int_{\mathbb{R}^d} p(z_t, y_{1:t-1}, z_{t-1})\mathrm{d}z_{t-1}. \tag{3.1}$$

By the definition of conditional probability the density function $p(z_t, y_{1:t-1})$ can be written as:

$$p(z_t, y_{1:t-1}) = p(z_t|y_{1:t-1})p(y_{1:t-1}) \tag{3.2}$$

and similarly the density function $p(z_t, y_{1:t-1}, z_{t-1})$ can be written as:

$$\begin{aligned} p(z_t, y_{1:t-1}, z_{t-1}) =& p(z_t|y_{1:t-1}, z_{t-1})p(y_{1:t-1}, z_{t-1}) \\ =& p(z_t|y_{1:t-1}, z_{t-1})p(z_{t-1}|y_{1:t-1})p(y_{1:t-1}). \end{aligned} \tag{3.3}$$

So the equation (3.1) can be rewritten as

$$p(z_t|y_{1:t-1})p(y_{1:t-1}) = \int_{\mathbb{R}^d} p(z_t|y_{1:t-1}, z_{t-1})p(z_{t-1}|y_{1:t-1})p(y_{1:t-1})\mathrm{d}z_{t-1}. \tag{3.4}$$

We divide the previous equation by $p(y_{1:t-1})$ and get:

$$p(z_t|y_{1:t-1}) = \int_{\mathbb{R}^d} p(z_t|y_{1:t-1}, z_{t-1})p(z_{t-1}|y_{1:t-1})\mathrm{d}z_{t-1}. \tag{3.5}$$

By the definition of linear-Gaussian state space model the conditional distribution of $Z_t|Y_{1:t-1}, Z_{t-1}$ is independent of $Y_{1:t-1}$ and equal to the conditional distribution of $Z_t|Z_{t-1}$. Therefore $p(z_t|y_{1:t-1}, z_{t-1}) = p(z_t|z_{t-1})$. Now we can write the conditional density (3.5) as:

$$\begin{aligned} p(z_t|y_{1:t-1}) &= \int_{\mathbb{R}^d} p(z_t|y_{1:t-1}, z_{t-1})p(z_{t-1}|y_{1:t-1})\mathrm{d}z_{t-1} \\ &= \int_{\mathbb{R}^d} p(z_t|z_{t-1})\phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1})\mathrm{d}z_{t-1}. \end{aligned} \tag{3.6}$$

As we proved before Z_{t-1} is Gaussian and $Z_{t-1} \sim \mathcal{N}(\mu_{t-1|0}, \mathbf{\Sigma}_{t-1|0})$. The system model (2.15) is $Z_t = \mathbf{A}_t Z_{t-1} + V_t$. Both Z_{t-1} and V_t have a normal distribution and $V_t \sim \mathcal{N}(0, \mathbf{Q}_t)$. Then according to Claim 1.1 the distribution of Z_t is also Gaussian and more precisely

$$Z_t \sim \mathcal{N}(\mathbf{A}_t \mu_{t-1|0}, \mathbf{A}_t \mathbf{\Sigma}_{t-1|0} \mathbf{A}_t^T + \mathbf{Q}_t). \tag{3.7}$$

Both Z_t and Z_{t-1} are Gaussian. Then the conditional distribution of $Z_t|Z_{t-1}=z_{t-1}$ is also Gaussian and more precisely

$$Z_t|Z_{t-1}=z_{t-1} \sim \mathcal{N}(\mathbf{A}_t z_{t-1}, \mathbf{Q}_t). \tag{3.8}$$

So the density function of $Z_t|Z_{t-1}=z_{t-1}$ is

$$p(z_t|z_{t-1}) = \phi(z_t|\mathbf{A}_t z_{t-1}, \mathbf{Q}_t). \tag{3.9}$$

We can now rewrite (3.6) as:

$$\begin{aligned} p(z_t|y_{1:t-1}) &= \int_{\mathbb{R}^d} p(z_t|z_{t-1}) \phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1}) \mathrm{d}z_{t-1} \\ &= \int_{\mathbb{R}^d} \phi(z_t|\mathbf{A}_t z_{t-1}, \mathbf{Q}_t) \phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1}) \mathrm{d}z_{t-1} \end{aligned} \tag{3.10}$$

We show, that $p(z_t|y_{1:t-1})$ has normal distribution. We take K to be

$$Z_{t-1}|Y_{1:t-1}=y_{1:t-1} \sim \mathcal{N}(\mu_{t-1}, \mathbf{\Sigma}_{t-1}). \tag{3.11}$$

We take L to be $Z_t = \mathbf{A}_t K + V_t$, where $V_t \sim \mathcal{N}(0, \mathbf{Q}_t)$. Then according to Claim 1.5 the following equation applies:

$$\begin{aligned} &\int_{\mathbb{R}^d} \phi(z_t|\mathbf{A}_t z_{t-1}, \mathbf{Q}_t) \phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1}) \mathrm{d}z_{t-1} \\ &= \phi(z_t|\mathbf{A}_t \mu_{t-1}, \mathbf{A}_t \mathbf{\Sigma}_{t-1} \mathbf{A}_t^T + \mathbf{Q}_t). \end{aligned} \tag{3.12}$$

Let us denote

$$\mu_{t|t-1} := \mathbf{A}_t \mu_{t-1}, \tag{3.13}$$

$$\mathbf{\Sigma}_{t|t-1} := \mathbf{A}_t \mathbf{\Sigma}_{t-1} \mathbf{A}_t^T + \mathbf{Q}_t. \tag{3.14}$$

Then (3.10) is

$$
\begin{aligned}
p(z_t|y_{1:t-1}) &= \int_{\mathbb{R}^d} \phi(z_t|\mathbf{A}_t z_{t-1}, \mathbf{Q}_t)\phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1})\mathrm{d}z_{t-1} \\
&= \phi(z_t|\mu_{t|t-1}, \mathbf{\Sigma}_{t|t-1}).
\end{aligned}
\tag{3.15}
$$

3.2 Measurement step

By the definition of state space models the conditional distribution of $Y_t|Y_{1:t-1} = y_{1:t}, Z_t = z_t$ is independent of $Y_{1:t-1}$ and equals to $Y_t|Z_t = z_t$. Then the density functions are also equal, i.e., $p(y_t|y_{1:t-1}, z_t) = p(y_t|z_t)$.

For the density function $p(z_t|y_{1:t})$ the Bayes' rule (1.3) becomes:

$$
p(z_t|y_{1:t}) = p(z_t|y_t, y_{1:t-1}) = \frac{p(y_t|z_t)p(z_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}. \tag{3.16}
$$

In the Bayes rule for linear-Gaussian systems, Claim 1.4, we take X to be $Z_t|Y_{1:t-1} = y_{1:t-1}$. The distribution of $Z_t|Y_{1:t-1} = y_{1:t-1}$ is Gaussian and

$$
Z_t|Y_{1:t-1} = y_{1:t-1} \sim \mathcal{N}(\mu_{t|t-1}, \mathbf{\Sigma}_{t|t-1}). \tag{3.17}
$$

We take Y to be Y_t. According to (2.16) Y_t can be expressed as:

$$
Y_t = \mathbf{C}_t Z_t + W_t, \tag{3.18}
$$

where

$$
W_t \sim \mathcal{N}(0, \mathbf{R}_t). \tag{3.19}
$$

Now according to the Bayes rule of the linear-Gaussian systems, Claim 1.4, the posterior is given as:

$$
p(z_t|Y_{1:t} = y_{1:t}) = \phi(z_t|\mu_t, \mathbf{\Sigma}_t) \tag{3.20}
$$

where

$$
\mathbf{\Sigma}_t^{-1} = \mathbf{\Sigma}_{t|t-1}^{-1} + \mathbf{C}_t^T \mathbf{R}_t^{-1} \mathbf{C}_t \tag{3.21}
$$

$$
\mu_t = \mathbf{\Sigma}_t \mathbf{C}_t^T \mathbf{R}_t^{-1} Y_t + \mathbf{\Sigma}_t \mathbf{\Sigma}_{t|t-1}^{-1} \mu_{t|t-1}. \tag{3.22}
$$

Using simple transformations [3, p. 643] the expectation μ_t and the covariance matrix $\mathbf{\Sigma}_t$ can be written as

$$\mu_t = \mu_{t|t-1} + \mathbf{K}_t(y_t - \hat{y}_t) \tag{3.23}$$

$$\mathbf{\Sigma}_t = (\mathbf{I} - \mathbf{K}_t\mathbf{C}_t)\mathbf{\Sigma}_{t|t-1} \tag{3.24}$$

where $\hat{y}_t$ is the predicted observation and

$$\hat{y}_t := \mathrm{E}[Y_t|Y_{1:t-1}=y_{1:t-1}] = \mathbf{C}_t\mu_{t|t-1} \tag{3.25}$$

and where $\mathbf{K}_t$ is the Kalman gain matrix given by

$$\mathbf{K}_t = \mathbf{\Sigma}_{t|t-1}\mathbf{C}_t^T\mathbf{S}_t^{-1} \tag{3.26}$$

where

$$\begin{aligned}\mathbf{S}_t : &= \mathrm{cov}[r_t|Y_{1:t-1}] \\ &= \mathbf{cov}[\mathbf{C}_t Z_t + \mathbf{w}_t - \mathbf{C}_t\mu_{t|t-1}|Y_{1:t-1}] \\ &= \mathbf{C}_t\mathbf{\Sigma}_{t|t-1}\mathbf{C}_t^T + \mathbf{R}_t.\end{aligned} \tag{3.27}$$

We have shown, that $Z_t|Y_{1:t}=y_{1:t}$ has normal distribution and that

$$p(z_t|Y_{1:t}=y_{1:t}) = \phi(z_t|\mu_t, \mathbf{\Sigma}_t). \tag{3.28}$$

Now we can estimate the state Z_t on the condition, that $Y_{1:t}=y_{1:t}$ with μ_t.

3.3 Example of Kalman Filter Estimation

Example 3.1. Let us look at a 1-dimensional linear-Gaussian state space model

$$Z_t = Z_{t-1} + V_t \tag{3.29}$$

$$Y_t = Z_t + W_t \tag{3.30}$$

where

$$V_t \sim \mathcal{N}(0, Q_t) \tag{3.31}$$

$$W_t \sim \mathcal{N}(0, R_t). \tag{3.32}$$

Now we take $Z_0 \sim \mathcal{N}(0, 0.1) =: \mathcal{N}(\mu_0, \sigma_0^2)$ and use the Kalman filter to find the density functions $p(z_t|Y_{1:t} = y_{1:t}) = \phi(z_t|\mu_t, \sigma_t^2)$ for every $t = 1, 2, \ldots$.

Recursively for every $t = 1, \ldots$ we first calculate $\mu_{t|t-1}$ and $\sigma_{t|t-1}^2$ from the prediction step, more precisely from the formulas (3.13) and (3.14). For our case it is:

$$\mu_{t|t-1} = \mu_{t-1}, \tag{3.33}$$

$$\sigma_{t|t-1}^2 = \sigma_{t-1}^2 + Q_t. \tag{3.34}$$

Then we calculate μ_t and σ_t^2 from the measurement step using the formulas (3.23) and (3.24), which for our case are:

$$\mu_t = \mu_{t|t-1} + K_t(y_t - \hat{y}_t) \tag{3.35}$$

$$\sigma_t^2 = (1 - K_t)\sigma_{t|t-1}^2, \tag{3.36}$$

where for our case according to (3.25) we have that $\hat{y}_t = \mu_{t|t-1}$ and according to (3.26) and (3.27) we have that $K_t = \sigma_{t|t-1}^2 * 1/(\sigma_{t|t-1}^2 + R_t)$.

Now from the Kalman filter measurement step we know, that $Z_t|Y_{1:t}{=}y_{1:t} \sim \mathcal{N}(\mu_t, \sigma_t^2)$.

We can write μ_t as

$$\mu_t = \mu_{t-1} + K_t(y_t - \mu_{t-1}) = \mu_{t-1}(1 - K_t) + K_t y_t \tag{3.37}$$

and K_t as:

$$K_t = \frac{\sigma_{t-1}^2 + Q_t}{\sigma_{t-1}^2 + Q_t + R_t}. \tag{3.38}$$

Then $K_t \in [0, 1]$. If $R_t \to \infty$, then $K_t \to 0$ and if $Q_t \to \infty$ or $R_t \to 0$, then $K_t \to 1$.

Now if $K_t \to 0$ the Kalman estimation $\mu_t \to \mu_{t-1}$. If $K_t \to 1$ then $\mu_t \to y_t$. That is the bigger R_t is, the more μ_t depends on μ_{t-1} and less on y_t. And the bigger Q_t is the more μ_t depends on the measurement y_t and less on μ_{t-1}).

For the general case for the Kalman filter according to (3.23), (3.13) and (3.25) μ_t can be written as:

$$\mu_t = \mathbf{A}_t\mu_{t-1} + \mathbf{K}_t(y_t - \mathbf{C}_t\mathbf{A}_t\mu_{t-1}). \tag{3.39}$$

Now if $\mathbf{C}_t = \mathbf{I}$ we can write μ_t as

$$\mu_t = (\mathbf{I} - \mathbf{K}_t)\mathbf{A}_t\mu_{t-1} + \mathbf{K}_t y_t \tag{3.40}$$

and the same relations hold as in the example. That is the bigger R_t is, the more μ_t depends on μ_{t-1} and the bigger Q_t is the more μ_t depends on the measurement y_t.

Now let us look at figures with the measurement values y_t, state values z_t and the expectation μ_t where $t = 0, \ldots, 100$ for different values of Q_t and R_t.

From the figures we can see, that the variance of the system noise Q_t affects the amplitude of the state function. The bigger Q_t is, the bigger is the amplitude of the state. The variance of the measurement noise R_t affects the amplitude of the measurements. The bigger R_t the bigger is the amplitude of the measurements. When R_t is small the measurement values are close to the state values and thus the extended Kalman filter estimation μ_t is close to both z_t and y_t.

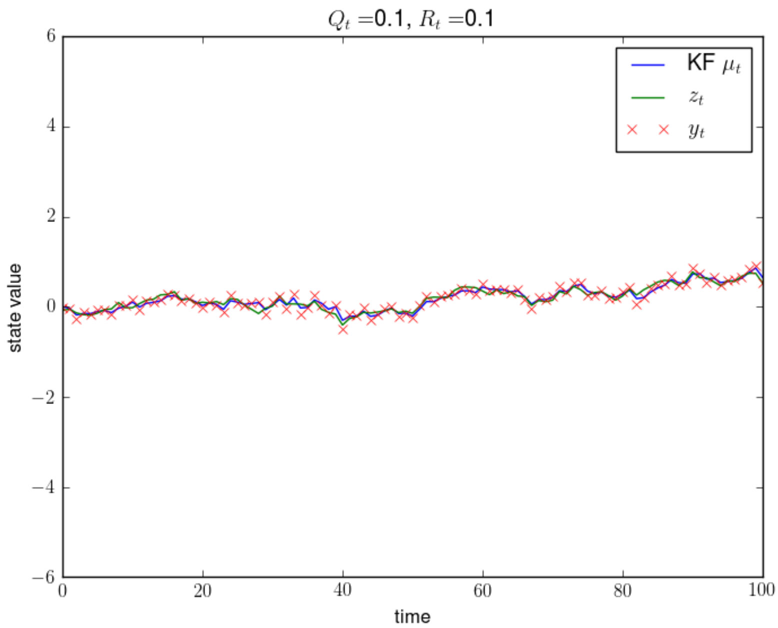

Figure 3.1: As both noise variances R_t and Q_t are small the expectation μ_t is close to the state value z_t.

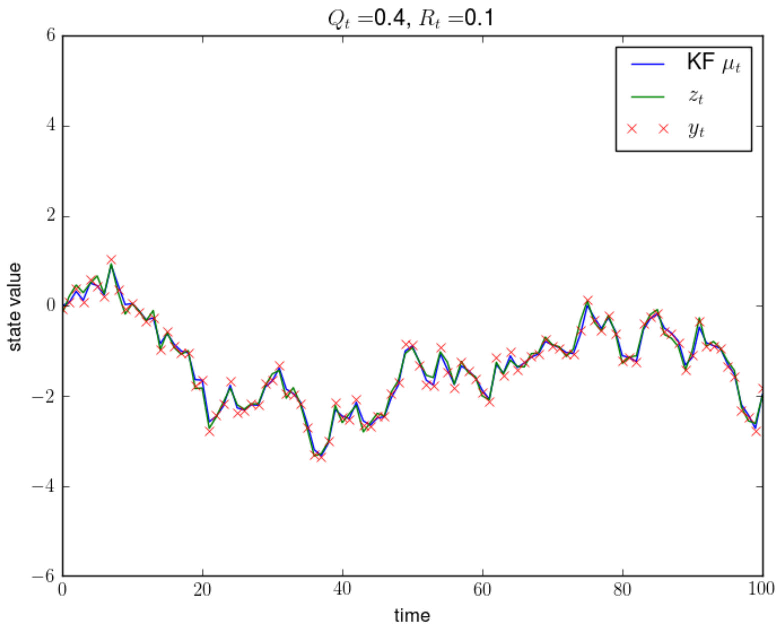

Figure 3.2: As the value of Q_t has increased from that in the previous figure the value of μ_t depends more on y_t then in figure 3.1.

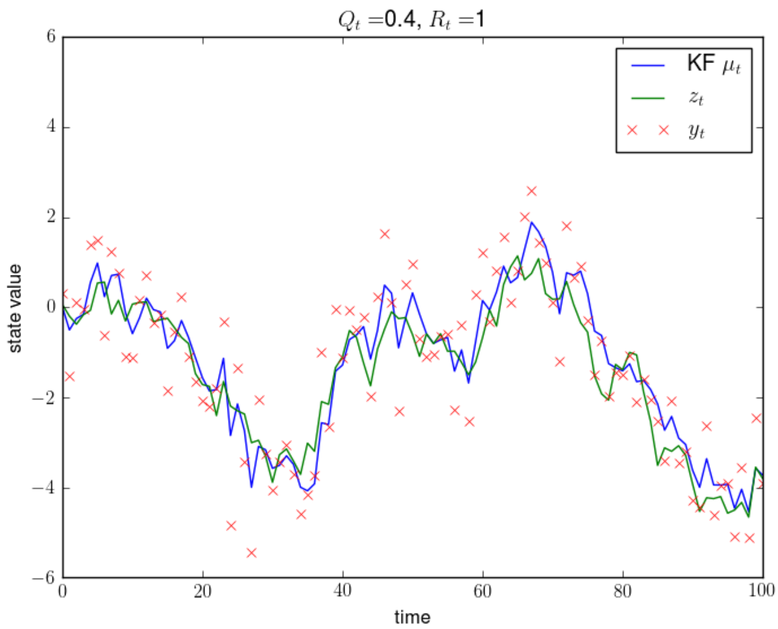

Figure 3.3: As the value of R_t has increased from that in the previous figure the value of μ_t depends more on μ_{t-1} and less on y_t then in figure 3.2.

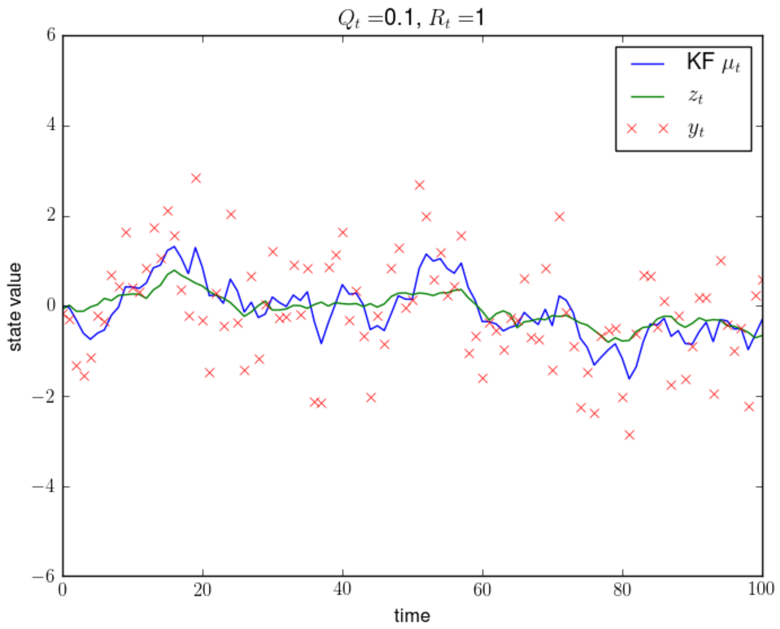

Figure 3.4: As the value of Q_t has decreased from that in the previous figure the value of μ_t depends more on y_t and less on μ_{t-1} then in figure 3.3.

4 The Extended Kalman Filter

Let us look at a state space model, where the system model (2.9) and the measurement model (2.10) are in the form:

$$Z_t = \mathbf{g}_t(Z_{t-1}) + V_t, \tag{4.1}$$

$$Y_t = \mathbf{h}_t(Z_t) + W_t, \tag{4.2}$$

where $\mathbf{g}_t, \mathbf{h}_t : \mathbb{R}^d \to \mathbb{R}^d$ are nonlinear but differentiable functions and $V_t \sim \mathcal{N}(0, \mathbf{Q}_t)$ and $W_t \sim \mathcal{N}(0, \mathbf{R}_t)$.

4.1 Prediction Step

Let us assume, that

$$Z_{t-1}|Y_{1:t-1}{=}y_{1:t-1} \sim \mathcal{N}(\mu_{t-1}, \mathbf{\Sigma}_{t-1}). \tag{4.3}$$

We know from the prediction step of the Kalman filter that

$$\begin{aligned} p(z_t|y_{1:t-1}) &= \int_{\mathbb{R}^d} p(z_t|y_{1:t-1}, z_{t-1}) p(z_{t-1}|y_{1:t-1}) \mathrm{d}z_{t-1} \\ &= \int_{\mathbb{R}^d} p(z_t|z_{t-1}) \phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1}) \mathrm{d}z_{t-1}. \end{aligned} \tag{4.4}$$

The expectation of $Z_t|Z_{t-1}=z_{t-1}$ is:

$$\mathrm{E}[Z_t|Z_{t-1}=z_{t-1}] = \mathrm{E}[\mathbf{g}_t(Z_{t-1}) + V_t|Z_{t-1}=z_{t-1}] = \mathbf{g}_t(z_{t-1}) \tag{4.5}$$

and the covariance matrix of $Z_t|Z_{t-1}=z_{t-1}$ is:

$$\begin{aligned} \mathrm{Var}[Z_t|Z_{t-1}=z_{t-1}] =& \mathrm{Var}[\mathbf{g}_t(Z_{t-1}) + V_t|Z_{t-1}=z_{t-1}] = \mathrm{Var}[\mathbf{g}_t(Z_{t-1})|Z_{t-1}=z_{t-1}] \\ &+\mathbf{Q}_t = \mathrm{E}[(\mathbf{g}_t(Z_{t-1}) - \mathrm{E}[\mathbf{g}_t(Z_{t-1})])\cdot \\ &\cdot(\mathbf{g}_t(Z_{t-1}) - \mathrm{E}[\mathbf{g}_t(Z_{t-1})])^T|Z_{t-1}=z_{t-1}] + \mathbf{Q}_t \\ &=\mathbf{Q}_t. \end{aligned} \tag{4.6}$$

We linearize the function $\mathbf{g}_t$ using the first order Taylor expansion, Definition 1.2, at the point μ_{t-1}, that is

$$\mathbf{g}_t(Z_{t-1}) \approx \mathbf{g}_t(\mu_{t-1}) + \mathbf{D}\mathbf{g}_t(\mu_{t-1})(Z_{t-1} - \mu_{t-1}). \tag{4.7}$$

Now let us denote the approximation of Z_t by $\tilde{Z}_t$, i.e.,

$$\tilde{Z}_t := \mathbf{g}_t(\mu_{t-1}) + \mathbf{D}\mathbf{g}_t(\mu_{t-1})(Z_{t-1} - \mu_{t-1}) + V_t. \tag{4.8}$$

We approximate

$$p(z_t|z_{t-1}) \approx \phi(z_t|\mathbf{D}\mathbf{g}_t(\mu_{t-1})z_{t-1} + (\mathbf{g}_t(\mu_{t-1}) - \mathbf{D}\mathbf{g}_t(\mu_{t-1})\mu_{t-1}), \mathbf{Q}_t) \tag{4.9}$$

We can now rewrite (4.4) as:

$$\begin{aligned} p(z_t|y_{1:t-1}) &= \int p(z_t|z_{t-1}) \phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1}) \mathrm{d}z_{t-1} \\ &\approx \int \phi(z_t|\mathbf{D}\mathbf{g}_t(\mu_{t-1})z_{t-1} + (\mathbf{g}_t(\mu_{t-1}) - \mathbf{D}\mathbf{g}_t(\mu_{t-1})\mu_{t-1}), \mathbf{Q}_t)\cdot \\ &\cdot \phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1}) \mathrm{d}z_{t-1} \end{aligned} \tag{4.10}$$

We take K to have the same distribution as $Z_{t-1}|Y_{1:t-1}=y_{1:t-1} \sim \mathcal{N}(\mu_{t-1}, \mathbf{\Sigma}_{t-1})$ and we take L to be $\tilde{Z}_t = \mathbf{Dg}_t(\mu_{t-1})K + (\mathbf{g}_t(\mu_{t-1}) - \mathbf{Dg}_t(\mu_{t-1})\mu_{t-1}) + V_t$, where $V_t \sim \mathcal{N}(0, \mathbf{Q}_t)$ is independent of K.

Then according to the Claim 1.5 the following equation applies:

$$\int \phi(z_t|\mathbf{Dg}_t(\mu_{t-1})z_{t-1} + (\mathbf{g}_t(\mu_{t-1}) - \mathbf{Dg}_t(\mu_{t-1})\mu_{t-1}), \mathbf{Q}_t)\phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1})\mathrm{d}z_{t-1}$$
$$=\phi(z_t|\mathbf{g}_t(\mu_{t-1}), \mathbf{Dg}_t(\mu_{t-1})\mathbf{\Sigma}_{t-1}\mathbf{Dg}_t(\mu_{t-1})^T + \mathbf{Q}_t). \tag{4.11}$$

Let us denote

$$\tilde{\mu}_{t|t-1} := \mathbf{g}_t(\mu_{t-1}), \tag{4.12}$$
$$\tilde{\mathbf{\Sigma}}_{t|t-1} := \mathbf{Dg}_t(\mu_{t-1})\mathbf{\Sigma}_{t-1}\mathbf{Dg}_t(\mu_{t-1})^T + \mathbf{Q}_t. \tag{4.13}$$

Then (4.4) is

$$\begin{aligned} p(z_t|y_{1:t-1}) \approx & \int_{\mathbb{R}^d} \phi(z_t|\mathbf{Dg}_t(\mu_{t-1})z_{t-1} + (\mathbf{g}_t(\mu_{t-1}) - \mathbf{Dg}_t(\mu_{t-1})\mu_{t-1}), \mathbf{Q}_t)\cdot \\ & \cdot \phi(z_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1})\mathrm{d}z_{t-1} \\ =& \phi(z_t|\tilde{\mu}_{t|t-1}, \tilde{\mathbf{\Sigma}}_{t|t-1}). \end{aligned} \tag{4.14}$$

Thus the approximation of $Z_t|Y_{1:t-1}=y_{1:t-1}$ is

$$\tilde{Z}_t|Y_{1:t-1}=y_{1:t-1} \sim \mathcal{N}(\tilde{\mu}_{t|t-1}, \tilde{\mathbf{\Sigma}}_{t|t-1}). \tag{4.15}$$

4.2 Measurement Step

We linearize the function $\mathbf{h}_t$ using the first order Taylor expansion, Definition 1.2, at the point $\tilde{\mu}_{t|t-1}$, that is

$$\mathbf{h}_t(Y_t) \approx \mathbf{h}_t(\tilde{\mu}_{t|t-1}) + \mathbf{Dh}_t(\tilde{\mu}_{t|t-1})(\tilde{Z}_t - \tilde{\mu}_{t|t-1}). \tag{4.16}$$

Now let us denote the approximation of Y_t by $\tilde{Y}_t$, i.e.,

$$\tilde{Y}_t := \mathbf{h}_t(\mu_{t|t-1}) + \mathbf{Dh}_t(\mu_{t|t-1})(\tilde{Z}_t - \mu_{t|t-1}) + W_t. \tag{4.17}$$

In the Bayes rule for linear-Gaussian systems, Claim 1.4, we take X to have the distribution

$$X \sim \mathcal{N}(\tilde{\mu}_{t|t-1}, \tilde{\mathbf{\Sigma}}_{t|t-1}). \tag{4.18}$$

We take Y to be $\tilde{Y}_t$ and

$$\begin{aligned}\tilde{Y}_t &= \mathbf{h}_t(\mu_{t|t-1}) + \mathbf{Dh}_t(\mu_{t|t-1})(X - \mu_{t|t-1}) + W_t \\ &= \mathbf{Dh}_t(\mu_{t|t-1})X + (\mathbf{h}_t(\mu_{t|t-1}) - \mathbf{Dh}_t(\mu_{t|t-1})\mu_{t|t-1}) + W_t\end{aligned} \tag{4.19}$$

where $W_t \sim \mathcal{N}(0, \mathbf{R}_t)$ is independent of X.

Now according to the Bayes rule for the linear-Gaussian systems, that is Claim 1.4, the following applies:

$$p(z_t|\tilde{Y}_{1:t} = y_{1:t}) = \phi(z_t|\tilde{\mu}_t, \tilde{\mathbf{\Sigma}}_t) \tag{4.20}$$

where

$$\tilde{\mathbf{\Sigma}}_t^{-1} = \tilde{\mathbf{\Sigma}}_{t|t-1}^{-1} + \mathbf{Dh}_t(\tilde{\mu}_{t|t-1})^T \mathbf{R}_t^{-1} \mathbf{Dh}_t(\tilde{\mu}_{t|t-1}) \tag{4.21}$$

$$\tilde{\mu}_t = \tilde{\mathbf{\Sigma}}_t[\mathbf{Dh}_t(\tilde{\mu}_{t|t-1})^T \mathbf{R}_t^{-1}(\mathbf{Dh}_t(\tilde{\mu}_{t|t-1})y_t + R_t) + \tilde{\mathbf{\Sigma}}_{t|t-1}^{-1}\tilde{\mu}_{t|t-1}] \tag{4.22}$$

Using simple transformations similar to [3, p. 643] the expectation μ_t and the covariance matrix $\mathbf{\Sigma}_t$ can be written as

$$\tilde{\mu}_t = \tilde{\mu}_{t|t-1} + \mathbf{K}_t(y_t - \hat{y}_t) \tag{4.23}$$

$$\tilde{\mathbf{\Sigma}}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{Dh}_t(\tilde{\mu}_{t|t-1}))\tilde{\mathbf{\Sigma}}_{t|t-1} \tag{4.24}$$

where $\hat{y}_t$ is the predicted observation and

$$\begin{aligned}\hat{y}_t &:= \mathrm{E}[\tilde{Y}_t|Y_{1:t-1}{=}y_{1:t-1}] = \mathrm{E}[\mathbf{Dh}_t(\tilde{\mu}_{t|t-1})\tilde{Z}_t + \mathbf{h}_t(\tilde{\mu}_{t|t-1}) \\ &\quad - \mathbf{Dh}_t(\tilde{\mu}_{t|t-1})\tilde{\mu}_{t|t-1} + W_t|Y_{1:t-1}{=}y_{1:t-1}] = \mathbf{Dh}_t(\tilde{\mu}_{t|t-1})\mathrm{E}[\tilde{Z}_t|Y_{1:t-1}{=}y_{1:t-1}] \\ &\quad + \mathbf{h}_t(\tilde{\mu}_{t|t-1}) - \mathbf{Dh}_t(\tilde{\mu}_{t|t-1})\tilde{\mu}_{t|t-1} + \mathrm{E}[W_t|Y_{1:t-1}{=}y_{1:t-1}] \\ &\quad = \mathbf{Dh}_t(\tilde{\mu}_{t|t-1})\tilde{\mu}_{t|t-1} + \mathbf{h}_t(\tilde{\mu}_{t|t-1}) - \mathbf{Dh}_t(\tilde{\mu}_{t|t-1})\tilde{\mu}_{t|t-1} = \mathbf{h}_t(\tilde{\mu}_{t|t-1}).\end{aligned} \tag{4.25}$$

and where $\mathbf{K}_t$ is the Kalman gain matrix given by

$$\mathbf{K}_t = \tilde{\mathbf{\Sigma}}_{t|t-1}\mathbf{Dh}_t(\tilde{\mu}_{t|t-1})^T \mathbf{S}_t^{-1} \tag{4.26}$$

where

$$
\begin{aligned}
\mathbf{S}_t := \mathbf{Cov}[\tilde{Y}_t - \hat{y}_t | Y_{1:t-1} = y_{1:t-1}] &= \mathbf{Cov}[\mathbf{Dh}_t(\tilde{\mu}_{t|t-1})\tilde{Z}_t + \mathbf{h}_t(\tilde{\mu}_{t|t-1}) \\
&- \mathbf{Dh}_t(\tilde{\mu}_{t|t-1})\tilde{\mu}_{t|t-1} + W_t - \mathbf{h}_t(\tilde{\mu}_{t|t-1}) | Y_{1:t-1} = y_{1:t-1}] \\
&= \mathbf{Dh}_t(\mu_{t|\tilde{t}-1})\tilde{\mathbf{\Sigma}}_{t|t-1}\mathbf{Dh}_t(\tilde{\mu}_{t|t-1})^T + \mathbf{R}_t.
\end{aligned} \tag{4.27}
$$

4.3 Examples of Extended Kalman Filter Estimation

Example 4.1. Let us look at a state space model where $g_t(Z_{t-1}) = Z_{t-1}^3 - 0.5Z_{t-1} + 0.2$ and $h_t(Z_t) = Z_t$, i.e., the following state space model

$$Z_t = Z_{t-1}^3 - 0.5Z_{t-1} + 0.2 + V_t \tag{4.28}$$

$$Y_t = Z_t + W_t \tag{4.29}$$

where

$$V_t \sim \mathcal{N}(0, Q_t) \tag{4.30}$$

$$W_t \sim \mathcal{N}(0, R_t). \tag{4.31}$$

Now we take $Z_0 \sim \mathcal{N}(0, 0.1) =: \mathcal{N}(\tilde{\mu}_0, \tilde{\sigma_0^2})$ and use the extended Kalman filter to estimate the density functions $p(z_t | Y_{1:t} = y_{1:t}) \approx \phi(z_t | \tilde{\mu}_t, \tilde{\sigma}_t^2)$ for every $t = 1, 2, \ldots$.

Recursively for every $t = 1, \ldots$ we first calculate $\tilde{\mu}_{t|t-1}$ and $\tilde{\sigma}_{t|t-1}^2$ from the prediction step, more precisely from the formulas (4.12) and (4.13). For our case it is:

$$\tilde{\mu}_{t|t-1} = \tilde{\mu}_{t-1}^3 - 0.5\tilde{\mu}_{t-1} + 0.2 = g_t(\tilde{\mu}_{t-1}), \tag{4.32}$$

$$\tilde{\sigma}_{t|t-1}^2 = (3\tilde{\mu}_{t-1}^2 + 0.5)^2 \tilde{\sigma}_{t-1}^2 + Q_t. \tag{4.33}$$

Then we calculate $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ from the measurement step. As the measurement model is linear we can use the Kalman filter formulas (3.23) and (3.24), which for our case are:

$$\tilde{\mu}_t = \tilde{\mu}_{t|t-1} + K_t(y_t - \hat{y}_t) \tag{4.34}$$

$$\tilde{\sigma}_t^2 = (1 - K_t)\tilde{\sigma}_{t|t-1}^2, \tag{4.35}$$

where for our case according to (3.25) we have that $\hat{y}_t = \tilde{\mu}_{t|t-1}$ and according to (3.26) and (3.27) we have that $K_t = \tilde{\sigma}^2_{t|t-1} * 1/(\tilde{\sigma}^2_{t|t-1} + R_t)$.

We can write μ_t as

$$\tilde{\mu}_t = g_t(\tilde{\mu}_{t-1}) + K_t(y_t - g_t(\tilde{\mu}_{t-1})) = g_t(\tilde{\mu}_{t-1})(1 - K_t) + K_t y_t. \tag{4.36}$$

Now K_t can be written as:

$$K_t = \frac{(3\tilde{\mu}^2_{t-1} - 0.5)^2 \tilde{\sigma}^2_{t-1} + Q_t}{(3\tilde{\mu}^2_{t-1} - 0.5)^2 \tilde{\sigma}^2_{t-1} + Q_t + R_t}. \tag{4.37}$$

Then $K_t \in [0, 1]$. If $R_t \to \infty$, then $K_t \to 0$ and if $Q_t \to \infty$ or $R_t \to 0$, then $K_t \to 1$.

Now if $K_t \to 0$ the extended Kalman estimation $\tilde{\mu}_t \to g_t(\tilde{\mu}_{t-1})$. If $K_t \to 1$ then $\tilde{\mu}_t \to y_t$. That is the bigger R_t is, the more $\tilde{\mu}_t$ depends on $g_t(\tilde{\mu}_{t-1})$ and less on y_t. And the bigger Q_t is the more $\tilde{\mu}_t$ depends on the measurement y_t and less on $g_t(\tilde{\mu}_{t-1})$.

For the general case for the extended Kalman filter according to (4.23), (4.12) and (4.25) μ_t can be written as:

$$\tilde{\mu}_t = \mathbf{g}_t(\tilde{\mu}_{t-1}) + \mathbf{K}_t(y_t - \mathbf{h}_t(\mathbf{g}_t(\tilde{\mu}_{t-1}))). \tag{4.38}$$

Now if $\mathbf{h}_t$ is an identity function we can write μ_t as

$$\tilde{\mu}_t = (\mathbf{I} - \mathbf{K}_t)\mathbf{g}_t(\tilde{\mu}_{t-1}) + \mathbf{K}_t y_t \tag{4.39}$$

and the same relations hold as in the example. That is the bigger R_t is, the more $\tilde{\mu}_t$ depends on $\tilde{\mu}_{t-1}$ and the bigger Q_t is the more μ_t depends on the measurement y_t.

Now let us look at figures with the measurement values y_t, state values z_t and the expectation $\tilde{\mu}_t$ where $t = 0, \ldots, 100$ for different values of Q_t and R_t.

The value of Q_t and R_t affects the state and measurement function amplitude in the same way as in Example 3.1. That is the bigger Q_t is, the bigger is the amplitude of the state and the bigger R_t the bigger is the amplitude of the measurements.

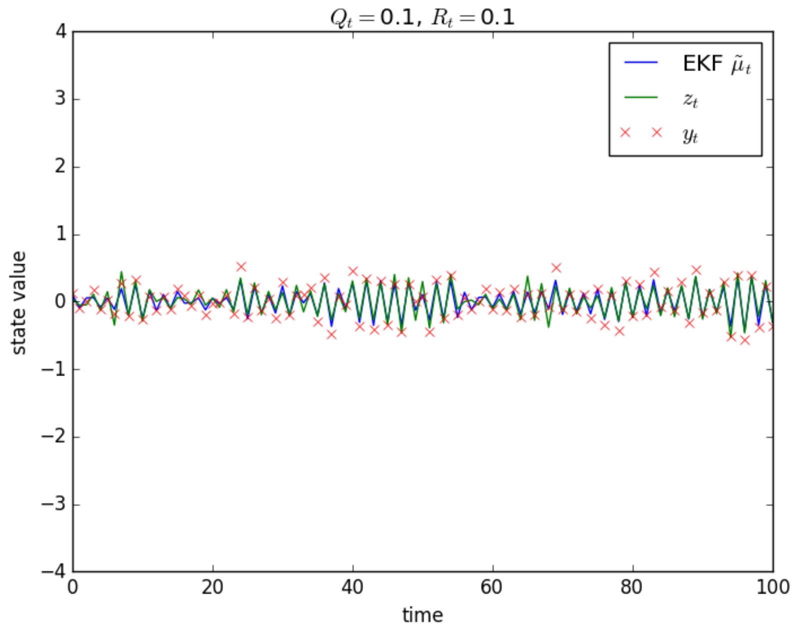

Figure 4.1: As both noise variances R_t and Q_t are small the expectation $\tilde{\mu}_t$ is close to the state value z_t.

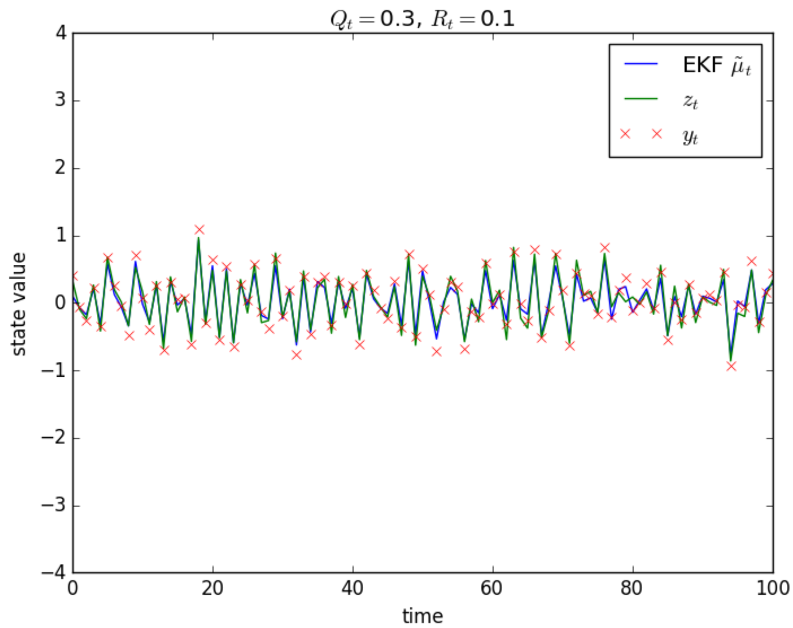

Figure 4.2: As the value of Q_t has increased from that in the previous figure the value of $\tilde{\mu}_t$ depends more on y_t then in figure 4.1.

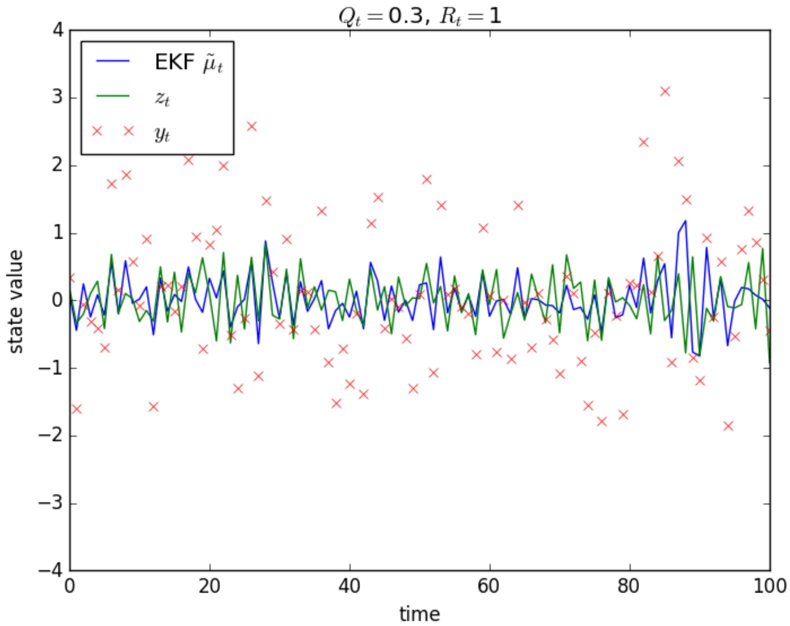

Figure 4.3: As the value of R_t has increased from that in the previous figure the value of $\tilde{\mu}_t$ depends more on $g_t(\tilde{\mu}_{t-1})$ and less on y_t then in figure 4.2.

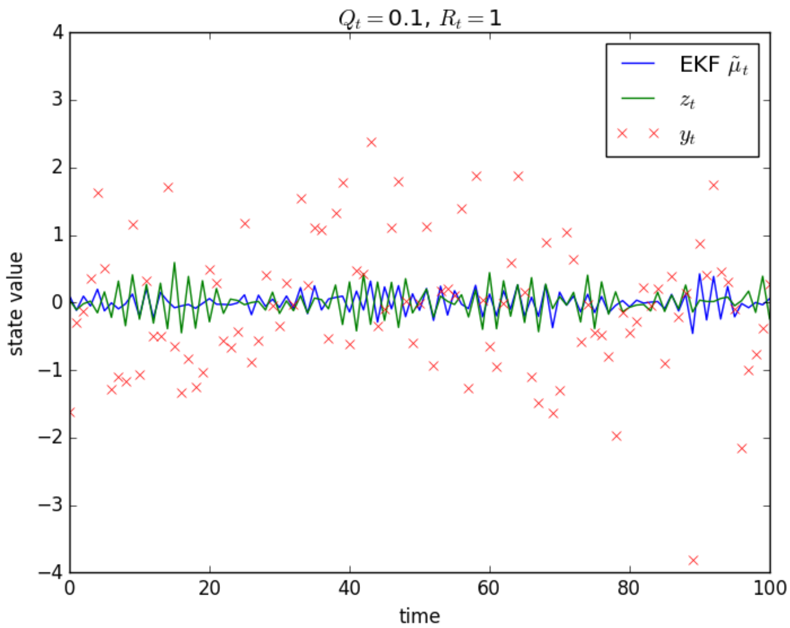

Figure 4.4: As the value of Q_t has decreased from that in the previous figure the value of $\tilde{\mu}_t$ depends more on y_t and less on $g_t(\tilde{\mu}_{t-1})$ then in figure 4.3.

Example 4.2. Let us look at a state space model where $g_t(Z_{t-1}) = Z_{t-1}^3 - 0.5Z_{t-1} + 0.2$ and $h_t(Z_t) = \exp(Z_t)$, i.e., the following state space model

$$Z_t = Z_{t-1}^3 - 0.5Z_{t-1} + 0.2 + V_t \tag{4.40}$$

$$Y_t = \exp(Z_t) + W_t \tag{4.41}$$

where

$$V_t \sim \mathcal{N}(0, Q_t) \tag{4.42}$$

$$W_t \sim \mathcal{N}(0, R_t). \tag{4.43}$$

Now we take $Z_0 \sim \mathcal{N}(0, 0.1) =: \mathcal{N}(\tilde{\mu}_0, \tilde{\sigma_0^2})$ and use the extended Kalman filter to estimate the density functions $p(z_t|Y_{1:t} = y_{1:t}) \approx \phi(z_t|\tilde{\mu}_t, \tilde{\sigma}_t^2)$ for every $t = 1, 2, \ldots$.

Recursively for every $t = 1, \ldots$ we first calculate $\tilde{\mu}_{t|t-1}$ and $\tilde{\sigma}_{t|t-1}^2$ from the prediction step, more precisely from the formulas (4.12) and (4.13). For our case it is:

$$\tilde{\mu}_{t|t-1} = \tilde{\mu}_{t-1}^3 - 0.5\tilde{\mu}_{t-1} + 0.2 = g_t(\tilde{\mu}_{t-1}), \tag{4.44}$$

$$\tilde{\sigma}_{t|t-1}^2 = (3\tilde{\mu}_{t-1}^2 - 0.5)^2 \tilde{\sigma}_{t-1}^2 + Q_t. \tag{4.45}$$

Then we calculate $\tilde{\mu}_t$ and $\tilde{\sigma}_t^2$ from (4.23) and (4.24), which for our case are:

$$\tilde{\mu}_t = \tilde{\mu}_{t|t-1} + K_t(y_t - \hat{y}_t) \tag{4.46}$$

$$\tilde{\sigma}_t^2 = (1 - K_t)\tilde{\sigma}_{t|t-1}^2, \tag{4.47}$$

where for our case according to (4.25) we have that $\hat{y}_t = \exp(\tilde{\mu}_{t|t-1})$. According to (3.26) and (3.27) we have that

$$K_t = \frac{\tilde{\sigma}_{t|t-1}^2 \exp(\tilde{\mu}_{t|t-1})}{\exp(2\tilde{\mu}_{t|t-1})\tilde{\sigma}_{t|t-1}^2 + R_t}. \tag{4.48}$$

We can write μ_t as

$$\tilde{\mu}_t = \tilde{\mu}_{t-1}^3 - 0.5\tilde{\mu}_{t-1} + 0.2 + K_t(y_t - \exp(\tilde{\mu}_{t-1}^3 - 0.5\tilde{\mu}_{t-1} + 0.2)). \tag{4.49}$$

Now K_t can be written as:

$$K_t = \frac{((3\tilde{\mu}_{t-1}^2 - 0.5)^2\tilde{\sigma}_{t-1}^2 + Q_t)\exp(\tilde{\mu}_{t-1}^3 - 0.5\tilde{\mu}_{t-1} + 0.2)}{\exp(2\tilde{\mu}_{t-1}^3 - \tilde{\mu}_{t-1} + 0.4)((3\tilde{\mu}_{t-1}^2 - 0.5)^2\tilde{\sigma}_{t-1}^2 + Q_t) + R_t}. \tag{4.50}$$

Then $K_t \in [0, 1]$. If $R_t \to \infty$, then $K_t \to 0$ and if $Q_t \to \infty$ or $R_t \to 0$, then $K_t \to \exp(-\tilde{\mu}_{t-1}^3 + 0.5\tilde{\mu}_{t-1} - 0.2)$.

Now if $K_t \to 0$ the extended Kalman estimation $\tilde{\mu}_t \to g_t(\tilde{\mu}_{t-1})$. If $K_t \to \exp(-\tilde{\mu}_{t-1}^3 + 0.5\tilde{\mu}_{t-1} - 0.2)$ then

$$\tilde{\mu}_t \to \tilde{\mu}_{t-1}^3 - 0.5\tilde{\mu}_{t-1} - 0.8 + y_t \exp(-\tilde{\mu}_{t-1}^3 + 0.5\tilde{\mu}_{t-1} - 0.2). \tag{4.51}$$

That is the bigger R_t is, the more $\tilde{\mu}_t$ depends on $g_t(\tilde{\mu}_{t-1})$ and less on y_t.

Now let us look at figures with the measurement values y_t, state values z_t and the expectation $\tilde{\mu}_t$ where $t = 0, \dots, 100$ for different values of Q_t and R_t.

The value of Q_t and R_t affects the state and measurement function amplitude in the same way as in Example 3.1 and Example 4.1. That is the bigger Q_t is, the bigger is the amplitude of the state and the bigger R_t the bigger is the amplitude of the measurements.

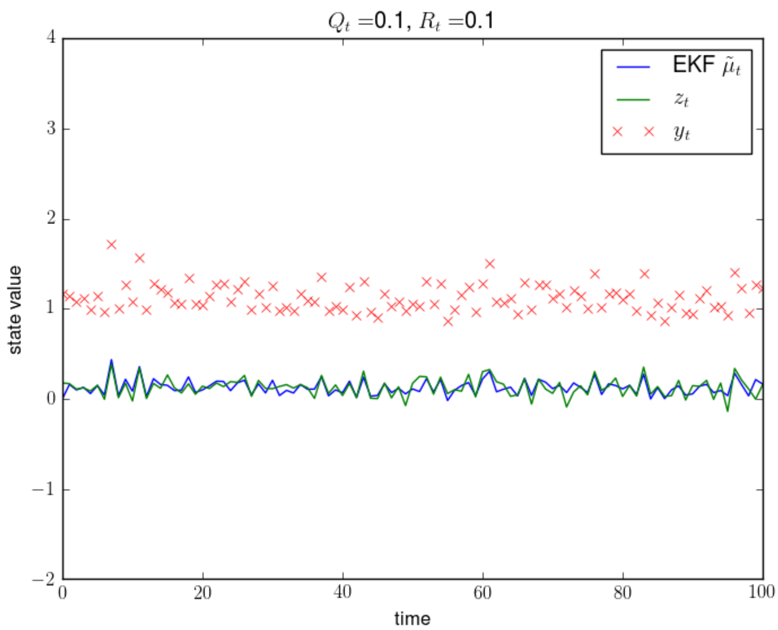

Figure 4.5: As R_t and Q_t are small the expectation $\tilde{\mu}_t$ is close to the state value z_t.

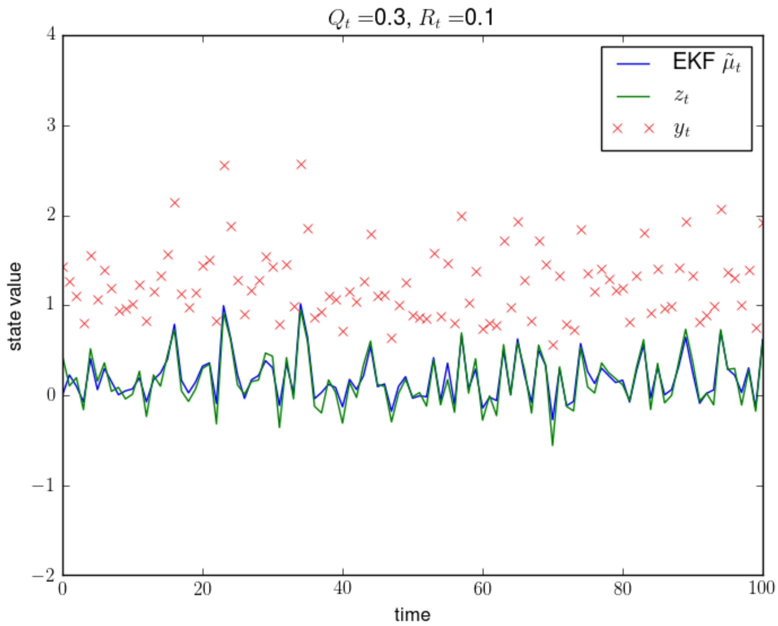

Figure 4.6: As the value of Q_t has increased from that in the previous figure the amplitude of the state function is bigger than in figure 4.5.

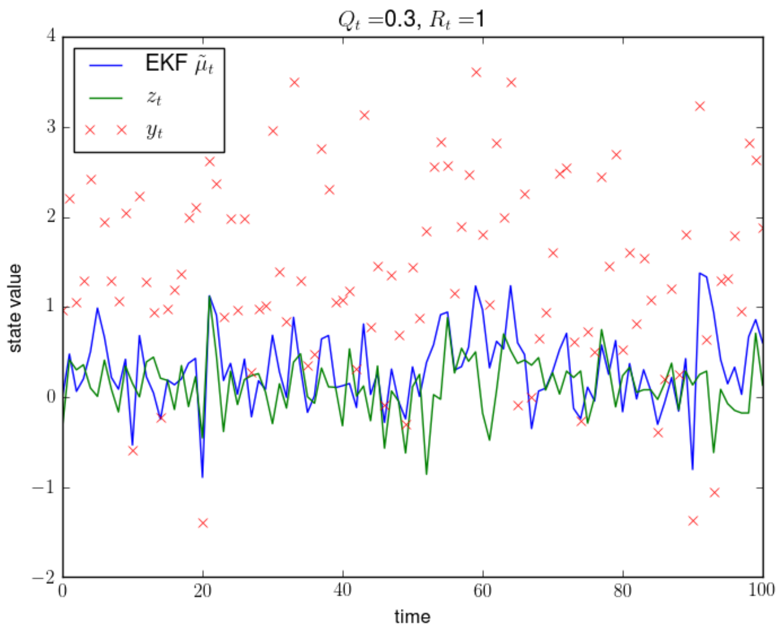

Figure 4.7: As the value of R_t has increased from that in the previous figure the amplitude of the measurements is bigger and the value of $\tilde{\mu}_t$ depends more on $g_t(\tilde{\mu}_{t-1})$ and less on y_t then in figure 4.6.

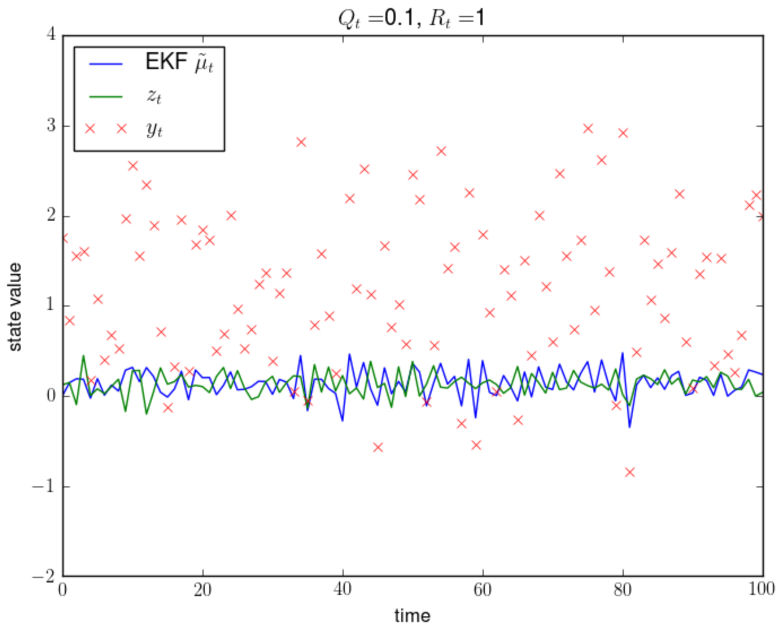

Figure 4.8: As the value of Q_t has decreased from that in the previous figure amplitude of the state function is smaller then in figure 4.7. The value of R_t is big so the value of $\tilde{\mu}_t$ depends more on $g_t(\tilde{\mu}_{t-1})$ and less on y_t.

4.4 Extended Kalman filter for Linear-Gaussian systems

Let us look at a state space model, where $\mathbf{g}_t$ and $\mathbf{h}_t$ are linear functions, that is

$$\mathbf{g}_t(Z_{t-1}) = \mathbf{A}_t Z_{t-1} \tag{4.52}$$

$$\mathbf{h}_t(Z_t) = \mathbf{C}_t Z_t \tag{4.53}$$

where $\mathbf{A}_t$ and $\mathbf{C}_t$ are matrices.

For linear $\mathbf{g}_t$ and $\mathbf{h}_t$ the matrix $\mathbf{Dg}_t(\mu_{t-1}) = \mathbf{A}_t$ and $\mathbf{Dh}_t(\tilde{\mu}_{t|t-1}) = \mathbf{C}_t$.

The prediction step now gives that

$$\tilde{Z}_t|Y_{1:t-1} = y_{1:t-1} \sim \mathcal{N}(\tilde{\mu}_{t|t-1}, \tilde{\mathbf{\Sigma}}_{t|t-1}) \tag{4.54}$$

where

$$\tilde{\mu}_{t|t-1} = \mathbf{g}_t(\mu_{t-1}) = \mathbf{A}_t \mu_{t-1} \tag{4.55}$$

$$\tilde{\mathbf{\Sigma}}_{t|t-1} = \mathbf{D}\mathbf{g}_t(\mu_{t-1})\mathbf{\Sigma}_{t-1}\mathbf{D}\mathbf{g}_t(\mu_{t-1})^T + \mathbf{Q}_t = \mathbf{A}_t \mathbf{\Sigma}_{t-1} \mathbf{A}_t^T + \mathbf{Q}_t. \tag{4.56}$$

The measurement step gives that

$$\tilde{Z}_t | \tilde{Y}_{1:t} = y_{1:t} \sim \mathcal{N}(\tilde{\mu}_t, \tilde{\mathbf{\Sigma}}_t) \tag{4.57}$$

where

$$\tilde{\mu}_t = \tilde{\mu}_{t|t-1} + \mathbf{K}_t(y_t - \mathbf{h}_t(\mu_{t|t-1})) = \tilde{\mu}_{t|t-1} + \mathbf{K}_t(y_t - \mathbf{C}_t \mu_{t|t-1}) \tag{4.58}$$

$$\tilde{\mathbf{\Sigma}}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{D}\mathbf{h}_t(\tilde{\mu}_{t|t-1}))\tilde{\mathbf{\Sigma}}_{t|t-1} = (\mathbf{I} - \mathbf{K}_t \mathbf{C}_t)\tilde{\mathbf{\Sigma}}_{t|t-1} \tag{4.59}$$

where the Kalman gain matrix is given by

$$\begin{aligned} \mathbf{K}_t &= \tilde{\mathbf{\Sigma}}_{t|t-1} \mathbf{D}\mathbf{h}_t(\tilde{\mu}_{t|t-1})^T (\mathbf{D}(\mu_{t|t-1})\tilde{\mathbf{\Sigma}}_{t|t-1}\mathbf{D}(\tilde{\mu}_{t|t-1})^T + \mathbf{R}_t)^{-1} \\ &= \tilde{\mathbf{\Sigma}}_{t|t-1} \mathbf{C}_t^T (\mathbf{C}_t \tilde{\mathbf{\Sigma}}_{t|t-1} \mathbf{C}_t^T + \mathbf{R}_t)^{-1}. \end{aligned} \tag{4.60}$$

Thus for a linear-Gaussian state space model the extended Kalman filter gives the same results as the Kalman filter.

References

[1] A. Gut. *Probability: A Graduate Course*. Springer, 2005.

[2] R. E. Kalman. “A New Approach to Linear Filtering and Prediction Problems”. In: *Transaction of the ASME–Journal of Basic Engineering* (Mar. 1960), pp. 35–45.

[3] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

5 Appendices

5.1 Appendix 1. Python Code for Kalman Filter Example

```
import numpy as np
import matplotlib.pylab as plt

plt.close('all')
#the expectation and variance of Z_0
mu0=0
sigmaruut0=.1
#the expectations of he noises V and W
muv=0
muw=0
#the length of time period
n=100
a=1
c=1

def g(z):
    return a*z

def h(z):
    return c*z

def KF(Qt,Rt):
    mu = np.zeros((n+1, n+1))
    mu[0, 0]=mu0

    sigmaruut = np.zeros((n+1, n+1))
    sigmaruut[0, 0]=sigmaruut0

    v=np.random.normal(muv, Qt, n+1)
    w=np.random.normal(muw, Rt, n+1)

    z = np.zeros(n+1)
    z[0]=z0+v[0]
    for k in np.arange(1,n+1):
        z[k]=g(z[k-1])+v[k]

    y = np.zeros(n+1)
    for j in np.arange(n+1):
        y[j]=h(z[j])+w[j]

    #KF algorithm
    for i in np.arange(1,n+1):
        #prediction step
        mu[i,i-1]=a*mu[i-1,i-1]
```

```
        sigmaruut[i,i-1]=a**2*sigmaruut[i-1,i-1]+Qt
        #measurement step
        Kt=sigmaruut[i,i-1]*c*(1/(c**2*sigmaruut[i,i-1]+Rt))
        mu[i,i]=mu[i,i-1]+Kt*(y[i]-c*mu[i,i-1])
        sigmaruut[i,i]=(1-Kt*c)*sigmaruut[i,i-1]
    return z, y, mu.diagonal()

def jon(z,y,mu,Qt,Rt):
    plt.plot(t, mu, label=r"KF $\mu_t$")
    plt.plot(t, z, label=r"$z_t$")
    plt.xlabel('time')
    plt.ylabel('state value')
    plt.title(r"$Q_t=$"+str(Qt)+r", $R_t=$"+str(Rt))
    plt.plot(t, y, 'x', label=r"$y_t$")
    plt.ylim([-6,6])
    plt.legend(loc='best')

z0= np.random.normal(mu0, sigmaruut0)

Qt1=.1
Rt1=.1
z1,y1,mu1=KF(Qt1,Rt1)
Qt2=.4
Rt2=.1
z2,y2,mu2=KF(Qt2,Rt2)
Qt3=.4
Rt3=1
z3,y3,mu3=KF(Qt3,Rt3)
Qt4=.1
Rt4=1
z4,y4,mu4=KF(Qt4,Rt4)

t = np.linspace(0, n, n+1)
plt.rc('text', usetex=True)

o=plt.figure(0)
jon(z1,y1,mu1,Qt1,Rt1)
o.show()

p=plt.figure(1)
jon(z2,y2,mu2,Qt2,Rt2)
p.show()

q=plt.figure(2)
jon(z3,y3,mu3,Qt3,Rt3)
q.show()

r=plt.figure(3)
jon(z4,y4,mu4,Qt4,Rt4)
r.show()
```

5.2 Appendix 2. Python Code for Extended Kalman Filter Examples

```
import numpy as np
import matplotlib.pylab as plt
from scipy.misc import derivative

plt.close('all')
#the expectation and variance of Z_0
mu0=0
sigmaruut0=.1
#the expectations of he noises V and W
muv=0
muw=0
#the length of time period
n=100

def g(z):
    return z**3-z

def h(z):
    return z

def EKF(Qt,Rt):
    mu = np.zeros((n+1, n+1))
    mu[0, 0]=mu0

    sigmaruut = np.zeros((n+1, n+1))
    sigmaruut[0, 0]=sigmaruut0

    v=np.random.normal(muv, Qt, n+1) #vead on normaaljaotusest
    w=np.random.normal(muw, Rt, n+1)

    z = np.zeros(n+1)
    z[0]=z0+v[0]
    for k in np.arange(1,n+1):
        z[k]=g(z[k-1])+v[k]

    y = np.zeros(n+1)
    for j in np.arange(n+1):
        y[j]=h(z[j])+w[j]

    #EKF algorithm
    for i in np.arange(1,n+1):
        #prediction step
        mu[i,i-1]=g(mu[i-1,i-1])
        sigmaruut[i,i-1]=(derivative(g,mu[i-1,i-1])**2*sigmaruut[i-1,i-1]+Qt)
        #measurement step
        Kt=sigmaruut[i,i-1]*derivative(h,mu[i,i-1])*(1/(derivative(h,mu[i,i-1])**2*sigmaruut[i,i-1]+
        mu[i,i]=mu[i,i-1]+Kt*(y[i]-h(mu[i,i-1]))
```

```
        sigmaruut[i,i]=(1-Kt*derivative(h,mu[i,i-1]))*sigmaruut[i,i-1]
    return z,y,mu.diagonal()

def jon(z,y,mu,Qt,Rt):
    plt.plot(t, mu, label=r"EKF $\tilde{\mu}_t$")
    plt.plot(t, z, label=r"$z_t$")
    plt.xlabel('time')
    plt.ylabel('state value')
    plt.title(r"$Q_t=$"+str(Qt)+r", $R_t=$"+str(Rt))
    plt.plot(t, y, 'x', label=r"$y_t$")
    plt.ylim([-3,3])
    plt.legend(loc='best')

z0=np.random.normal(mu0, sigmaruut0)

Qt1=.1
Rt1=.1
z1,y1,mu1=EKF(Qt1,Rt1)
Qt2=.3
Rt2=.1
z2,y2,mu2=EKF(Qt2,Rt2)
Qt3=.3
Rt3=1
z3,y3,mu3=EKF(Qt3,Rt3)
Qt4=.1
Rt4=1
z4,y4,mu4=EKF(Qt4,Rt4)

t = np.linspace(0, n, n+1)
plt.rc('text', usetex=True)

o=plt.figure(0)
jon(z1,y1,mu1,Qt1,Rt1)
o.show()

p=plt.figure(1)
jon(z2,y2,mu2,Qt2,Rt2)
p.show()

q=plt.figure(2)
jon(z3,y3,mu3,Qt3,Rt3)
q.show()

r=plt.figure(3)
jon(z4,y4,mu4,Qt4,Rt4)
r.show()
```

Non-exclusive licence to reproduce thesis and make thesis public

I, Johanna Adele Järvsoo (date of birth: 26.09.1995),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

   1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

   1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

   ”Kalman Filter and Extended Kalman Filter”, supervised by Jüri Lember and Andris Slavinskis,

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 12.06.2017