Adaptive Nonlinear Kalman Filters for Non-stationary Observation Errors

Akio NAKABAYASHI

Doctor of Philosophy

Department of Statistical Science

School of Multidisciplinary Sciences

SOKENDAI (The Graduate University for

Advanced Studies)

SOKENDAI
(THE GRADUATE UNIVERSITY FOR ADVANCED STUDIES)

DOCTORAL THESIS

Adaptive Nonlinear Kalman Filters for Non-stationary Observation Errors

Author:
Akio NAKABAYASHI

Supervisor:
Associate Prof. Genta UENO

Department of Statistical Science,
School of Multidisciplinary Sciences

Abstract

It is natural that the observation error model is hard to develop due to unknowns in both stationary and non-stationary characteristics. If there is discrepancy between an assumed model and the actual behavior of the observation error, the accuracy of state estimation in filtering is severely decreased. Therefore, we have studied adaptive filtering methods for the observation errors. We have addressed the theme from two points of views: one is the case for the long-term temporal change of characteristics of the observation error, and another is the case for the impulsive change of the characteristics. In this thesis, we presents two methods for these cases. The key idea of the methods is to use an adaptive observation error model, which includes time-indexed parameters, and estimate the parameters in filtering at each time step. The enabler is the variational Bayes method which approximates the joint filtered distribution of the state vector and the time-index parameters with another tractable distribution. We have examined the methods in several numerical experiments and found that the methods can improve the accuracy of state estimation as a result of appropriately handling the observation errors. The presented methods are expected to contribute to increase the value of the applications of filtering.

Acknowledgements

I would like to express the deepest appreciation to my supervisor, Associate Prof. Genta Ueno. Completion of this work is greatly due to his proper and sincere guidance.

I would like to show appreciation to my sub supervisor, Associate Prof. Ryo Yoshida. His feedback for research progress encouraged me.

I would like to show appreciation to the examiners, Prof. Yoshinori Kawasaki, Associate Prof. Daichi Mochihashi, Associate Prof. Shinya Nakano, and Dr. Yousuke Fujii. Thanks to their comments and advices, this thesis was greatly improved.

I would like to offer my special thanks to colleagues in my company. They gave me motivations to continue the research by sharing me the practical issues.

Finally, I greatly appreciate my family. My parents cared about my research and supported me. My two sons gave me the energy to study on a day I got tired from work. And my wife let me concentrate on my research at the expense of her time.

Contents

List of Figures

List of Tables

List of Abbreviations

KF	Kalman Filter
EKF	Extended Kalman Filter
UKF	Unscented Kalman Filter
EnKF	Ensemble Kalman Filter
PF	Particle Filter
VB method	Variational Bayese method
SSM	State Space Model

List of Symbols

<u>valiables</u>

t	discrete time step	≥ 0
T	total time steps	$\gg 0$
d	dimensionality of a state vector	> 0
m	dimensionality of a observation vector	> 0
$\mathbf{x}_t$	state vector	$\in \mathbb{R}^d$
$\mathbf{F}_t$	time transition matrix	$\mathbb{R}^{d \times d}$
$\mathbf{v}_t$	system noise	$\mathbb{R}^d$
$\mathbf{Q}_t$	model error covariance matrix	$\mathbb{R}^{d \times d}$
$\mathbf{y}_t$	observation vector	$\in \mathbb{R}^m$
$\mathbf{H}_t$	observation matrix	$\mathbb{R}^{m \times d}$
$\mathbf{w}_t$	observation error	$\mathbb{R}^d$
$\mathbf{R}_t$	observation error covariance matrix	$\in \mathbb{R}^{m \times m}$
$\mathbf{I}$	identity matrix	

<u>functions or functionals</u>

$\mathbf{h}_t$	nonlinear observation function	$\mathbb{R}^d \to \mathbb{R}^m$
$\mathbf{f}_t$	nonlinear time-evolution function	$\mathbb{R}^d \to \mathbb{R}^d$
$\mathcal{F}$	variational lower bound	
$\mathcal{H}$	entropy of probability distribution	

<u>probability (density) distributions</u>

$\mathcal{N}$	Gaussian (normal) distribution
$\mathcal{W}$	Wishart distribution
$\mathcal{IW}$	inverse Wishart distribution
$Bern$	Bernoulli distribution
$\mathcal{G}$	gamma distribution
$\mathcal{IG}$	inverse gamma distribution

Chapter 1

Introduction

1.1 Motivation

We concern a simulation technology of plants, which are production systems in process industry such as petrochemical, paper, and steel industries.

Simulation of a plant with a physical model receives much attentions of manufacturers. The simulation has a potential for valuable applications in production such as what-if analysis for redesign of control system or operational condition, or inner-inspection for diagnosis. Although a physical model is typically used in design phase of a plant, the applicable range is basically limited to a certain operational condition. The accuracy of simulation will be decreased in production phase because the operational condition is changed as time passes [1]. To improve the accuracy of simulation, attention to be noted is data collected in production. In a plant, a number of measurement values from sensors or analyzers are consolidated for automated control. Since the measurement system is carefully designed and implemented to control the process, the data is expected to have much information for refining simulation.

The integration of simulation and data has greatly advanced in the filed of earth science, and it is referred to as the data assimilation. The key technology is a large-scale and nonlinear filtering method, where the number of dimensionality of system state can exceed millions. Many applications have been successfully reported in the ocean, atmosphere, magnetosphere, and so on. In data assimilation, a deterministic physical simulation model is extended to a state-space model (SSM) by introducing probabilistic models typically represented by system noise and observation error discussed in the following section. Probabilistic models compensate mismatch between simulation and observed data. Even though the original model is imperfect, it is possible to increase the accuracy of time-evolution of the system state by incorporating information in the observed data. Furthermore, it is also possible to estimate parameters of physical models, which determine behavior of concerned system, for prediction in what-if analysis.

Although techniques of data assimilation is expected to be effective for plants, there is an issue to import the technology. Contrary to the system of earth science, that is, the nature itself, statistical property of a plant is artificially and frequently changed [2]. For examples, there are change of feed material, malfunction or maintenance of equipments. Change of operational condition due to production adjustment is also common. From a statistical perspective, the data to be modeled as an SSM is non-stationary. Even if an SSM is appropriately designed and developed at a certain time point, the accuracy of simulation will decrease after such changes.

Therefore, an SSM for a plant needs to be maintained continuously. Of course, the necessity of continuous maintenance is usually unacceptable in terms of cost. In addition, although practitioners who consider applications with simulation are

assumed to be great chemical engineering and/or process control experts, they may not be statisticians. It is required to provide a convenient way for developing SSMs.

To overcome the issue, we combine data assimilation and machine learning techniques. The machine learning has prominently evolved since 1990s along with the evolution of computational capability and data storage. Even If, in the case where an analytical solution does not exist, the machine learning can appropriately estimate a complicated probabilistic model, for example, mixture models, hierarchical models, deep neural networks, and so on. By adopting the machine learning techniques, the above issue is expected to be resolved without statistical expertise.

On the basis of this concept, in this thesis, we develop two nonlinear filtering methods. The two methods adaptively model unknown and non-stationary behavior without pre-design by using a machine learning technique. The outcome enables an accurate simulation without spending effort regarding the design of additional probabilisitic models. In other words, it can focus practitioners on developing the applications to make the production efficient, and safe. This will lead innovations in production, and it is the final aim of our research.

1.2   Background: Kalman filters

Filtering for estimating an unobserved system state from the observations using a state-space model (SSM) is one of the most important data analytical techniques in engineering. An SSM consists of a system model that defines time-evolution of the system state and an observation model that defines how the state is observed. With the two models, the technique predicts a next observation and correct the estimate of the state based on the prediction error of an obtained observation, which is referred to as innovation, alternately. There are a number of applications of filtering: for examples, position tracking for mobile navigation, state estimation for feedback control, system identification for simulation or optimization, and so on.

The Kalman filter (KF) [3] is one of the fundamental algorithms for a linear Gaussian SSM. The original KF have been proposed as the minimum variance estimator for the system state using the orthogonal projection. It is also known to a sequential Bayesian estimator of the posterior distribution, which is referred to as the filtered distribution in this context. The algorithm of the KF can be expressed with basic linear algebra concept and is easy to implement. This is the reason why the KF is widely applied in actual applications.

To deal with nonlinearity often appearing in actual applications, a number of variants of the KF for nonlinear SSMs have been proposed; and the extended Kalman filter (EKF) was firstly considered. The EKF linearizes a nonlinear SSM and adopts the KF algorithm at each time step. For the cases where the linearized form can be obtained analytically, the EKF is the very promising solution. However, in reality, since it is hard to use the analytical form, the algorithm of the EKF needs to depend on the numerical differentiation. As a result of including the numerical differentiation, the EKF tends to require high computational cost and can be unstable.

Recently, derivative-free Kalman filters that can deal with nonlinear SSMs have received considerable attentions of practitioners. The unscented Kalman filter (UKF) [4] uses the sigma points, which are temporarily and deterministically drawn samples from a distribution to be transformed as shown in Fig. 1.1 (b). The UKF performs nonlinear calculus for each of the sigma points and approximates the nonlinearly transformed distribution with the points. As detailed in Chapter 2, since the UKF requires sigma points proportional to the dimensionality of the system state,

(a) KF   (b) UKF   (c) EnKF
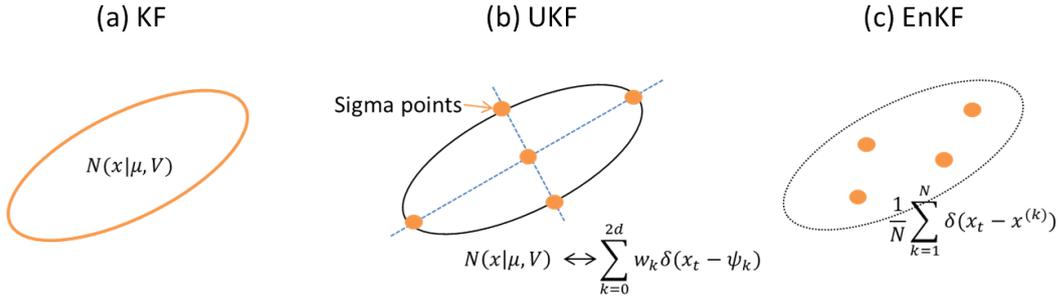
FIGURE 1.1: Distributions for the state vector of KF, UKF, and EnKF

it is computationally expensive and seems to be suitable for up to moderate-scale problems. The ensemble Kalman filter (EnKF) [5], [6] is a filtering method originated in geophysics, which handles a very large problem with, sometimes, more than a million variables in the state. The EnKF avoids the linearization by using an ensemble-based distribution, which consists of a finite number of samples as shown in Fig. 1.1 (c). The algorithm of the EnKF is constructed so that the moments of the filtered distribution up to the second order are equal to those of the KF in the limit as the number of samples approaches to infinity. Although the EnKF only deals with a linear Gaussian observation model, it is successfully applied to actual large-scale problems. The particle filter (PF) is also an ensemble-based filter. Since the estimation of the filtered distribution is based on the sequential Monte Carlo (SMC) method, it can even deal with non-Gaussian SSMs. Although the PF seems to be an universal filtering technique, it is well known that the PF has the problem of curse of dimensionality [7]. For a small problem relative to the available computation resource, the PF could be a promising solution.

In this thesis, we address problems regarding such derivative-free nonlinear Kalman filters, especially, the UKF and the EnKF. The formal definition of these filtering methods are given in Chapter 2. In the following section, we present our problem statement.

1.3   Problem statement: observation errors

Let us define a filtering problem with nonlinear SSMs. At this moment, we use a nonlinear SSM as follows:

$$\mathbf{x}_t | \mathbf{x}_{t-1} \sim \mathbf{f}_t(\mathbf{x}_{t-1}) + \mathbf{v}_t \tag{1.1}$$

$$\mathbf{y}_t | \mathbf{x}_t \sim \mathbf{h}_t(\mathbf{x}_t) + \mathbf{w}_t \tag{1.2}$$

where t is the discrete time index, $\mathbf{x}_t \in \mathbb{R}^d$ is the state vector that represents the system state, $\mathbf{y}_t \in \mathbb{R}^m$ is the observation vector that is observation of the state, $\mathbf{f}_t : \mathbb{R}^d \to \mathbb{R}^d$ and $\mathbf{h}_t : \mathbb{R}^d \to \mathbb{R}^m$ are referred to as the transition function and the observation function, respectively, $\mathbf{v}_t \in \mathbb{R}^d$ is a random error that represents the uncertainty in the transition, which is referred to as the model error, $\mathbf{w}_t \in \mathbb{R}^m$ is also a random error that represents the uncertainty in the observation, which is referred to as the observation error. Expressions (1.1) and (1.2) are referred to as the system model and the observation model, respectively. From a Bayesian perspective, the filtering problem is to recursively estimate the tth posterior, that is, filtered distribution for the system state $p(\mathbf{x}_t | \mathbf{y}_{1:t})$ from the $(t-1)$th filtered distribution $p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1})$

and the tth observation $\mathbf{y}_t$ using the given SSM (1.1) and (1.2), where the notation $\mathbf{y}_{1:t}$ denotes a set of observations $\{\mathbf{y}_1, \cdots, \mathbf{y}_t\}$.

Of course, although appropriate design of each building block of Eqs. (1.1) and (1.2) is crucial for filtering, our research of interest is the observation error $\mathbf{w}_t$. Design of $\mathbf{w}_t$, that is, what kind of a probabilistic model is assumed for $\mathbf{w}_t$ is directly connected to the performance of filtering. For a model with small variance, that is, assuming low uncertainty for observation, state estimation will be performed to mimic observations as possible. In this case, if observations are actually uncertain, state estimation will be biased as a result of over-fitting to observations. While, for a model with large variance, that is, assuming high uncertainty for observation, the observations might be neglected in the state estimation. If observations are actually certain, under-fitting to observations will happen.

However, since there are essential unknowns in characteristics of $\mathbf{w}_t$, the design is not trivial task. To discuss in depth, we decompose $\mathbf{w}_t$ on the basis of the concept of representativeness error [8]. The observation error can be regarded as consisting of two kinds of errors: one is referred to as a measurement error $\mathbf{w}_{t,meas.}$, and another is referred to as a representativeness error $\mathbf{w}_{t,rep.}$. That is,

$$\mathbf{w}_t = \mathbf{w}_{t,meas.} + \mathbf{w}_{t,rep.}. \tag{1.3}$$

The error $\mathbf{w}_{t,meas.}$ comes from an uncertainty of measurement instruments. The sources are, for example, precision of calibration curve, noises in transmission, quantization error of AD converter, and so on. To know the statistics, a specification including the precision provided by manufacturers of measurement instruments is helpful. However, there are cases where self-build or experimental instruments, whose specifications are inaccurate, are installed.

While, $\mathbf{w}_{t,rep.}$ originates from deviation between an assumed SSM and an ideal SSM, and thus the characteristics is inherently unknown. It is because a system model and an observation function are usually developed from partial but best-effort understanding of the target system.

So far, although we have discussed the difficulties in designing a model for $\mathbf{w}_t$ from the stationary characteristics, there are mainly two other difficulties related to the non-stationary characteristics of $\mathbf{w}_t$: one is the long-term temporal change, and another is the impulsive change.

For the long-term temporal change of characteristics, in the case of $\mathbf{w}_{t,meas.}$, the sources could be deterioration or maintenance of measurement instruments. For example, the deterioration of a measurement instruments increases the error scale in general. In the case of $\mathbf{w}_{t,rep.}$, the source could be increase of the discrepancy of the assumed SSM on a particular time or spatial region. For example, if an empirical model identified with experimental data is incorporated in a system model, the precision might be deteriorated in extrapolation.

For the impulsive change of characteristics, the problem of outliers is known. Outliers are unexpectedly gross observation error, which are unpredictably present. Appearance of outliers is very common in actual systems. There are variety of the sources: e.g., a large disturbance to the target system, temporal malfunction of measurement instruments, human error in data manipulation, and so on. Especially, for a Gaussian observation error model, which is often assumed in actual applications, it is well known that the outliers severely decrease the accuracy of the state estimation [9].

It is natural that the observation error model is hard to develop due to unknowns in both stationary and non-stationary characteristics. As already discussed, if there

is discrepancy between the assumed model and the actual behavior of $\mathbf{w}_t$, the accuracy of state estimation is decreased. Therefore, we have focused our research on developing adaptive filtering methods for the observation errors. By estimating the error characteristics appropriately, the accuracy of state estimation can be improved. We have addressed the theme from two points of views: one is the case for the long-term temporal change of the characteristics, and another is the case for the impulsive change, that is, outliers. The research is expected to contribute to obtain better estimates of the system state in filtering and improve the value of the applications.

1.4 Literature reviews and our specific research themes

In this section, we review the former studies related with our goal. On the basis of the review, we clarify specific research themes which are to be addressed.

1.4.1 Kalman filters for time-variant observation error statistics

To deal with the long-term temporal change in the characteristics of $\mathbf{w}_t$, a basic idea is to estimate a state vector and parameters that specify the statistics of observation errors simultaneously. Let us assume that the observation error is modeled as

$$\mathbf{w}_t \sim p(\mathbf{w}_t|\theta_t), \tag{1.4}$$

where θ_t is a vector that represents the parameters. With this expression, the filtering problem to estimate $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is redefined as a problem to estimate $p(\mathbf{x}_t, \theta_t|\mathbf{y}_{1:t})$.

A straightforward idea is to use an augmented SSM, which is also referred to as a self-organizing SSM, as discussed in [6], [10], [11], by incorporating θ_t into $\mathbf{x}_t$ as $\tilde{\mathbf{x}}_t = [\mathbf{x}_t^{\mathrm{T}} \; \theta_t^{\mathrm{T}}]^{\mathrm{T}}$. The approach is common in the UKF, e.g. [12], or the PF, e.g. [11]. However, this approach does not work for the EnKF as discussed in [13]. From a Bayesian perspective, this is because a distribution for $\tilde{\mathbf{x}}_t$ is no longer conjugate to a Gaussian observation model assumed in the EnKF.

Moreover, there is another problem using an augmented SSM within the UKF or the PF. For the correlated observation error, the covariance parameters of $\mathbf{w}_t$ are required to be estimated. If incorporating such θ_t into $\tilde{\mathbf{x}}_t$, the dimensionality could be very large because the size is increased proportional to the square of the dimensionality of $\mathbf{w}_t$. As already mentioned, the computational cost of the PF is proportional to the exponential of the dimensionality of $\tilde{\mathbf{x}}_t$. That of the UKF is proportional to the dimensionality of $\tilde{\mathbf{x}}_t$. In nonlinear SSMs, calculation of a transition functions is often very time-consuming due to, for example, solving the numerical integration for differential equations that defines the time-evolution of $\mathbf{x}_t$. As a result, the computational cost of the UKF or the PF could be massive. Therefore, we can say that the state augmentation is only available for relatively small problems.

Recently, methods for separately estimating the state vector $\mathbf{x}_t$ and θ_t for the variances or covariances of Gaussian distributed $\mathbf{w}_t$, have been reported and seem to be promising [13]–[16]. That is, $\mathbf{w}_t$ is assumed to follow the Gaussian given by,

$$\mathbf{w}_t \sim \mathcal{N}(\mathbf{w}_t|\mathbf{0}, \mathbf{R}(\theta_t)), \tag{1.5}$$

where $\mathbf{R}$ denotes the observation error covariance matrix parameterized with θ_t. The methods use a separated parametric distribution for θ_t in order to perform sequential Bayesian estimation. Since the dimensionality of $\mathbf{x}_t$ does not increase, computationally efficient algorithms could be constructed. Moreover, most of the methods

assume a dynamics for θ_t in addition to that of $\mathbf{x}_t$. Due to the dynamics for θ_t, it is possible to estimate time-variant θ_t successfully.

However, there are shortcomings in those studies. First, there is no discussion for designing the dynamics for θ_t; practitioners have to determine the optimal value by trial and error. Second, the long-term stability of filtering is not discussed. The methods [14]–[16] adopt an approximation method for estimating $\mathbf{x}_t$ and θ_t, separately. The effect of approximation error in long-term filtering has not been discussed: for examples, how much the approximation error is accumulated, or whether the filtering is diverged by the error.

We have presented a study for addressing the shortcomings with a new method by extending the EnKF. In the study, we have proposed a maximum likelihood method for specifying the dynamics for θ_t. In addition, we experimentally demonstrate the stability of our method in a long-run filtering when there are unmodeled disturbances. The study is detailed in Chapter 3.

1.4.2   Kalman filters for outliers

The problem of outliers in filtering is over-fitting to the gross observation error as a result of using an error model with a small variance. To overcome this problem, a basic idea is to use a heavy-tailed error model. From this aspect, the former studies can be classified into two approaches: one is a static approach and another is a dynamic approach.

The static approach uses an inherently heavy-tailed distribution as an observation error model, such as the t-distribution[17] or the Cauchy distribution [18], to avoid over-fitting to outliers by reducing the sensitivity for gross errors. Kalman filters based on robust estimation, as discussed in e.g., [19]–[21], that replaces a quadratic loss function for innovations with ad-hoc loss functions to reduce the sensitivity for gross errors, can be also regarded as one of the static approaches. Although the approach can be applied with a reasonable computational cost, it is required to tune parameters for each filtering problem. In addition, application of a heavy-tailed model to all observations might cause under-fitting because the heavy tail also reduces the sensitivity to observations with regular errors.

The dynamic approach uses an error model with time-indexed parameters and estimates the parameters at each time step. By adopting the dynamic approach, only when outliers occur, over-fitting to outliers is expected to be avoided, for example, by enlarging the variance parameter temporarily. There are methods that estimate an observation error covariance matrix of a Gaussian error model [14], [22] by conducting the Bayesian estimation at each time step. Although effective in avoiding over-fitting to outliers, the methods tend to be fluctuated even for outlier-free observations because the prior is non-informative and easily causes over-fitting. This leads a deterioration in state estimation because of under- or over-fitting to observations.

Therefore, as one of the dynamic approaches, we have proposed a new method for dealing with outliers in nonlinear filtering problem by extending the UKF, and we present the method in 4. The key idea is to switch two Gaussian observation models that have distinct observation error covariance matrices: one is for observations with regular error, and another, which has larger variance, is for outliers. In addition to a state vector, the method estimates an indicator variable that switches the two models and a scale parameter of the covariance matrix for outliers. As a result of estimating the indicator variable and the scale parameter, estimate of the observation error covariance matrix that can handle both regular observations and

outliers is obtained at each time step. Furthermore, by estimating the scale parameter, the proposed method is applicable to any problem without tuning regarding outlier characteristics. Through numerical experiments, we demonstrate that the proposed method is able to estimate the state better than existing methods for both datasets with and without outliers.

Chapter 2

Preliminaries

2.1 Kalman filters

2.1.1 Formulation of filtering problem

From a Bayesian perspective, the purpose of a filtering problem is to recursively estimate the tth filtered distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ from the $(t-1)$th filtered distribution $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$, The distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ can be obtained from $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ in the following two steps:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1}, \tag{2.1}$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1}), \tag{2.2}$$

where $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and $p(\mathbf{y}_t|\mathbf{x}_t)$ are a system model and an observation model that constitute the general SSM, and the distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ is referred to as the tth predictive distribution. The calculation of Eq. (2.1) and (2.2) are referred to as *prediction* and *update*, respectively.

Filtering is a technique that recursively estimates the filtered distribution starting from given distribution of the state vector by using a SSM.

2.1.2 Kalman filter (KF)

The Kalman filter (KF) is the special case of calculation for Eqs. (2.1) and (2.2) by assuming following distributions. That is,

$$p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{x}_{t-1}|\mu_{t-1|t-1}, \mathbf{V}_{t-1|t-1}), \tag{2.3}$$

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t|\mathbf{F}\mathbf{x}_t, \mathbf{Q}), \tag{2.4}$$

$$p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t|\mathbf{H}_t\mathbf{x}_t, \mathbf{R}), \tag{2.5}$$

where $\mathcal{N}$ denotes a Gaussian distribution, $\mu \in \mathbb{R}^d$ and $\mathbf{V} \in \mathbb{R}^{d \times d} \succ 0$ are the mean and the covariance matrix of the state vector, the subscript $T_1|T_2$ is used for denoting a variable at time step T_1 inferred from given observations at time steps from 1 to T_2, and $\mathbf{F} \in \mathbb{R}^{d \times d}$, $\mathbf{Q} \in \mathbb{R}^{d \times d} \succ 0$, $\mathbf{H}_t \in \mathbb{R}^{m \times d}$, and $\mathbf{R} \in \mathbb{R}^{m \times m} \succ 0$ are given parameters. Especially, $\mathbf{Q}$ is referred to as the model error covariance matrix, and $\mathbf{R}$ is referred to as the observation error covariance matrix.

By substituting Eqs. (2.3) and (2.4) into Eq. (2.1), the tth predictive distribution is obtained as

$$
\begin{aligned}
&p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \\
&= \int \mathcal{N}(\mathbf{x}_t|\mathbf{F}\mathbf{x}_{t-1}, \mathbf{Q})\mathcal{N}(\mathbf{x}_{t-1}|\mu_{t-1|t-1}, \mathbf{V}_{t-1|t-1})d\mathbf{x}_{t-1}, \\
&= \mathcal{N}(\mathbf{x}_t|\mu_{t|t-1}, \mathbf{V}_{t|t-1}), \\
&\mu_{t|t-1} = \mathbf{F}\mu_{t-1|t-1}, \\
&\mathbf{V}_{t|t-1} = \mathbf{F}\mathbf{V}_{t-1|t-1}\mathbf{F}^{\mathrm{T}} + \mathbf{Q},
\end{aligned}
\tag{2.6}
$$

By substituting Eqs. (2.6) and (2.5) into Eq. (2.2), the tth filtered distribution is obtained as

$$
\begin{aligned}
&p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto \mathcal{N}(\mathbf{x}_t|\mathbf{H}_t\mathbf{x}_t, \mathbf{R})\mathcal{N}(\mathbf{x}_t|\mu_{t|t-1}, \mathbf{V}_{t|t-1}), \\
&= \mathcal{N}(\mathbf{x}_t|\mu_{t|t}, \mathbf{V}_{t|t}), \\
&\mu_{t|t} = \mu_{t|t-1} + \mathbf{K}(\mathbf{y}_t - \mathbf{H}_t\mu_{t|t-1}), \\
&\mathbf{V}_{t|t} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{V}_{t|t-1} \\
&\mathbf{K} = \mathbf{V}_{t|t-1}\mathbf{H}_t(\mathbf{H}_t\mathbf{V}_{t|t-1}\mathbf{H}_t^{\mathrm{T}} + \mathbf{R})^{-1},
\end{aligned}
\tag{2.7}
$$

Since tth filtered distribution is obtained as a Gaussian, this calculation can be applied recursively. The successive calculation constitutes the algorithm of the KF.

As already discussed, in actual applications, since the system model or the observation model includes nonlinear functions, such an exact calculation of Eqs. (2.1) and (2.2) is intractable in general. Therefore, the following algorithms as the variant of the KF were developed.

2.1.3   Unscented Kalman filter (UKF)

The UKF is a filtering algorithm for nonlinear SSMs by using the unscented transformation [4]. The unscented transformation is a method for approximating the statistics up to the second moment of a nonlinearly-transformed probability distribution by using the sigma points and the corresponding weights detailed in Appendix A.

In this thesis, we assume a nonlinear SSM with additive Gaussian error for the UKF. The SSM is defined as

$$
p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t|\mathbf{f}(\mathbf{x}_{t-1}), \mathbf{Q}),
\tag{2.8}
$$
$$
p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t|\mathbf{h}(\mathbf{x}_t), \mathbf{R}),
\tag{2.9}
$$

where $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$ and $\mathbf{h} : \mathbb{R}^d \to \mathbb{R}^m$ are multivariate nonlinear functions.

By assuming the $(t-1)$th filtered distribution as Eq. (2.3) and applying the unscented transformation, the UKF approximates the tth predictive distribution Eq.

(2.1) as a Gaussian with the system model, Eq. (2.8). That is,

$$
\begin{aligned}
p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) &= \int \mathcal{N}(\mathbf{x}_t|\mathbf{f}(\mathbf{x}_{t-1}),\mathbf{Q})\mathcal{N}(\mathbf{x}_{t-1}|\mu_{t-1|t-1},\mathbf{V}_{t-1|t-1})d\mathbf{x}_{t-1} \\
&\approx \mathcal{N}(\mathbf{x}_t|\mu_{t|t-1},\mathbf{V}_{t|t-1}), \\
\mu_{t|t-1} &= \sum_{i=0}^{2d} W_{i,t-1|t-1}^{(m)}\mathbf{f}(\chi_{t-1|t-1}^{(i)}), \\
\mathbf{V}_{t|t-1} &= \sum_{i=0}^{2d} W_{i,t-1|t-1}^{(c)}(\mathbf{f}(\chi_{t-1|t-1}^{(i)})-\mu_{t|t-1})(\mathbf{f}(\chi_{t-1|t-1}^{(i)})-\mu_{t|t-1})^{\mathrm{T}}+\mathbf{Q},
\end{aligned} \tag{2.10}
$$

where $\chi_{t-1|t-1}^{(i)}$ are the sigma points of $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$, and $W_{i,t-1|t-1}^{(m)}$ and $W_{i,t-1|t-1}^{(c)}$ are the corresponding weights. Then, the tth filtered distribution Eq. (2.2) is also approximated as a Gaussian. That is,

$$
\begin{aligned}
p(\mathbf{x}_t|\mathbf{y}_{1:t}) &\propto \mathcal{N}(\mathbf{x}_t|\mathbf{h}(\mathbf{x}_t),\mathbf{R})p(\mathbf{x}_t|\mu_{t|t-1},\mathbf{V}_{t|t-1}) \\
&\approx \mathcal{N}(\mathbf{x}_t|\mu_{t|t},\mathbf{V}_{t|t}), \\
\mu_{t|t} &= \mu_{t|t-1}+\mathbf{K}(\mathbf{y}_t-\bar{\mathbf{y}}_t), \\
\mathbf{V}_{t|t} &= \mathbf{V}_{t|t-1}-\mathbf{K}\mathbf{V}_y\mathbf{K}^{\mathrm{T}}, \\
\mathbf{K} &= \mathbf{V}_{xy}\mathbf{V}_y^{-1}, \\
\bar{\mathbf{y}}_t &= \sum_{i=0}^{2d} W_{i,t|t-1}^{(m)}\mathbf{h}(\chi_{t|t-1}^{(i)}), \\
\mathbf{V}_y &= \sum_{i=0}^{2d} W_{i,t|t-1}^{(c)}(\mathbf{h}(\chi_{t|t-1}^{(i)})-\bar{\mathbf{y}}_t)(\mathbf{h}(\chi_{t|t-1}^{(i)})-\bar{\mathbf{y}}_t)^{\mathrm{T}}+\mathbf{R}, \\
\mathbf{V}_{xy} &= \sum_{i=0}^{2d} W_{i,t|t-1}^{(c)}(\chi_{t|t-1}^{(i)}-\mu_{t|t-1})(\mathbf{h}(\chi_{t|t-1}^{(i)})-\bar{\mathbf{y}}_t)^{\mathrm{T}},
\end{aligned} \tag{2.11}
$$

where $\chi_{t|t-1}^{(i)}$ are the sigma points of $\mathcal{N}(\mu_{t|t-1},\mathbf{V}_{t|t-1})$, and $W_{i,t|t-1}^{(m)}$ and $W_{i,t|t-1}^{(c)}$ are the corresponding weights.

Since the tth filtered distribution is obtained as a Gaussian distribution, these equations can be applied recursively and constitute the algorithm of the UKF.

2.1.4 Ensemble Kalman filter (EnKF)

The original ensemble Kalman filter (EnKF) is a filtering algorithm for nonlinear SSMs by using a ensemble-based distribution [5], [6], [23]. In EnKF the $(t-1)$th filtered distribution is given as

$$
p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) = \frac{1}{N}\sum_{n=1}^{N}\delta(\mathbf{x}_{t-1}-\mathbf{x}_{t-1|t-1}^{(n)}), \tag{2.12}
$$

where δ is Dirac's delta function, N is the ensemble size, and $\mathbf{x}_{t-1|t-1}^{(n)}$ is the n-th ensemble member constituting the ensemble-based distribution.

In EnKF, the following SSM is assumed. That is,

$$\mathbf{x}_t \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}), \tag{2.13}$$

$$\mathbf{y}_t \sim p(\mathbf{y}_t|\mathbf{x}_t) \equiv \mathcal{N}(\mathbf{y}_t|\mathbf{H}_t\mathbf{x}_t, \mathbf{R}). \tag{2.14}$$

Note that, while the system model is an arbitrary distribution, the observation model is limited to a Gaussian distribution with the mean $\mathbf{H}_t\mathbf{x}_t$, which is a linear transformation of $\mathbf{x}_t$.

For the given SSM, the tth predictive distribution is obtained by substituting Eqs. (2.13) and (2.12) into Eq. (2.1) as

$$
\begin{aligned}
p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) &= \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) \cdot \frac{1}{N} \sum_{n=1}^{N} \delta(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1}^{(n)}) \\
&= \frac{1}{N} \sum_{n=1}^{N} \delta(\mathbf{x}_t - \mathbf{x}_{t|t-1}^{(n)}), \\
\mathbf{x}_{t|t-1}^{(n)} &\sim p(\mathbf{x}_t|\mathbf{x}_{t-1|t-1}^{(n)}),
\end{aligned}
\tag{2.15}
$$

where $\mathbf{x}_{t|t-1}^{(n)}$ is the n-th member constituting the tth predictive distribution. By substituting Eqs. (2.14) and (2.15) into Eq. (2.2), the tth filtered distribution is approximated as

$$
\begin{aligned}
p(\mathbf{x}_t|\mathbf{y}_{1:t}) &\propto \mathcal{N}(\mathbf{y}_t|\mathbf{H}_t\mathbf{x}_t, \mathbf{R}) \cdot \frac{1}{N} \sum_{n=1}^{N} \delta(\mathbf{x}_t - \mathbf{x}_{t|t-1}^{(n)}) \\
&\approx \frac{1}{N} \sum_{n=1}^{N} \delta(\mathbf{x}_t - \mathbf{x}_{t|t}^{(n)}), \\
\mathbf{x}_{t|t}^{(n)} &= \mathbf{x}_{t|t-1}^{(n)} + \mathbf{K}_t(\mathbf{y}_t + \mathbf{w}_t^{(n)} - \mathbf{H}_t\mathbf{x}_{t|t-1}^{(n)}), \\
\mathbf{K}_t &= \hat{\mathbf{V}}_{t|t-1}\mathbf{H}_t^{\mathrm{T}}(\mathbf{H}_t\hat{\mathbf{V}}_{t|t-1}\mathbf{H}_t^{\mathrm{T}} + \hat{\mathbf{R}})^{-1}, \\
\mathbf{w}_t^{(n)} &\sim \mathcal{N}(\mathbf{w}_t|0, \mathbf{R}),
\end{aligned}
\tag{2.16}
$$

where $\hat{\mathbf{V}}_{t|t-1}$ is the ensemble covariance matrix of $\{\mathbf{x}_{t|t-1}^{(n)}\}$, $\mathbf{x}_{t|t}^{(n)}$ is the n-th member constituting the tth filtered distribution, $\mathbf{w}_t^{(n)}$ is the n-th realization of the observation error, and $\hat{\mathbf{R}}$ is the ensemble covariance matrix of $\{\mathbf{w}_t^{(n)}\}$. Successive application of Eqs. (2.15)-(2.16) constitutes the EnKF algorithm.

2.2 Variational Bayese (VB) Method

In this section, we briefly introduce the variational Bayes (VB) method which is used throughout this thesis.

2.2.1 The VB method in genral

Formulation

The VB method is an approximation method for the posterior distribution. For a given dataset $\mathcal{D}$, the VB method approximates the posterior distribution of latent random variables $Z = \{Z_1, \cdots, Z_L\}$. The graphical model is shown in Fig. 2.1. In
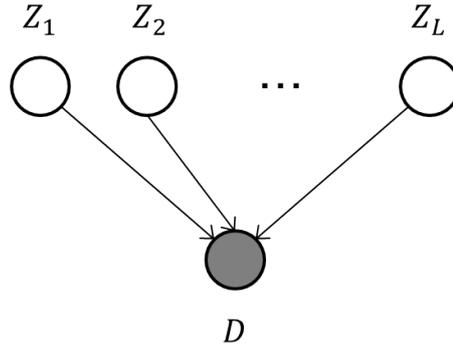
FIGURE 2.1: Graphical model representation for D and Z. Gray circles represent observable variables, and white circles represent latent variables that are not observable.

the prior distribution $p(Z)$, we assume $\{Z_l\}_{l=1}^{L}$ are independent of each other. That is,

$$p(Z) = \prod_{l=1}^{L} p(Z_l). \tag{2.17}$$

Let us assume a marginal likelihood $p(\mathcal{D})$ and, by introducing an arbitrary distribution $q(Z)$, the likelihood can be decomposed as

$$
\begin{aligned}
& \ln p(\mathcal{D}) \\
&= \ln p(\mathcal{D}) \int q(Z) dZ \\
&= \int q(Z) \ln \frac{p(\mathcal{D}|Z)p(Z)}{p(Z|\mathcal{D})} dZ \\
&= \int q(Z) \{ \ln p(\mathcal{D}|Z) + \ln p(Z) - \ln p(Z|\mathcal{D}) + \ln q(Z) - \ln q(Z) \} dZ \\
&= \mathcal{F}[q(Z)] + \mathcal{KL}[q(Z) \parallel p(Z|\mathcal{D})],
\end{aligned}
\tag{2.18}
$$

where

$$\mathcal{KL}[q(Z) \parallel p(Z|\mathcal{D})] \equiv \int q(Z) \ln \frac{q(Z)}{p(Z|D)} dZ, \tag{2.19}$$

$$\mathcal{F}[q(Z)] \equiv \int q(Z) \ln \frac{p(\mathcal{D}|Z)p(Z)}{q(Z)} dZ. \tag{2.20}$$

This decomposition is illustrated in Fig. 2.2.

Here, a functional $\mathcal{KL}$ is referred to as the Kullback-Leibler (KL) divergence, which measures dissimilarity between two distributions. The KL divergence is always non-negative and becomes zero if and only if the two input distributions are identical. The example is illustrated in Fig. 2.3.

According to the properties of the KL divergence, we can say that the minimization of the divergence with respect to $q(Z)$ yields the best estimate of $p(Z|\mathcal{D})$. That
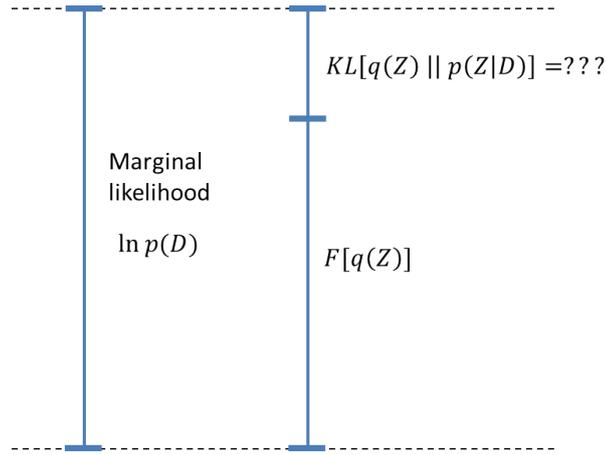
FIGURE 2.2: Decomposition of the marginal likelihood
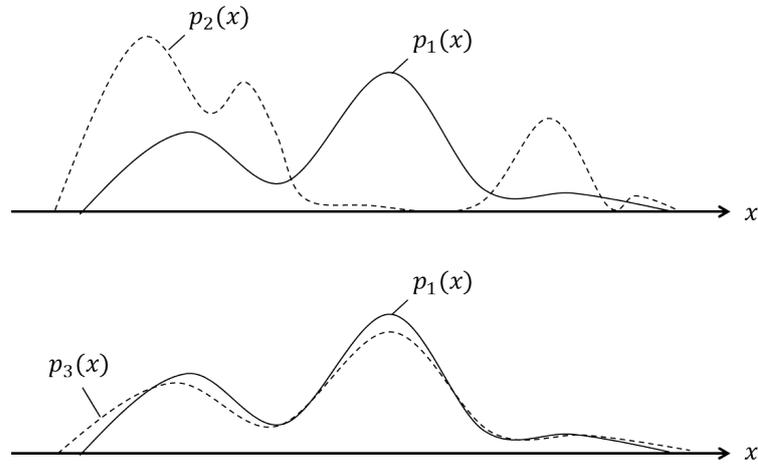
FIGURE 2.3: Example of the Kullback-Leibler divergence; $\mathcal{KL}[p_1(x) \parallel p_2(x)] > \mathcal{KL}[p_1(x) \parallel p_3(x)]$

is,

$$p(Z|\mathcal{D}) \approx \arg\min_{q(Z)} \mathcal{KL}[q(Z) \parallel p(Z|\mathcal{D})]. \tag{2.21}$$

However, the KL divergence cannot be evaluated because it depends on the true posterior distribution $p(Z|\mathcal{D})$, which is unknown and currently trying to be estimated. Here, since the left-hand side of Eq. (2.18) is constant, minimization of the $\mathcal{KL}$ divergence is equivalent to maximization of the functional $\mathcal{F}$ with respect to $q(Z)$. That is,

$$p(Z|\mathcal{D}) \approx \arg\max_{q(Z)} \mathcal{F}[q(Z)]. \tag{2.22}$$

The meaning of maximizing $\mathcal{F}$ with respect to $q(Z)$ can be understood as follows. The functional $\mathcal{F}$ can be rearranged as

$$\mathcal{F}[q(Z)] = \langle \ln p(D|Z)p(Z) \rangle_{q(Z)} + \mathcal{H}[q(Z)], \qquad (2.23)$$

$$\mathcal{H}[q(Z)] \equiv \int q(Z) \ln \frac{1}{q(Z)} dZ, \qquad (2.24)$$

where $\langle \cdot \rangle_q$ denotes an expectation operator on a probability distribution q, and a functional $\mathcal{H}$ is referred to as the entropy which measures the lack of information of a probability distribution. For the uniform distribution, the entropy takes maximum. The first term of Eq. (2.23) measures the fitness of a generative model $p(D|Z)$ regularized by the prior distribution of latent variables in the model, $p(Z)$. If we only consider the maximization of the first term, we obtain the maximum a posteriori (MAP) estimate. By simultaneously maximizing $\mathcal{F}$ and $\mathcal{H}$, we can obtain the distribution $q(Z)$ as an approximation of the true posterior $p(Z|D)$.

Solution

Maximization of the functional $\mathcal{F}$ can be performed by using the variational method under the following assumption. That is,

$$q(Z) = q(Z_1)q(Z_2) \cdots q(Z_L), \qquad (2.25)$$

where $q(Z_1), \cdots, q(Z_L)$ are referred to as variational posterior distributions. Note that the assumption means that latent random variables are also independent of each other in the posterior distribution.

Given the space constraints, we will only describe the resultant maximizer for Eq. (2.22). The details for applying the variational method to the VB method can be found in [24]. The maximizer should satisfy

$$\ln q(Z_l) = \langle \ln p(\mathcal{D}|Z) \rangle_{-q(Z_l)} + \ln p(Z_l) + const. \quad l = 1, \cdots, L, \qquad (2.26)$$

where $\langle \cdot \rangle_{-q(Z_l)}$ is an expectation operator on all variational distributions except $q(Z_l)$, and $const.$ is a normalization constant. Since the variational posterior distributions depend on each other, the maximizer can be obtained by calculating them alternately from an initial guess of variational distributions. The iterative calculation until convergence of the variational distributions constitutes an algorithm of the VB method. Although, as far as we know, the convergence is not theoretically guaranteed, it was confirmed through a number of successful applications of the VB method [25].

Only when $\ln p(\mathcal{D}|Z)$ is a distribution of the exponential family regarding Z and $p(Z_l), l = 1, \cdots, L$ are the conjugate prior, the iterative calculation to obtain the posterior becomes simple by just updating their sufficient statistics.

2.2.2 The VB method for filtering

To adopt the VB method in filtering, we derive a recursive form for observations $\mathbf{y}_1, \cdots, \mathbf{y}_t$. In the context of machine learning, it is known as the online VB method [26], which process data samples one by one, and our derivation is based on the discussion.

Here, as a dataset, we assume a single observation $\mathbf{y}_t$. Furthermore, we use time-indexed notation $Z_t = \{Z_{t,1}, \cdots, Z_{t,L}\}$ as the independent latent random variables.

In Chapter 3 and 4, we intend to use Z_t as $Z_t = [\mathbf{x}_t \ \mathbf{R}_t]^{\mathrm{T}}$ and $Z_t = [\mathbf{x}_t \ s_t \ \gamma_t]^{\mathrm{T}}$, respectively. The relation between $\mathbf{y}_{1:t}$ and Z_t is illustrated in Fig. 2.4. Starting from a
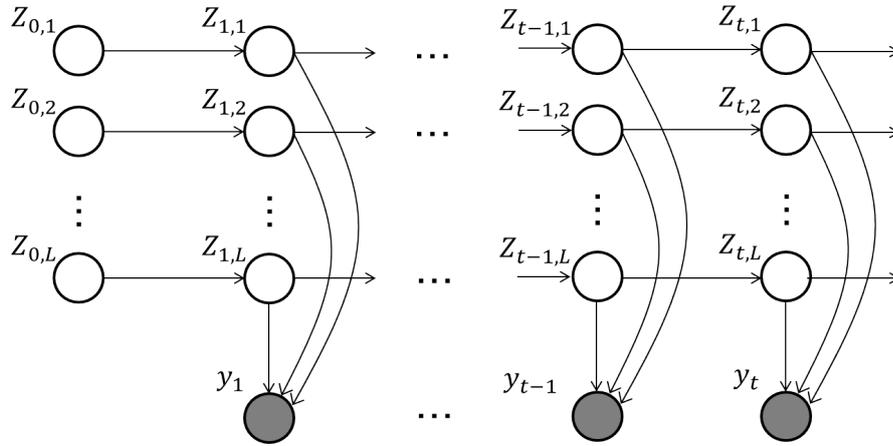
FIGURE 2.4: Graphical model representation for $\mathbf{y}_t$ and Z_t. Gray circles represent observable variables, and white circles represent latent variables that are not observable.

marginal likelihood of $\mathbf{y}_t$ conditioned on $\mathbf{y}_{1:t-1}$, which is represented by $p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$, the derivation approach in section 2.2.1 again leads to an approximation of posterior distribution, that is, the filtered distribution $p(Z_t|\mathbf{y}_{1:t})$ as

$$
p(Z_t|\mathbf{y}_{1:t}) \approx \arg\max_{q(Z_t)} \mathcal{F}[q(Z_t)],
$$
$$
\mathcal{F}[q(Z_t)] \equiv \int q(Z_t) \ln \frac{p(\mathbf{y}_t|Z_t) \prod_{l=1}^{L} p(Z_{t,l}|\mathbf{y}_{1:t-1})}{q(Z_t)} dZ_t, \tag{2.27}
$$

where $p(Z_{t,l}|\mathbf{y}_{1:t-1})$ is referred to as the predictive distribution of $Z_{t,l}$ at time step t. Note that we assume $\mathbf{y}_t$ is conditionally independent of $\mathbf{y}_{1:t-1}$ for given Z_t. Similarly to section 2.2.1, by assuming the independency of $Z_{t,1}, \cdots, Z_{t,L}$ in variational posterior distributions and applying the variational method, the maximizer of Eq. (2.27) can be immediately obtained as

$$
\ln q(Z_{t,l}) = \langle \ln p(\mathbf{y}_t|Z_t) \rangle_{-q(Z_{t,l})} + \ln p(Z_{t,l}|\mathbf{y}_{1:t-1}) + const. \quad l = 1, \cdots, L. \tag{2.28}
$$

By substituting $Z_t = \mathbf{x}_t$ into Eq. (2.28) and taking the exponential, the equation is equivalent to Eq.(2.2). Thus, by using the VB method for filtering, we can perform filtering for other random variables in addition to $\mathbf{x}_t$.

2.2.3   Consideration of the approximation in the VB method

In this section, we consider the effect of the independence assumption in the VB method by using a probability density distribution that has a correlation between the random variables.

Let us assume two dimensional Gaussian distribution as

$$
p(x_1, x_2) = \mathcal{N}((\begin{array}{c} x_1 \\ x_2 \end{array})|\mathbf{0}, \Lambda^{-1} \equiv \left(\begin{array}{cc} 1 & r \\ r & 1 \end{array} \right)), \tag{2.29}
$$

where x_1 and x_2 are random variables, Λ is a precision matrix of the Gaussian distribution, and r is the covariance between x_1 and x_2.

The variational approximation of this distribution with the two independent distributions $q(x_1)$ and $q(x_2)$ was discussed in [27] and given as

$$
\begin{aligned}
p(x_1, x_2) &\approx q(x_1)q(x_2), \\
q(x_1) &= \mathcal{N}(x_1 | m_1 = \Lambda_{1,1}^{-1}\Lambda_{1,2}m_2, \Lambda_{1,1}^{-1}), \\
q(x_2) &= \mathcal{N}(x_2 | m_2 = \Lambda_{2,2}^{-1}\Lambda_{2,1}m_1, \Lambda_{2,2}^{-1}),
\end{aligned}
\tag{2.30}
$$

where $\Lambda_{i,j}$ is the (i,j)th element of Λ. Note that, since $q(x_1)$ and $q(x_2)$ depend on each other via the means m_1 and m_2, the solution can be obtained by an iterative procedure starting from randomly generated m_1.

For $r = 0, 0.12, 0.24, 0.37, 0.49, 0.61, 0.74, 0.86, 0.99$, the true and approximated distributions are plotted in Fig. 2.5. As can be seen in the figure, the means are appropriately estimated, while the variances tend to be underestimated especially for the correlated cases. Since the independence is assumed, the VB method cannot estimate the correlation between x_1 and x_2 in all cases. In [27], the problem of underestimation of the variances is discussed in the comparison with the expectation propagation (EP) method, which is also an approximation method of the posterior.

In Fig. 2.6, the number of iterations until convergence is shown. For the low value of r, that is, x_1 and x_2 are independent, the required number of iterations is very small. When r approaches to 0.99, that is, x_1 and x_2 become correlated, the required iterations increases prominently.

Although there is no theoretical guarantee, this experiment shows three implications for applying the VB method to dependent random variables.

- it can estimate the posterior mean appropriately

- it may underestimate the variance in the posterior

- it may take more iterations to get the converged distributions
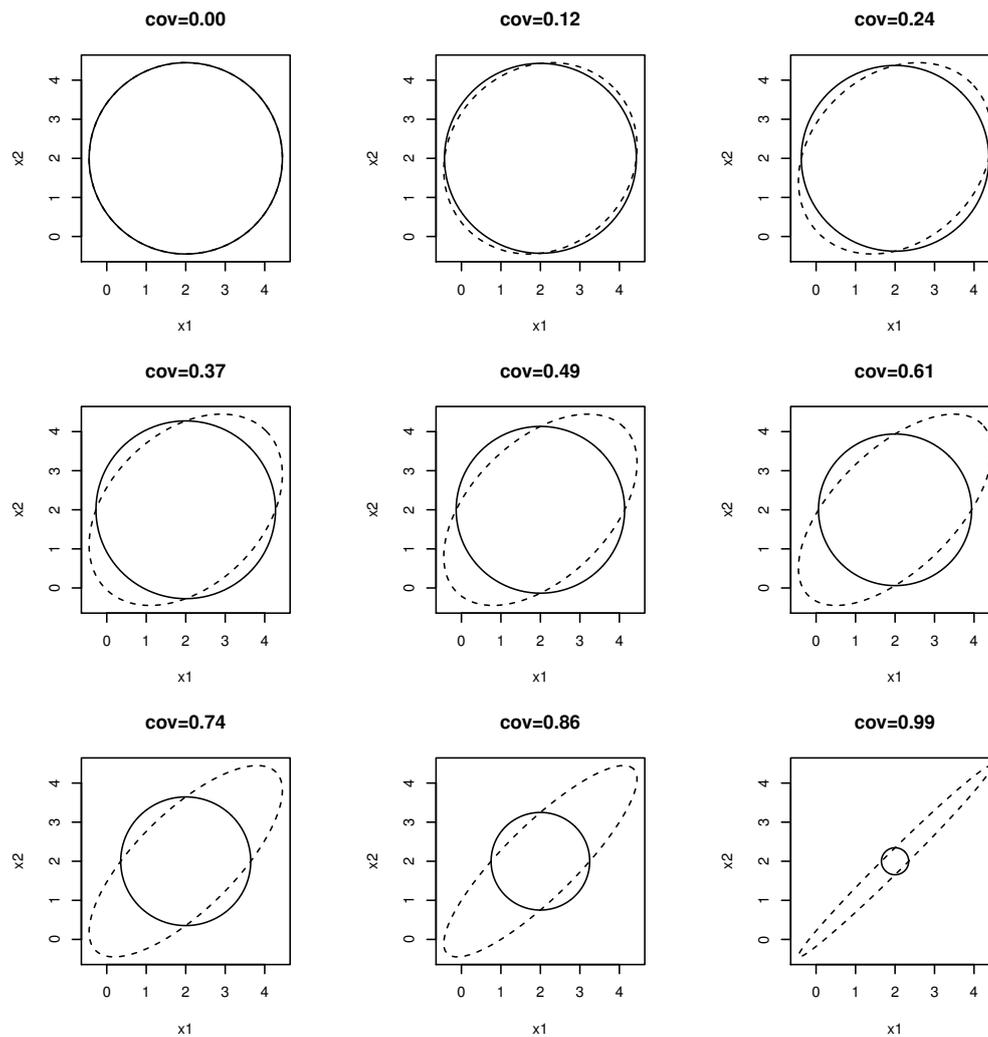
FIGURE 2.5: Contour plot for 2σ of the true and approximated distributions. The dashed lines denote plots of the true distributions, and the solid lines denote plots of the approximated distributions.
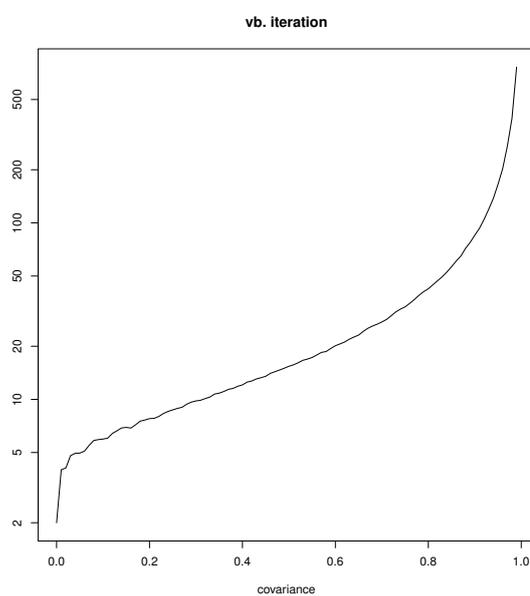
FIGURE 2.6: The number of iterations until convergence. The horizontal axis denotes the values of r, and the vertical axis denotes the number of iterations in the log-scale.

Chapter 3

Nonlinear Kalman filters for time-variant observation error statistics

This chapter is based on the published article,

Akio Nakabayashi and Genta Ueno, "An Extension of the Ensemble Kalman Filter for Estimating the Observation Error Covariance Matrix Based on the Variational Bayes's Method", Monthly Weather Review, 145.1 (2017):199-213. ©American Meteorological Society. Used with permission.

3.1 Introduction

The EnKF requires that the observation error covariance matrix $\mathbf{R}_t$ is set prior to execution. It is known that a poor choice of $\mathbf{R}_t$ can result in under- or overfitting [28]. An appropriate estimation of $\mathbf{R}_t$ is, therefore, a crucial problem in applications of the EnKF. Although $\mathbf{R}_t$ is typically estimated as a time-invariant parameter [e.g. 29], in some cases, it may be reasonable to estimate $\mathbf{R}_t$ as a time-variant parameter. One such example is a system in which the measurement instruments deteriorate over time. We thus note that there is a demand for a method for estimating a time-variant $\mathbf{R}_t$ that works with the EnKF.

The most straightforward idea for estimating $\mathbf{R}_t$ is to use an augmented state-space model that includes $\mathbf{R}_t$ in the state vector as $\tilde{\mathbf{x}}_t = [\mathbf{x}_t \ \mathbf{R}_t]^{\mathrm{T}}$, as discussed in [10] and [6]; this is also known as the self-organizing state-space model [11]. However, this approach does not work in the framework of the EnKF. The reason is that the filtered estimate of $\tilde{\mathbf{x}}_t$ cannot be represented in a closed form, because the Kalman gain, which is used to filter the augmented state vector that includes $\mathbf{R}_t$, depends on $\mathbf{R}_t$ itself. From a Bayesian perspective, the problem comes from the fact that the Gaussian distribution approximating the ensemble for $\tilde{\mathbf{x}}_t$, which is assumed in the EnKF, is no longer conjugate to the observation model, that is, the likelihood function. [13] also discussed this issue from another perspective.

Therefore, estimation of $\mathbf{R}_t$ in the EnKF is an active area of research. [13] have proposed a filtering method for $\mathbf{x}_t$ and a scaling factor α for the covariance matrices of the model error (referred to as background error or system noise), $\mathbf{x}_t$, and the observation error. In their method, however, the basis matrix of $\mathbf{R}_t$ to be scaled needs to be determined in advance. In addition, the estimation of the time-variant $\mathbf{R}_t$ is difficult because α approaches to a constant value as a result of its time-invariant assumption. [30] and [31] have proposed hybrid methods that combine the EnKF for $\mathbf{x}_t$ and the particle filter (PF) for the parameter θ_t of $\mathbf{R}(\theta_t)$. Although not discussed

explicitly, their methods are applicable to a time-variant $\mathbf{R}_t$. In addition, these approaches [30], [31] can, in principle, be applied to an arbitrary parameterized $\mathbf{R}(\theta_t)$, which may include the case in which all the elements of $\mathbf{R}_t$ are assigned as θ_t. However, in a practical sense, the cost of estimating all the elements is prohibitive due to the limitations of the PF for high-dimensional problems [7], [32]. As a result, the dimensionality of θ_t needs to be restricted to be quite low, and strong assumptions about the structure of $\mathbf{R}_t$ are necessary.

In the field of signal processing, [15], [33], and [16] have proposed filtering methods for a time-evolving $\mathbf{x}_t$ and a time-variant $\mathbf{R}_t$. While the methods of [15] and [33] were limited to linear system models, the method of [16] dealt with the nonlinear case. The key idea of [16] was to combine the Gaussian filter for $\mathbf{x}_t$ and the Bayesian estimation of $\mathbf{R}_t$ with the inverse Wishart distribution [34], [35], which is the conjugate prior distribution for $\mathbf{R}_t$. As detailed in [16], the combination is realized by using the variational Bayes (VB) method, which is a method for approximating a joint posterior distribution for multiple random variables with their independent distributions. With the VB method, the Gaussian filter for $\mathbf{x}_t$ and the Bayesian estimation for $\mathbf{R}_t$ can be applied separately [16]. However, the Gaussian filter is not available for high-dimensional systems, since it explicitly estimates the mean and the covariance matrix of $\mathbf{x}_t$. The cost of the estimation is proportional to the dimensionality of $\mathbf{x}_t$. In addition, [16] did not provide a way to optimize the hyperparameters used in their method.

In this chapter, we propose an extension of the ensemble Kalman filter (EnKF) that can simultaneously estimate the state vector and the observation error covariance matrix by using the variational Bayes (VB) method. In numerical experiments, we examine this capability for a time-variant observation error covariance matrix, and it is noteworthy that our method works well even when the true observation error covariance matrix is nondiagonal. We also present two complementary studies. First, we demonstrate the stability of a long-run assimilation when there are unmodeled disturbances. Second, we derive and demonstrate a maximum-likelihood (ML) method for optimizing the hyperparameters used in our method.

3.2   The proposed method

3.2.1   Formulation

We assume an observation error $\mathbf{w}_t \in \mathbb{R}^m$ is a random variable that has time-variant characteristics. To estimate the characteristics, we assume that the error can be represented as a Gaussian defined as

$$\mathbf{w}_t \sim p(w_t) = \mathcal{N}(\mathbf{w}_t|\mathbf{0}, \mathbf{R}_t), \tag{3.1}$$

where $\mathbf{R}_t \in \mathbb{R}^{m \times m} \succ 0$ is the time-indexed observation error covariance matrix. We estimate $\mathbf{R}_t$ in addition to $\mathbf{x}_t$ within the EnKF. To achieve the goal, we redefine a filtering problem for both $\mathbf{x}_t$ and $\mathbf{R}_t$ with an SSM for $\mathbf{x}_t$ and $\mathbf{R}_t$ by modifying Eqs. (2.1)-(2.2) as also discussed in [16]. Similarly to $\mathbf{x}_t$, we assume $\mathbf{R}_t$ is a realization of the first-order Markov process. Thus, we obtain,

$$p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{y}_{1:t-1}) =$$
$$\int p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{x}_{t-1}, \mathbf{R}_{t-1})p(\mathbf{x}_{t-1}, \mathbf{R}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1}d\mathbf{R}_{t-1}, \tag{3.2}$$

$$p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}_t)p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{y}_{1:t-1}), \tag{3.3}$$
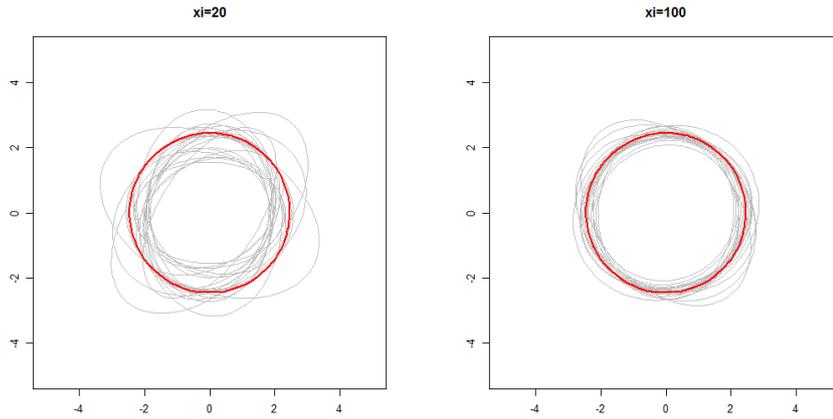
FIGURE 3.1: Contour plot of 2σ of $\mathcal{N}(\mathbf{0}, \mathbf{R}_t)$, where $\mathbf{R}_t$ is sampled from
Eq. (3.5) for $\mathbf{R}_{t-1} = \mathbf{I}$ and $\xi = 20, 100$. The red circle denote the mean,
that is, the contour plot of $\mathcal{N}(0, \mathbf{I})$.

where $\mathbf{x}_t \in \mathbb{R}^d$ is the state vector, $\mathbf{y}_t \in \mathbb{R}^m$ is the observation vector, $p(\mathbf{x}_{t-1}, \mathbf{R}_{t-1}|\mathbf{y}_{1:t-1})$, $p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{y}_{1:t-1})$, and $p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{y}_{1:t})$ are the $(t-1)$th joint filtered distribution, tth joint predictive distribution, and tth joint filtered distribution for $\mathbf{x}_t$ and $\mathbf{R}_t$. By specifying models in Eqs. (3.2)-(3.3), we have derived an extended EnKF algorithm.

3.2.2 State-space model

In this subsection, we define a system model for $\mathbf{x}_t$ and $\mathbf{R}_t$. In the time-evolution, we assume independency between $\mathbf{x}_t$ and $\mathbf{R}_t$ as

$$p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{x}_{t-1}, \mathbf{R}_{t-1}) = p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{R}_t|\mathbf{R}_{t-1}). \tag{3.4}$$

Due to the assumption, for $\mathbf{x}_t$, we can use the same system model as the EnKF, which is defined by Eq. (2.13).

For $\mathbf{R}_t$, we define a conditional distribution that represents the time-evolution of $\mathbf{R}_t$ for given $\mathbf{R}_{t-1}$ as,

$$p(\mathbf{R}_t|\mathbf{R}_{t-1}) = \mathcal{IW}(\mathbf{R}_t|\xi, (\xi - 2m - 2)\mathbf{R}_{t-1}), \tag{3.5}$$

where $\mathcal{IW}$ denotes an inverse Wishart distribution [34], and $\xi > 2m + 2$ is the degree of freedom parameter. This system model has two characteristics. First, for any ξ, the system model preserves $\mathbf{R}_{t-1}$ as the mean. Second, ξ controls the uncertainty in the time evolution of $\mathbf{R}_t$. At the limit as $\xi \to \infty$, the system model gives a point mass on $\mathbf{R}_{t-1}$, which implies a null transition. On the other hand, at the limit as $\xi \to 2m + 2$, the system model becomes uniform, which means that the transition is totally uncertain. Examples of covariance matrices which are sampled from Eq. (3.5) are illustrated in Fig. 3.1.

We define an observation model by assuming additive error given by Eq. (3.1). That is,

$$p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t|\mathbf{H}\mathbf{x}_t, \mathbf{R}_t), \tag{3.6}$$

where $\mathbf{y}_t \in \mathbb{R}^m$ is the observation vector, and $\mathbf{H} \in \mathbb{R}^{m \times d}$. Note that a linear model is assumed for mapping $\mathbf{x}_t$ onto $\mathbf{y}_t$ to use the EnKF algorithm.
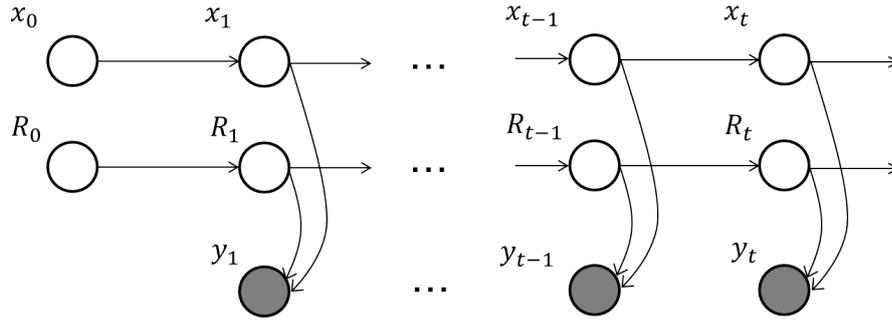
FIGURE 3.2: Graphical model representation of the SSM of the proposed method. Gray circles represent observable variables, and white circles represent latent variables that are not observable.

Therefore, the assumed SSM is summarized as follows:

$$
\begin{aligned}
p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{R}_t|\mathbf{R}_{t-1}) &= p(\mathbf{x}_t|\mathbf{x}_{t-1})\mathcal{IW}(\mathbf{R}_t|\xi, (\xi - 2m - 2)\mathbf{R}_{t-1}), \\
p(\mathbf{y}_t|\mathbf{x}_t) &= \mathcal{N}(\mathbf{y}_t|\mathbf{H}\mathbf{x}_t, \mathbf{R}_t).
\end{aligned}
\tag{3.7}
$$

The SSM is illustrated as a graphical model in Fig. 3.2. With the SSM and given $(t-1)$th filtered distribution, we derive tractable forms of Eqs. (3.2) and (3.3).

3.2.3 Prediction step

At first, we prescribe the $(t-1)$th filtered distribution for $\mathbf{x}_t$ and $\mathbf{R}_t$. Since we will extend the EnKF in this chapter, we use an ensemble-based distribution for $\mathbf{x}_t$. For $\mathbf{R}_t$, we use an inverse Wishart distribution, which is a conjugate prior to covariance matrix of a Gaussian likelihood. That is, the $(t-1)$th filtered distribution is defined as

$$
\begin{aligned}
& p(\mathbf{x}_{t-1}, \mathbf{R}_{t-1}|\mathbf{y}_{1:t-1}) \\
&= p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})p(\mathbf{R}_{t-1}|\mathbf{y}_{1:t-1}) \\
&= \frac{1}{N}\sum_n \delta(\mathbf{x}_{t-1} - \mathbf{x}_{t-1}^{(n)}) \cdot \mathcal{IW}(\mathbf{R}_{t-1}|\eta_{t-1|t-1}, \mathbf{B}_{t-1|t-1}),
\end{aligned}
\tag{3.8}
$$

where δ is Dirac's delta function, $\mathbf{x}^{(n)} \in \mathbb{R}^d$ is the nth member of an ensemble, and $\eta > 2m + 2$ and $\mathbf{B} \in \mathbb{R}^{m \times m} \succ 0$ are the degree of freedom parameter and the scale matrix, respectively. The properties of the inverse Wishart distribution is detailed in Appendix B.

Since we assume the independency between $\mathbf{x}_t$ and $\mathbf{R}_t$, the prediction can be performed separately as

$$
\begin{aligned}
p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{y}_{1:t-1}) &= \int p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{x}_{t-1}, \mathbf{R}_{t-1})p(\mathbf{x}_{t-1}, \mathbf{R}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1}d\mathbf{R}_{t-1}, \\
&= \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \times \int p(\mathbf{R}_t|\mathbf{R}_{t-1})p(\mathbf{R}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{R}_{t-1}, \\
&\equiv p(\mathbf{x}_t|\mathbf{y}_{1:t-1})p(\mathbf{R}_t|\mathbf{y}_{1:t-1}),
\end{aligned}
\tag{3.9}
$$

where $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ and $p(\mathbf{R}_t|\mathbf{y}_{1:t-1})$ are the tth predictive distribution of $\mathbf{x}_t$ and $\mathbf{R}_t$, respectively.

The calculation of $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ is the same as the EnKF. Thus, again, the distribution is given by

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \frac{1}{N}\sum_{n=1}^{N}\delta(\mathbf{x}_t - \mathbf{x}_{t|t-1}^{(n)}),$$

$$\mathbf{x}_{t|t-1}^{(n)} \sim p(\mathbf{x}_t|\mathbf{x}_{t-1|t-1}^{(n)}).\tag{3.10}$$

For $\mathbf{R}_t$, by substituting the second term in Eq.(3.8) and Eq. (3.5), $p(\mathbf{R}_t|\mathbf{y}_{1:t-1})$ becomes

$$
\begin{aligned}
&p(\mathbf{R}_t|\mathbf{y}_{1:t-1})\\
&= \int \mathcal{IW}(\mathbf{R}_t|\xi, (\xi - 2m - 2)\mathbf{R}_{t-1}) \cdot \mathcal{IW}(\mathbf{R}_{t-1}|\eta_{t-1|t-1}, \mathbf{B}_{t-1|t-1})d\mathbf{R}_t
\end{aligned}\tag{3.11}
$$

Since the exact integration is intractable, we approximate the predictive distribution for $\mathbf{R}_t$ with another inverse Wishart distribution that has the same moments up to the second order. The approximated predictive distribution for $\mathbf{R}_t$ is thus

$$p(\mathbf{R}_t|\mathbf{y}_{1:t-1}) \approx \mathcal{IW}(\mathbf{R}_t|\eta_{t|t-1}, \mathbf{B}_{t|t-1}),\tag{3.12}$$

where

$$
\begin{aligned}
\eta_{t|t-1} &= \frac{2}{\left(1 + \frac{2}{\xi - 2m - 4}\right)\left(1 + \frac{2}{\eta_{t-1|t-1} - 2m - 4}\right) - 1}\\
&\quad + 2m + 4,\\
\mathbf{B}_{t|t-1} &= \frac{\eta_{t|t-1} - 2m - 2}{\eta_{t-1|t-1} - 2m - 2}\mathbf{B}_{t-1|t-1}.
\end{aligned}\tag{3.13}
$$

The derivation of this approximation is detailed in Appendix C.

Here, we have the tth joint predictive distribution for $\mathbf{x}_t$ and $\mathbf{R}_t$ as follows:

$$p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{y}_{1:t-1}) = \frac{1}{N}\sum_{n=1}^{N}\delta(\mathbf{x}_t - \mathbf{x}_{t|t-1}^{(n)}) \cdot \mathcal{IW}(\mathbf{R}_t|\eta_{t|t-1}, \mathbf{B}_{t|t-1}).\tag{3.14}$$

3.2.4 Update step

Since the exact calculation of Eq. (3.3) by substituting Eqs. (3.6) and (3.14) is intractable, we derive the approximated form by adopting the VB method.

In conformity to section 2.2, by defining $Z_t = \{\mathbf{x}_t, \mathbf{R}_t\}$ and assuming the independency of $\mathbf{x}_t$ and $\mathbf{R}_t$ in the tth joint filtered distribution, the approximated joint filtered distribution is immediately given as follows:

$$\ln q(\mathbf{x}_t) = \langle \ln p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}_t)\rangle_{q(\mathbf{R}_t)} + \ln p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) + const.,\tag{3.15}$$

$$\ln q(\mathbf{R}_t) = \langle \ln p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}_t)\rangle_{q(\mathbf{x}_t)} + \ln p(\mathbf{R}_t|\mathbf{y}_{1:t-1}) + const.,\tag{3.16}$$

where $\langle\cdot\rangle_q$ is the expectation operator on the distribution q. In the following, we calculate the actual form of these two distributions, $q(\mathbf{x}_t)$ and $q(\mathbf{R}_t)$.

First, by calculating the expectation with respect to $\mathbf{R}_t$, Eq. (3.15) becomes

$$
\begin{aligned}
\ln q(\mathbf{x}_t) &= -\frac{1}{2}\langle(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)\rangle_{q(\mathbf{R}_t)} + \ln p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) + const. \\
&= -\frac{1}{2}\langle\mathrm{tr}[(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)^{\mathrm{T}}\mathbf{R}_t^{-1}]\rangle_{q(\mathbf{R}_t)} + \ln p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) + const. \\
&= -\frac{1}{2}\mathrm{tr}[(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)^{\mathrm{T}}\langle\mathbf{R}_t^{-1}\rangle_{q(\mathbf{R}_t)}] + \ln p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) + const. \\
&= -\frac{1}{2}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)^{\mathrm{T}}\langle\mathbf{R}_t^{-1}\rangle_{q(\mathbf{R}_t)}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t) + \ln p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) + const. \\
&= \ln \mathcal{N}(\mathbf{y}_t|\mathbf{H}\mathbf{x}_t, \bar{\mathbf{R}}_t) + \ln p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) + const.,
\end{aligned}
\tag{3.17}
$$

where

$$
\bar{\mathbf{R}}_t \equiv \langle\mathbf{R}_t^{-1}\rangle_{q(\mathbf{R}_t)}^{-1}.
\tag{3.18}
$$

Since the exponential of Eq. (3.17) is the same form as that of the EnKF by substituting $\mathbf{R}_t = \bar{\mathbf{R}}_t$, $q(\mathbf{x}_t)$ can be calculated based on the EnKF update. Here, $\bar{\mathbf{R}}_t$ is the estimate of the random variable $\mathbf{R}_t$ based on $q(\mathbf{R}_t)$, which is also an estimating tth filtered distribution for $\mathbf{R}_t$. The obtained distribution is the following ensemble:

$$
\begin{aligned}
q(\mathbf{x}_t) &= \frac{1}{N}\sum_{n=1}^{N}\delta(\mathbf{x}_t - \mathbf{x}_{t|t}^{(n)}), \\
\mathbf{x}_{t|t}^{(n)} &= \mathbf{x}_{t|t-1}^{(n)} + \mathbf{K}_t(\mathbf{y}_t + \mathbf{w}_t^{(n)} - \mathbf{H}_t\mathbf{x}_{t|t-1}^{(n)}), \\
\mathbf{K}_t &= \hat{\mathbf{V}}_{t|t-1}\mathbf{H}_t^{\mathrm{T}}(\mathbf{H}_t\hat{\mathbf{V}}_{t|t-1}\mathbf{H}_t^{\mathrm{T}} + \hat{\mathbf{R}}_t)^{-1}, \\
\mathbf{w}_t^{(n)} &\sim \mathcal{N}(\mathbf{w}_t|0, \bar{\mathbf{R}}_t).
\end{aligned}
\tag{3.19}
$$

Similarly, Eq. (3.19) can be replaced with another EnKF variant, such as the ensemble square root filter (EnSRF) [36].

Next, by calculating the expectation with respect to $\mathbf{x}_t$, Eq. (3.16) becomes

$$
\ln q(\mathbf{R}_t) = -\frac{1}{2}\ln|\mathbf{R}_t| - \frac{1}{2}\langle(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)\rangle_{q(\mathbf{x}_t)} + \ln p(\mathbf{R}_t|\mathbf{y}_{1:t-1}) + const.
\tag{3.20}
$$

By substituting Eq. (3.19) into Eq. (3.20), we obtain $q(\mathbf{R}_t)$ as, again, an inverse Wishart distribution. That is,

$$
q(\mathbf{R}_t) = \mathcal{IW}(\mathbf{R}_t|\eta_{t|t}, \mathbf{B}_{t|t}),
\tag{3.21}
$$

where

$$
\begin{aligned}
\eta_{t|t} &= \eta_{t|t-1} + 1, \\
\mathbf{B}_{t|t} &= \mathbf{B}_{t|t-1} + (\mathbf{y} - \mathbf{H}\hat{\mu}_{t|t})(\mathbf{y} - \mathbf{H}\hat{\mu}_{t|t})^{\mathrm{T}} + \mathbf{H}\hat{\mathbf{V}}_{t|t}\mathbf{H}^{\mathrm{T}}.
\end{aligned}
\tag{3.22}
$$

We use $\hat{\mu}_{t|t}$ and $\hat{\mathbf{V}}_{t|t}$ as the mean vector and the covariance matrix of $\{\mathbf{x}_{t|t}^{(n)}\}$, respectively. Now, since we have found the actual form of $q(\mathbf{R}_t)$, $\bar{\mathbf{R}}_t$, defined by Eq. (3.18), can be calculated as

$$
\bar{\mathbf{R}}_t = \frac{1}{\eta_{t|t} - m - 1}\mathbf{B}_{t|t}.
\tag{3.23}
$$

Algorithm 1 Algorithm for the VBEnKF

Require: an initial distribution, $p(\mathbf{x}_0, \mathbf{R}_0) = \frac{1}{N} \sum_{n=1}^{N} \delta(\mathbf{x}_0 - \mathbf{x}_{0|0}^{(n)}) \cdot \mathcal{IW}(\mathbf{R}_0|\eta_0, \mathbf{B}_0)$.
 1: **for** $t = 1$ to T **do**
 2:　　calculate $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ with Eq. (3.10)
 3:　　calculate $p(\mathbf{R}_t|\mathbf{y}_{1:t-1})$ with Eq. (3.12)
 4:　　let $q(\mathbf{R}_t) = p(\mathbf{R}_t|\mathbf{y}_{1:t-1})$
 5:　　**while** true **do**
 6:　　　calculate $q(\mathbf{x}_t)$ with Eq. (3.19) (or with other EnKF variants)
 7:　　　calculate $q(\mathbf{R}_t)$ with Eq. (3.21)
 8:　　　**if** the update of $q(\mathbf{x}_t)$ and $q(\mathbf{R}_t)$ have converged **then**
 9:　　　　exit while loop
10:　　　**end if**
11:　　**end while**
12:　　let $p(\mathbf{x}_t|\mathbf{y}_{1:t}) = q(\mathbf{x}_t)$, $p(\mathbf{R}_t|\mathbf{y}_{1:t}) = q(\mathbf{R}_t)$
13: **end for**

Here, we have tth joint filtered distribution for $\mathbf{x}_t$ and $\mathbf{R}_t$ as follows:

$$p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{y}_{1:t}) \approx q(\mathbf{x}_t)q(\mathbf{R}_t) = \frac{1}{N} \sum_{n=1}^{N} \delta(\mathbf{x}_t - \mathbf{x}_{t|t}^{(n)}) \cdot \mathcal{IW}(\mathbf{R}_t|\eta_{t|t}, \mathbf{B}_{t|t}). \qquad (3.24)$$

The two distributions, $q(\mathbf{x}_t)$ and $q(\mathbf{R}_t)$, depend on each other via the expectations in Eqs. (3.19) and (3.21). Therefore, we calculate them alternately, beginning with an initial guess for $q(\mathbf{R}_t)$ and continuing until convergence. To obtain an initial guess for $q(\mathbf{R}_t)$, one option is to use the predictive distribution, Eq. (3.14). As the convergence metric, we will use the L2 norm of $\hat{\mu}_{t|t}$.

3.2.5 Algorithm

Here, since tth joint filtered distritbution, Eq. (3.24), is the same form as the $(t-1)$th joint filtered distribution, Eq. (3.8), we can construct an algorithm which is recursively applied for estimating $\mathbf{x}_t$ and $\mathbf{R}_t$. This algorithm is shown in Algorithm 1, which we will refer to as the variational Bayesian ensemble Kalman filter (VBEnKF).

3.3 Tips for usage

3.3.1 Selection of ξ for the system model of $\mathbf{R}_t$

We present an idea for selecting ξ, which is a hyperparameter that must to be specified in advance of assimilation, based on the relation between ξ and $\eta_{t|t}$.

From Eqs. (3.12) and (3.22), the temporal variation of $\eta_{t|t}$ can be simulated with the initial value $\eta_{0|0}$ and ξ. The results obtained by starting with $\eta_{0|0} = \{50, 75, 100\}$ are shown in Fig. 3.3, where it can be seen that the converged value of $\eta_{t|t}$ depends only on ξ.
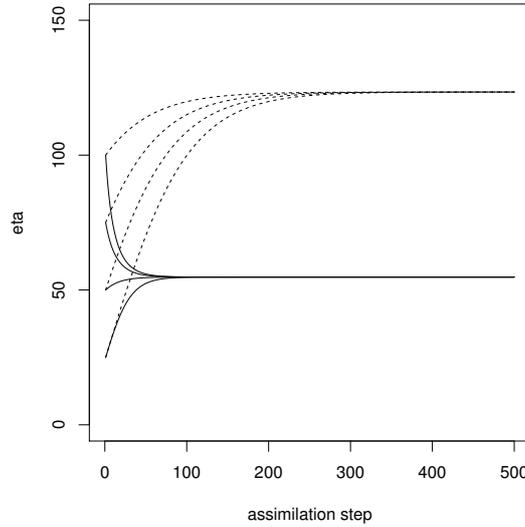
FIGURE 3.3: Temporal variation of η from different initial values for a ten-dimensional observation problem. The horizontal axis represents the assimilation step, the vertical axis represents $\eta_{t|t}$, the solid line is the case of $\xi = 1e3$, and the dashed line is the case of $\xi = 1e4$.

By rearranging Eqs. (3.12) and (3.22) such that the temporal change of $\eta_{t|t}$ becomes zero, we obtain a relation between ξ and the converged value of $\eta_{t|t}$ as follows:

$$\xi = \eta_c^2 - (7 + 4m)\eta_c + 4m^2 + 16m + 14, \tag{3.25}$$

where η_c is the converged value of $\eta_{t|t}$.

This relation is illustrated in Fig. 3.4, from which we can see that the relation between ξ and η_c is monotonic and depends on the dimensionality of the observation. Therefore, using this relation, ξ can be replaced by η_c.

The parameter η_c can be more easily interpreted than can ξ, because it represents the number of degrees of freedom of $\mathcal{IW}$, that is, it is the effective number of samples. For example, in the case of $\xi = 1e3$ shown in Fig. 3.3, we can say that information contained in the $\eta_c = 54.8$ samples is retained in the filtered distribution of $\mathbf{R}_t$ in the limit as $t \to \infty$.

We recommend first selecting η_c and then using Eq. (3.25) to convert η_c into ξ. For convenience, in the following discussion, we will regard η_c (instead of ξ) as the hyperparameter of the VBEnKF.

3.3.2   Covariance tapering for $\mathbf{R}_t$

Estimating $\mathbf{R}_t$ from limited observations might result in spurious correlations and might harm the estimation of $\mathbf{x}_t$. When an uncertain system model or an uncertain initial distribution is used, this problem can become significant in the early stages of data assimilation.

If we have spatial information about the observation vector which indicates an inverse correlation between the covariance and the distances to two observation sites, it can be helpful to use the covariance tapering approach for $\mathbf{R}_t$; this approach
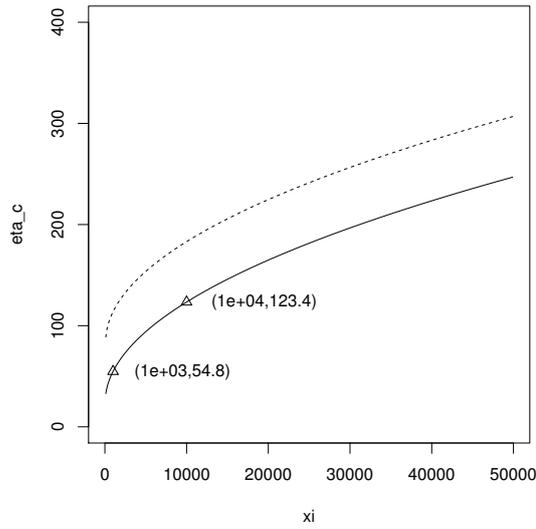
FIGURE 3.4: Relation between ξ and η_c as given by Eq. (3.25). The solid line is the case of $m = 10$, and the dashed line is the case of $m = 40$. The triangles indicate the points at which $\xi = \{1e3, 1e4\}$ for $m = 10$.

has been widely used for the ensemble covariance matrix of the state vector. For example, with the correlation matrix **C** discussed in [37], $\mathbf{R}_t$ is tapered as

$$\mathbf{R}_t \leftarrow \mathbf{R}_t \circ \mathbf{C}, \tag{3.26}$$

where $\circ$ is the operator for the Hadamard product. When we can assume the elements of the observation error are mutually independent, a reasonable tapering operation might be to let the nondiagonal elements be zero.

3.3.3 Hyperparameter estimation

Here, we introduce an ML method for optimizing a hyperparaemter, η_c, that need to be specified in advance of the filtering. For implementing the ML method, we derive the likelihood function for the time-series data $\mathbf{y}_{1:T} = \{\mathbf{y}_1, \cdots, \mathbf{y}_T\}$. The likelihood function is composed of the predictive likelihood functions as follows:

$$p(\mathbf{y}_{1:T}) = \prod_{t=1}^{T} p(\mathbf{y}_t | \mathbf{y}_{1:t-1}). \tag{3.27}$$

Note that the symbols of the hyperparameters are omitted in this derivation. The predictive likelihood for a single observation $\mathbf{y}_t$ can be represented as

$$
\begin{aligned}
&p(\mathbf{y}_t|\mathbf{y}_{1:t-1})\\
&= \int p(\mathbf{y}_t, \mathbf{x}_t, \mathbf{R}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t d\mathbf{R}_t,\\
&= \int p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}_t)p(\mathbf{x}_t, \mathbf{R}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t d\mathbf{R}_t,\\
&= \int p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})p(\mathbf{R}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t d\mathbf{R}_t.
\end{aligned}
\tag{3.28}
$$

The matrix $\mathbf{R}_t$ can be integrated out as follows:

$$
\begin{aligned}
&\int p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}_t)p(\mathbf{R}_t|\mathbf{y}_{1:t-1})d\mathbf{R}_t\\
&= \int \mathcal{N}(\mathbf{y}_t|\mathbf{H}_t\mathbf{x}_t, \mathbf{R}_t)\mathcal{IW}(\mathbf{R}_t|\eta_{t|t-1}, \mathbf{B}_{t|t-1})d\mathbf{R}_t\\
&= \mathcal{T}\left(\mathbf{y}_t|\mathbf{H}_t\mathbf{x}_t, \frac{\mathbf{B}_{t|t-1}}{\eta_{t|t-1}-2D}, \eta_{t|t-1}-2m\right).
\end{aligned}
\tag{3.29}
$$

The notation $\mathcal{T}$ represents the multivariate t-distribution [38], defined as

$$
\mathcal{T}(\mathbf{x}|\mu, \mathbf{V}, \nu) \propto [1 + \frac{1}{\nu}(\mathbf{x}-\mu)^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{x}-\mu)]^{-\frac{\nu+d}{2}},
\tag{3.30}
$$

where μ is the location vector, $\mathbf{V}$ is the scale matrix, and ν is the number of degrees of freedom. Then, the integral of Eq. (3.28) can be calculated as follows:

$$
\begin{aligned}
&p(\mathbf{y}_t|\mathbf{y}_{1:t-1})\\
&= \int \mathcal{T}\left(\mathbf{y}_t|\mathbf{H}_t\mathbf{x}_t, \frac{\mathbf{B}_{t|t-1}}{\eta_{t|t-1}-2m}, \eta_{t|t-1}-2m\right)\times\\
&\qquad p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t\\
&= \frac{1}{N}\sum_{n=1}^{N}\mathcal{T}\left(\mathbf{y}_t|\mathbf{H}_t\mathbf{x}_{t|t-1}^{(n)}, \frac{\mathbf{B}_{t|t-1}}{\eta_{t|t-1}-2m}, \eta_{t|t-1}-2m\right).
\end{aligned}
\tag{3.31}
$$

In the second line, $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ is replaced by the predictive ensemble given by Eq. (3.10). By substituting Eq. (3.31) into Eq. (3.27), the ML estimation of the hyperparameters becomes possible.

3.4 Numerical Experiments

In this section, we present two numerical experiments that show the properties of the VBEnKF. First, we show the estimation of a time-variant $\mathbf{R}_t$ in a linear system. In the experiment, we varied the hyperparameters in order to clarify their characteristics. Second, for a nonlinear system, we demonstrate the capabilities of our method and show that it is superior to an existing method.

3.4.1 Experiment with a linear system

System and synthetic data

In the first experiment, we used the following 10-dimensional system:

$$
\begin{aligned}
\mathbf{x}_t &\sim \mathcal{N}(\mathbf{x}_t|\mathbf{F}\mathbf{x}_{t-1}, 0.1\mathbf{I}), \\
\mathbf{y}_t &\sim \mathcal{N}(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}_{true,t}), \\
\mathbf{x}_{0,true} &= [10, \cdots, 10]^{\mathrm{T}},
\end{aligned}
\tag{3.32}
$$

where $\mathbf{F} \in \mathbb{R}^{10 \times 10}$ is a tridiagonal matrix with 0.3 on the main diagonal, 0.6 on the superdiagonal, and 0.1 on the subdiagonal, as presented in [39]. The matrix $\mathbf{R}_{true,t}$ is defined as

$$
\mathbf{R}_{true,t} =
\begin{cases}
\mathbf{S}_0 & (t = 1, \cdots, \frac{T}{2}) \\
2\mathbf{S}_0 & (t = \frac{T}{2} + 1, \cdots, T),
\end{cases}
\tag{3.33}
$$

where T is a total number of assimilation steps, and $\mathbf{S}_0$ is a tridiagonal matrix with 2.0 on the main diagonal and 0.8 on the super- and subdiagonals. Note that the true observation matrix is doubled from $t = \frac{T}{2} + 1$.

In actual case, such a step change of the observation error covariance could happen when the measurement instruments encounter a malfunction. We have intension to demonstrate the performance deterioration in estimating the state vector if the change is not followed.

With this system, we generated synthetic data for $T = 2000$ steps, $\mathbf{x}_{1:2000}$ and $\mathbf{y}_{1:2000}$.

Experimental setting

We conducted twin experiments with the VBEnKF. To obtain a baseline for the estimation, we also conducted two additional experiments with the EnKF for a fixed $\mathbf{R}_t$. The first experiment used $\mathbf{R}_{true,0}$ for all of the assimilation steps, and the second experiment used $\mathbf{R}_{true,t}$; they are referred to as Oracle1 and Oracle2, respectively.

The settings for the EnKF that were used in all experiments are as follows. The EnKF was implemented using the EnSRF [36] in order to reduce the Monte Carlo error in the estimation of $\mathbf{x}_t$. The ensemble size was set to $N = 32$. The initial distribution for $\mathbf{x}_0$ was $\mathcal{N}(\mathbf{x}_0|\mathbf{x}_{true,0}, 10\mathbf{I})$.

For the VBEnKF, we applied the following settings. We used $\eta_c = \{100, 200, 400\}$; the corresponding ξ was obtained by substituting η_c into Eq. (3.25). The parameters of the initial distribution $p(\mathbf{R}_0)$ were $(\eta_0, \mathbf{B}_0) = (25, 15\mathbf{I})$, $(50, 140\mathbf{I})$, and $(100, 390\mathbf{I})$. Since the domain of η is $(2m + 2, \infty]$, and $m = 10$, the combination $(25, 15\mathbf{I})$ can be regarded as a relatively uncertain initial distribution. The initial distributions specified by the above three parameter combinations have different variances but have the same mean, $E(\mathbf{R}_0) = 5\mathbf{I}$.

We compare the results from three points of view: 1) the root mean square error (RMSE) of the predicted mean of $\mathbf{x}_t$, 2) the predicted $\mathbf{R}_t$ at each assimilation step, and 3) the required number of iterations for convergence in the VBEnKF. The RMSE is given as

$$
RMSE = \sqrt{\frac{1}{Td} \sum_{t=1}^{T} (\mathbf{x}_{true,t} - \hat{\mu}_{t|t-1})^{\mathrm{T}} (\mathbf{x}_{true,t} - \hat{\mu}_{t|t-1})},
\tag{3.34}
$$
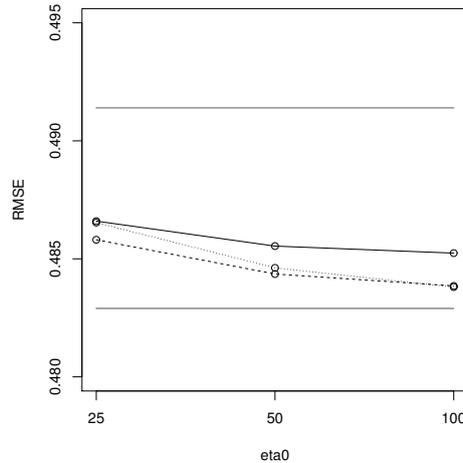
FIGURE 3.5: RMSE of the predicted mean of $\mathbf{x}_t$. The horizontal axis represents η_0, and the vertical axis represents the RMSE. Gray lines at $RMSE = 0.492$ and $RMSE = 0.483$ are the results of Oracle1 and Oracle2, respectively, both of which are independent of η_0. The other lines are the results of the VBEnKF: solid for $\eta_c = 100$, dashed for $\eta_c = 200$, and dotted for $\eta_c = 400$.

where $\hat{\mu}_{t|t-1}$ is the mean of the predictive ensemble.

Results

We conducted the experiments 128 times, each with a distinct random seed used to initialize the ensemble. The results of each trial were averaged. The RMSEs of the predicted mean of $\mathbf{x}_t$ are shown in Fig. 3.5. Temporal variations of the predicted mean of $\mathbf{R}_t$ for the combinations of $\{\eta_c, \eta_0\}$ are shown in Fig. 3.6. The average numbers of iterations required for convergence in the VBEnKF are shown in Fig. 3.7.

We can see in Fig. 3.5 that the best estimation was achieved by Oracle2, which uses the true $\mathbf{R}_t$. It should be noted that the VBEnKF outperformed Oracle1, which uses the true $\mathbf{R}_t$ for the first half in the assimilation steps. With the VBEnKF, the initial distribution for $\mathbf{R}_0$ with $\eta_0 = 100$ outperformed the other settings, $\eta_0 = \{25, 50\}$. In other words, the most certain distribution for $\mathbf{R}_0$, the one with the smallest variance, showed the best result. The setting $\eta_c = 200$ outperformed the other settings, $\eta_c = \{100, 400\}$. We assume that this result comes from a trade-off between stability and speed of convergence to the estimation of $\mathbf{R}_t$, as discussed in the following paragraphs.

Next, we consider Fig.3.6. Although there were differences in the speed of convergence and the variance, all of the settings successfully estimated the time-variant $\mathbf{R}_t$. We also observe the following two properties. First, the effect of the initial distribution is limited to the earliest stages of the assimilation. For example, from Fig. 3.6 (a), (b), (c), we note that the temporal variations became similar after 500 steps. Second, η_c controls the convergence speed and the variance, and there is a trade-off between them. From the top three panels of Fig. 3.6, we see that when η_c was small, it gave estimates that converged quickly to the true value of $\mathbf{R}_t$ but had large variance; the opposite was true when η_c was large (bottom three panels).
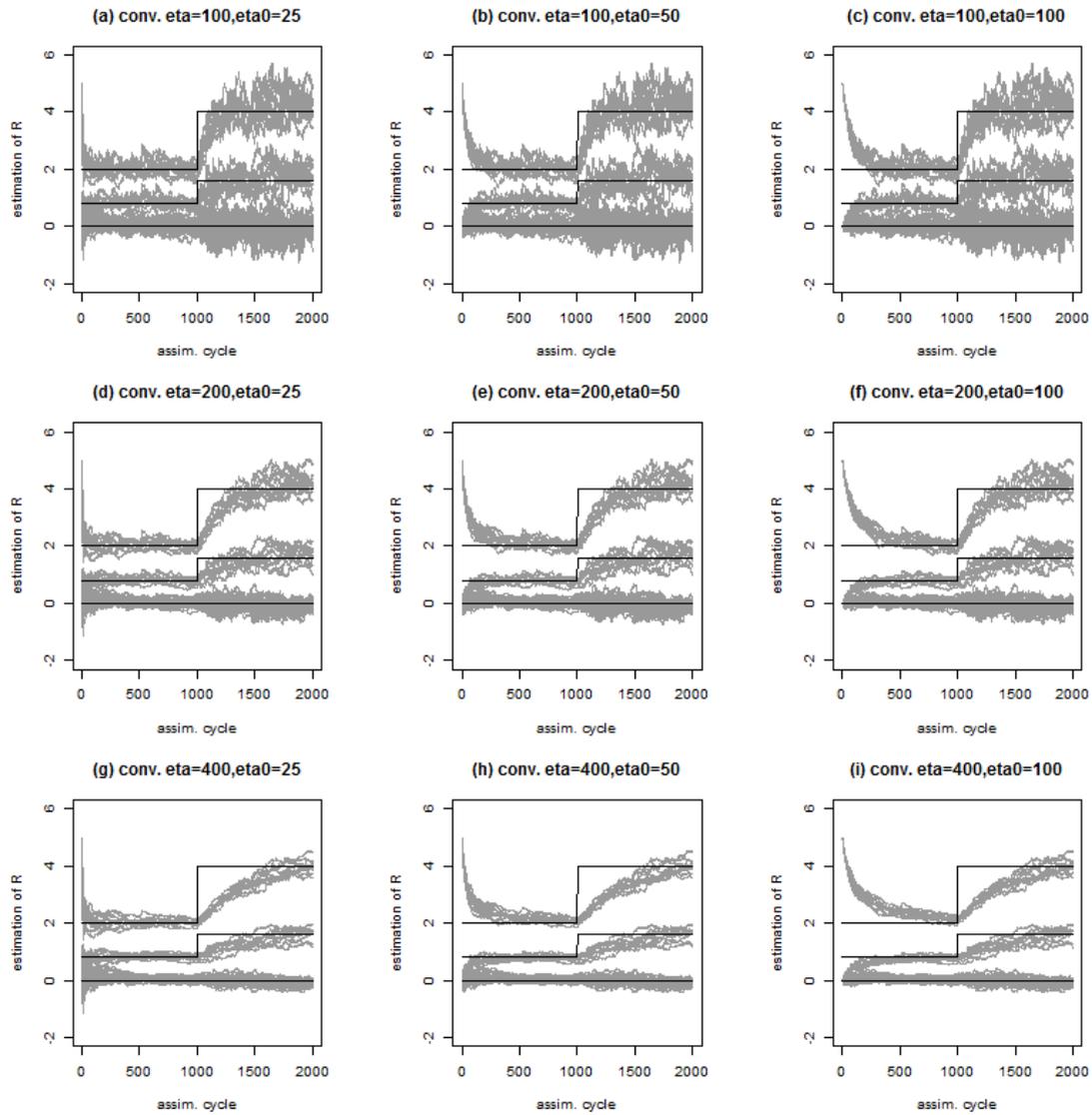
FIGURE 3.6: Temporal variations of the predicted mean of $\mathbf{R}_t$ as a function of $\{\eta_c, \eta_0\}$. The horizontal axis represents the assimilation step, and the vertical axis represents the predicted $\mathbf{R}_t$. The gray lines are the mean of the predictive distribution for $\mathbf{R}_t$ given by Eq. (3.12), and the solid lines are $\mathbf{R}_{true,t}$.
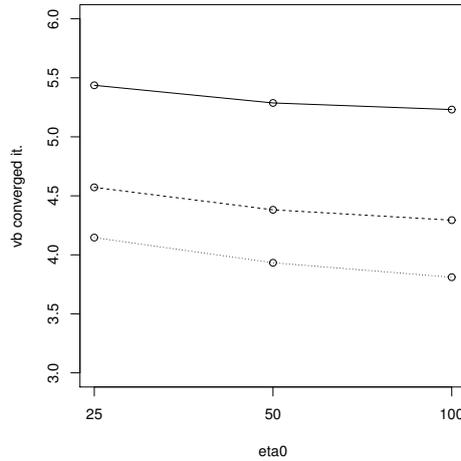
FIGURE 3.7: Time-averaged number of iterations required for convergence in the VBEnKF. The horizontal axis represents η_0, and the vertical axis represents the required number of iterations. The case of $\eta_c = 100$ is shown by a solid line, the case of $\eta_c = 200$ is shown by a dashed line, and the case of $\eta_c = 400$ is shown by a dotted line.

The best performance occurred when $\eta_c = 200$ had a moderate value, as shown in Fig. 3.5; this is because this estimation of $\mathbf{R}_t$ is balanced in terms of convergence speed and variance.

Finally, we consider Fig. 3.7. The required average number of iterations ranged from 3.5 to 5.5. This result depends on η_0 and η_c, which control the uncertainty of $\mathbf{R}_t$, that is, the variance of the initial distribution and the time evolution. The combination of the largest parameters, $\eta_0 = 100$ and $\eta_c = 400$, required the smallest number of iterations, and vice versa.

3.4.2 Nonlinear system

System and synthetic data

In Experiment 2, we used the *Lorenz96* system [40]. *Lorenz96* is a nonlinear and chaotic dynamical system that is used as a benchmark for data assimilation and is given in differential equations on continuous time space τ as

$$
\begin{aligned}
\frac{dx_{\tau,k}}{d\tau} &= (x_{\tau,k+1} - x_{\tau,k-2})x_{\tau,k-1} - x_{\tau,k} + F, \\
k &= 1, \cdots, d, \\
x_{t,0} &= x_{t,d}, x_{t,-1} = x_{t,d-1}, x_{t,d+1} = x_{t,1},
\end{aligned}
\tag{3.35}
$$

where F is the forcing parameter, and k is the position index of $\mathbf{x}_t$. In this experiment, we used $F = 8$ and $d = 40$.

From the equations, we used a discretized system $f : \mathbf{x}_{t-1} \rightarrow \mathbf{x}_t$ on time space $\{0.05t | t = 0, 1, \cdots, T\}$. The time evolution was calculated by the fourth-order Runge-Kutta method with time step $\Delta = 0.005$.

We assumed that every other element of $\mathbf{x}_t$ are observed with Gaussian noise as

$$\mathbf{y}_t \sim \mathcal{N}(\mathbf{y}_t | \mathbf{H}\mathbf{x}_t, \mathbf{R}_{true,t}). \tag{3.36}$$

The dimensionality of the observation vector was set to $m = d/2 = 20$. The elements of the observation matrix $\mathbf{H} \in \mathbb{R}^{m \times d}$ are $H_{i,j} = 1(j = 2i - 1)$ and $H_{i,j} = 0(otherwise)$ for $i = 1, \cdots, m$ and $j = 1, \cdots, d$. The true observation error covariance matrix is assumed to be

$$\mathbf{R}_{true,t} = \begin{cases} \mathbf{S}_z & (t = 1, \cdots, \frac{T}{2}) \\ 2\mathbf{S}_z & (t = \frac{T}{2} + 1, \cdots, T) \end{cases}, \tag{3.37}$$

$$z = \{a, b, c\},$$

where $\mathbf{S}_z, z = \{a, b, c\}$ are the three types of matrices: $\mathbf{S}_a = \mathbf{I}, \mathbf{S}_b = \text{diag}(1, 2, 1, 2, \cdots, 1, 2)$, and $\mathbf{S}_c = \mathbf{S}_0$, which was used in Experiment 1. Using this system, we generated synthetic data for $T = 4000$ steps, $\mathbf{x}_{1:4000}$ and $\mathbf{y}_{1:4000}$

Experimental setting

For comparison, we conducted experiments with a method proposed by Frei and Künsch [31]. They also developed a filtering method that can estimate joint filtered distribution of the state vector and the observation error covariance matrix by combining the EnKF and the PF. The EnKF is used for estimating $\mathbf{x}_t$, which is a high-dimensional vector, and the PF is used for θ_t, which is a parameter vector of $\mathbf{R}(\theta_t)$.

Since their method does not specify the parametrization of $\mathbf{R}_t$ (it can be used with any arbitrary parametrization of $\mathbf{R}_t$), we designed two experiments with different parameterizations, referred to as FK1 and FK2, respectively. In FK1, $\mathbf{R}_t$ was parametrized with the scale parameter $\alpha_t > 0$ as $\mathbf{R}_t(\alpha_t) = \alpha_t \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. In FK2, the parametrization $\mathbf{R}_t(\alpha_{t,1}, \cdots, \alpha_{t,20}) = \text{diag}(\alpha_{t,1}, \cdots, \alpha_{t,20}), \alpha_{t,1} > 0, \cdots, \alpha_{t,20} > 0$ was used to estimate the diagonal elements of $\mathbf{R}_t$. To obtain a baseline, we also repeated Oracle1 and Oracle2 in this setting.

The common settings for the EnKF were the same as in Experiment 1, but in the present experiment, there were some additional settings, as follows. To avoid the degeneration of the ensemble for $\mathbf{x}_t$, we added system noise as in Eq. (2.8). The matrix $\mathbf{Q}$ was set to $0.01\mathbf{I}$. The ensemble covariance matrix of $\mathbf{x}_t$ was tapered with the correlation function $\mathbf{C}$ given by [37], where the correlation radius was set to 4.

For the VBEnKF, we selected $(\eta_c, \eta_0) = (200, 100)$ based on the results of the first experiment. The scale matrix $\mathbf{B}_0$ of $p(\mathbf{R}_0)$ was set such that $E(\mathbf{R}_0) = 5\mathbf{I}$. The estimated $\mathbf{R}_t$ was tapered with the correlation function $\mathbf{C}$, as discussed in Section 3.d, and the radius was set to 4.

For FK1 and FK2, we used algorithm 2 from [31]. To avoid the degeneration of the ensemble for the parameters of $\mathbf{R}_t$, we used the kernel resampling technique. From preliminary experiments, we selected $h = 0.2$ that controls the variance of resampled ensemble of the parameters. This is because the suggested value in [31], $h \approx 0.74$ for FK2, had resulted in highly fluctuated estimates of $\mathbf{R}_t$. The size of the ensemble for parameters of $\mathbf{R}_t$ was boosted as MN, where M is the boosting parameter and was set to 8. The initial ensemble for the parameters of $\mathbf{R}_0$ was generated from the same distribution $p(\mathbf{R}_0)$ as used for the VBEnKF.

TABLE 3.1: RMSE of the predicted $\mathbf{x}_t$ for each filtering method and the form of the true $\mathbf{R}_t$

	$\mathbf{S}_a$	$\mathbf{S}_b$	$\mathbf{S}_c$
Oracle1	0.548	0.636	0.603
Oracle2	0.532	0.621	0.585
FK1	0.535	0.650	0.692
FK2	0.567	0.659	0.714
VBEnKF	0.538	0.629	0.599

Results

We repeated the experiments 128 times, each with a distinct random seed for initializing the ensemble and generating the system noise. The results of each trial were then averaged.

The predicted RMSE of $\mathbf{x}_t$ for each method and $\mathbf{S}_z, z = \{a, b, c\}$ are summarized in Table 3.1. The temporal variation of the predicted mean of $\mathbf{R}_t$ is shown in Fig. 3.8.

Oracle2 with $\mathbf{R}_{true,t}$ gave the best results for all $\mathbf{S}_z$ (see the second row of Table 3.1), as was expected. This implies that an appropriate setting of $\mathbf{R}_t$ is crucial for estimating $\mathbf{x}_t$. In the following paragraphs, we will compare the results to the baseline of Oracle2.

First, we consider the results of $\mathbf{R}_a$. Note that VBEnKF and FK1 showed similar and better results than Oracle1, as can be seen in the first column of Table 3.1. These results came from the appropriate estimation of the time-variant $\mathbf{R}_t$ as shown in Fig. 3.8 (a) and (c). However, the RMSE of FK2 was worse than that of the others because FK2 failed to estimate $\mathbf{R}_t$ in the first half of the assimilation, as shown in Fig. 3.8 (b). In addition, the fluctuation of the estimated $\mathbf{R}_t$ was larger than those of the VBEnKF and FK1.

Second, we consider the results of $\mathbf{R}_b$ and $\mathbf{R}_c$. As shown in the second and third columns of Table 3.1, the VBEnKF had the best results, other than those of Oracle2. In particular, the difference between the VBEnKF and the others is prominent for $\mathbf{R}_c$, where the true observation error covariance matrix is correlated. Comparing the results in Fig. 3.8 (g), (h), and (i), we see that the VBEnKF is the only method that can estimate the nondiagonal $\mathbf{R}_t$. Moreover, as shown in Fig. 3.8 (i), the VBEnKF was able to estimate the time-variant $\mathbf{R}_t$. As shown in Fig. 3.8 (g) and (h), although FK1 and FK2 detected the step change of $\mathbf{R}_{true,t}$ at $t = 2000$, it failed to estimate the nondiagonal $\mathbf{R}_t$.

For the FK methods, the number of particles is a crucial design parameter especially in FK2, which estimates the twenty diagonal elements of $\mathbf{R}_t$. Therefore, we conducted additional experiments for FK2 by changing the boosting parameter M. The RMSE of the FK2 was improved by increasing M as shown in Table 3.2. For $\mathbf{R}_a$ and $\mathbf{R}_b$, although the RMSE of FK2 approached to those of the VBEnKF (see Table 3.1), it required a large number of particles. The results of estimating $\mathbf{R}_t$ for $M = 1024$ are shown in Fig. 3.9. Compared to Fig. 3.8 (b) and (e), FK2 succeeded in tracking $\mathbf{R}_{true,t}$, but the fluctuations were much larger than those of the VBEnKF (see Fig. 3.8 (c) and (f)).

Finally, we found that the average number of iterations required to reach convergence in the VBEnKF was < 5.2 for all $\mathbf{S}_z$.
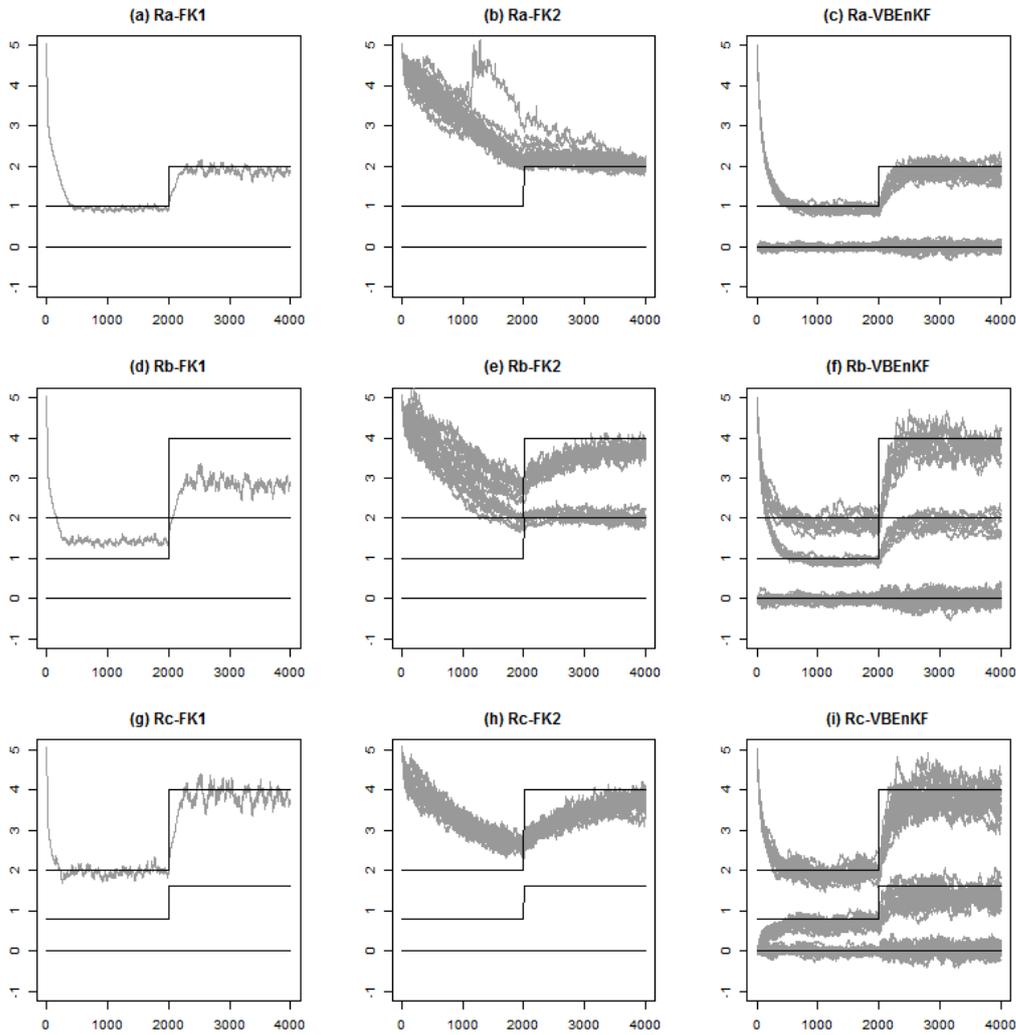
FIGURE 3.8: Temporal variation of the predicted mean of $\mathbf{R}_t$. The panels (a) to (l) show the various combinations of $\mathbf{S}_z$, $z = \{a, b, c\}$ and the filtering methods. The horizontal axis represents the assimilation step, and the vertical axis represents the elements of $\mathbf{R}_t$. The gray lines show the predicted means, and the solid lines show $\mathbf{R}_{true,t}$.

TABLE 3.2: RMSE of the predicted $\mathbf{x}_t$ in the case of FK2 for $M = 8, 64, 128, 256, 512,$ and 1024.

	$\mathbf{S}_a$	$\mathbf{S}_b$	$\mathbf{S}_c$
FK2, $M = 8$	0.567	0.659	0.714
FK2, $M = 64$	0.549	0.639	0.705
FK2, $M = 128$	0.545	0.636	0.703
FK2, $M = 256$	0.543	0.634	0.703
FK2, $M = 512$	0.542	0.635	0.703
FK2, $M = 1024$	0.542	0.633	0.704
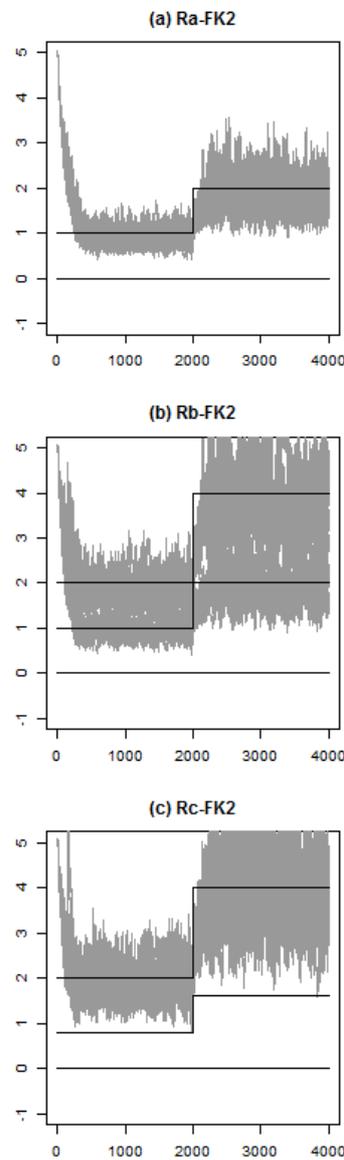
FIGURE 3.9: Temporal variation of the predicted mean of $\mathbf{R}_t$ in FK2 for $\mathbf{S}_z$, $z = \{a, b, c\}$, where $M = 1024$. The horizontal axis represents the assimilation step, and the vertical axis represents the elements of $\mathbf{R}_t$. The gray lines show the predicted means, and the solid lines show $\mathbf{R}_{true,t}$.

3.4.3   Stability when unmodeled disturbances are present

The question of stability of the estimation obtained with the VB method is not trivial. One reason for this is that the VB method may accumulate estimation errors because it approximates the joint posterior distribution. In this discussion, we will show experimental evidence that the VBEnKF estimate is stable for long runs when there are unmodeled disturbances.

With the Lorenz96 system, we examined two types of unmodeled disturbances: 1) model error, and 2) outliers in the observations. For the model error, we created an incorrect system model by assuming a forcing parameter $F = 8.2, 8.4, \cdots, 9.0$ in Eq. (3.35). Note that the synthetic data were generated for $F = 8.0$, as in Experiment 2. For the outliers in the observations, we contaminated the synthetic data in two ways. First, we used the following contaminated observations:

$$
\mathbf{y}_{cont.,t} = \begin{cases} \begin{bmatrix} 10 & 10 & \cdots & 10 \end{bmatrix}^{\mathrm{T}} & (r < 0.01) \\ \mathbf{y}_t & (otherwise) \end{cases}, \tag{3.38}
$$

where $\mathbf{y}_{cont.,t}$ is the contaminated observation vector, and r is a realization from the uniform distribution bounded in $[0, 1]$. Second, we randomly increased $\mathbf{R}_t$ when generating the synthetic data as follows:

$$
\mathbf{R}_{t,true} = \begin{cases} 100\mathbf{S}_0 & (r < 0.01) \\ \mathbf{S}_0 & (otherwise) \end{cases}, \tag{3.39}
$$

where $\mathbf{S}_0$ is as defined in Experiment 2. To obtain a baseline, we conducted an Oracle experiment that used the fixed $\mathbf{S}_0$ for each of the assimilation steps. The number of assimilation steps was set to $T = 10^5$.

The resultant time-averaged RMSE and $\mathrm{tr}\,\mathbf{R}_t$ are summarized in Table 3.3. For the RMSE (see the fourth and fifth columns of Table 3.3), we note that the VBEnKF was stable for long runs because in no case was the RMSE notably larger than that obtained by Oracle. Indeed, in some cases (see cases (v)–(viii) in Table 3.3), the RMSE of the VBEnKF was smaller than that of Oracle. In these cases, a large $\mathrm{tr}\,\mathbf{R}_t$ was estimated by the VBEnKF (see sixth and seventh columns of Table 3.3). We assume that with the VBEnKF, over-fitting of the unmodeled disturbances was avoided due to the large $\mathbf{R}_t$.

Here, we can see that $\mathbf{R}_t$ is correctly estimated for $F = 8.6$, where there is an medium magnitude of the model error, but that is underestimated in the case of the true $F\,(= 8)$. The reason is considered in Section 3.4.7.

3.4.4   Estimation of the hyperparameter ξ of the system model for $\mathbf{R}_t$

We demonstrate the ML estimation of η_c with the Lorenz96 system. We examined four cases by changing $E(\mathbf{R}_0)$ and $\mathbf{R}_{t,true}$, as summarized in Table 3.4. For $E(\mathbf{R}_0)$, we examined $5\mathbf{I}$(far from $\mathbf{R}_{0,true}$) and the true value at the first assimilation step $\mathbf{R}_{1,true}$. For $\mathbf{R}_{t,true}$, we examined time-invariant and time-variant cases. The matrix was set to $\mathbf{R}_{t,true} = \mathbf{S}_0 (t = 1, \cdots, T)$ for the time-invariant case, and for the time-variant case, $\mathbf{R}_t$ was doubled from $t = T/2$, as in the previous experiments.

The purpose of evaluating these cases is to show that an appropriate η_c can be estimated for different problems. For example, a large η_c is expected when it is not necessary to change the estimated $\mathbf{R}_t$.

TABLE 3.3: RMSE of the predicted $\mathbf{x}_t$ and $\operatorname{tr}\mathbf{R}_t$. Case (i): without an unmodeled disturbance; Cases (ii)–(vi): with model error; and Cases (vii)–(viii): with outliers in the observations.

	case	disturbance	RMSE		averaged $\operatorname{tr}\mathbf{R}_t$	
			Oracle	VBEnKF	Oracle	VBEnKF
None	(i)	-	0.486	0.494	40	38.3
Model error	(ii)	F=8.2	0.500	0.507	40	38.4
	(iii)	F=8.4	0.544	0.549	40	38.9
	(iv)	F=8.6	0.618	0.619	40	39.8
	(v)	F=8.8	0.722	0.717	40	41.2
	(vi)	F=9.0	0.862	0.843	40	43.5
Outliers	(vii)	large variance	0.880	0.580	40	78.7
	(viii)	constant value	0.589	0.542	40	51.6

TABLE 3.4: Experimental case number

	$\mathbf{R}_{t,true}$	
	time invariant	time variant
$E(\mathbf{R}_0) = 5\mathbf{I}$	(i)	(ii)
$E(\mathbf{R}_0) = \mathbf{R}_{1,true}$	(iii)	(iv)

The experimental setting was the same as in Section 4.b. We calculate the likelihood and the RMSE for $\eta_c = \{50, 65, 80, 100, 200, \cdots, 1000\}$.

The results are summarized in Table 3.5. In case (iii), the ML estimate was $\eta_c = 1000$. This is because there is no initial difference and no temporal change in $\mathbf{R}_t$. As a result, the largest value of η_c, which indicates the certainty of the system model of $\mathbf{R}_t$, resulted in the maximum value of the likelihood. In cases (i), (ii), and (iv), the ML estimate was $\eta_c = 200$ in order to correct the initial difference or the temporal changes in $\mathbf{R}_t$ during the assimilation.

Roughly speaking, we can say that there are correspondences between the predictive log-likelihood and the RMSE. Although the true $\mathbf{x}_t$ is, of course, unknown in actual applications, we can optimize η_c with the predictive likelihood.

3.4.5 Comparison with the exact filtered distribution

The proposed method depends on the variational Bayes (VB) method that approximates the joint posterior distribution with independent distributions. A natural concern is the effect of the assumption to the filtered estimate. In this subsection, we compare the proposed method with the PF. The PF can numerically calculate the exact joint posterior because it does not use any assumption in the filtering algorithm.

Experimental setting

For the experiment, we use a 1-dimensional system for generating synthetic data defined as follows:

$$y_t \sim \mathcal{N}(y_t | 5, 5), \tag{3.40}$$

With this system, we generate a synthetic dataset with $T = 100$ steps.

TABLE 3.5: Time-average of the predictive log-likelihood (lnL) and root mean square error (RMSE) of the predicted $\mathbf{x}_t$. The best values in each column are marked with bold font.

η_c	case (i)		case (ii)		case (iii)		case (iv)	
	lnL	RMSE	lnL	RMSE	lnL	RMSE	lnL	RMSE
40	-46.406	2.922	-50.152	3.238	-46.670	2.970	-50.483	3.301
60	-35.822	0.529	-39.235	0.641	-35.808	0.529	-39.221	0.640
80	-35.240	0.510	-38.661	0.608	-35.197	0.508	-38.618	0.607
100	-35.116	0.506	-38.543	0.601	-35.046	0.503	-38.474	0.598
200	**-35.081**	**0.504**	**-38.542**	**0.599**	-34.921	0.497	**-38.382**	0.593
300	-35.123	0.505	-38.610	0.600	-34.907	0.496	-38.395	**0.592**
400	-35.159	0.506	-38.672	0.601	-34.903	**0.495**	-38.420	**0.592**
500	-35.190	0.507	-38.724	0.602	-34.902	**0.495**	-38.448	**0.592**
600	-35.215	0.507	-38.768	0.603	-34.901	**0.495**	-38.477	**0.592**
700	-35.237	0.508	-38.807	0.603	-34.901	**0.495**	-38.506	0.593
800	-35.255	0.508	-38.841	0.604	-34.901	**0.495**	-38.534	0.593
900	-35.271	0.509	-38.871	0.605	**-34.900**	**0.495**	-38.563	0.593
1000	-35.284	0.509	-38.898	0.605	**-34.900**	**0.495**	-38.590	0.593

The initial distribution and the SSM are as follows:

$$
\begin{aligned}
p(x_0)p(R_0) &= \mathcal{N}(0, 10)\mathcal{IW}(6.1, 21), \\
p(x_t|x_{t-1})p(R_t|R_{t-1}) &= \delta(x_t - x_{t-1})\delta(R_t - R_{t-1}), \\
p(y_t|x_t, R_t) &= \mathcal{N}(y_t|x_t, R_t).
\end{aligned}
\tag{3.41}
$$

Note that $\delta(R_t - R_{t-1})$ is the limit of Eq. (3.5) as $\xi \to \infty$. The intention is that we only calculate the update step to clearly compare the methods with and without independence assumption in estimating the joint filtered (posterior) distribution. For the PF, the number of particles was set to 2×10^5, and the ensemble size is set to 32 for the EnKF in the proposed method.

Results

Using the synthetic data and the SSM, we conduct a filtering experiment with the PF and the proposed method.

The resultant filtered distributions $p(x_t, R_t|y_{1:t-1})$ for the first nine assimilation cycles are shown in Figs. 3.10 and 3.11.

For the result of the PF, as can be seen in Fig. 3.10 (a) prominently, there is a relation of inverse proportion between x_t and R_t. That is, x_t far from the true x_t is justified with large R_t. At the fourth step shown in Fig. 3.10 (d), the estimated mean values are very close to the true values.

For the result of the proposed method, since there is the independence assumption between x_t and R_t in the filtering, the relation estimated by the PF cannot be seen in Fig. 3.11. In addition, the estimated joint distribution is broader than that of the PF. Although the mean of the filtered distribution approaches to the true values, the speed is slower than that of the PF.

In Fig. 3.12, the marginalized filtered distributions for all assimilation cycles are plotted for x_t and R_t separately. For the estimation of x_t, except the earliest steps, the trajectories are almost the same for both methods as shown in Fig. 3.12 (a) and

(c). The RMSEs between the mean of x_t and the true x_t by the PF and the proposed method were 0.302 and 0.331, respectively. For R_t, the PF could estimate the true value in the early few steps as shown in Fig. 3.12 (b). However, in the case of the proposed method, the estimate slowly approaches to the true mean as shown in Fig. 3.12 (d). Although the proposed method takes longer time steps and overlooks the relation between x_t and R_t, the mean values converge to the true value.
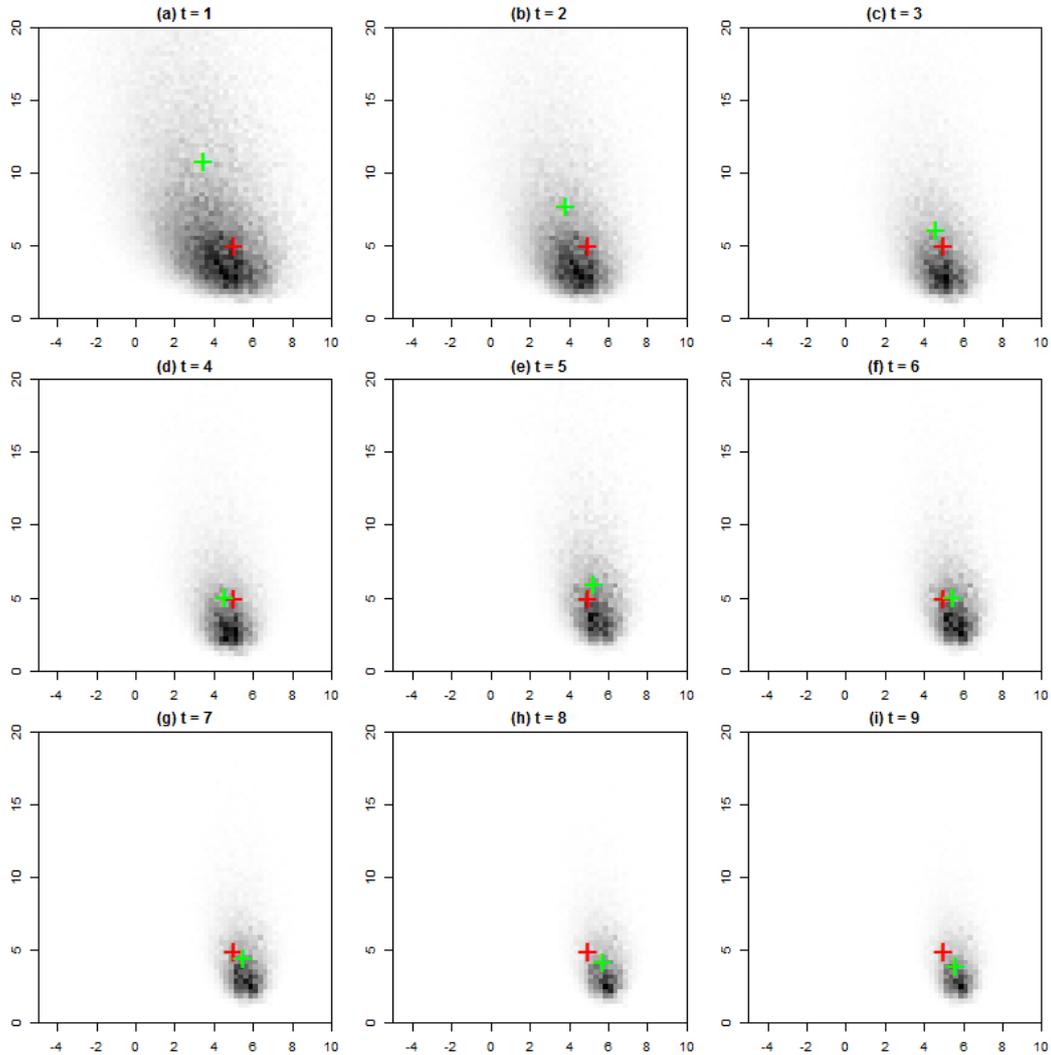
FIGURE 3.10: Filtered distributions for x_t and R_t by the PF. The horizontal axis denotes x_t, and the vertical axis denotes R_t. The green cross represents the mean of the filtered distribution, and the red cross represents the true values.

3.4.6 Estimation of various $\mathbf{R}_t$

Experimental setting

So far, we have only dealt with $\mathbf{R}_t$ with simple structure. In this subsection, we demonstrate capability of the proposed method for $\mathbf{R}_t$ with complex structure. For
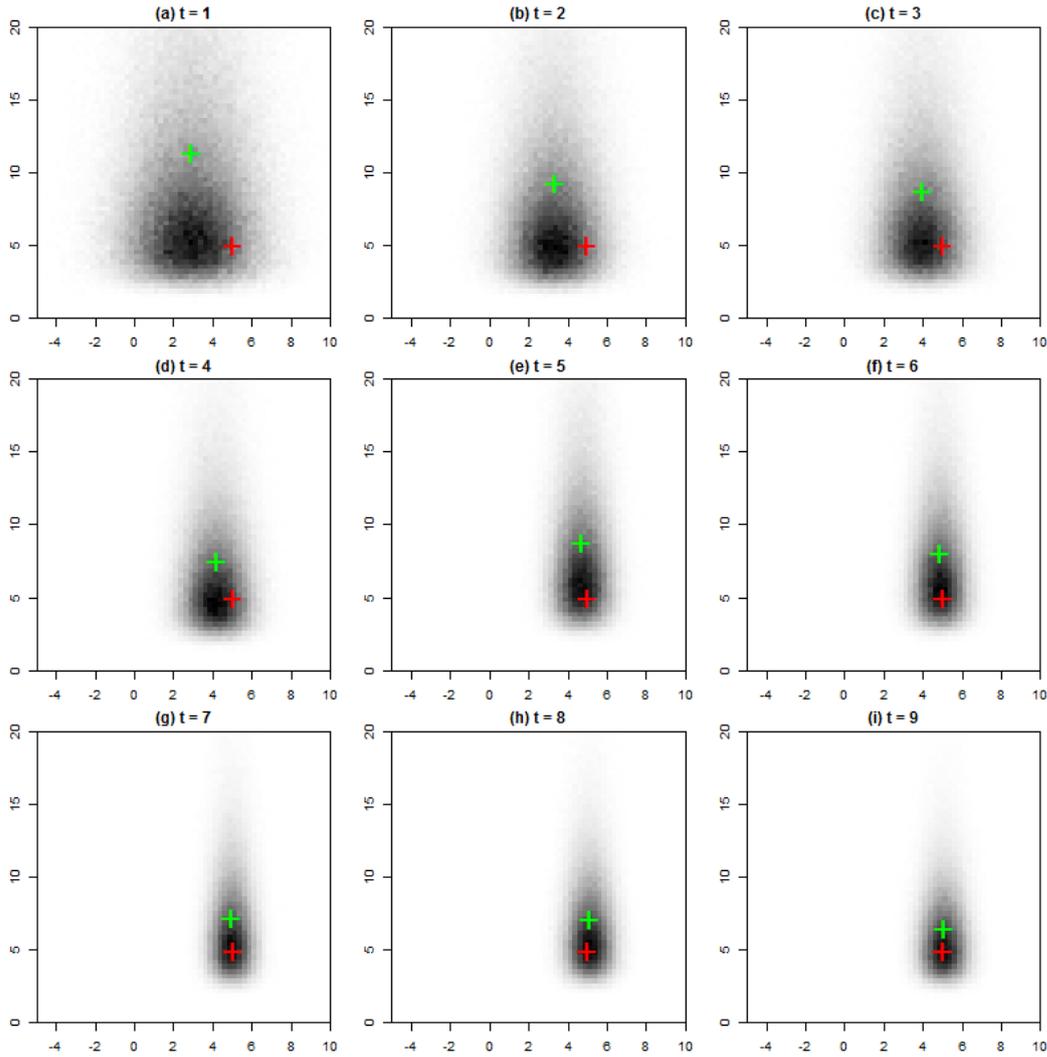
FIGURE 3.11: Filtered distributions for x_t and R_t by the proposed method. The horizontal axis denotes x_t, and the vertical axis denotes R_t. The green cross represents the mean of the filtered distribution, and the red cross represents the true values.

the experiment, we use 100-dimensional synthetic datasets generated by the following model.

$$\begin{aligned} \mathbf{y}_t &\sim \mathcal{N}(\mathbf{y}_t | \mathbf{0}, \mathbf{R}_{true}) \\ \mathbf{R}_{true} &\sim \mathcal{IW}(\mathbf{R}_{true} | 203, \mathbf{I}), \end{aligned} \qquad (3.42)$$

where $\mathbf{R}_{true}$ is the true observation error covariance matrix.

Note that the degree of freedom parameter of $\mathcal{IW}$ is set to the minimum integer that satisfy $> 2m + 2 = 202$, and the scale matrix is selected such that the expectation become $\mathbf{I}$. For the proposed method, we use the following initial distributions and
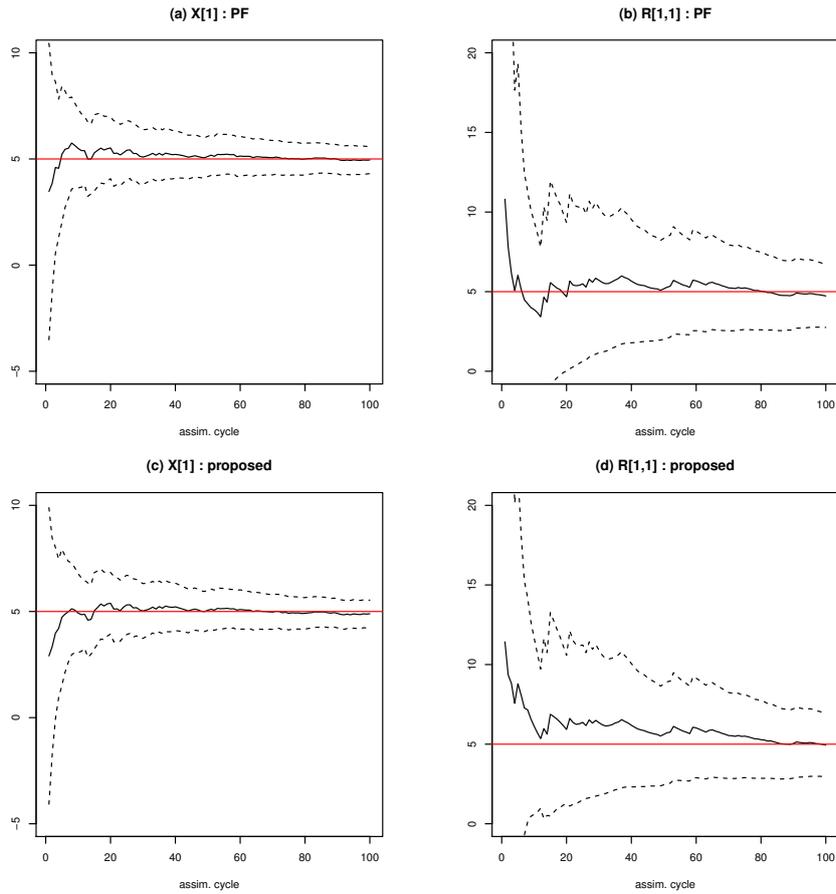
FIGURE 3.12: Temporal variation of the marginalized predictive distribution for x_t and R_t. Panels (a) and (b) are the mean and the 3-sigma band for x_t and R_t by the PF, respectively. Panels (c) and (d) are those of the proposed method. Center red lines denote the true values.

SSM as,

$$
\begin{aligned}
p(\mathbf{x}_0) &= \mathcal{N}(\mathbf{x}_0|[0.5, \cdots, 0.5]^{\mathrm{T}}, \mathbf{I}), \\
p(\mathbf{R}_0) &= \mathcal{IW}(\mathbf{R}_0|203, \mathbf{I}), \\
p(\mathbf{x}_t|\mathbf{x}_{t-1}) &= \mathcal{N}(\mathbf{x}_t|, \mathbf{x}_{t-1}, 0.1^2\mathbf{I}), \\
p(\mathbf{y}_t|\mathbf{x}_t) &= \mathcal{N}(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}_t).
\end{aligned}
\tag{3.43}
$$

The hyperparameters are as follows: the ensemble size and η_c are set to 256 and 50 ($\xi = 23764$), respectively.

Results

For three synthetic datasets with 500 obseravtions generated from different random seeds, we conduct filtering experiments with the proposed method. The results for three cases are shown in Figure 3.13.

In each row in Figure 3.13, there are two panels: colored plots of estimates of $\mathbf{R}_{t=500}$, and colored plots of $\mathbf{R}_{true}$. Although the three cases use observation error

covariance with complex structures, the proposed method almost correctly estimate the true matrix in the filtering.

3.4.7 The effect of misspecified model and system noise

In Section 3.4.3, we have observed that $\mathbf{R}_t$ was underestimated in the case of the true forcing parameter F of *Lorenz96*, but that $\mathbf{R}_t$ was correctly estimated in the case of larger F. To clarify the phenomenon, we show further consideraton with a simple problem.

Analytical property

The proposed method is derived as a result of maximizing $\mathcal{F}$, which is the lower bound of the logarithm of conditional marginal likelihood $\ln p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$. Thus, we consider a maximization problem of $\ln p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$.

For 1-dimensional system, Let us assume a 1-dimensional system with the following generative models as

$$
\begin{aligned}
x_{t,true} &= x_{t-1,true} = 1, \\
y_t &\sim \mathcal{N}(y_t|x_{t,true}, R_{true} = 1),
\end{aligned}
\tag{3.44}
$$

where $x_{t,true}$ and R_{true} is the true system state and observation error variance. For filtering, we assume the following $(t-1)$th filtered distribution for x_{t-1} and SSM as

$$
p(x_{t-1}|y_{1:t-1}) = \mathcal{N}(x_{t-1}|\mu_{t-1|t-}, V_{t-1|t-1}),
\tag{3.45}
$$

$$
p(x_t|x_{t-1}) = \mathcal{N}(x_t|\alpha x_{t-1}, Q),
\tag{3.46}
$$

$$
p(y_t|x_t) = \mathcal{N}(y_t|x_t, R),
\tag{3.47}
$$

where α is a parameter that define time-evolution of the state vector, Q and R are the model error and observation error covariance matrices, respectively.

With Eqs. 3.45-3.47, the log marginal likelihood of y_t conditioned on $y_{1:t-1}$ with respect to x_t is given as

$$
\begin{aligned}
\ln p(y_t|y_{1:t-1}) = &-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln(R + V_{t|t-1}) \\
&-\frac{1}{2}(y_t - \mu_{t|t-1})^2(R + V_{t|t-1})^{-1},
\end{aligned}
\tag{3.48}
$$

Thus, a likelihood function $L(Q, R, \alpha, \mu_{t-1|t-1}, V_{t-1|t-1}|x_{t,true}, R_{true})$ that is defined as the expectation of Eq. 3.48 on Eq. 3.44 is given as

$$
\begin{aligned}
&L(Q, R, \alpha, \mu_{t-1|t-1}, V_{t-1|t-1}|x_{t,true}, R_{true}) \\
&= \langle \ln p(y_t|y_{1:t-1}) \rangle_{\mathcal{N}(y_t|1,1)} \\
&= -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln(R + \alpha^2 V_{t-1|t-1} + Q) \\
&\quad -\frac{1}{2}\{(x_{true} - \alpha\mu_{t-1|t-1})^2 + R_{true}\}(R + \alpha^2 V_{t-1|t-1} + Q)^{-1}.
\end{aligned}
\tag{3.49}
$$

Simulation

With the likelihood function Eq. (3.49), we simulate the optimization of R for some cases. Here, we assume that $(t-1)$th filtered distribution is correctly estimated without uncertainty, that is, $\mu_{t-1|t-1} = x_{t-1,true}$ and $V_{t-1|t-1} = 0$.
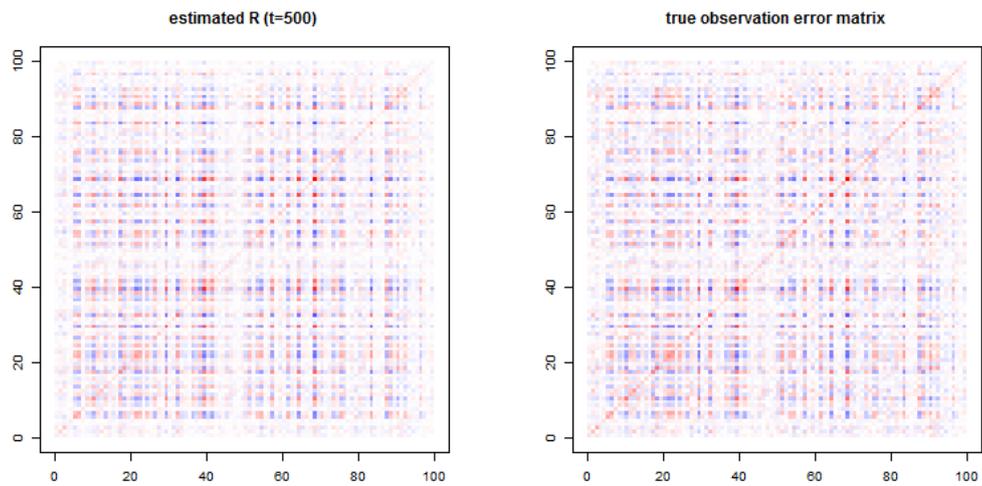
(A) dataset1 1
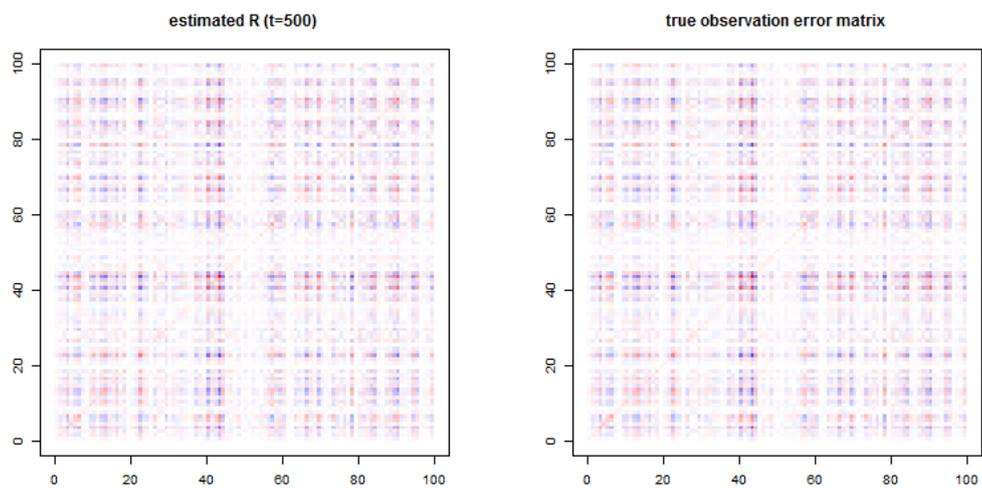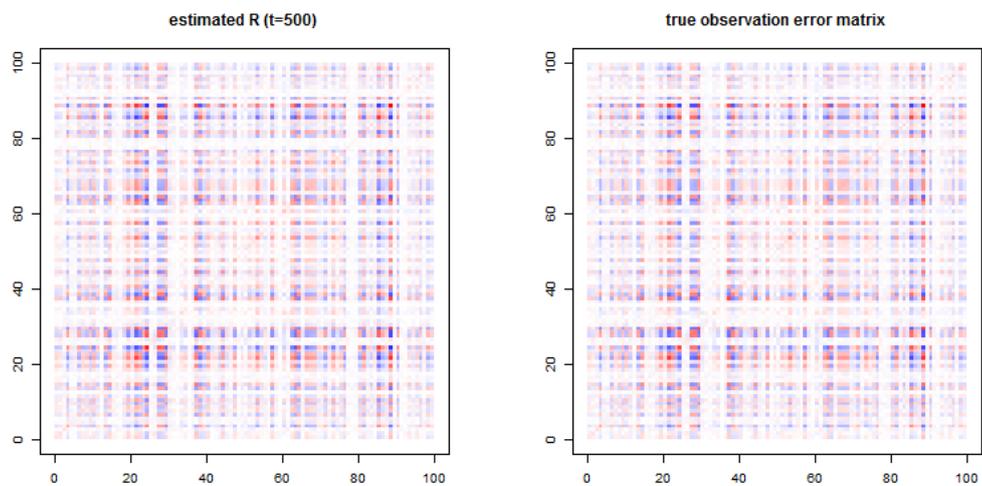
(B) dataset 2

(C) dataset 3

FIGURE 3.13: Examples of estimation for complex observation error covariance matrices. In each row, RGB(0,0,1) and RGB(1,0,0) correspond to the minimum and the maximum values of elements of covariance matrices.
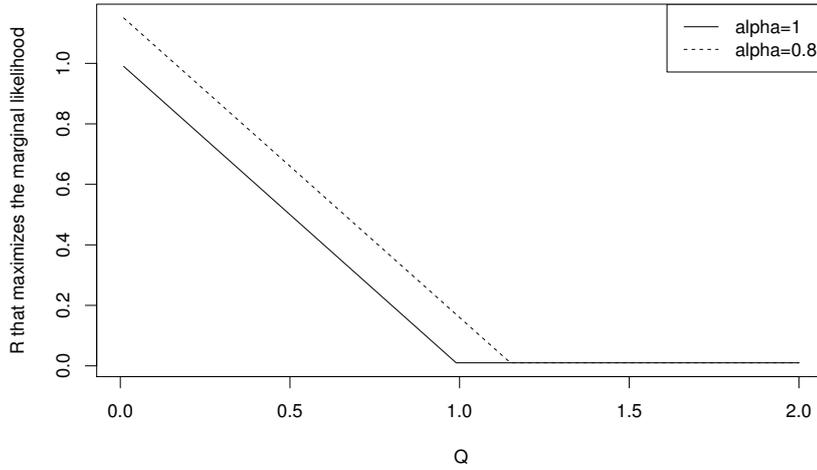
FIGURE 3.14: The maximizer R for Eq. (3.49)

For each $(Q, \alpha) \in \{0.01, 0.02, , \cdots, 2.00\} \times \{1.0, 0.8\}$, we optimize Eq. (3.49) with respect to $R \in \{0.01, 0.02, \cdots, 2.00\}$. Note that $\alpha = 1$ represents the case without model error, and $\alpha = 0.8$ represents the case with model error.

The maximizer R and the maximized likelihood are shown in Figs. 3.14 and 3.15. The implications in the maximization of the likelihood Figs. 3.14 and 3.15 are as follows:

- The log-maximum marginal likelihood takes the same and maximum values for small Q as shown in Fig. 3.15

- For the maximum values, there is a linear dependency $Q + R = const.$ as shown in Fig. 3.14. Especially, in the case $\alpha = 1$, the dependency seems to be $Q + R = R_{true}(= 1)$.

- Misspecified model (represented by $\alpha = 0.8$) increases the sum $Q + R$ with the decrease of the log-maximum marginal likelihood.

- For large Q, the maximizer R approaches to 0 (the lowest value 0.01 in this case) and the log-maximum likelihood decreases along with the increase of Q.

Although this is a result of a simple 1-dimensional and linear example, the tendency is expected to be presented in a nonlinear system as discussed in Section 3.4.3. From the second implication that constraints Q and R, we can guess that $\mathbf{R}_t$ was underestimated in the case without model error in Table 3.3 because Q is set to 10^{-2} to prevent the degeneration of the ensemble and make the filtering stable. Furthermore, from the third implication, in the case where $F = 8.6$, $\mathbf{R}_t$ was correctly estimated by incorporating the effect of model error.

3.5 Conclusions

We have proposed an extension of the EnKF for estimating the state vector $\mathbf{x}_t$ and the observation error covariance matrix $\mathbf{R}_t$ by using the VB method. In numerical
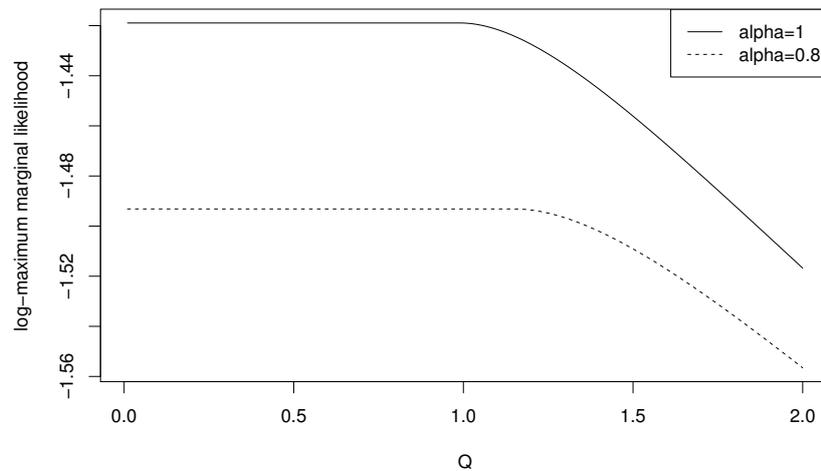
FIGURE 3.15: The log-maximum marginal likelihood Eq. (3.49)

experiments, we have examined the capabilities of our method, especially for a time-variant observation error covariance matrix. We note that the experiments show that our method works well even when the true observation error covariance matrix is nondiagonal.

We have also presented two complementary studies. First, we have demonstrated the stability of our method for long-run assimilations when there are unmodeled disturbances. In some such cases, our method outperformed the EnKF with true $\mathbf{R}_t$, because it avoided over-fitting by involving the unmodeled disturbance as the observation error. Second, we have derived an ML method for optimizing the hyperparameters, and, using this method, we have experimentally examined the possibility of estimating one of the hyperparameters.

Chapter 4

Nonlinear Kalman filters for outliers

This chapter is based on an article,

Akio Nakabayashi, Genta Ueno, "Nonlinear Filtering Method Using a Switching Error Model for Outlier-Contaminated Observations."

The article is submitted to *IEEE Transactions on Automatic Control Technical Note* on April 27, 2017. At the time of August 1, 2017, the article is kindly being reviewed.

4.1 Introduction

It is well known that the accuracy of estimates of the system state by such methods is greatly reduced when outliers are present in the observations. For outliers, the resulting estimates of the system state may be largely biased as a result of over-fitting to the outliers. This is a common issue when assuming Gaussian errors, that is, evaluating errors with a quadratic loss criterion [9].

The occurrence of outliers in observations is very common in real applications due to, for example, sensor malfunctions, large disturbances to the target system, and human errors when inputting data. Therefore, how to handle outliers in the filtering is an active area of research. Former studies can be classified into two approaches: one that adopts a static heavy-tailed observation model and another that adopts a dynamic observation model.

In the static approach, a stationary heavy-tailed distribution is assumed for observation errors, such as the t-distribution[17] or the Cauchy distribution [18]. By using such heavy-tailed distributions, over-fitting to outliers can be avoided because the sensitivity to large deviations is reduced. Although a probability distribution is not explicitly assumed, studies on the basis of robust estimation that redefines the loss function for the innovation can also be regarded as static approaches and have been widely reported. For example, Durovic and Kovacevic [19] used the Huber function [41], which combines the $l1$ and $l2$ criterions. Although the method reported in [19] is for linear SSMs, there are studies that have applied robust estimation methods for nonlinear SSMs using the UKF [20], [21]. Because they only require a few design parameters for representing a heavy-tail, robust filtering algorithms are straightforward to implement. However, the parameters must be tuned for each filtering problem. In addition, the application of a heavy-tailed distribution to all observations might cause under-fitting because the heavy tail also reduces the sensitivity to observations with regular errors.

In the dynamic approach, a time-variant observation model, with non-constant variance, is assumed in order to reduce the influence of outliers. There are broadly two kinds of methods: one uses a single observation model, and another uses multiple observation models with distinct error variances and switches the models.

Among the methods using a single observation model, there are studies that assume a Gaussian observation error and estimate the covariance matrix simultaneously with the system state. Ting *et al.* have proposed a method for estimating the scale parameter of the observation error covariance matrix using Bayesian estimation from a gamma prior [14]. For correlated outliers, Agamennoni *et al.* estimated the covariance elements using the Wishart prior [22]. As well as the methods for linear SSMs [14], [22], Särkkä and Hartikainen proposed a method for nonlinear SSMs using the Gaussian filter [16]. However, since the method in [16] assumes that the observation error statistics change slowly, it is not appropriate for handling outliers, which are randomly occurring gross errors. Although effective for avoiding overfitting to outliers, the methods in [14], [22] use a non-informative prior at each time step, which can result in fluctuations in the estimated error covariance matrix even for outlier-free observations. This leads to a decrease in the accuracy in the state estimation because of under- or over-fitting to observations.

Among the methods using several observation models, switching techniques that select one model from a set of SSMs at each time step have been proposed [42], [43]. The switching techniques can handle outliers by including an observation model with a large variance in the set. In addition, since the variance of observation errors is determined from several candidates, large fluctuations in the variance can be avoided in contrast to methods that freely estimate the variance parameter at each time step. However, in general, we cannot know the properties of outliers before conducting the filtering, and thus it is necessary to determine the appropriate set of SSMs by trial and error. Studies have formulated the estimation of parameters of a set of SSMs as a batch optimization problem for a time series of observations [44], [45]. However, in real applications, it is often infeasible to solve the batch optimization problem because of limited computing and data storage resources, or demands for real-time responses. Moreover, in the context of switching techniques, filtering with nonlinear SSMs has not yet been considered.

Against this background, we propose a new dynamic method for handling outliers in the nonlinear filtering problem by extending the UKF. The key idea is to switch between models with recursive estimation of a parameter in the model. We use two Gaussian observation models that have distinct observation error covariance matrices: one is for observations with regular errors, and another, which has a larger variance specified by a scale parameter, is for outliers. In addition to the system state, the method estimates an indicator variable that determines which of the two models to use and the scale parameter for outliers. By estimating the indicator variable and the scale parameter, estimates of the observation error covariance matrix that can handle both regular observations and outliers are obtained at each time step. Furthermore, due to the estimation of the scale parameter, the proposed method can be applied to a problem without requiring tuning to account for outlier characteristics. Through numerical experiments, we demonstrate that our proposed method can estimate the system state better than existing methods for datasets with and without outliers.

4.2 The proposed method

4.2.1 Formulation

We assume an observation error $\mathbf{w}_t \in \mathbb{R}^m$ is a random variable that has two different characteristics. That is,

$$\mathbf{w}_t \sim \begin{cases} p_{regular}(\mathbf{w}_t) = \mathcal{N}(\mathbf{w}_t|\mathbf{0}, \mathbf{R}) & (regular\ observation) \\ p_{outlier}(\mathbf{w}_t) & (outlier) \end{cases}, \qquad (4.1)$$

where $\mathcal{N}$ denotes a Gaussian distribution, and $\mathbf{R} \in \mathbb{R}^{m \times m} \succ 0$ is an observation error covariance matrix.

Our purpose is to detect outliers based on the difference in observation error characteristics. For outliers, we additionally assume two following characteristics: (1) the occurrence of outliers at each time step is decides by i.i.d. Bernoulli trials, and (2) $Var[p_{outlier}(w_t)] \gg Var[p_{regular}(w_t)]$. By introducing two other random variables, the observation error model in this chapter is formally assumed as

$$p(\mathbf{w}_t) = \mathcal{N}(\mathbf{w}_t|\mathbf{0}, \mathbf{R})^{1-s_t} \mathcal{N}(\mathbf{w}_t|\mathbf{0}, \gamma_t \mathbf{R})^{s_t}, \qquad (4.2)$$

where $s_t \in \{0 : notoutlier, 1 : outlier\}$ is the outlier indicator that switches the two Gaussians and is defined at each time step, and $\gamma_t > 0$ is the scale parameter of $\mathbf{R}$. Note that this observation error model is reduced to a Gaussian error assumed in Eq. (2.5) when $s_t = 0$. If $s_t = 1$ and $\gamma_t \gg 1$, this model becomes Gaussian that has a larger covariance matrix than $\mathbf{R}$. We expect that the first Gaussian is used for the regular observations and the second Gaussian is used for outliers by appropriately estimating s_t and γ_t.

Since the occurrence of outlier is not predictable and the characteristics is not trivial, we have developed a filtering algorithm for $\mathbf{x}_t$, s_t, and γ_t by calculating the posterior of these random variables. Similarity to $\mathbf{x}_t$, we assume that γ_t is a realization of the first-order Markov process, while s_t at each time step are independent of each other due to the i.i.d. assumption which is already mentioned. Thus, the exact calculation is given in two steps as follows:

$$p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{x}_{t-1}, s_{t-1}, \gamma_{t-1}) p(\mathbf{x}_{t-1}, s_{t-1}, \gamma_{t-1}|\mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1} ds_{t-1} d\gamma_{t-1}, \qquad (4.3)$$

$$p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t, s_t, \gamma_t) p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{y}_{1:t-1}), \qquad (4.4)$$

where $p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{x}_{t-1}, s_{t-1}, \gamma_{t-1})$ and $p(\mathbf{y}_t|\mathbf{x}_t, s_t, \gamma_t)$ are the system model and the observation model, respectively, $p(\mathbf{x}_{t-1}, s_{t-1}, \gamma_{t-1}|\mathbf{y}_{1:t-1})$ is the $(t-1)$th joint filtered distribution, $p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{y}_{1:t-1})$ is the tth joint predictive distribution, and $p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{y}_{1:t})$ is the tth joint filtered distribution. By specifying models in Eqs. (4.3)-(4.4), we have derived a nonlinear filtering algorithm for dealing with outliers.

4.2.2 State-space model

In this subsection, we define a system model for $\mathbf{x}_t$, s_t, and γ_t. In the time evolution, we assume the independency among the random variables, that is,

$$p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{x}_{t-1}, s_{t-1}, \gamma_{t-1}) = p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(s_t|s_{t-1}) p(\gamma_t|\gamma_{t-1}). \qquad (4.5)$$
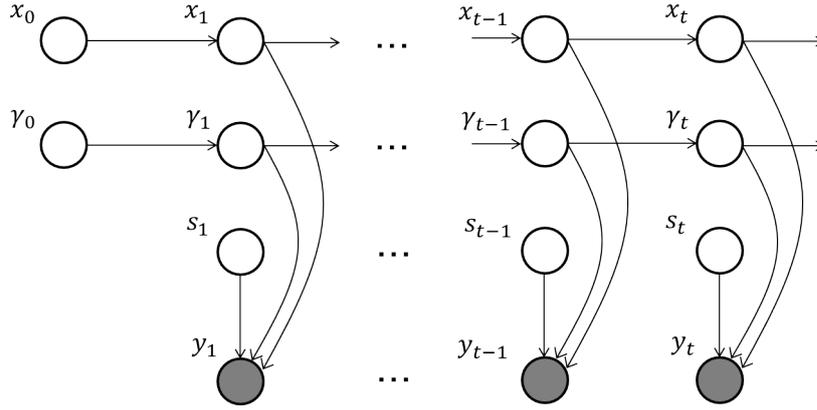
FIGURE 4.1: Graphical model representation of the SSM of the proposed method. Gray circles represent observable variables, and white circles represent latent variables that are not observable.

Due to the independence, we can construct system models for each random variables, separately. For $\mathbf{x}_t$, we use the same model as the UKF. Again, that is,

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t|\mathbf{f}(\mathbf{x}_{t-1}), \mathbf{Q}), \tag{4.6}$$

where $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$ is the nonlinear function, and $\mathbf{Q} \in \mathbb{R}^{d \times d} \succ 0$ is the covariance matrix of the system noise.

For s_t, by assuming that outliers occur at random, s_t is independent of s_{t-1}. Thus, we assume the system model of s_t as,

$$p(s_t|s_{t-1}) = Bern(s_t|\lambda_{prior}), \tag{4.7}$$

where $Bern$ is the Bernoulli distribution, and $0 < \lambda_{prior} < 1$ is a prior probability of outlier occurrence, which is a constant. Since we do not have any idea about outlier occurrence, we will use a neutral value $\lambda_{prior} = 0.5$ throughout this chapter.

For γ_t, which is the outlier scale against the regular observation error, we assume that the characteristics is time-invariant. That is, the system model of γ_t is given as

$$p(\gamma_t|\gamma_{t-1}) = \delta(\gamma_t - \gamma_{t-1}), \tag{4.8}$$

where δ is Dirac's delta function.

Finally, we define an observation model by assuming the additive error defined by Eq. (4.1) as

$$p(\mathbf{y}_t|\mathbf{x}_t, s_t, \gamma_t) = \mathcal{N}(\mathbf{y}_t|\mathbf{h}(\mathbf{x}_t), \mathbf{R})^{1-s_t} \mathcal{N}(\mathbf{y}_t|\mathbf{h}(\mathbf{x}_t), \gamma_t \mathbf{R})^{s_t}. \tag{4.9}$$

Similar to the observation error, Eq. (4.1), this model reduces to a typical nonlinear observation model given by Eq. (2.9) for $s_t = 0$. If $s_t = 1$, this model reduces to an nonlinear observation model with large variance scaled by γ_t.

The assumed SSM is summarized by as follows:

$$
\begin{aligned}
p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{x}_{t-1}, s_{t-1}, \gamma_{t-1}) &= \mathcal{N}(\mathbf{x}_t|\mathbf{f}(\mathbf{x}_{t-1}), \mathbf{Q}) Bern(s_t|\lambda_{prior}) \delta(\gamma_t - \gamma_{t-1}), \\
p(\mathbf{y}_t|\mathbf{x}_t, s_t, \gamma_t) &= \mathcal{N}(\mathbf{y}_t|\mathbf{h}(\mathbf{x}_t), \mathbf{R})^{1-s_t} \mathcal{N}(\mathbf{y}_t|\mathbf{h}(\mathbf{x}_t), \gamma_t \mathbf{R})^{s_t}.
\end{aligned}
\tag{4.10}
$$

The SSM is illustrated as a graphical model in Fig. 4.1. Note that, as seen in Fig.

4.1, s_t does not depend on s_{t-1}, while $\mathbf{x}_t$ and γ_t depends on the previous random variables. With this SSM and given $(t-1)$th filtered distribution, we derive tractable forms of Eqs. (4.3) and (4.4).

4.2.3 Prediction

At first, we define the $(t-1)$th filtered distribtuion. Since we will extend the algorithm of the UKF, we use a Gaussian distribution for $\mathbf{x}_t$. For s_t, which is an indicator whether $\mathbf{y}_t$ is outlier or not, we assume a Bernoulli distribution that is defined for $\{0, 1\}$. For γ_t, which is the scale of covariance matrix $\mathbf{R}$, we assume a gamma distribution defined for non-negative real numbers. Thus, the $(t-1)$th joint filtered distribution for $\mathbf{x}_t$, s_t, and γ_t is

$$
\begin{aligned}
p(\mathbf{x}_{t-1}, s_{t-1}, \gamma_{t-1}|\mathbf{y}_{1:t-1}) = \\
\mathcal{N}(\mathbf{x}_{t-1}|\mu_{t-1|t-1}, \mathbf{V}_{t-1|t-1})) \cdot Bern(s_{t-1}|\lambda_{t-1}) \cdot IG(\gamma_{t-1}|a_{t-1|t-1}, b_{t-1|t-1}),
\end{aligned}
\tag{4.11}
$$

where $\mu \in \mathbb{R}^d$ is the mean vector of the state vector, $\mathbf{V} \in \mathbb{R}^{d \times d} \succ 0$ is the covariance matrix of the state vector, $Bern$ denotes a Bernoulli distribution, $0 < \lambda < 1$ is a parameter of a Bernoulli distribution which is the probability of outlier occurrence in this context, IG denotes an inverse gamma distribution, and $a > 0$ and $b > 0$ are shape and scale parameters of an inverse gamma distribution, respectively.

Since we assume the independence among random variables as Eq. (4.5), the prediction can be performed separately as

$$
\begin{aligned}
p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{y}_{1:t-1}) \\
= \int p(\mathbf{x}_t, s_t, \gamma_t|\mathbf{x}_{t-1}, s_{t-1}, \gamma_{t-1})p(\mathbf{x}_{t-1}, s_{t-1}, \gamma_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1}ds_{t-1}d\gamma_{t-1}, \\
= \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \cdot \int p(s_t|s_{t-1})p(s_{t-1}|\mathbf{y}_{1:t-1})ds_{t-1} \cdot \\
\int p(\gamma_t|\gamma_{t-1})p(\gamma_{t-1}|\mathbf{y}_{1:t-1})d\gamma_{t-1}, \\
\equiv p(\mathbf{x}_t|\mathbf{y}_{1:t-1})p(s_t|\mathbf{y}_{1:t-1})p(\gamma_t|\mathbf{y}_{1:t-1}),
\end{aligned}
\tag{4.12}
$$

where $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$, $p(s_t|\mathbf{y}_{1:t-1})$, and $p(\gamma_t|\mathbf{y}_{1:t-1})$ are the tth predictive distributions of each random variable, respectively.

The predictive distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ is given by the prediction step of the UKF as Eq. (2.10). For the distribution, $p(s_t|\mathbf{y}_{1:t-1})$, since s_t is independent of s_{t-1}, the forth term can be obtained immediately as

$$
p(s_t|\mathbf{y}_{1:t-1}) = p(s_t|\lambda_{prior}).
\tag{4.13}
$$

For the distribution, $p(\gamma_t|\mathbf{y}_{1:t-1})$, due to Dirac's delta function, it again produce a inverse gamma distribution. By renaming the parameters, we obtain the actual form of the predictive distribution for γ_t as,

$$
\begin{aligned}
p(\gamma_t|\mathbf{y}_{1:t-1}) &= IG(\gamma_t|a_{t|t-1}, b_{t|t-1}), \\
a_{t|t-1} &= a_{t-1|t-1} \\
b_{t|t-1} &= b_{t-1|t-1}
\end{aligned}
\tag{4.14}
$$

Here, we have the tth joint predictive distribution for $\mathbf{x}_t$, s_t, and γ_t as follows:

$$
\begin{aligned}
p(\mathbf{x}_t, s_t, \gamma_t | \mathbf{y}_{1:t-1}) = \\
\mathcal{N}(\mathbf{x}_t | \mu_{t|t-1}, \mathbf{V}_{t|t-1}) \cdot Bern(s_t | \lambda_{prior}) \cdot IG(\gamma_t | a_{t|t-1}, b_{t|t-1}).
\end{aligned}
\tag{4.15}
$$

4.2.4 Update

Since the exact calculation of Eq. (4.4) by substituting Eqs. (4.9) and (4.15) is intractable, we derive the approximated form by adopting the VB method. In conformity to section 2.2, by defining $Z_t = \{\mathbf{x}_t, s_t, \gamma_t\}$ and assuming the independency in the tth joint filtered distribution, the approximated joint filtered distribution is immediately given as follows:

$$
p(\mathbf{x}_t, s_t, \gamma_t | \mathbf{y}_{1:t}) \approx q(\mathbf{x}_t) q(s_t) q(\gamma_t),
\tag{4.16}
$$

where

$$
\ln q(\mathbf{x}_t) = \langle \ln p(\mathbf{y}_t | \mathbf{x}_t, s_t, \gamma_t) \rangle_{q(s_t)q(\gamma_t)} + \ln p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) + const.,
\tag{4.17}
$$

$$
\ln q(s_t) = \langle \ln p(\mathbf{y}_t | \mathbf{x}_t, s_t, \gamma_t) \rangle_{q(\mathbf{x}_t)q(\gamma_t)} + \ln p(s_t) + const.,
\tag{4.18}
$$

$$
\ln q(\gamma_t) = \langle \ln p(\mathbf{y}_t | \mathbf{x}_t, s_t, \gamma_t) \rangle_{q(\mathbf{x}_t)q(s_t)} + \ln p(\gamma_t | \mathbf{y}_{1:t-1}) + const.
\tag{4.19}
$$

We will derive the actual forms of Eqs. (4.17)-(4.19).

First, by calculating the expectation, Eq. (4.17) becomes

$$
\ln q(\mathbf{x}_t) = \ln \mathcal{N}(\mathbf{y}_t | \mathbf{h}(\mathbf{x}_t), \bar{\mathbf{R}}_t) + \ln p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) + const.,
\tag{4.20}
$$

$$
\bar{\mathbf{R}}_t \equiv (1 - \langle s_t \rangle_{q(s_t)} + \langle s_t \rangle_{q(s_t)} \langle \gamma_t^{-1} \rangle_{q(\gamma_t)})^{-1} \mathbf{R}_t.
\tag{4.21}
$$

Since the exponential of Eq. (4.20) is the same form as a distribution which is approximated by the UKF in the update step by substituting $\mathbf{R}_t = \bar{\mathbf{R}}_t$. Thus, $q(\mathbf{x}_t)$ is reasonably approximated by the UKF as

$$
\begin{aligned}
q(\mathbf{x}_t) &= \mathcal{N}(\mathbf{x}_t | \mu_{t|t}, \mathbf{V}_{t|t}), \\
\mu_{t|t} &= \mu_{t|t-1} + \mathbf{K}(\mathbf{y}_t - \bar{\mathbf{y}}_t), \\
\mathbf{V}_{t|t} &= \mathbf{V}_{t|t-1} - \mathbf{K}\mathbf{V}_y\mathbf{K}^{\mathrm{T}}, \\
\mathbf{K} &= \mathbf{V}_{xy}\mathbf{V}_y^{-1}, \\
\bar{\mathbf{y}}_t &= \sum_{i=0}^{2d} W_{i,t|t-1}^{(m)} \mathbf{h}(\chi_{t|t-1}^{(i)}), \\
\mathbf{V}_y &= \sum_{i=0}^{2d} W_{i,t|t-1}^{(c)} (\mathbf{h}(\chi_{t|t-1}^{(i)}) - \bar{\mathbf{y}}_t)(\mathbf{h}(\chi_{t|t-1}^{(i)}) - \bar{\mathbf{y}}_t)^{\mathrm{T}} + \bar{\mathbf{R}}_t, \\
\mathbf{V}_{xy} &= \sum_{i=0}^{2d} W_{i,t|t-1}^{(c)} (\chi_{t|t-1}^{(i)} - \mu_{t|t-1})(\mathbf{h}(\chi_{t|t-1}^{(i)}) - \bar{\mathbf{y}}_t)^{\mathrm{T}}.
\end{aligned}
\tag{4.22}
$$

Here, since the observation model, Eq. (4.9), is reduced to a single Gaussian as the first term in Eq. (4.20), we can say that $\bar{\mathbf{R}}_t$ is the estimate of observation error covariance matrix at time step for an additive Gaussian observation model with time-variant statistics as discussed in Chapter 3. When outlier is detected, that is, $\langle s_t \rangle_{q(s_t)}$ approaches unity, the estimate is enlarged by the scale parameter γ_t.
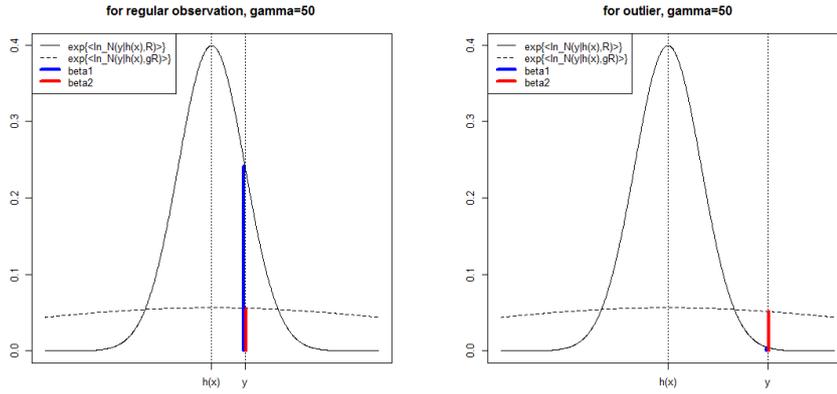
FIGURE 4.2: Example of β_1 and β_2 for two observations: an observation near its prediction(regular observation) and an observation far from its prediction (outlier)

Second, the distribution Eq. (4.18) becomes

$$
\begin{aligned}
\ln q(s_t) =\ & (1 - s_t)\langle \ln \mathcal{N}(\mathbf{y}_t|\mathbf{h}(\mathbf{x}_t), \mathbf{R}_t)\rangle_{q(\mathbf{x}_t)} + s_t\langle \ln \mathcal{N}(\mathbf{y}_t|\mathbf{h}(\mathbf{x}_t), \langle \gamma_t^{-1}\rangle_{q(\gamma_t)}^{-1}\mathbf{R}_t)\rangle_{q(\mathbf{x}_t)} + \\
& (1 - s_t)\ln(1 - \lambda_{prior}) + s_t \ln \lambda_{prior} + const. \\
=\ & (1 - s_t)\ln\beta_1 + s_t \ln\beta_2 + const. \\
\ln\beta_1 \equiv\ & -\frac{1}{2}\langle (\mathbf{y} - \mathbf{h}(\mathbf{x}_t))^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}_t))\rangle_{q(\mathbf{x}_t)}, \\
\ln\beta_2 \equiv\ & -\frac{m}{2}\langle \ln\gamma_t\rangle_{q(\gamma_t)} - \frac{\langle\gamma_t^{-1}\rangle_{q(\gamma_t)}}{2}\langle (\mathbf{y} - \mathbf{h}(\mathbf{x}_t))^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}_t))\rangle_{q(\mathbf{x}_t)}.
\end{aligned}
\tag{4.23}
$$

Note that λ_{prior} vanishes because we set the value to 0.5. We can see that the form given by Eq. (4.23) is the logarithm of the Bernoulli distribution given as follows:

$$
\begin{aligned}
q(s_t) &= Bern(s_t|\lambda_t), \\
\lambda_t &= \frac{\beta_2}{\beta_1 + \beta_2},
\end{aligned}
\tag{4.24}
$$

Here, λ_t can be regarded as the posterior probability of outlier occurrence, while λ_{prior} is the prior probability. For regular observations, the posterior probability becomes small because β_1, that is, likelihood for a regular observation takes larger value than β_2 as shown in the left panel of Fig. 4.2, and vice versa for outliers as shown in the right panel of Fig. 4.2. Due to the two likelihood with the same mean but different variances, both regular and outlier observations are handled. 4

Finally, the distribution Eq. (4.19) becomes

$$
\begin{aligned}
\ln q(\gamma_t) =\ & \langle s_t\rangle_{q(s_t)}\langle \ln \mathcal{N}(\mathbf{y}_t|\mathbf{h}(\mathbf{x}_t), \gamma_t\mathbf{R}_t)\rangle_{q(\mathbf{x}_t)} + \ln IG(\gamma_t|a_{t|t-1}, b_{t|t-1}), \\
=\ & -\frac{m}{2}\langle s_t\rangle_{q(s_t)}\ln\gamma_t - \gamma_t^{-1}\frac{\langle s_t\rangle_{q(s_t)}}{2}\langle (\mathbf{y} - \mathbf{h}(\mathbf{x}_t))^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}_t))\rangle_{q(\mathbf{x}_t)} \\
& - a_{t|t-1}\ln\gamma_t - \gamma_t^{-1}b_{t|t-1} + const.
\end{aligned}
\tag{4.25}
$$

Therefore, $q(\gamma_t)$ is an inverse gamma distribution given by

$$
\begin{aligned}
q(\gamma_t) &= IG(\gamma_t | a_{t|t}, b_{t|t}), \\
a_{t|t} &= a_{t|t-1} + \frac{m}{2} \langle s_t \rangle_{q(s_t)}, \\
b_{t|t} &= b_{t|t-1} + \frac{\langle s_t \rangle_{q(s_t)}}{2} \langle (\mathbf{y} - \mathbf{h}(\mathbf{x}_t))^{\mathrm{T}} \mathbf{R}_t^{-1} (\mathbf{y} - \mathbf{h}(\mathbf{x}_t)) \rangle_{q(\mathbf{x}_t)}.
\end{aligned}
\tag{4.26}
$$

Since we know the actual forms of $q(\mathbf{x}_t)$, $q(s_t)$ and $q(\gamma_t)$, we can calculate the expectations as

$$
\langle s_t \rangle_{q(s_t)} = \lambda_t,
\tag{4.27}
$$

$$
\langle \gamma_t^{-1} \rangle_{q(\gamma_t)} = \frac{a_{t|t}}{b_{t|t}},
\tag{4.28}
$$

$$
\langle \ln \gamma_t \rangle_{q(\gamma_t)} = \ln b_{t|t} - \psi(a_{t|t}),
\tag{4.29}
$$

$$
\langle (\mathbf{y}_t - \mathbf{h}(\mathbf{x}_t))^{\mathrm{T}} \mathbf{R}_t^{-1} (\mathbf{y}_t - \mathbf{h}(\mathbf{x}_t)) \rangle_{q(\mathbf{x}_t)} = \sum_{i=0}^{L} w_{t|t}^{(i)} (\mathbf{y}_t - \mathbf{h}(\mathbf{x}_{t|t}^{(i)}))^{\mathrm{T}} \mathbf{R}_t^{-1} (\mathbf{y}_t - \mathbf{h}(\mathbf{x}_{t|t}^{(i)})),
$$

$$
\tag{4.30}
$$

where ψ is the digamma function, $\mathbf{x}_{t|t}^{(i)}$ is the i-th sigma point of $q(\mathbf{x}_t)$, and $w_{t|t}^{(i)}$ is the corresponding weight.

Here, we have tth joint filtered distribution for $\mathbf{x}_t$, s_t, and γ_t as follows:

$$
\begin{aligned}
p(\mathbf{x}_t, s_t, &\gamma_t | \mathbf{y}_{1:t}) \\
&\approx q(\mathbf{x}_t) q(s_t) q(\gamma_t), \\
&= \mathcal{N}(\mathbf{x}_t | \mu_{t|t}, \mathbf{V}_{t|t}) \cdot Bern(s_t | \lambda_t) \cdot IG(\gamma_t | a_{t|t}, b_{t|t}).
\end{aligned}
\tag{4.31}
$$

The distributions $q(\mathbf{x}_t)$, $q(s_t)$ and $q(\gamma_t)$ depend on each other via the expectations in Eqs. (4.22), (4.23), and (4.26). Therefore, we calculate the distributions alternately starting from an initial guess and repeat the calculation until convergence. As the convergence metric, we use the $l2$ norm of $\mathbf{x}_t$.

4.2.5   Algorithm

Here, since we have the approximated posterior distributions $q(\mathbf{x}_t)$ and $q(\gamma_t)$ with the same form as the given $(t-1)$-th distributions Eq.(4.11), we can construct a recursive estimation algorithm as shown in Algorithm 2. Note that, as the initial guess of $q(s_t)$, and $q(\gamma_t)$, we use the prior of s_t and the previous filtered distribution of $q(\gamma_t)$ (see line 5 of Algorithm 1). As the initial distribution $p(\gamma_t | a_0, b_0)$, we recommend to use non-informative distribution such that the expectation becomes $\gg 1$. In the following experiments, we use $a_0 = 0.1$ and $b_0 = 1$; with this setting, $\langle \gamma_t^{-1} \rangle^{-1}$ becomes 10.

4.3   Tips for usage

4.3.1   Estimation of time-variant outlier characteristics

In the system model of γ_t, we have specified it as Dirac's delta function, which is the time-invariant assumption about γ_t. Thus, in the limit as $t \to \infty$, the estimate of γ could be biased and disrupt the estimation of $\mathbf{x}_t$ and s_t if the outlier statistics is

---

Algorithm 2 Algorithm for the proposed method

Require: the system model Eq. (4.5), the error model for regular observations as in Eq. (4.1), and then, construct the observation model Eq. (4.9).
Require: initial distributions, $p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0|\mu_0, \mathbf{V}_0)$ and $p(\gamma_0) = IG(\gamma_0|a_0, b_0)$, and a hyperparameter ρ for $p(\gamma_t|\gamma_{t|t-1})$.
 1: **for** $t = 1$ to T **do**
 2:    **prediction step:**
 3:    calculate $p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{x}_t|\mu_{t|t-1}, \mathbf{V}_{t|t-1})$ with Eq. (2.10).
 4:    calculate $p(\gamma_t|\mathbf{y}_{1:t-1}) = IG(\gamma_t|a_{t|t-1}, b_{t|t-1})$ with Eq. (4.14).
 5:    **update step:**
 6:    let $q(s_t) = Bern(s_t|\lambda_{prior})$, and $q(\gamma_t) = IG(\gamma_t|a_{t|t-1}, b_{t|t-1})$.
 7:    **while** true **do**
 8:       calculate $q(\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_t|\mu_{t|t}, \mathbf{V}_{t|t})$ with Eq. (4.22).
 9:       calculate $q(s_t) = Bern(s_t|\lambda_t)$ with Eq. (4.24).
10:       calculate $q(\gamma_t) = IG(\gamma_t|a_t, b_t)$ with Eq. (4.26).
11:       exit while if the update of $q(\mathbf{x}_t)$ has converged.
12:    **end while**
13:    let $p(\mathbf{x}_t|\mathbf{y}_{1:t}) = q(\mathbf{x}_t)$ and $p(\gamma_t|\mathbf{y}_{1:t}) = q(\gamma_t)$
14: **end for**

---

changed in filtering. Therefore, we introduce another system model for γ_t to estimate time-variant γ_t.

The system model for γ_t is

$$p(\gamma_t|\gamma_{t-1}) = IG(\gamma_t|\rho, (\rho - 1)\gamma_{t-1}), \tag{4.32}$$

where IG denotes an inverse gamma distribution, and ρ is a control parameter for time-evolution of γ_t. Since inverse Wishart distribution is reduced to inverse gamma distribution when the dimensionality is unity, the meaning of ρ follows the discussion of system model for $\mathbf{R}_t$ in section 3.2.2. In the limit as $\rho \to \infty$, this system model approaches to Eq. (4.8).

Following the discussion of the inverse Wishart distribution as discussed in Appendix. C, it can be approximated, again, with an inverse gamma distribution, which has the same moments up to second order. That is,

$$p(\gamma_t|\mathbf{y}_{1:t-1}) = \int p(\gamma_t|\gamma_{t-1})p(\gamma_{t-1}|\mathbf{y}_{1:t-1})d\gamma_{t-1} \approx IG(\gamma_t|a_{t|t-1}, b_{t|t-1}),$$

$$a_{t|t-1} = 2 + \frac{1}{(1 + \frac{1}{\rho-2})(1 + \frac{1}{a_{t-1|t-1}-2}) - 1}, \tag{4.33}$$

$$b_{t|t-1} = \frac{a_{t-1|t-1} - 1}{a_{t|t-1} - 1}b_{t-1|t-1}.$$

In Algorithm 2, by specifying the value of ρ and replacing Eq. (4.14) with Eq. (4.33), estimation of scale parameter γ_t for a time-variant outlier statistics is available.

4.4 Numerical Experiments

In this section, we examine the proposed method through two numerical experiments. In Experiment 1, by using a univariate linear system, we detail the filtering

of the proposed method. In Experiment 2, we compare the proposed method to
related existing methods for three multivariate systems.

4.4.1   Experiment 1

Synthetic data

We assume a univariate linear system for generating synthetic data as

$$
\begin{aligned}
p(x_t|x_{t-1}) &= \mathcal{N}(x_t|0.9x_{t-1} + u_t, 1), \\
x_0 &\sim \mathcal{N}(0, 0.1^2), \\
u_t &\sim 0.2\mathcal{N}(u_t|0, 5^2) + 0.8\delta(u_t),
\end{aligned}
\tag{4.34}
$$

where u_t is a known input to the system, and δ is Dirac's delta function as the point-
mass probability density function.

   To generate outlier-contaminated observations, we use the following mixture
model as

$$
\begin{aligned}
p(y_t|x_t) &= 0.9p_{regular}(y_t|x_t) + 0.1p_{outlier}(y_t|x_t), \\
p_{regular}(y_t|x_t) &\equiv \mathcal{N}(y_t|x_t, 1), \\
p_{outlier}(y_t|x_t) &\equiv \mathcal{N}(y_t|x_t, 128),
\end{aligned}
\tag{4.35}
$$

where $p_{regular}$ is the generative model for not-contaminated observations, and $p_{outlier}$
is the generative model for outliers.

Experimental Setting

With 200 steps synthetic data generated by the above models, we conducted a filter-
ing experiment for given $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and $p_{regular}(\mathbf{y}_t|\mathbf{x}_t)$. Note that we do not use any
information about outliers: $p_{outliers}(y_t|x_t)$ and when outliers occurred. We construct
an observation model as Eq.(4.9) on the basis of $p_{regular}(\mathbf{y}_t|\mathbf{x}_t)$. That is,

$$
p(y_t|x_t, s_t, \gamma_t) = \mathcal{N}(y_t|x_t, 1)^{1-s_t}\mathcal{N}(y_t|x_t, \gamma_t \cdot 1)^{s_t}.
\tag{4.36}
$$

As explained in Section 3, we estimate s_t and γ_t in addition to $\mathbf{x}_t$.

   The setting of the proposed method is as follows. The implementation of the
UKF is that of [12], and the design parameters for generating sigma points are se-
lected based on the guidance (see section 7.3 in [12]). Along with the notation to
define sigma points in Appendix A, the parameters κ and ν are set to $3 - d$ and 2,
respectively. Initial distribution for $\mathbf{x}_0$ and γ_0 is $\mathcal{N}(\mathbf{x}_0|\mathbf{x}_{true,0}, \mathbf{I})$ and $\mathcal{IG}(\gamma_0|1, 10)$, re-
spectively, where $\mathbf{x}_{true,0}$ is the true system state at time step 0. Note that the mean
of γ_0 on the initial distribution is $10 \gg 1$ to make the outlier scale larger than the
regular scale.

   For comparison, we also run the ordinary UKF. The estimation accuracy for the
state variable by the UKF is expected to be deteriorated by outliers. Moreover, we
will clarify how the proposed method avoid the deterioration.

Results

The temporal variations of the true state and the outlier-contaminated observation
are shown in Fig. 4.3. We can see that outliers largely deviate from the true state,
which are marked with crosses at the bottom of the panel.
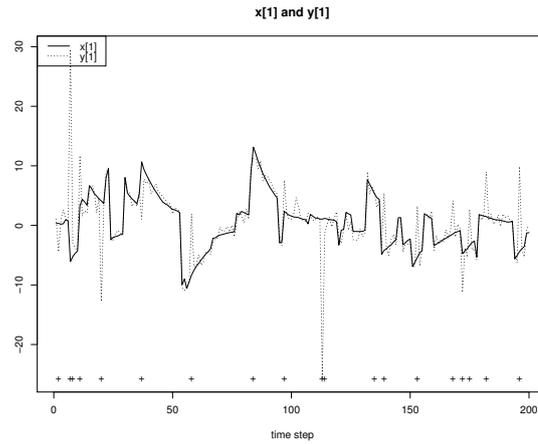
FIGURE 4.3: The synthetic data of the state (solid line) and the outlier-contaminated observation (dotted line). The crosses represents time-steps when outliers occur.

The predicted state by the UKF and the proposed method are shown in Figs. 4.4 (a) and (b), respectively. For outlier-contaminated observations, applying the UKF causes deterioration in the estimation of the state. For example, around the 113th step in Fig. 4.4 (a), the estimation errors increased after filtering the outlier. This is because the estimation of the UKF was biased as a result of over-fitting to the outlier. While, the proposed method handled such outliers without deterioration in the estimation as can be seen in Fig. 4.4 (b). The root mean square error (RMSE) of the estimation of the state variable was improved from 0.47 (that of the UKF) to 0.22.

To detail the behavior of the proposed method, temporal variations of the posterior mean of the outlier indicator, $\mathrm{tr}\, \bar{\mathbf{R}}_t$, and the posterior mean of γ_t are shown in Fig. 4.5 (a)-(c), respectively.

In Fig. 4.5 (a), for the most of outliers, which are marked with vertical gray lines, the posterior probabilities of outlier indicators were estimated to be unity. That is, the proposed method successfully detected outliers. For some outliers which were not estimated as outliers, we can see the amplitude of the observation errors were small, for example, the one at the 87th step. This is because the outliers in this experiment are also zero-mean Gaussian random variable and can be small error. Due to the same reason, there were cases where regular samples are estimated as outliers. It is because such samples largely deviated from its mean accidentally.

Once an observation is estimated as outlier, since the expectation of s_t approaches unity, $\bar{\mathbf{R}}_t$ becomes large due to Eq. (4.21) as shown in Fig. 4.5 (b). By using the enlarged $\bar{\mathbf{R}}_t$, over-fitting to the outlier can be avoided. Since variances were enlarged for most of outliers, we can say that the proposed method successfully neglected outlier-contaminated observations.

4.4.2 Experiment 2: Comparison with existing methods

In the second experiment, we compare the proposed method with existing methods through three systems: (1) a 5-dimensional linear system (5DLS), which is a direct extension of the linear model examined in Experiment 1, (2) *Mackey-Glass* [46], which is a nonlinear time-delay system, and (3) *Lorenz 63* [47], which is a chaotic nonlinear model.
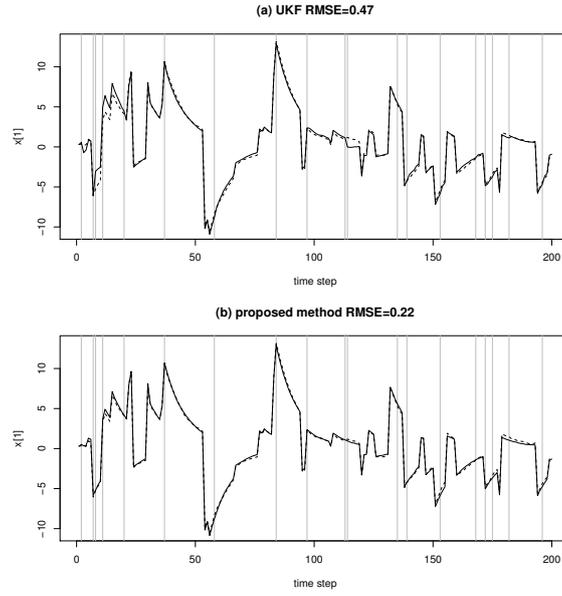
FIGURE 4.4: The prediction of the state by (a) the UKF, and (b) the proposed filter plotted as solid lines. The dotted line denotes the true state. The RMSE are 0.47 and 0.22, respectively. The vertical gray lines represents the time steps when outliers occur.

Synthetic data

At first, we declare generative models for synthetic data. The nonlinear functions $\mathbf{f}$ and $\mathbf{h}$ of each system are defined as

$$
\begin{aligned}
&5DLS \\
&\mathbf{f}(\mathbf{x}_{t-1}) = 0.9\mathbf{x}_{t-1} + \mathbf{u}_t, \\
&\mathbf{u}_t \sim 0.2\mathcal{N}(\mathbf{u}_t|\mathbf{0}, 5^2\mathbf{I}) + 0.8\delta(\mathbf{u}_t), \\
&\mathbf{h}(\mathbf{x}_t) = \mathbf{x}_t,
\end{aligned}
\tag{4.37}
$$

$$
\begin{aligned}
&Mackey - Glass \\
&\mathbf{f}(\mathbf{x}_{t-1}) = [x_{t-1,1} + 0.2\frac{x_{t-1,17}}{1 + x_{t-1,17}^{10}} - 0.1x_{t-1,1}, \\
&\qquad x_{t-1,1}, \cdots, x_{t-1,16}]^{\mathrm{T}}, \\
&h(\mathbf{x}_t) = 10x_{t,1},
\end{aligned}
\tag{4.38}
$$

$$
\begin{aligned}
&Lorenz63 \\
&\mathbf{f}_t(\mathbf{x}_{t-1}) = \phi^{10}(\mathbf{x}_{t-1}), \\
&\phi(\mathbf{x}_{t-1}) = \\
&\mathbf{x}_{t-1} + 0.01
\begin{pmatrix}
-10(x_{t-1,1} - x_{t-1,2}) \\
28x_{t-1,1} - x_{t-1,2} - x_{t-1,1}x_{t-1,3} \\
x_{t-1,1}x_{t-1,2} - \frac{8}{3}x_{t-1,3}
\end{pmatrix}, \\
&\mathbf{h}(\mathbf{x}_t) = \mathbf{x}_t,
\end{aligned}
\tag{4.39}
$$
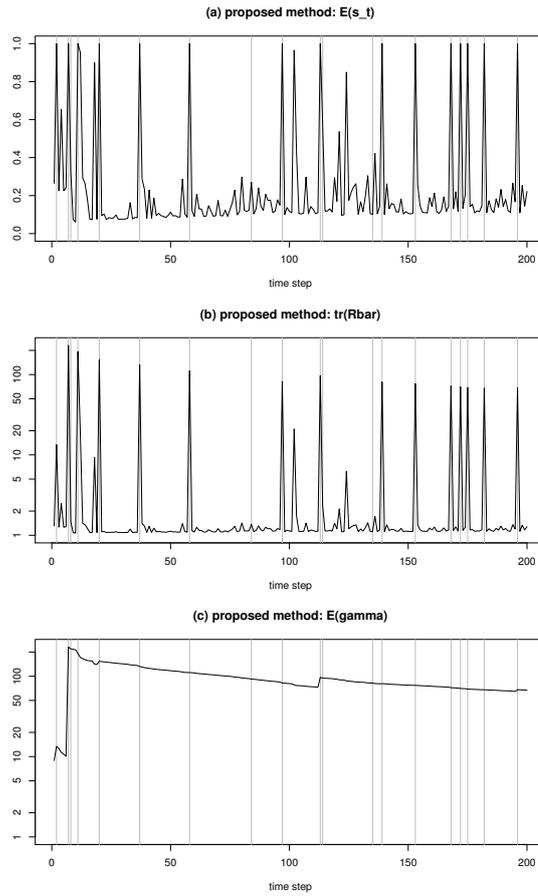
FIGURE 4.5: The posterior mean of s_t is plotted in panel (a), the trace of the net $\mathbf{R}_t$ used in the VB iteration is plotted in panel (b), and the posterior mean of γ_t is plotted in panel (c). The vertical gray lines represents the time steps when outliers occur.

where the notation $\mathbf{I}$ denotes the identity matrix. Dimensionality of $\mathbf{x}_t$ and $\mathbf{y}_t$, parameters of the initial distribution $\mathcal{N}(\mathbf{x}_0|\mu_0, \mathbf{V}_0)$, $\mathbf{Q}_{t,true}$, and $\mathbf{R}_{t,regular}$ are summarized in Table 4.1.

To generate outlier-contaminated observations, we again use a mixture model as

$$
\begin{aligned}
p(\mathbf{y}_t|\mathbf{x}_t) &= (1 - \xi)p_{regular}(\mathbf{y}_t|\mathbf{x}_t) + \xi p_{outlier}(\mathbf{y}_t|\mathbf{x}_t), \\
p_{regular}(\mathbf{y}_t|\mathbf{x}_t) &= \mathcal{N}(\mathbf{y}_t|\mathbf{h}(\mathbf{x}_t), \mathbf{R}_{t,regular}),
\end{aligned}
\tag{4.40}
$$

where ξ is the probability of outlier occurrence. We assume two outlier models: one is an additive Gaussian error model with a large variance, $16\mathbf{R}_t$, and the other is a constant model independent of $\mathbf{x}_t$:

$$
p_{outlier}(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{h}_t(\mathbf{x}_t), 16\mathbf{R}_{t,regular}),
\tag{4.41}
$$

$$
p_{outlier}(\mathbf{y}_t|\mathbf{x}_t) = \delta(\mathbf{y}_t - [10, \cdots, 10]^{\mathrm{T}}).
\tag{4.42}
$$

In practice, as the sources of Eq. (4.41), for example, the impulsive disturbance to the system can be possible. In this case, the mean is almost the same, but the observation will be highly deviated from its mean. As the sources of Eq. (4.42), for

TABLE 4.1: The parameters of initial distribution for the state vector, and **Q** and **R**

	5D-LS	Mackey-Glass	Lorenz63
(d,m)	(5,5)	(17,1)	(3,3)
μ_0	**0**	**0**	**0**
$\mathbf{V}_0$	**I**	**I**	**I**
$\mathbf{Q}_{t,true}$	$10^{-2}\mathbf{I}$	$10^{-4}\mathbf{I}$	$10^{-2}\mathbf{I}$
$\mathbf{R}_{t,regular}$	**I**	**I**	**I**

example, exceeding the preset range could happen. In that case, although it depends on the system, upper or lower limit is output as the current value of measurement. Another case is the connection error in transmission of measured signal, especially using a wide area network. In that case, when the failure in connectivity happen, the previous valid value might be output.

For each system, we examined three kinds of datasets generated by the following settings: (1) no outliers by setting $\xi = 0$, (2) Gaussian outlier given by Eq.(4.41) with $\xi = 0.3$, and (3) constant outlier given by Eq.(4.42) with $\xi = 0.3$. We refer to the datasets as OD(Outlier contaminated Dataset)-None, OD-16R, and OD-Const10, respectively.

Experimental Setting

We conducted filtering experiments with the proposed method and existing methods. As in Experiment 1, we assume that $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and $p_{regular}(\mathbf{y}_t|\mathbf{x}_t)$ are given. We construct an observation model as Eq. (4.9) by substituting $\mathbf{R}_t = \mathbf{R}_{t,regular}$.

We compare the proposed method with three existing methods referred to as TTS07 [14], ANN11 [22], and SH13 [16]. The three are the methods that estimate the observation error covariance matrix simultaneously with the state vector.

TTS07 estimate the scale of the observation error covariance matrix from a fixed prior at each time step. As the prior for the scale, a gamma distribution $\mathcal{G}$, is assumed.

ANN11 estimate the all elements of the observation error covariance matrix also from a fixed prior at each time step. To estimate all elements, they have applied a Wishart distribution $\mathcal{W}$ as the prior, which is the conjugate prior for the precision matrix of Gaussian likelihood. Although there is no recommendation for setting of the prior, to estimate impulsive change of the observation error, we guess that the non-informative prior will be required.

SH13 also estimate the all elements of the observation error covariance matrix by using the inverse Wishart distribution. The essential difference compared to ANN11 is the first order Markov assumption for the observation error covariance matrix, that is, the current estimate is based on the previous estimate. The method use a forgetting parameter to erase the past information. For large value of the forgetting parameter, this method approaches to ANN11.

Since TTS07 and ANN11 are designed for linear systems using KF, we extended them with the UKF as in Section 3. The setting of the UKF, which is used in all above methods and the proposed method, is same as that of Experiment 1.

For TTS07 and ANN11, we use the non-informative priors for the observation error covariance matrix such that the expectation becomes $\mathbf{R}_{t,regular}$. For TTS07, we use the recommended setting $\mathcal{G}(1,1)$. For ANN11, we use a non-informative prior

TABLE 4.2: Median of the resultant RMSE

	UKF	TTS07	ANN11	SH13	Proposed
5DLS OD-None	**0.212**	0.214	0.214	**0.212**	**0.212**
5DLS OD-16R	0.271	0.219	0.219	0.222	**0.217**
5DLS OD-Const10	1.010	0.220	0.220	0.219	**0.218**
Mackey-Glass OD-None	**0.042**	0.044	0.043	**0.042**	**0.042**
Mackey-Glass OD-16R	0.066	**0.050**	**0.050**	0.053	**0.050**
Mackey-Glass OD-Const10	0.451	**0.063**	0.064	0.087	**0.063**
Lorenz63 OD-None	**0.513**	0.550	0.555	0.539	0.514
Lorenz63 OD-16R	1.190	0.810	4.120	1.063	**0.800**
Lorenz63 OD-Const10	9.612	1.489	6.620	5.377	**0.835**

$\mathcal{W}(m + 2, (m + 2)\mathbf{I})$. For SH13, we use a non-informative initial distributions for $\mathbf{R}_t$ as $\mathcal{IW}(m, m\mathbf{I})$. Note that the definition of $\mathcal{IW}$ used in SH13 [16] is different from the one used in this thesis, and we follow the definition in [16] in this case. A forgetting parameter, ρ in SH13 [16], is determined by preliminary experiments for each dataset from $[0.5, 1)$.

To make the discussion consistent, we use the notation $\bar{\mathbf{R}}_t$ in Eq. (4.21) to represent the estimate of $\mathbf{R}_t$ by TTS07, ANN11, and SH13.

Results

We conducted each experiment 1024 times with distinct random seed for generating synthetic data, and we compared the RMSE and the computational cost.

The results are shown in Figs. 4.6-4.8 for 5DLS, Mackey-Glass, and Lorenz63, respectively. In the cases of 5DLS and Mackey-Glass, RMSE for OD-None are almost same for all methods except TTS07 and ANN11 as shown in Fig. 4.6 (a) and Fig. 4.7 (a). The results of the UKF was significantly deteriorated for OD-16R, and especially for OD-Const10 as shown in Figs. 4.6 (b)-(c) and Figs. 4.7 (b)-(c). For these outlier datasets, the results of TTS07, ANN11 and the proposed method are almost same and minimum.

For Lorenz63, RMSE of TTS07 and SH13 are relatively fluctuated even for OD-None as shown in Fig. 4.8 (a). For OD-16R, TTS07 and the proposed method dealt with the outliers better than ANN11 and SH13 as shown Figs. 4.8 (b)-(c). ANN11 was highly fluctuated for OD-16R. For the results of OD-Const10, as can be seen in Fig. 4.8 (c), although the fluctuation of TTS07 increased, the result of the proposed method is still stable. The median of RMSE is summarized in Table 4.2. As shown in Table 4.2, totally, the proposed method handled outliers well.

Furthermore, we will detail the differences among the proposed method and the existing methods. Temporal variations of $\operatorname{tr} \bar{\mathbf{R}}_t$ for 5DLS OD-16R are plotted in Fig. 4.9. As can be seen in Figs. 4.9 (a), (b) and (d), the estimates of $\mathbf{R}_t$ by TTS07, ANN11 and the proposed method are spiky at outlier-contaminated observations. Contrary to the three methods, SH13 shows saw-toothed variations as shown in Fig. 4.9 (c). This is because SH13 is designed for estimating time-variant observation error covariance matrix by assuming that the matrix is a realization of the first-order Markov process. The assumption of the Markov process is effective for estimating gradually changing observation error statistics. However, under the presence of outliers, although over-fitting to an outlier can be avoided by enlarging the estimate of $\mathbf{R}_t$,

(a) 5DLS OD−None

(b) 5DLS OD−16R

(c) 5DLS OD−Const10
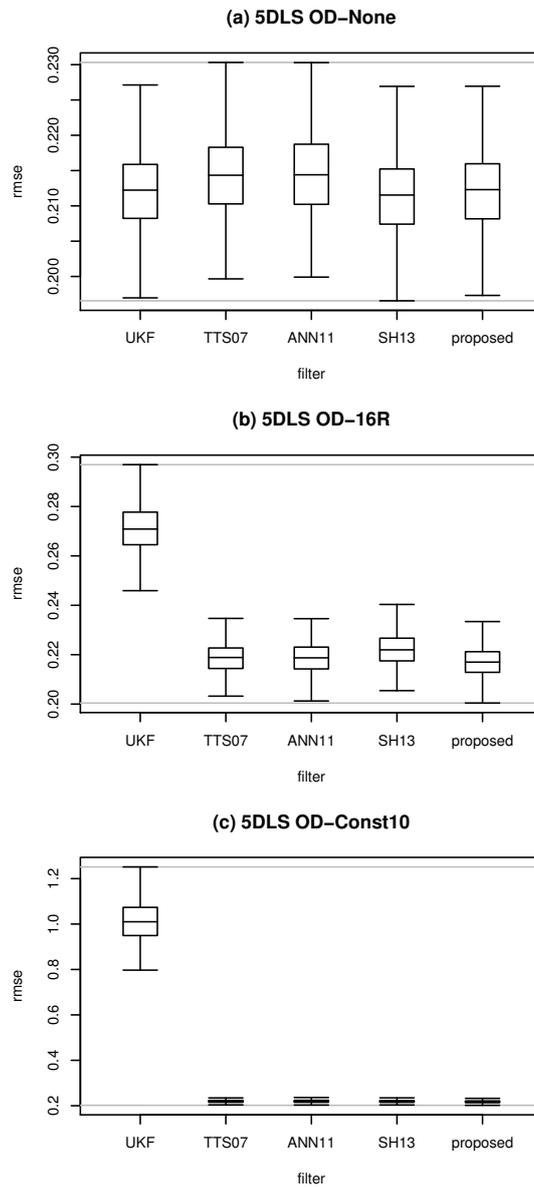
FIGURE 4.6: The box-plot of RMSE (5DLS) for 1024 trials
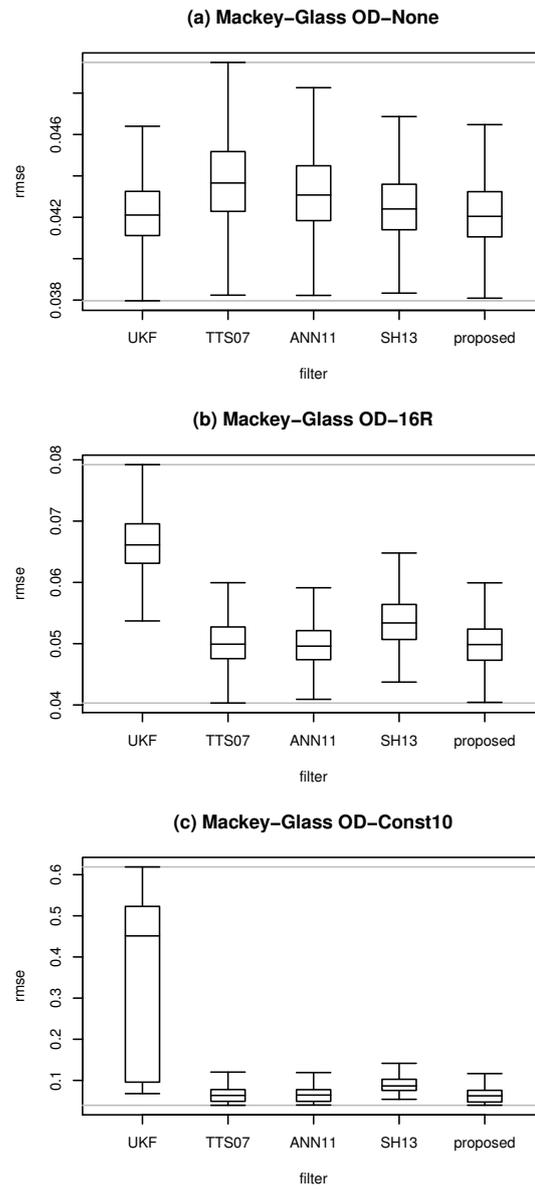
FIGURE 4.7: The box-plot of RMSE (Mackey-Glass) for 1024 trials
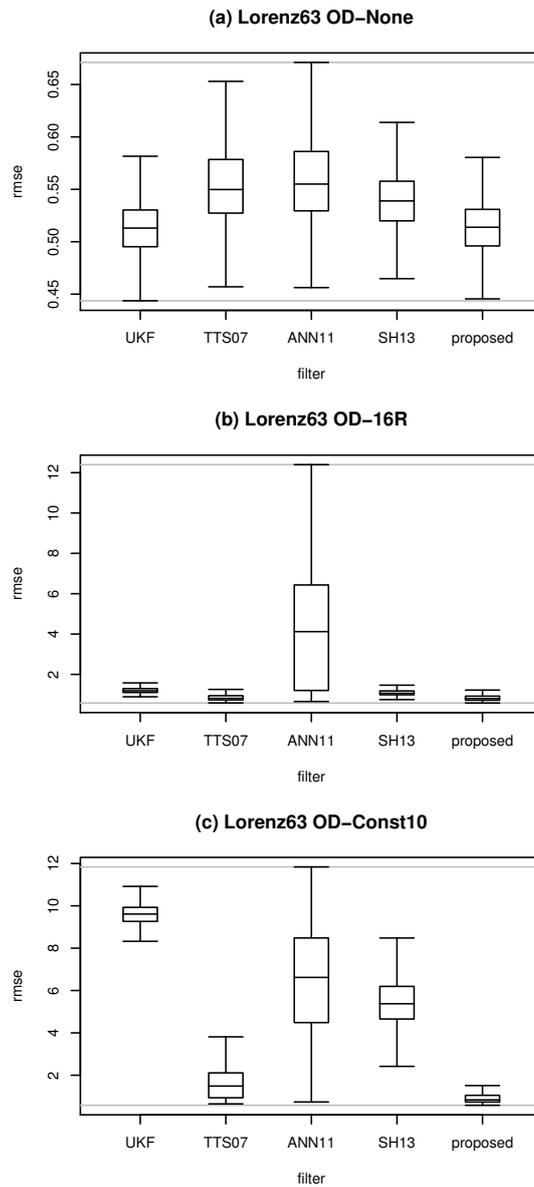
FIGURE 4.8: The box-plot of RMSE (Lorenz63) for 1024 trials
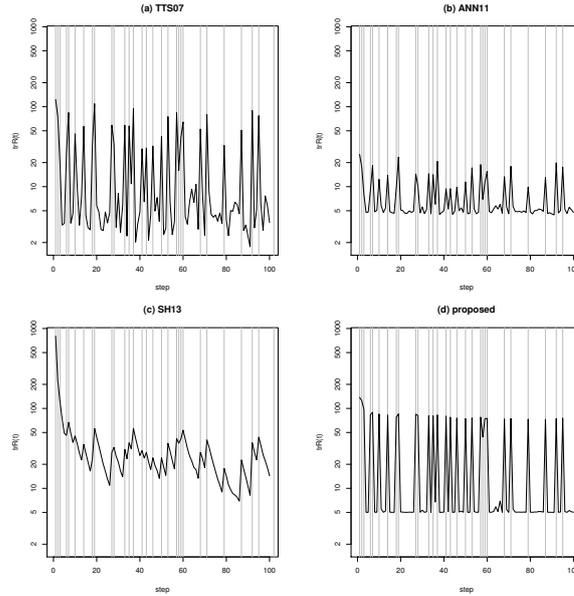
FIGURE 4.9: Example temporal variations of the trace of estimated **R** in 5DLS OD-16R. The vertical gray lines denote the positions where samples are outlier-contaminated.

under-fitting to the following regular samples happens because the enlarged estimate of $\mathbf{R}_t$ continues for a while as shown in Fig. 4.9 (c). In other words, estimation of $\mathbf{R}_t$ at later steps by SH13 is still contaminated by the outlier. The contamination can be confirmed significantly, for example, around the 19th step in Fig. 4.9 (c).

For TTS07, ANN11, and the proposed method, the difference can be seen clearly in the histograms of $\mathrm{tr}\,\bar{\mathbf{R}}_t$ as shown in Fig. 4.10. We can see that, from Fig. 4.10 (d), the estimates of $\mathbf{R}_t$ by the proposed method became polarized around the two variances. As intended, this is a result of estimating outlier indicator s_t, which switches the two observation models that have distinct variances. While, those by TTS07 and ANN11 became fluctuated as shown in Figs. 4.10 (a) and (b). The difference comes from that TTS07 and ANN11 are methods that estimate $\mathbf{R}_t$ from the innovation at each time step. The estimates of $\mathbf{R}_t$ by ANN11 were more concentrated around true $\mathrm{tr}\,\mathbf{R}_{t,regular}$ than those by TTS07. This is because the assumed prior information for $\mathbf{R}_t$ of ANN11 has to be larger than that of TTS07 as a result of using the Wishart distribution, whose degree of freedom parameter is defined for $> m - 1$.

Although TTS07 can be regarded as a flexible method for estimating $\mathbf{R}_t$ from Fig. 4.10 (a), the flexibility can cause performance deterioration in estimation of $\mathbf{x}_t$. As shown in the results of OD-None in Experiment 2, the RMSE of TTS07 for OD-None was worse than those of the proposed method in all systems. This can be explained from the estimated $\mathbf{R}_t$. The histograms of $\mathrm{tr}\,\bar{\mathbf{R}}_t$ for 5DLS OD-None are shown in Fig. 4.11. From this figure, we can see that the result by TTS07 was highly fluctuated. That is, over- and under-fitting to observations occur in TTS07 because the observation model with a static observation error covariance matrix is optimal for OD-None. Contrary to this, as can be seen in Fig. 4.11 (d), the estimates of $\mathbf{R}_t$ by the proposed method are concentrated on $\mathrm{tr}\,\mathbf{R}_{t,regular}$.
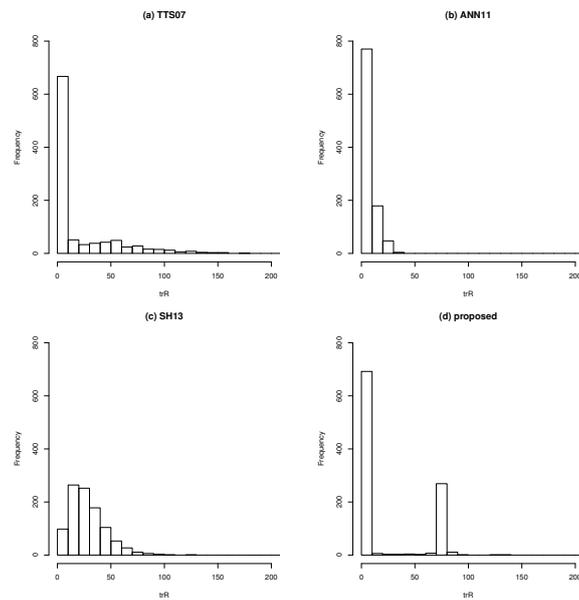
FIGURE 4.10: Example histograms of the trace of estimated $\mathbf{R}_t$ for
5DLS OD-16R.
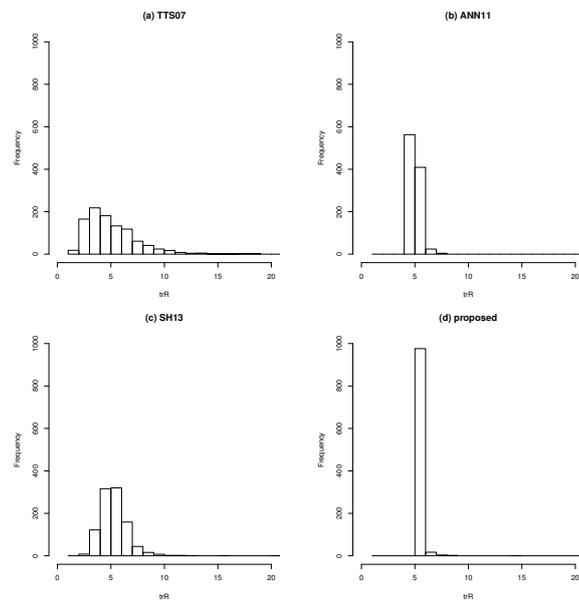
FIGURE 4.11: Example histograms of the trace of estimated $\mathbf{R}_t$ for
5DLS OD-None.

4.5 Conclusion

We have proposed a new filtering method for dealing with outlier-contaminated observations. The method switches two Gaussian observation error models that have distinct covariance matrix for both observations with and without outliers.

Through numerical experiments, we have examined the proposed method. First, with a univariate system, we have shown the detail of filtering. Second, we have compared the proposed method with existing methods by using three multivariate systems. In the experiments, we have shown that the state estimation performance by the proposed method is the same or better than those by the existing methods. Especially, in the case of outlier-free observations, estimates of the observation error covariance matrix by the existing methods were fluctuated, while that of the proposed method was stable. As a result, the state estimation performance for outlier-free observations was improved by the proposed method. In addition, in the case of a highly nonlinear system with outlier-contaminated observations, the state estimation performance by the proposed method prominently outperformed the others. Thus, we can say that the proposed method is a better choice as a nonlinear filtering method for both datasets with and without outliers.

Chapter 5

Conclusion

Through out this thesis, we have considered filtering methods with adaptive error models for observation errors. In Chapter 3, we have presented an extension of the EnKF for simultaneously estimating the system state and the observation error covariance matrix for dealing with long-term temporal change of the observation error characteristics. The method has been examined in a nonlinear system under the presence of unmodeled disturbance and succeeded in keeping the accuracy of state estimation. In addition, for a design parameter of the presented method, we have derived a ML method. Using the ML method, it is possible to adopt the presented method to a problem without trial and error.

In Chapter 4, we have presented a filtering method for dealing with outliers as an extension of the UKF. The method switches between two observation models adaptively: one is for regular observations, and another is for observations containing outliers. Since a parameter in the models is simultaneously estimated with the system state, it is possible to adopt the method to a problem without further tuning that accounts for outlier characteristics. We have compared the accuracy of state estimation with existing methods in three multivariate systems and confirmed that the accuracy is equal to or better than those of the existing methods.

With the methods presented in this thesis, applicability of filtering technique has been enhanced. The key idea is to use an adaptive observation error model, which includes time-indexed parameters to be estimated at each time step. The enabler is the VB method which approximates the joint filtered distribution of the state vector and the time-index parameters with another tractable distribution. The enhanced adaptivity in filtering is valuable for improving the accuracy of state estimation. Furthermore, an effort in starting to use filtering techniques can be decreased due to the adaptivity for unknown error statistics in practical problems, and thus the front for practitioners can be widened.

Appendix A

Unscented Transformation

A.1   Basic idea

The unscented transformation is a technique for calculating the statistics of non-linearly transformed Gaussian random variable. Julier and Uhlmann applied the technique for nonlinear Kalman filtering [4], which needs to deal with nonlinear time-evolution and observation function for a Gaussian-distributed state vector.

We show the schematic diagram in Fig. A.1. In Fig. A.1 (A), an example of exact nonlinear transformation of a Gaussian is illustrated. Generally, nonlinear transformation does not reproduce a Gaussian distribution. In addition, it is even hard to obtain the tractable distribution. This motivates the linearization used in the EKF as shown in Fig. A.1 (B). However, as discussed in Chapter 1, the linearization is often problematic because of computational stability and cost for calculating the Jacobian matrix.

The unscented transformation is a derivative-free approach that directly approximates the transformed probability distribution with a Gaussian by using sigma points $\{\chi_i\} \in \mathbb{R}^d$. Sigma points $\{\chi_i\}$ are deterministically sampled realizations of a Gaussian. A nonlinear function is applied for the sigma points as $\{\mathbf{f}(\chi_i)\}$ to obtain the realizations of transformed random variables. By calculating the mean and the covariance matrix with $\{\mathbf{f}(\chi_i)\}$, a Gaussian is calculated as shown in Fig. A.1 (C). In the next section, we detail the algorithm.

A.2   Algorithm

In this section, we detail the unscented transformation of a Gaussian $\mathcal{N}(\mathbf{x}|\mu, \mathbf{V})$ for a nonlinear function $\mathbf{f}(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^d$, $\mu \in \mathbb{R}^d$, and $\mathbf{V} \in \mathbb{R}^{d \times d}$.

Generating sigma points

Sigma points are weighted samples for efficiently approximate original Gaussian random variable. According to [12], the sigma points and the weights of a d-dimensional

(A) Exact transformation

(B) Linearized transformation

(C) Unscented transformation
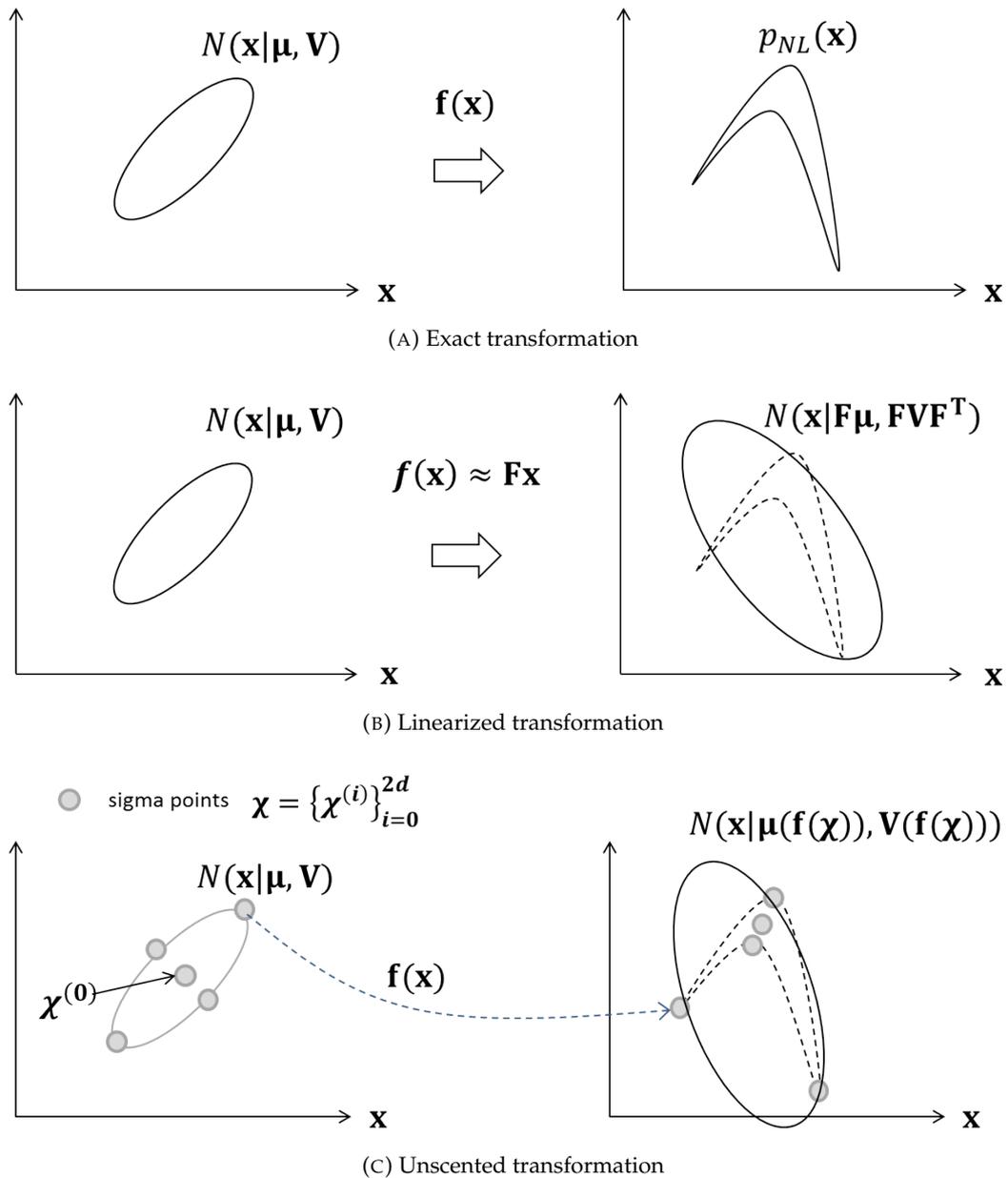
FIGURE A.1: Nonlinear transformation of Gaussian distribution

random variable that has the mean μ and the covariance $\mathbf{V}$ are calculated as

$$
\begin{aligned}
\chi^{(0)} &= \mu, \\
\chi^{(i)} &= \mu + (\sqrt{(d+\kappa)\mathbf{V}})_i, \quad i = 1, \cdots, d, \\
\chi^{(i)} &= \mu - (\sqrt{(d+\kappa)\mathbf{V}})_{i-d}, \quad i = d+1, \cdots, 2d, \\
W_0^{(m)} &= \frac{\kappa}{d+\kappa}, \\
W_0^{(c)} &= \frac{\kappa}{d+\kappa} + \nu, \\
W_i^{(m)} &= W_i^{(c)} = \frac{1}{2(d+\kappa)}, \quad i = 1, \cdots, 2d,
\end{aligned}
\tag{A.1}
$$

where $\chi^{(i)}, i = 0, \cdots, 2d$ is the sigma point, the notation $(\cdot)_i$ represents the i-th column vector of the matrix, W_i is the weight of the sigma point $\chi^{(i)}$, and κ and ν are design parameters for sigma points.

Estimating the transformed probability distribution as a Gaussian

Although the analytical transformation of a Gaussian distribution for a nonlinear function is intractable, the nonlinear mapping of a sample can be calculated. We can say that the nonlinearly mapped samples are realization of transformed random variable. Therefore, the unscented transformation estimates the transformed probability distribution $p_{NL}(\mathbf{x})$ from the sigma points as follows:

$$
\begin{aligned}
p_{NL}(\mathbf{x}) &\approx \mathcal{N}(\mathbf{x}|\mu_{NL}, \mathbf{V}_{NL}), \\
\mu_{NL} &= \sum_{i=0}^{2d} W_i^{(m)} \mathbf{f}(\chi^{(i)}), \\
\mathbf{V}_{NL} &= \sum_{i=0}^{d} W_i^{(c)} (\mathbf{f}(\mathbf{x}) - \mu_{NL})(\mathbf{f}(\mathbf{x}) - \mu_{NL})^{\mathrm{T}}.
\end{aligned}
\tag{A.2}
$$

Appendix B

The inverse Wishart distribution

B.1 Definition

The inverse Wishart distribution [34] is defined as

$$\mathcal{IW}(\mathbf{R}) = C(\eta, \mathbf{B})|\mathbf{R}|^{-\frac{1}{2}\eta} \exp\{-\frac{1}{2}\operatorname{tr}(\mathbf{B}\mathbf{R}^{-1})\},$$

$$C(\eta, \mathbf{B}) \equiv |\mathbf{B}|^{\frac{1}{2}(\eta-m-1)} 2^{-\frac{1}{2}(\eta-m-1)m} \Gamma_m(\frac{\eta-m-1}{2})^{-1},$$

(B.1)

where Γ_m is the m-dimensional multivariable gamma function, $\eta > 2m + 2$ is the number of degrees of freedom, and $\mathbf{B} \in \mathbb{R}^{m \times m} \succ 0$ is the scale matrix. The mean and variance are given as follows:

$$Mean(\mathbf{R}) = \frac{\mathbf{B}}{\eta - 2m - 2},$$

(B.2)

$$Var(R_{ii}) = \frac{2B_{ii}}{(\eta - 2m - 2)^2(\eta - 2m - 4)}.$$

(B.3)

B.2 Expectation of $\mathbf{R}^{-1}$

In Chapter 3, we need to calculate the expectation of $\mathbf{R}^{-1}$ on the inverse Wishart distribution. As far as we know, since the direct reference does not exist, we show how to calculate the expectation.

The concerned expectations is,

$$\langle \mathbf{R}^{-1} \rangle = \int \mathbf{R}^{-1} \mathcal{IW}(\mathbf{R}|\eta, \mathbf{B}) d\mathbf{R}.$$

(B.4)

By transforming $\Lambda = \mathbf{R}^{-1}$, the above equation becomes

$$\langle \mathbf{R}^{-1} \rangle = \int \Lambda \cdot C(\eta, \mathbf{B}) |\Lambda|^{\frac{1}{2}\eta} \exp(-\frac{1}{2}\mathbf{B}\Lambda) |\Lambda|^{-(m+1)} d\Lambda.$$

(B.5)

Note that, since $\Lambda = \mathbf{R}^{-1}$ is the transformation of the probability density function, the Jacobian of transformation, $\mathbf{J}(\mathbf{R}^{-1} \rightarrow \Lambda) = |\Lambda|^{-(m+1)}$, needs to be considered. By reparameterizing $\kappa = \eta - m - 1$ and $\mathbf{C} = \mathbf{B}^{-1}$, Eq. (B.5) becomes

$$\langle \mathbf{R}^{-1} \rangle = \int \Lambda \cdot \mathcal{W}(\Lambda|\kappa, \mathbf{C}) d\Lambda,$$

(B.6)

where $\mathcal{W}$ denotes the Wishart distribution defined as

$$\mathcal{W}(\Lambda|\kappa, \mathbf{C}) \propto |\Lambda|^{\frac{1}{2}(\kappa-m-1)} \exp\{-\frac{1}{2}tr\mathbf{C}^{-1}\Lambda\}. \tag{B.7}$$

The concerned expectation $\langle \mathbf{R}^{-1} \rangle$ is the mean of the Wishart distribution. That is,

$$\begin{aligned} \langle \mathbf{R}^{-1} \rangle &= \kappa^{-1}\mathbf{C}, \\ &= (\eta - m - 1)\mathbf{B}^{-1}. \end{aligned} \tag{B.8}$$

Appendix C

Approximation of predictive distribution for $\mathbf{R}_t$

C.1 Formulation

In this appendix, we derive the approximation of Eq. (3.11) that denotes the tth predictive distribution for $\mathbf{R}_t$. Again, the approximating distribution is given by

$$p(\mathbf{R}_t|\mathbf{y}_{1:t-1}) = \int \mathcal{IW}(\mathbf{R}_t|\xi, (\xi - 2m - 2)\mathbf{R}_{t-1}) \cdot \mathcal{IW}(\mathbf{R}_{t-1}|\eta_{t-1|t-1}, \mathbf{B}_{t-1|t-1})d\mathbf{R}_t,$$

(C.1)

Since the exact integration is intractable, we approximate $p(\mathbf{R}_t|\mathbf{y}_{1:t-1})$ with another inverse Wishart distribution as

$$p(\mathbf{R}_t|\mathbf{y}_{1:t-1}) \approx \mathcal{IW}(\mathbf{R}_t|\eta_{t|t-1}, \mathbf{B}_{t|t-1}),$$

(C.2)

In the following, we derive this approximation, Eq. (C.2).

C.2 Derivation

The basic idea of the approximation is to find another inverse Wishart distribution that has the same mean and variance as the exact distribution. Although it is not possible to obtain a exact form of Eq. (C.1), the moments can be calculated exactly. The mean and the variance of Eq. (C.1) are

$$Mean(\mathbf{R}_t)$$
$$= \int \mathbf{R}_t \cdot \Big[\int p(\mathbf{R}_t|\mathbf{R}_{t-1})p(\mathbf{R}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{R}_{t-1} \Big] d\mathbf{R}_t$$
$$= \frac{\mathbf{B}_{t-1|t-1}}{\eta_{t-1|t-1} - 2m - 2},$$

(C.3)

$$Var(R_{t,ii})$$
$$= \langle R_{t,ii}^2 \rangle - \langle R_{t,ii} \rangle^2$$
$$= \frac{B_{t-1|t-1,ii}^2}{(\eta_{t-1|t-1} - 2m - 2)^2} \times$$
$$\Big[(1 + \frac{2}{\xi - 2m - 4})(1 + \frac{2}{\eta_{t-1|t-1} - 2m - 4}) - 1 \Big].$$

(C.4)

The mean and the variance of RHS of Eq. (C.2) are immediately obtained from Eqs. (B.2) and (B.3) as

$$Mean(\mathbf{R}_t) = \frac{\mathbf{B}_{t|t-1}}{\eta_{t|t-1} - 2m - 2}, \tag{C.5}$$

$$Var(R_{t,ii}) = \frac{2B_{t|t-1,ii}}{(\eta_{t|t-1} - 2m - 2)^2(\eta_{t|t-1} - 2m - 4)}. \tag{C.6}$$

Then, we solve Eqs. (C.3)–(C.6) with respect to $\eta_{t|t-1}$ and $\mathbf{B}_{t|t-1}$. The resultant parameters are

$$\begin{aligned} \eta_{t|t-1} &= \frac{2}{\left(1 + \frac{2}{\xi - 2m - 4}\right)\left(1 + \frac{2}{\eta_{t-1|t-1} - 2m - 4}\right) - 1} \\ &\quad + 2m + 4, \\ \mathbf{B}_{t|t-1} &= \frac{\eta_{t|t-1} - 2m - 2}{\eta_{t-1|t-1} - 2m - 2}\mathbf{B}_{t-1|t-1}, \end{aligned} \tag{C.7}$$

which correspond to Eqs. (3.12) and (3.13).

Bibliography

[1] M. Nakaya, "Mirror plant for innovating plant operation", *Yokogawa Technical Report*, vol. 53, no. 1, pp. 7–10, 2013.

[2] M. Kano and M. Ogawa, "The state of the art in chemical process control in japan: good practice and questionnaire survey", *Journal of Process Control*, vol. 20, no. 9, pp. 969–982, 2010.

[3] R. E. Kalman *et al.*, "A new approach to linear filtering and prediction problems", *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[4] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems", in *AeroSense'97*, International Society for Optics and Photonics, 1997, pp. 182–193.

[5] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics", *Journal of Geophysical Research*, vol. 99, no. C5, pp. 10 143–10 162, 1994.

[6] ——, *Data assimilation, second edition*. Springer, 2009.

[7] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, "Obstacles to high-dimensional particle filtering", *Monthly Weather Review*, vol. 136, no. 12, pp. 4629–4640, 2008.

[8] S. E. Cohn, "An introduction to estimation theory (gtspecial issueltdata assimilation in meteology and oceanography: theory and practice)", *Journal of the Meteorological Society of Japan. Ser. II*, vol. 75, no. 1B, pp. 257–288, 1997.

[9] C. Masreliez and R Martin, "Robust Bayesian estimation for the linear model and robustifying the Kalman filter", *IEEE transactions on Automatic Control*, vol. 22, no. 3, pp. 361–371, 1977.

[10] J. L. Anderson, "An ensemble adjustment Kalman filter for data assimilation", *Monthly weather review*, vol. 129, no. 12, pp. 2884–2903, 2001.

[11] G. Kitagawa, "A self-organizing state-space model", *Journal of the American Statistical Association*, pp. 1203–1215, 1998.

[12] S. S. Haykin *et al.*, *Kalman filtering and neural networks*. Wiley Online Library, 2001.

[13] J. R. Stroud and T. Bengtsson, "Sequential state and variance estimation within the ensemble Kalman filter", *Monthly Weather Review*, vol. 135, pp. 3194–3208, 2007.

[14] J.-A. Ting, E. Theodorou, and S. Schaal, "A Kalman filter for robust outlier detection", in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2007, pp. 1514–1519.

[15] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations", *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 596–600, 2009.

[16] S. Särkkä and J. Hartikainen, "Non-linear noise adaptive Kalman filtering via variational Bayes", in *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, 2013.

[17] M West, "Robust sequential approximate Bayesian estimation", *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 157–166, 1981.

[18] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models", *Journal of computational and graphical statistics*, vol. 5, no. 1, pp. 1–25, 1996.

[19] Z. M. Durovic and B. D. Kovacevic, "Robust estimation with unknown noise statistics", *IEEE Transactions on Automatic Control*, vol. 44, no. 6, pp. 1292–1296, 1999.

[20] X Wang, N Cui, and J Guo, "Huber-based unscented filtering and its application to vision-based relative navigation", *IET radar, sonar & navigation*, vol. 4, no. 1, pp. 134–141, 2010.

[21] L. Chang, B. Hu, G. Chang, and A. Li, "Robust derivative-free Kalman filter based on Huber's m-estimation methodology", *Journal of Process Control*, vol. 23, no. 10, pp. 1555–1561, 2013.

[22] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "An outlier-robust Kalman filter", in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2011, pp. 1551–1558.

[23] G. Burgers, P. J. Van Leeuwen, and G. Evensen, "Analysis scheme in the ensemble Kalman filter", *Monthly weather review*, vol. 126, no. 6, pp. 1719–1724, 1998.

[24] M. J. Beal, "Variational algorithms for approximate Bayesian inference", *PhD thesis University of Cambridge, UK*, 2003.

[25] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[26] M.-A. Sato, "Online model selection based on the variational Bayes", *Neural Computation*, vol. 13, no. 7, pp. 1649–1681, 2001.

[27] C. M. Bishop, *Pattern recognition and machine learning (information science and statistics)*. Springer-Verlag New York, Inc., 2006.

[28] I. Fukumori, "Data assimilation by models", *International Geophysics*, vol. 69, pp. 237–265, 2001.

[29] G. Ueno, T. Higuchi, T. Kagimoto, and N. Hirose, "Maximum likelihood estimation of error covariances in ensemble based-filters and its application to a coupled atmosphere ocean model", *Quarterly Journal of the Royal Meteorological Society*, vol. 136, pp. 1316–1343, 2010.

[30] V. Šmídl and R. Hofman, "Marginalized particle filtering framework for tuning of ensemble", *Monthly Weather Review*, vol. 139, pp. 3589–3599, 2010.

[31] M. Frei and H. R. Künsch, "Sequential state and observation noise covariance estimation using combined ensemble Kalman and particle filters", *Monthly Weather Review*, vol. 140, no. 5, pp. 1476–1495, 2012.

[32] S. Nakano, G. Ueno, and T. Higuchi, "Merging particle filter for sequential data assimilation", *Nonlinear Processes in Geophysics*, vol. 14, no. 4, pp. 395–408, 2007.

[33] G. Agamennoni, J. Nieto, E. M. Nebot, *et al.*, "Approximate inference in state-space models with heavy-tailed noise", *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5024–5037, 2012.

[34] S. J. Press, *Applied multivariate analysis: using Bayesian and frequentist methods of inference*. Courier Dover Publications, 2005.

[35] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*. CRC Press, 1999.

[36] J. S. Whitaker and T. M. Hamill, "Ensemble data assimilation without perturbed observations", *Monthly Weather Review*, vol. 130, no. 7, pp. 1913–1924, 2002.

[37] G. Gaspari and S. E. Cohn, "Construction of correlation functions in two and three dimensions", *Quarterly Journal of the Royal Meteorological Society*, vol. 125, no. 554, pp. 723–757, 1999.

[38] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis, second edition*. Chapman and Hall/CRC, 2003.

[39] K. Xu and C. K. Wikle, "Estimation of parametrized spatio-temporal dynamic models", *Journal of Statistical Planning and Inference*, vol. 137, pp. 567–688, 2007.

[40] E. N. Lorenz and K. A. Emanuel, "Optimal sites for supplementary weather observations: simulation with a small model", *Journal of the Atmospheric Sciences*, vol. 55, no. 3, pp. 399–414, 1998.

[41] P. J. Huber *et al.*, "Robust estimation of a location parameter", *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[42] G Ackerson and K Fu, "On state estimation in switching environments", *IEEE Transactions on Automatic Control*, vol. 15, no. 1, pp. 10–17, 1970.

[43] J. Tugnait, "Adaptive estimation and identification for discrete systems with Markov jump parameters", *IEEE Transactions on Automatic control*, vol. 27, no. 5, pp. 1054–1065, 1982.

[44] R. H. Shumway and D. S. Stoffer, "Dynamic linear models with switching", *Journal of the American Statistical Association*, vol. 86, no. 415, pp. 763–769, 1991.

[45] Z. Ghahramani and G. E. Hinton, "Variational learning for switching state-space models", *Neural computation*, vol. 12, no. 4, pp. 831–864, 2000.

[46] M. C. Mackey, L. Glass, *et al.*, "Oscillation and chaos in physiological control systems", *Science*, vol. 197, no. 4300, pp. 287–289, 1977.

[47] E. N. Lorenz, "Deterministic nonperiodic flow", *Journal of the atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.