Title
Non-linear Filtering for State Space Models - High-Dimensional Applications and Theoretical Results

Author
Lei, Jing

Publication Date
2010

Peer reviewed|Thesis/dissertation

Non-linear Filtering for State Space Models – High-Dimensional Applications and Theoretical Results

by

Jing Lei

B.S. (Peking University) 2005

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Statistics

and the Designated Emphasis

in

Communication, Computation, and Statistics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Peter Bickel, Chair
Professor Bin Yu
Professor John Rice
Professor Inez Fung

Spring 2010

The dissertation of Jing Lei is approved:

Chair Date

Date

Date

Date

University of California, Berkeley

Spring 2010

Non-linear Filtering for State Space Models – High-Dimensional Applications and Theoretical Results

by

Jing Lei

Abstract

Non-linear Filtering for State Space Models – High-Dimensional Applications and Theoretical Results

by

Jing Lei

Doctor of Philosophy in Statistics

and the Designated Emphasis in

Communication, Computation, and Statistics

University of California, Berkeley

Professor Peter Bickel, Chair

State space models are powerful modeling tools for stochastic dynamical systems and have been an important research area in the statistics community in the last several decades. This thesis makes contributions to the filtering problem, a key inference problem in general state space models. Our work in this area is motivated by both high-dimensional, nonlinear applications such as numerical weather forecasting and fundamental theoretical problems such as the convergence of filters.

First we study the ensemble Kalman filters (EnKF), a popular class of filtering methods in geophysics because they are easy to implement in large systems. However, their behavior in non-Gaussian situations is only partially understood. We compare two

common versions of EnKF's under non-Gaussianity from a robustness perspective. The results support previous empirical studies on the same issue and provide additional insight in choosing a free parameter in the EnKF algorithms. Second, we consider the filtering problem in high dimensional situations such as numerical weather forecasting. We review the EnKF from a statistical perspective and analyze its sources of bias. Then we propose a new method to reduces the bias, namely the non-linear ensemble adjustment filter (NLEAF). The one-step consistency of the NLEAF is studied and the performance is examined through simulations in two common testbeds in the weather forecasting literature. Finally we look at the theoretical properties of another popular class of filtering methods, the sequential Monte Carlo (SMC) filter. The convergence of SMC filters has been a challenging problem in both probability and statistics. The previous results either depend on strong mixing conditions which only hold in compact spaces or provide no rates of convergence or are under weak notions of distance, limiting the application of their practical use. We provide checkable sufficient conditions under which explicit rates of convergence of the SMC filter can be derived. The conditions essentially requires the regularity of the tail behavior of the process and they are general enough to include a wide class of autoregressive models as well as Gaussian linear models.

Professor Peter Bickel
Dissertation Committee Chair

To my parents,

Chunying Cao,

Haijun Lei.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my adviser Professor Peter Bickel, who led me into such an interesting area and made my Ph.D. study a wonderful experience. I would also like to thank Chris Snyder and Thomas Bengtsson for inspiring discussions and helpful comments, and Kehui Chen for being so understanding and supportive over these years.

Chapter 1

Introduction

A state space model consists of an underlying unobservable process – the (hidden) state process, and a sequence of incomplete and noisy functions of the state – the observation process. Usually the state process is assumed to be Markovian therefore it is also known as the hidden Markov model especially when the state space is discrete. State space models originated in engineering in the early sixties with the most famous names being the Kalman–Bucy filter (Kalman, 1960; Kalman & Bucy, 1961) and Baum with an early version of EM algorithm (Baum & Petrie, 1966; Baum et al., 1970). After their appearance, state space models have been continuously used in control engineering and speech recognition. In the last two decades it has become an important area in probability and statistics because of its wide application in engineering, computer sciences, biology, econometrics and geophysics. For a thorough introduction of state space models and its applications we refer the reader to the book chapter by Künsch (2001).

Specifically, a state space model features a Markovian state process $\{X_i \in \mathcal{X} =$

$\mathbb{R}^p, i \geq 0\}$ with transition kernel $q(\cdot,\cdot)$:

$$(X_{i+1}|X_i = x) \sim q(x,\cdot), \quad i \geq 0,$$

and an observation sequence $\{Y_i \in \mathcal{Y} = \mathbb{R}^d, i \geq 1\}$, where Y_i's are conditionally independent given X_i's, with likelihood $g(\cdot;\cdot)$:

$$(Y_i|X_i = x) \sim g(\cdot;x), \quad i \geq 1.$$

The joint distribution of $(X_i, Y_i)_{i=1}^t$ is determined by q, g and ϕ_0, the initial distribution of X_0. Typical inference tasks in state space models include: 1) estimation of parameters in the dynamics $q(\cdot,\cdot)$ and/or the observation mechanism $g(\cdot;\cdot)$ Bickel et al. (1998); Olsson & Rydén (2008); and 2) calculating the conditional distribution, $\phi_{i|s}$, of state variables X_i given the observations Y_1^s Liu & Chen (1998), where $Y_1^s = (Y_1, \ldots, Y_s)^T$. Calculating $\phi_{i|s}$ for $s = i$, $s > i$ and $s < i$ are called filtering, smoothing and predicting, respectively. This thesis focuses on the filtering problem. Therefore from now on we assume that q and g are known.

1.1 The filtering recursion

Let $p_Z(\cdot)$ denote the density function of a random variable Z. The conditional density of X_i given Y_1^s is specially written as $\phi_{i|s}(\cdot)$.

The dependence structure of a state space model can be described by the following diagram:

$$\begin{array}{ccccccccc} \ldots & \longrightarrow & X_{i-1} & \longrightarrow & X_i & \longrightarrow & X_{i+1} & \longrightarrow & \ldots \\ & & \downarrow & & \downarrow & & \downarrow & & \\ \ldots & & Y_{i-1} & & Y_i & & Y_{i+1} & & \ldots \end{array}$$

This graph representation leads to some basic recursive formulas which we state without proof (see Künsch (2001)).

Suppose at time $i \geq 1$ we have obtained $\phi_{i-1|i-1}$, then the one-step forecast distribution of X_i giving Y_1^{i-1} is obtained by applying the Markov transition kernel q on the density function $\phi_{i-1|i-1}$:

$$\phi_{i|i-1}(x_i) = \int \phi_{i-1|i-1}(x_{i-1}) q(x_{i-1}, x_i) dx_{i-1}. \tag{1.1}$$

When the new observation $Y_i = y_i$ is available, the distribution of X_i given $Y_1^i = y_1^i$ (the filtering distribution) is obtained by applying Bayes rule to the forecast density $\phi_{i|i-1}$ with likelihood function g_i :

$$\phi_{i|i}(x_i) = \frac{\phi_{i|i-1}(x_i) g_i(x_i)}{\int \phi_{i|i-1}(x) g_i(x) dx}, \tag{1.2}$$

where

$$g_i(\cdot) := g(y_i; \cdot).$$

The most famous example of recursive filtering is the Kalman filter. Consider a Gaussian linear State space model:

$$X_t = F X_{t-1} + U_t, \tag{1.3}$$

$$Y_t = H X_t + V_t, \tag{1.4}$$

for all $t \geq 1$ and $U_t \overset{iid}{\sim} N(0, \Sigma)$, $V_t \overset{iid}{\sim} N(0, R)$ are Gaussian random variables. Then $\phi_{t|t-1} \sim N(\mu_{t|t-1}, \Sigma_{t-1})$, with

$$\mu_{t|t-1} = F \mu_{t-1}, \tag{1.5}$$

$$\Sigma_{t|t-1} = F\Sigma_{t-1}F^T + R. \tag{1.6}$$

Moreover, $\phi_{t|t} \sim N(\mu_{t|t}, \Sigma_{t|t})$, with

$$\mu_{t|t} = \mu_{t|t-1} + K_t(y_t - H\mu_{t|t-1}), \tag{1.7}$$

$$\Sigma_{t|t} = (I - K_tH)\Sigma_{t|t-1}, \tag{1.8}$$

where

$$K_t = \Sigma_{t|t-1}H^T \left(H\Sigma_{t|t-1}H^T + R\right)^{-1} \tag{1.9}$$

is the Kalman gain.

Unfortunately, except a few special cases such as the Gaussian linear model mentioned above and the finite state hidden Markov chain (Baum & Petrie, 1966), in general the prediction (1.1) and Bayes update (1.2) do not permit any close-form solutions. Usually this difficulty is tackled by ensemble filtering methods which use Monte Carlo methods to approximate the conditional distributions. A general ensemble filtering algorithm can be outlined as following:

A general ensemble filtering algorithm

1. (Initialize) Generate random sample $\{x_0^j\}_{j=1}^n$, from initial distribution ϕ_0. Set time $t = 0$.

2. $t \to t + 1$.

   (a) (One-step forecasting) Generate random sample

$$\{x_{t|t-1}\}_{j=1}^n \overset{iid}{\sim} \hat{\phi}_{t|t-1}(x) := \frac{1}{n}\sum_{j=1}^n q(x_{t-1}^j, x).$$

(b) (Update) Generate random sample

$$\{x_t^j\}_{j=1}^n \overset{iid}{\sim} \hat{\phi}_{t|t}(x) \propto \hat{\phi}_{t|t-1} g(y_t; x).$$

The forecasting step is usually straightforward because the Markov kernel is known. However the update step is much more sophisticated because of the unknown normalization constant in the target distribution in step (b) and the Bayes operator used in the update step is non-linear and hence intractable. Moreover, if the dimensionality of X_t or Y_t is high, the method might face the "curse of dimensionality". That is, the sample size n has to be prohibitively large to ensure convergence. There have been both practical update algorithms which work for high-dimensional data such as the ensemble Kalman filter (EnKF, Evensen (1994)) and methods with nice theoretical properties but with poor scalability in dimensionality (sequential Monte Carlo –SMC– filter, Gordon et al. (1993)). Major open problems in ensemble filtering include, but are not limited to, 1) understanding the behavior of ensemble Kalman filters under non-Gaussian situations; 2) designing of better filtering algorithms that are less biased than the EnKF but more stable than SMC filters; and 3) develop convergence results for SMC filters. In the rest of this chapter we review the related works in each of these directions and highlight our contributions.

1.2 The ensemble Kalman filter and the issue of non-Gaussianity

The ensemble Kalman filter (EnKF),an empirical version of the Kalman filter, is mostly used in geophysical data assimilation and has performed successfully in high dimensional models (Evensen, 2007). Like other ensemble filtering algorithms, the EnKF

also uses a random sample (ensemble) to represent the forecasting and filtering distributions. From now on, we will adopt the ensemble filtering terminology. The random sample and a single sample point will be called "the ensemble" and "a particle", respectively.

In the update step, the EnKF pretends that $\phi_{t|t-1}$ is Gaussian and the observation is linear with Gaussian noise at in (1.4). Therefore the ensemble is updated to have the updated mean and variance, as in the Kalman filter. There have been two major version of EnKF's, the stochastic update and the deterministic update. Stochastic methods (Houtekamer & Mitchell, 1998; Evensen, 2003) directly use the Kalman gain together with random perturbations:

$$x_t^j = x_{t|t-1}^j + K_t(y_t - Hx_t - \epsilon_t^j), \tag{1.10}$$

with $\epsilon_t^j \overset{iid}{\sim} N(0, R)$, where the matrices H and K_t are defined as in Equations 1.3 – 1.9. This is an extension of the Kalman filter update of the mean (Eq. 1.7), where the additive error term is used to adjust the updated covariance.

On the other hand, deterministic methods (Anderson, 2001; Bishop et al., 2001) use a deterministic transformation on the forecast ensemble, which is also known as a special case of the Kalman square-root filter (Tippett et al., 2003):

$$x_t^j = \mu_t + A_t(x_{t|t-1}^j - \mu_{t|t-1}), \tag{1.11}$$

where the matrix A_t satisfies $A_t\Sigma_{t|t-1}A_t^T = \Sigma_t$. Again, μ_t, $\mu_{t|t-1}$, Σ_t, and $\Sigma_{t|t-1}$ are defined as in Equations 1.3 – 1.9.

It is obvious that both methods are asymptotically unbiased under Gaussian linear models. An important open problem is their behavior under non-Gaussian situations. Following the direction of Lawson & Hansen (2004), we compare the asymptotic behavior

of the two versions of EnKF's under non-Gaussianity through a robustness perspective. Our question is: which method is more stable against non-Gaussianity? Here "stability" is a statistical notion which refers to the analysis being not seriously biased when the forecast distribution is slightly non-Gaussian. Another notion of "stability" is introduced by Sacher & Bartello (2009) which refers to the size of analysis error covariance being large enough to cover the true analysis center.

In Chapter 2 we give a rigorous analysis of the sensitivity of the two EnKFs to non-Gaussianity of the forecasting ensemble based on the notion of *robustness* in statistics. We show that the stochastic filter is more robust than the deterministic filter especially when the position of outliers is wild and/or the observation is accurate. Simulation results support our calculation not only for the L_2 distance but also for other quantities such as the third moment. These findings are consistent with those in Lawson & Hansen (2004). Moreover, such a comparison can be extended to many other types of model violations, such as the modeling error in the observation and the observation model. On the other hand, we also show that such a stability criterion leads to a natural choice of the orthogonal matrix in the unbiased ensemble square root filter Sakov & Oke (2007); Livings et al. (2008). Chapter 2 is based on Lei et al. (2009).

1.3 Improving the EnKF: the NLEAF algorithm

The EnKF update formulas (1.10) and (1.11) are both variants of the Kalman filter in the sense that they adjust only the first and second moments of the forecasting ensemble $\{x^j_{t|t-1}\}_{j=1}^n$. A simpler way to generate a sample with the desired mean and

variance would be to sample directly from the corresponding Gaussian distribution. However, this method is not successful at least in the weather forecasting literature because the forecasting distribution is strongly non-Gaussian. For example, the forecasting sample may have clusters or may be tilted, sampling from a Gaussian will lose this information. Using proper transformation on the sample points as in Equations (1.10) and (1.11) may partially retain such information.

When the forecast distribution $\phi_{t|t-1}$ is non-Gaussian, adjusting only the first two moments will inevitably introduce bias in the updated ensemble. On the other hand, one needs some simple update procedures for very large data sets such as those in numerical weather forecasting. Therefore it is highly desirable to find a method with less bias for non-Gaussian distributions while being scalable in dimensionality.

A natural approach to obtain higher accuracy for general non-linear non-Gaussian filtering problem is to sample from the whole conditional distribution, rather than focusing only on the first two moments. This approach, known as the particle filter (SMC filter), was introduced by Gordon et al. (1993)[1]. The basic idea of the update from $\phi_{t-1|t-1}$ to $\phi_{t|t}$ is directly generating independent random samples from the target distribution

$$\hat{\phi}_{t|t}(x) \propto \frac{1}{n}\sum_{j=1}^{n} q(x_{t-1}^{j}, x)g(y_t; x), \tag{1.12}$$

where $\{x_{t-1}^{j}\}_{j=1}^{n} \overset{iid}{\sim} \hat{\phi}_{t-1|t-1}$, and the recursion starts from $\hat{\phi}_{0|0} = \phi_0$. Here the notation $\hat{\phi}$ means the estimated function ϕ. For details about sampling schemes, we refer the reader to Künsch (2005).

It is straightforward to show that the particle filter is consistent for general mod-

[1]In this thesis, the terms "SMC filter" and "particle filter" will be used interchangeably.

els provided that the Markov kernel q and likelihood function g are smooth. However, as many other non-parametric methods, it suffers from the curse of dimensionality. Bengtsson et al. (2008) showed that under simple Gaussian models, the particle filter collapse in a single step unless the sample size n grows exponentially in the dimension of Y_t. Here the filter collapse means all particles become identical and the updated distribution becomes a point mass.

Particle filters have been used in numerical weather forecasting in low dimensional problems (Anderson & Anderson, 1999). Some early efforts toward high-dimensional cases include Bengtsson et al. (2003), who extended the EnKF to Gaussian mixtures. Their method can also be viewed as a hybrid of the EnKF and PF, combining the computational advantage of ensemble Kalman filters and the accuracy of particle filters.

In Chapter 3 we develop the NonLinear Ensemble Adjustment Filter (NLEAF) as another combination of the advantages of both the EnKF and the PF. The basic idea is to view the EnKF as a regression of the hidden state on the observations and then instead of using the Kalman filter update, one can use more accurate techniques such as importance sampling to update the moments to avoid serious bias in non-Gaussian nonlinear models. It is theoretically justifiable under certain (possibly strong) conditions. In numerical experiments, it gives very competitive performance in a 40 dimensional synthesized atmospheric model.

1.4 Convergence of Sequential Monte Carlo Filters

Sequential Monte Carlo filters (particle filters, Gordon et al. (1993); Liu & Chen (1998); Künsch (2005)) have been a major breakthrough in non-linear state space models. This is probably due to both the fast development of accurate physical models and the rapid growth of computing power. Using empirical approximation, the infinite-dimensional object $\phi_{t|t}$ can be represented by a random sample (ensemble) of sample points (particles) independently drawn from it. In principle, such an ensemble allows one to approximate the conditional expectation of any function of X_t given the observations:

$$\hat{E}(f(X_t)|y_1^t) = \frac{1}{n} f(x_{t|t}^j).$$

Two major issues in the study of SMC filters are 1) the development of accurate and efficient algorithms for sampling problem of form (1.12) since the target distribution is known only up to a normalizing constant; and 2) the relationship between the approximation error and the ensemble size n since the errors are propagated non-linearly over time. Our interest is in the second problem. That is, we assume that one can generate exact random samples from (1.12). The sampling issue is of course very important and we refer the reader to Künsch (2005) for a nice discussion.

To formalize the question, we look at the distance between the approximation and the truth:

$$||\hat{\phi}_{t|t} - \phi_{t|t}||,$$

where $||\cdot||$ is some norm of continuous functions. In particular, the L_1 norm (or total

variation) norm

$$||f||_{\mathrm{TV}} = \frac{1}{2}\int |f(x)|dx$$

is commonly studied and is also considered in this thesis.

At a first glance, by the continuity of the Markov operator and Bayes operator, it is easy to establish consistency

$$||\hat{\phi}_{t|t} - \phi_{t|t}|| \to 0, \text{ as } n \to 0,$$

by induction on t. However, such an argument indicates that in order to keep the approximation error small, n needs to grow exponentially in t, the length of the time range. It is more interesting to ask the relationship between ensemble size n and the time-uniform approximation error:

$$\sup_{1\le i\le t} ||\hat{\phi}_{i|i} - \phi_{i|i}||,$$

or

$$\sup_{t\ge 0} E||\hat{\phi}_{t|t} - \phi_{t|t}||.$$

Then the problem is how large n needs to be in order to have the time-uniform approximation error be small? Controlling the above quantities requires conditions that the sampling error does not grow over time even with a finite sample size.

Time-uniform convergence of SMC filters has been a challenging topic in both probability and statistics because the Bayes operator is non-linear and usually intractable. The first breakthrough is due to Del Moral & Guionnet (2001), who established uniform convergence under strong mixing conditions which typically hold in compact state spaces.

The key assumption is

$$\sigma_- a(x) \leq q(x', x) \leq \sigma^+ a(x), \quad \forall\ x', x,$$

where $0 < \sigma_- \leq \sigma_+$ are positive constants and $a(\cdot)$ is a fixed density function. This condition is originally proposed to show the stability of the optimal filter (true filter). That is, the conditional chain forgets its initial distribution geometrically fast, which is an important building block in the proof of uniform convergence of SMC filters.

Recently, Douc et al. (2009b) suggested a possible extension of the results from compact spaces to general spaces, provided that the observation is informative enough so that the hidden state stays within a compact set with overwhelming probability. This idea is used to establish the filter stability instead of convergence of SMC filters. Heine & Crisan (2008) and van Handel (2009) tried to make the similar idea work for SMC filters. However, their results are under weaker notions of norm instead of the commonly used total variation norm and do not provide rates of convergence.

In Chapter 4 we obtain convergence results for non-compact state spaces by analyzing the tail behavior of the Markov kernel and observation likelihood function. Our conditions cover a much broader family of models including a general class of nonlinear autoregressive models and stationary Gaussian linear models.

Chapter 2

Ensemble Kalman Filters under Non-Gaussianity

The ensemble Kalman filter (EnKF, Evensen (1994, 2003, 2007)) has become a popular tool for data assimilation because of its computational efficiency and flexibility (Anderson, 2001; Whitaker & Hamill, 2002; Ott et al., 2004; Bengtsson et al., 2003; Evensen, 2007). In various versions of EnKFs, one major difference is how to get the updated ensemble after obtaining the updated mean and variance. Stochastic methods (Houtekamer & Mitchell, 1998; Evensen, 2003) directly use the Kalman gain together with random perturbations. On the other hand, deterministic methods (Anderson, 2001; Bishop et al., 2001) use a non-random transformation on the forecast ensemble, which is also known as a special case of the Kalman square-root filter (Tippett et al., 2003).

The analysis error of EnKF consists of two parts: the use of a linear analysis algorithm that is suboptimal for all except Gaussian distributions; and the variance caused

by using only a finite sample. The latter is studied for the stochastic filter by Sacher & Bartello (2008, 2009). In this chapter we study the first part of error, that is, the error caused by non-Gaussianity.

Following the direction of Lawson & Hansen (2004), who did empirical comparison of the stochastic and deterministic filters, in this work we attempt to quantify the difference between these two methods under non-Gaussianity, through the perspective of *robustness*. It is known that in a Gaussian linear model both methods are consistent (Furrer & Bengtsson, 2007). However, when the forecasting distribution is non-Gaussian both methods are biased even asymptotically, where the bias refers to the deviation from the true conditional distribution or equivalently the distribution given by the Bayes rules. Suppose the previous updated ensemble is approximately Gaussian. After propagation through the non-linear dynamics, the resulting forecast ensemble will be slightly non-Gaussian if the time interval is short. Figure 2.1 gives such an example by looking at the first two coordinates of the Lorenz 63 3-dimensional system[1], where the previous update ensemble is Gaussian but the forecasting ensemble has some outliers. Therefore one would expect some bias in EnKF update due to the non-Gaussianity, and the bias could be different for different implementation of EnKF. Our question is: which method is more stable against non-Gaussianity? Here "stability" is a statistical notion which refers to the analysis being not seriously biased when the forecast distribution is slightly non-Gaussian. Another notion of "stability" is introduced by Sacher & Bartello (2009) which refers to the size of analysis error covariance being large enough to cover the true analysis center.

[1]The Lorenz 63 system (Lorenz, 1963) is a three dimensional continuous chaotic system, which is very sensitive to initial conditions in the discrete-step form. It has been used to test filtering methods in many data assimilation research works (see Anderson & Anderson, 1999; Bengtsson, Snyder, & Nychka, 2003).

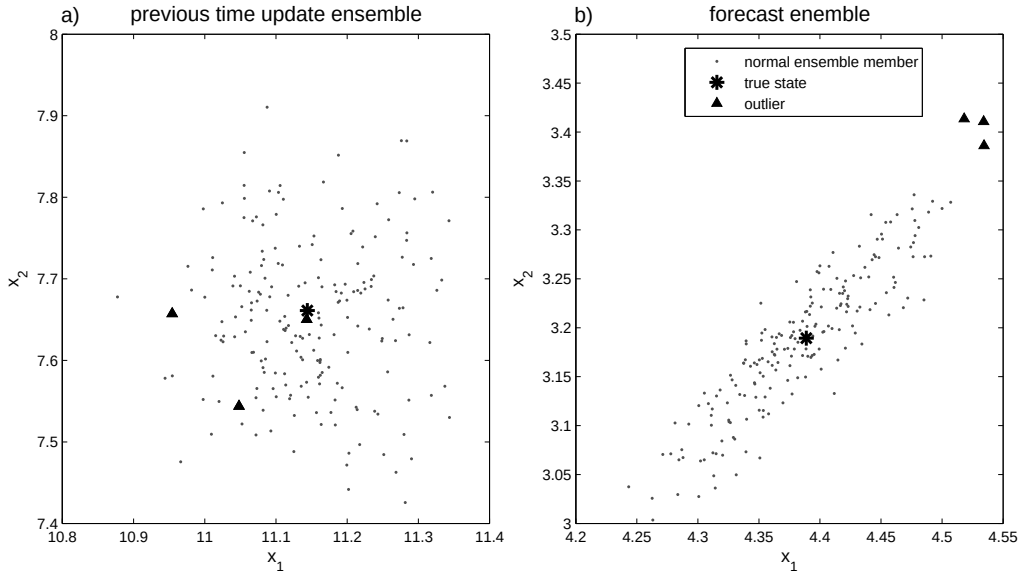

Figure 2.1: The scatter plots of the previous updated ensemble (left) and the forecast ensemble (right) in the Lorenz 63 system (simulated using fourth order Runge-Kutta method with step size 0.05, propagated 4 steps).

We give a rigorous analysis of the sensitivity of the two EnKFs to non-Gaussianity of the forecasting ensemble based on the notion of *robustness* in statistics.

We show that the stochastic filter is more robust than the deterministic filter especially when the position of outliers is wild and/or the observation is accurate. Simulation results support our calculation not only for the L_2 distance but also for other quantities such as the third moment. These findings are consistent with those in Lawson & Hansen (2004). Moreover, such a comparison can be extended to many other types of model violations, such as the modeling error in the observation and the observation model. On the other hand, we also show that such a stability criterion leads to a natural choice of the orthogonal matrix in the unbiased ensemble square root filter Sakov & Oke (2007);

Livings et al. (2008).

In Section 2.1 we introduce the ensemble Kalman filters, with a brief discussion on the large-ensemble behavior of the EnKF. Section 2.2 contains the main part of our comparison, beginning with some intuition in Section 2.2.1; The basic concepts of asymptotic robustness can be found in Hampel et al. (1986), and we give a brief summary in Section 2.2.2; In Section 2.2.3 we state our analytical results. Finally, in Section 2.3, we present various numerical experiments.

2.1 Ensemble Kalman filters

2.1.1 The Kalman filter

Consider a Gaussian linear model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \epsilon,$$

where $\mathbf{x} \in \mathbb{R}^p$ is the hidden state variable, $\mathbf{y} \in \mathbb{R}^q$ the observation, $\epsilon \in \mathbb{R}^q$ an independent random noise, and $\mathbf{H} \in \mathbb{R}^{q\times p}$ the observation matrix. Assuming all the variables are Gaussian:

$$\mathbf{x} \sim N(\mu^{\mathrm{f}}, \mathbf{P}^{\mathrm{f}}), \quad \epsilon \sim N(0, \mathbf{R}),$$

then the updated state variable $\mathbf{x}|\mathbf{y}^{\mathrm{o}}$ given a specific observation $\mathbf{y}^{\mathrm{o}}$ is still Gaussian[2]:

$$\mathbf{x}|\mathbf{y}^{\mathrm{o}} \sim N(\mu^{\mathrm{a}}, \mathbf{P}^{\mathrm{a}}),$$

with

$$\mu^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mu^{\mathrm{f}} + \mathbf{K}\mathbf{y}^{\mathrm{o}}, \quad \mathbf{P}^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^{\mathrm{f}}, \tag{2.1}$$

[2]Throughout this chapter we use superscript "f" and "a" to denote "forecast" and "analysis (update)" respectively.

where $\mathbf{K} = \mathbf{P}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$ is the *Kalman Gain*. Throughout this chapter we always assume that $\mathbf{P}^{\mathrm{f}}$ and $\mathbf{R}$ are positive definite.

Several practical issues arise in geophysics. First, the state variable is driven by non-linear geophysical dynamics, so its exact distribution is unknown and certainly is non-Gaussian. Usually only a random sample from the distribution is available. Second, the linear form of the observation is, again, only an approximation. The true observation model $\mathbf{y} = h(\mathbf{x}) + \varepsilon$ might involve a nonlinear $h(\cdot)$, or $h(\cdot)$ might even have no explicit functional form (e.g., a black-box function). These problems are partially addressed, as described below, by the ensemble Kalman filter.

2.1.2 The ensemble Kalman filter

Suppose $(\mathbf{x}^{\mathrm{f}(i)})_{i=1}^{n}$ is an i.i.d (independent, identically distributed) sample from the forecast distribution of the state variable $\mathbf{x}^{\mathrm{f}}$. The ensemble Kalman filter update consists of the following steps:

1. Let $\hat{\mu}^{\mathrm{f}}$ and $\hat{\mathbf{P}}^{\mathrm{f}}$ be the sample mean and covariance.

2. Estimate the Kalman gain: $\hat{\mathbf{K}} = \hat{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\hat{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$.

3. Update the mean and covariance according to the Kalman filter:

$$< \hat{\mu}^{\mathrm{a}} >= (\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\hat{\mu}^{\mathrm{f}} + \hat{\mathbf{K}}\mathbf{y}^{\mathrm{o}}, \quad < \hat{\mathbf{P}}^{\mathrm{a}} >= \left(\mathbf{I} - \hat{\mathbf{K}}\mathbf{H}\right)\hat{\mathbf{P}}^{\mathrm{f}},$$

   where $< \cdot >$ denotes the expectation over the randomness of the update procedure. If the update is deterministic, then $< \hat{\mu}^{\mathrm{a}} >= \hat{\mu}^{\mathrm{a}}$ and $< \hat{\mathbf{P}}^{\mathrm{a}} >= \hat{\mathbf{P}}^{\mathrm{a}}$.

4. Update the ensemble $(\mathbf{x}^{\mathrm{f}(i)})_1^n \to (\mathbf{x}^{\mathrm{a}(i)})_1^n$, so that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}^{\mathrm{a}(i)} = \hat{\mu}^{\mathrm{a}}, \quad \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}^{\mathrm{a}(i)} - \hat{\mu}^{\mathrm{a}})(\mathbf{x}^{\mathrm{a}(i)} - \hat{\mu}^{\mathrm{a}})^{\mathrm{T}} = \hat{\mathbf{P}}^{\mathrm{a}}. \tag{2.2}$$

It is worth noting that in practice, the sample covariance matrix $\hat{\mathbf{P}}^{\mathrm{f}}$ is not computed explicitly. Instead, it is sufficient to compute $\hat{\mathbf{P}}^{\mathrm{f}}\mathbf{H}^T = \frac{1}{n-1}\sum(\mathbf{x}^{\mathrm{f}(i)})\left(\mathbf{H}\mathbf{x}^{\mathrm{f}(i)}\right)^{\mathrm{T}}$, which is computationally more efficient if p is much larger than q.

The stochastic and the deterministic filters differ in step 4. In the stochastic filter,

$$\mathbf{x}_{\mathrm{s}}^{\mathrm{a}(i)} = \mathbf{x}^{\mathrm{f}(i)} + \hat{\mathbf{K}}(\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{f}(i)} + \epsilon^{(i)}), \quad \forall 1 \le i \le n, \tag{STO.}$$

where $\epsilon^{(i)} \overset{iid}{\sim} N(0, R)$. The intuition is to use directly the Kalman gain to combine the forecast ensemble member $\mathbf{x}^{\mathrm{f}(i)}$ and the observation $\mathbf{y}^{\mathrm{o}}$, using additive noise $\epsilon^{(i)}$ to adjust the total variance of the updated ensemble, as if the perturbed observation associated with $\mathbf{x}^{\mathrm{f}(i)}$ is another possible value of random variable $\mathbf{y}$. In some applications in order to reduce the sampling error of the noise, $\epsilon^{(i)}$'s are adjusted by a shifting and rescaling to ensure one of the following:

- $\epsilon^{(i)}$'s have zero mean.
- $\epsilon^{(i)}$'s have zero mean and covariance $\mathbf{R}$.
- $\epsilon^{(i)}$'s have zero mean, covariance $\mathbf{R}$ and zero covariance with $X_f^{(i)}$'s.

When the ensemble size n is large, such a shifting and rescaling is negligible and all these variants are equivalent to the update given by (STO.). Therefore the analysis in this chapter is applicable to these variants too.

The deterministic filter works in a different way:

$$\mathbf{x}_{\mathrm{d}}^{\mathrm{a}(i)} = \hat{\mu}^{\mathrm{a}} + \hat{\mathbf{A}}(\mathbf{x}^{\mathrm{f}(i)} - \hat{\mu}^{\mathrm{f}}), \quad \forall 1 \leq i \leq n, \tag{DET.}$$

where $\hat{\mathbf{A}}$ satisfies $\hat{\mathbf{A}}\hat{\mathbf{P}}^{\mathrm{f}}\hat{\mathbf{A}}^{\mathrm{T}} = \hat{\mathbf{P}}^{\mathrm{a}}$. Loosely speaking, the matrix $\hat{\mathbf{A}}$ can be viewed as the square root of the difference between $\hat{\mathbf{P}}^{\mathrm{a}}$ and $\hat{\mathbf{P}}^{\mathrm{f}}$. The matrix $\hat{\mathbf{A}}$ is not unique in the multivariate case. Suppose $n > p$ and $\hat{\mathbf{P}}^{\mathrm{f}}$ is full rank, then $\hat{\mathbf{A}}$ has the general form:

$$\hat{\mathbf{A}} = (\hat{\mathbf{P}}^{\mathrm{a}})^{\frac{1}{2}}\mathbf{U}(\hat{\mathbf{P}}^{\mathrm{f}})^{-\frac{1}{2}}, \tag{2.3}$$

where $\mathbf{U}$ is any $p \times p$ orthogonal matrix chosen by the user. See Tippett et al. (2003); Sakov & Oke (2007) for further discussion on the choice of $\mathbf{U}$. If $n \leq p$ and $\hat{\mathbf{P}}^{\mathrm{f}}$ is not full rank, (2.3) no longer holds but one can work on the principal components of the state space instead of the whole state space as described in Ott et al. (2004).

There is another formula for the update step of the deterministic filter using the right-multiplication:

$$\mathbf{x}_{\mathrm{d}}^{\mathrm{a}(i)} = \hat{\mu}^{\mathrm{a}} + \sum_{j=1}^{n} \hat{a}'_{ij}(\mathbf{x}^{\mathrm{f}(j)} - \hat{\mu}^{\mathrm{f}}). \tag{2.4}$$

This formula can be shown to be closely related to (DET.) when the filter is unbiased, i.e., $\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_{\mathrm{d}}^{\mathrm{a}(i)} = \hat{\mu}^{\mathrm{a}}$ (Tippett et al., 2003; Livings et al., 2008). We will use the left-multiplication throughout this chapter because: 1) it has a clear geometrical interpretation; 2) we assume that n is large.

In practical applications, good performance of the EnKFs defined by (STO.) and (DET.) depends on a sufficiently large ensemble and on system dynamics and observation models that are sufficiently close to linear. For example, the EnKF will dramatically underestimate $\mathbf{P}^{\mathrm{a}}$ with small ensembles as it is analytically described by Sacher & Bartello

(2008). As a result, covariance localization and covariance inflation have been widely used to overcome such practical difficulties (Whitaker & Hamill, 2002; Ott et al., 2004; Anderson, 2003, 2007).

2.1.3 The large-ensemble behavior of the EnKF

If $n \to \infty$, then by law of large numbers, everything converges to its population counterpart. That is, $\hat{\mu}^{\mathrm{f}} \xrightarrow{P} \mu^{\mathrm{f}}$, $\hat{\mathbf{P}}^{\mathrm{f}} \xrightarrow{P} \mathbf{P}^{\mathrm{f}}$, $\hat{\mathbf{K}} \xrightarrow{P} \mathbf{K}$, $\hat{\mu}^{\mathrm{a}} \xrightarrow{P} \mu^{\mathrm{a}}$, $\hat{\mathbf{P}}^{\mathrm{a}} \xrightarrow{P} \mathbf{P}^{\mathrm{a}}$, and $\hat{\mathbf{A}} \xrightarrow{P} \mathbf{A}$ where $\mathbf{A} = (\mathbf{P}^{\mathrm{a}})^{\frac{1}{2}}\mathbf{U}(\mathbf{P}^{\mathrm{f}})^{-\frac{1}{2}}$ is the population counterpart of $\hat{\mathbf{A}}$. Here $\xrightarrow{P}$ denotes convergence in probability[3]. Let $\delta_{\mathbf{x}}$ denote the point mass at $\mathbf{x}$ (i.e., a probability distribution that puts all its mass at $\mathbf{x}$), then intuitively the empirical updated distributions $\hat{F}_{\mathrm{s}} = \frac{1}{n}\sum \delta_{\mathbf{x}_{\mathrm{s}}^{\mathrm{a}(i)}}$ and $\hat{F}_{\mathrm{d}} = \frac{1}{n}\sum \delta_{\mathbf{x}_{\mathrm{d}}^{\mathrm{a}(i)}}$ should converge weakly to the distribution of the random variables $(\mathbf{I} - \mathbf{KH})\mathbf{x} + \mathbf{K}(\mathbf{y} + \epsilon)$ and $\mu^{\mathrm{a}} + \mathbf{A}(\mathbf{x} - \mu^{\mathrm{f}})$, respectively. In fact it can be shown that the above intuition is true (Proposition 2.5.1). As a result, our comparison between the stochastic filter and the deterministic filter will be based on the comparison between these two limiting distributions.

2.2 Comparing the stochastic and the deterministic filters

2.2.1 Intuition and the contaminated Gaussian model

A simple and natural deviation from Gaussianity is a contaminated Gaussian model:

$$\mathbf{x}^{\mathrm{f}} \sim F_r = (1 - r)F + rG, \tag{2.5}$$

[3]For a sequence of random variables α_n, $n \geq 1$, and constant β, $\alpha_n \xrightarrow{P} \beta$ means that for any $\delta > 0$, $\lim_{n\to\infty} P(|\alpha_n - \beta| > \delta) = 0$.

where, without loss of generality, $F = N(0, \mathbf{P})$, $G = N(t, \mathbf{S})$, where $\mathbf{P}$ and $\mathbf{S}$ are positive definite, and $0 \leq r < 1$ is the amount of contamination. The interpretation of model (2.5) is that we assume a proportion of $(1 - r)$ of the forecast ensemble are drawn from a Gaussian distribution centered at 0, with covariance $\mathbf{P}$, while the rest are outliers coming from another Gaussian distribution centered at t with covariance $\mathbf{S}$. Since we use the Gaussian distribution $G = N(t, \mathbf{S})$ to model the outliers, we would expect G to be much different from $F = N(0, \mathbf{P})$, the majority of the forecast ensemble. That is, we expect (t, S) to be somewhat extreme: $||t||_2 >> 0$ and/or $||S||_2 >> ||\mathbf{P}||_2$. For example, a large[4] t and small $\mathbf{S}$ mean that the outliers forms a small cluster far away from the majority, while a small t and a large $\mathbf{S}$ mean that the outliers are widely dispersed. Also, denote $F_{\mathrm{o},r}(\cdot|\mathbf{y})$ the true distribution of $\mathbf{x}^{\mathrm{a}}$, here the subindex "o" stands for "optimal". Again, the optimal updated distribution refers to the one given by the Bayes rule. Similarly, the corresponding limiting updated distributions of EnKFs are denoted by $F_{\mathrm{s},r}(\cdot|\mathbf{y})$ and $F_{\mathrm{d},r}(\cdot|\mathbf{y})$, respectively. Here we keep in mind that t and $\mathbf{S}$ are fixed. For simplicity, we focus on the case $q = p$ and $\mathbf{P} = \mathbf{I}_p$.

The merit of a filter can be characterized naturally in terms of the distance between the updated density and the optimal density $f_{\mathrm{o},r}$. Recall that if $\mathbf{x}^{\mathrm{f}}$ is Gaussian, i.e., $r = 0$, then $F_{\mathrm{s},0}$ and $F_{\mathrm{d},0}$ are both Gaussian, with the same mean and covariance agreeing with the optimal conditional distribution: $F_{\mathrm{s},0} = F_{\mathrm{d},0} = F_{\mathrm{o},0} = N(\mu_{\mathrm{o}}^{\mathrm{a}}, \mathbf{P}_{\mathrm{o}}^{\mathrm{a}})$. Now the question is, when $r \neq 0$, i.e., $\mathbf{x}^{\mathrm{f}}$ is non-Gaussian, which one is closer to $F_{\mathrm{o},r}$?

We take a quick look at the densities of $F_{\mathrm{o},r}$, $F_{\mathrm{s},r}$ and $F_{\mathrm{d},r}$ in a simple one-

[4]Here and throughout this chapter, by saying a vector or matrix is large we mean its L_2 norm is large.

dimensional setup similar to Lawson & Hansen (2004), but with $r = 0.05$ (right column of Figure 2.2). The original figure in Lawson & Hansen (2004) with $r = 0.5$ are included in the left column for comparison. We choose $t = 8$, $\mathbf{S} = 1$, and $\mathbf{y} = 0.5$, which makes $\mathbf{y}$ a plausible observation from F_r. We consider three values of $\mathbf{R}$: In the top row, $\mathbf{R} = \mathbf{P}_r^{\mathrm{f}}/4$, where $\mathbf{P}_r^{\mathrm{f}}$ is the variance of F_r. In this case the observation is accurate, which indicates that the likelihood function is highly unimodal (with a single high peak). As a result, the stochastic filter approximates the true density better because adding Gaussian perturbations to the bimodal ensemble will make the distribution more unimodal. In the middle row $\mathbf{R} = \mathbf{P}_r^{\mathrm{f}}$, where the accuracy is modest and it is hard to tell which filter gives better approximation to the truth. Finally, in the bottom row we have $\mathbf{R} = 4\mathbf{P}_r^{\mathrm{f}}$, a relatively inaccurate observation. Now when the two components are equally weighted (left column), the stochastic incorrectly populates the middle part because of the random perturbation while the deterministic retains the bimodal structure. In the right column, when the weights of two components are very unbalanced, the deterministic update is closer to the optimal for a wide range of $\mathbf{x}$ near the origin. However, it carries more outliers due to the small bump at $+7$, which might cause a larger bias in the higher moments.

2.2.2 The robustness perspective

Robustness (Hampel et al., 1986) is a natural notion of the stability of an inference method against small model violation. Intuitively, a "good" method should give stable outcomes when the true underlying distribution deviates slightly from the ideal dis-

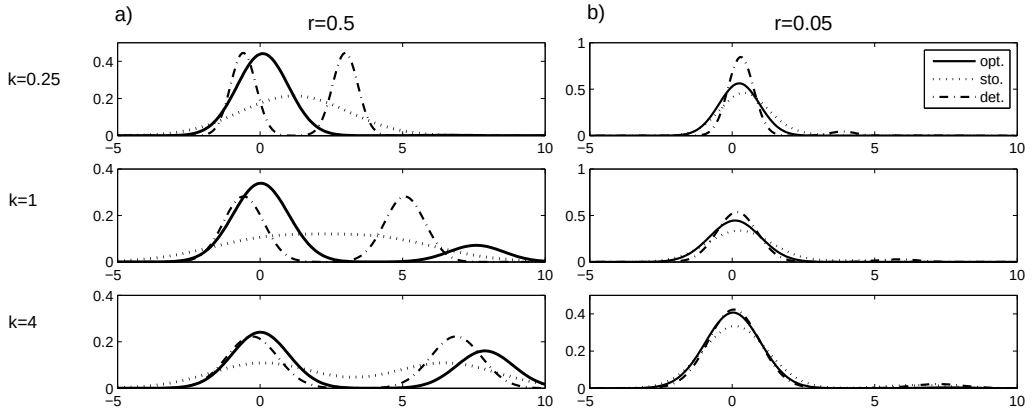

Figure 2.2: The density plots for $F_{\mathrm{o},r}$ (solid); $F_{\mathrm{s},r}$ (dotted) and $F_{\mathrm{d},r}$ (dash-dotted). Parameters: $t = 8$, $\mathbf{S} = 1$, $\mathbf{R} = k\mathbf{P}_r^{\mathrm{f}}$. $k = 0.25$ (top row); $k = 1$ (middle row); $k = 4$ (bottom row). $r = 0.5$ (left column); $r = 0.05$ (right column).

tribution. In the context of EnKF, the ideal distribution refers to the Gaussian forecast distribution under which the EnKF gives unbiased analysis. In parameter estimation, let $g(\hat{F}_n)$ be the estimator of parameter from the empirical distribution $\hat{F}_n$, and $g(F)$ denotes its population counterpart, which is usually the large-sample limit of $g(\hat{F}_n)$. Suppose the true distribution is $(1-r)F + rG$, a contaminated version of F, for some small $r > 0$. Then the estimator becomes $g((1-r)F + rG)$. The robustness of g at F means that no matter what G looks like, $g((1-r)F + rG)$ should be close to $g(F)$ as long as r is small. The quantification of this idea leads to the *Gâteaux derivative* and the *influence function*.

The Gâteaux derivative and the influence function Following the above notation, the estimator can be viewed as a function of r, the amount of contamination. The *Gâteaux derivative* of g at F in the direction of G is defined by

$$\nu(G, F; g) = \lim_{r \to 0+} \frac{g((1-r)F + rG) - g(F)}{r}. \tag{2.6}$$

Intuitively, the Gâteaux derivative measures approximately how g is affected by an infinitesimal contamination of shape G on F.

If $G = \delta_t$ is a point mass at t, then one can define

$$\mathrm{IF}(t; F, g) = \nu(\delta_t, F; g),$$

which is the *influence function* of g at F. There is a close analogy between the influence function and Green's function. In both cases, the general solution to a linear problem is a superposition of the solution to point mass problems. It can be shown that, under appropriate conditions, (see Bickel and Doksum, ch. 7.3),

$$\nu(G, F; g) = \int \mathrm{IF}(t; F, g) dG(t). \tag{2.7}$$

As a result, the function $\mathrm{IF}(\cdot; F, g)$ reflects the robustness of g at F. An important criterion in designing robust estimators is a bounded influence function:

$$\sup_t |\mathrm{IF}(t; F, g)| < \infty.$$

Intuitively, this means that distorting any small proportion of the data can not have a big impact on the outcome.

2.2.3 Comparison from the robustness perspective: analytical results

In our study, the parameter, and hence the estimator, is a distribution. For any fixed $\mathbf{x}$, $\mathbf{y}$, the Gâteaux derivatives of the conditional densities at $\mathbf{x}$ are[5], under Model (2.5),

$$\nu(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y})) = \lim_{r \to 0+} \frac{f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y}) - f_{\mathrm{s},0}(\mathbf{x}|\mathbf{y})}{r} = \frac{\partial}{\partial r} f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y}) \Big|_{r=0} \tag{2.8}$$

[5]In this chapter we use $f(\cdot) = F'(\cdot)$ as the density function of $F(\cdot)$, whenever possible. E.g., $f_{\mathrm{s},r}(\cdot|\mathbf{y})$ is the density function of $F_{\mathrm{s},r}(\cdot|\mathbf{y})$. For succinctness, we will use $f_{\mathrm{s},r}$ instead of $f_{\mathrm{s},r}(\cdot|\mathbf{y})$ without confusion.

for the stochastic filter, and

$$\nu(G,F;f_{\mathrm{d}}(\mathbf{x}|\mathbf{y}))=\lim_{r\to 0+}\frac{f_{\mathrm{d},r}(\mathbf{x}|\mathbf{y})-f_{\mathrm{d},0}(\mathbf{x}|\mathbf{y})}{r}=\frac{\partial}{\partial r}f_{\mathrm{d},r}(\mathbf{x}|\mathbf{y})\bigg|_{r=0} \tag{2.9}$$

for the deterministic filter. In our contaminated Gaussian model, the ideal distribution is $F=N(0,\mathbf{I})$ and $G=N(t,\mathbf{S})$ is the contamination distribution. Recall again that $f_{\mathrm{s},0}=f_{\mathrm{d},0}=f_{\mathrm{o},0}$, then equations (2.8) and (2.9) are comparing $f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})$ and $f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})$ with $f_{\mathrm{o},0}(\mathbf{x}|\mathbf{y})$ respectively.

However, the quantities in (2.8) and (2.9) involve not only $\mathbf{x}$ but also $\mathbf{y}$, the random observation. In order to take all $\mathbf{x}$ as well as the randomness of $\mathbf{y}$ into account, we integrate the square of the Gâteaux derivatives and take expectation over $\mathbf{y}$ under its marginal distribution when $r=0$, which is $N(0,\mathbf{I}+\mathbf{R})$. Finally, the quantities indicating the robustness of the EnKFs are

$$E_{\mathbf{y}}\left(\int\nu^2(G,F;f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}\right)=E_{\mathbf{y}}\left[\int\left(\frac{\partial}{\partial r}f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})\bigg|_{r=0}\right)^2 d\mathbf{x}\right] \tag{2.10}$$

for the stochastic filter, and

$$E_{\mathbf{y}}\left(\int\nu^2(G,F;f_{\mathrm{d}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}\right)=E_{\mathbf{y}}\left[\int\left(\frac{\partial}{\partial r}f_{\mathrm{d},r}(\mathbf{x}|\mathbf{y})\bigg|_{r=0}\right)^2 d\mathbf{x}\right] \tag{2.11}$$

for the deterministic filter.

On the other hand, note that

$$\frac{\partial}{\partial r}\left[\int(f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})-f_{\mathrm{s},0}(\mathbf{x}|\mathbf{y}))^2 d\mathbf{x}\right]=2\int(f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})-f_{\mathrm{s},0}(\mathbf{x}|\mathbf{y}))\frac{\partial}{\partial r}f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})d\mathbf{x},$$

and

$$\frac{\partial^2}{\partial r^2}\left[\int(f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})-f_{\mathrm{s},0}(\mathbf{x}|\mathbf{y}))^2 d\mathbf{x}\right]$$

$$= 2\int \left(\frac{\partial}{\partial r} f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})\right)^2 d\mathbf{x} + 2\int (f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y}) - f_{\mathrm{s},0}(\mathbf{x}|\mathbf{y}))\frac{\partial^2}{\partial r^2} f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})d\mathbf{x}.$$

Evaluate the above derivatives at $r = 0$, we have

$$\frac{\partial}{\partial r}\left[E_{\mathbf{y}} \int (f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y}) - f_{\mathrm{s},0}(\mathbf{x}|\mathbf{y}))^2 \, d\mathbf{x}\right]\Big|_{r=0} = 0.$$

and

$$\frac{\partial^2}{\partial r^2}\left[\int (f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y}) - f_{\mathrm{s},0}(\mathbf{x}|\mathbf{y}))^2 d\mathbf{x}\right]\Big|_{r=0} = 2\int \left(\frac{\partial}{\partial r} f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})\Big|_{r=0}\right)^2 d\mathbf{x}.$$

Taking expectation over $\mathbf{y}$,

$$E_{\mathbf{y}}\left[\int \left(\frac{\partial}{\partial r} f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y})\Big|_{r=0}\right)^2 d\mathbf{x}\right] = \frac{1}{2}\,\frac{\partial^2}{\partial r^2}\left[E_{\mathbf{y}} \int (f_{\mathrm{s},r}(\mathbf{x}|\mathbf{y}) - f_{\mathrm{s},0}(\mathbf{x}|\mathbf{y}))^2 \, d\mathbf{x}\right]\Big|_{r=0},$$

As a result, the quantity defined in (2.10) has a straightforward interpretation: It is the second derivative of the expected square of L_2 distance between $f_{\mathrm{s},r}$ and $f_{\mathrm{s},0}$. The same argument also holds for the deterministic filter. So a smaller value in (2.10) (or (2.11)) indicates a slower change in the updated distribution when r changes from zero to non-zero.

Our main theoretical results are summarized in the following theorems:

Theorem 2.2.1. *In model (2.5), we have*

(i) For all $\mathbf{R}, \mathbf{S}$

$$\lim_{||t||_2 \to \infty} E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x} = \infty, \tag{2.12}$$

and

$$0 < \lim_{||t||_2 \to \infty} \frac{E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}}{E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{d}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}} < 1; \tag{2.13}$$

(ii) For all $\mathbf{R}, t$*,*

$$\lim_{||\mathbf{S}||_2 \to \infty} E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x} = \infty, \tag{2.14}$$

and

$$0 < \lim_{||\mathbf{S}||_2 \to \infty} \frac{E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}}{E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{d}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}} < 1; \tag{2.15}$$

(iii) For all $t, \mathbf{S}$*,*

$$\lim_{||\mathbf{R}||_2 \to 0} E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x} = \infty, \tag{2.16}$$

and

$$\lim_{||\mathbf{R}||_2 \to 0} \frac{E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}}{E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{d}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}} = 0. \tag{2.17}$$

Proof. The proof is included in Section 2.6. □

Parts (i) and (ii) of Theorem 2.2.1 indicate that neither of the two filters has bounded Gâteaux derivative over all possible contaminations. However, when the contamination is wild, the stochastic filter is more stable than the deterministic filter. Loosely speaking, when there are outliers in the forecast ensemble, the Kalman filter will suffer from its non-robustness due to the use of the sample mean and sample covariance matrix. The deterministic filter is affected more because its rigid shifting and re-scaling (in order to make the exact covariance) leaves no chance to correct the outliers, while the stochastic filter uses a "softer" method to adjust the ensemble mean and covariance by using random perturbations. It is thus more resilient to outliers because there is some chance that the outliers are partially corrected by the random perturbations. This effect can also be seen

in the top right plot of Figure 2.2. Moreover, it also implies that, in the multivariate case, when the contamination is wild, the deviation in the updated density is largely determined by the magnitude, not the orientation, of t and/or $\mathbf{S}$. As shown later in Section 2.3, the asymptotic result also holds even for moderately large choices of $||t||_2$ and $||\mathbf{S}||_2$.

Part (iii) indicates that stochastic filter is more stable when the observation is accurate. This result nicely supports the intuitive argument in Lawson & Hansen (2004): the convolution with a Gaussian random perturbation in the stochastic filter makes the updated ensemble closer to Gaussian while the deterministic might push the edge-members in the ensemble to be far-outliers and have the major component in the mixture overly tight.

The case that $||\mathbf{R}||_2 \to \infty$ is particularly interesting. Intuitively, a very large $||\mathbf{R}||_2$ indicates a very non-informative observation. Thus the conditional distribution should be close to the forecast distribution. As a result, one should expect little change on the forecast ensemble when $||\mathbf{R}||_2$ is large. This intuition suggests choosing the orthogonal matrix $\mathbf{U} = \mathbf{I}$ in the deterministic filter, the benefit of which can be seen through Theorem 2.2.2:

Theorem 2.2.2. *If in (2.3) we choose* $\mathbf{U} = \mathbf{I}$*, then for all* $t, \mathbf{S}$*,*

$$0 < \lim_{||\mathbf{R}||_2 \to \infty} E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x} < \infty, \tag{2.18}$$

and

$$\lim_{||\mathbf{R}||_2 \to \infty} \frac{E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}}{E_{\mathbf{y}} \int \nu^2(G, F; f_{\mathrm{d}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}} = 1. \tag{2.19}$$

Otherwise, we have

$$\lim_{||t||_2\to\infty}\lim_{||\mathbf{R}||_2\to\infty}\frac{E_{\mathbf{y}}\int\nu^2(G,F;f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}}{E_{\mathbf{y}}\int\nu^2(G,F;f_{\mathrm{d}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}}=0, \tag{2.20}$$

and

$$\lim_{||\mathbf{S}||_2\to\infty}\lim_{||\mathbf{R}||_2\to\infty}\frac{E_{\mathbf{y}}\int\nu^2(G,F;f_{\mathrm{s}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}}{E_{\mathbf{y}}\int\nu^2(G,F;f_{\mathrm{d}}(\mathbf{x}|\mathbf{y}))d\mathbf{x}}=0. \tag{2.21}$$

Proof. See Section 2.6. □

Theorem 2.2.2 is easy to understand. Intuitively, when $\mathbf{R}$ is large, we have $\mu^{\mathrm{a}} \approx \mu^{\mathrm{f}}$ and $\mathbf{P}^{\mathrm{a}} \approx \mathbf{P}^{\mathrm{f}}$ in the Kalman filter. Here $\mathbf{U} = \mathbf{I}$ implies $\mathbf{A} \approx \mathbf{I}$, which means making little change on the forecast ensemble. In Section 2.3 we will see that the choice of $\mathbf{U} = \mathbf{I}$ does beat other choices even for moderately large $\mathbf{R}$, $\mathbf{S}$ and t. The issue of choosing the orthogonal matrix in the square root filter has been discussed in Sakov & Oke (2007), which mainly focuses on the right-multiplication case. Theorem 2.2.2 suggests a stable choice of the left-multiplying orthogonal matrix which means the corresponding right-multiplying orthogonal matrix is stable due to the correspondence between the left and right-multiplication in unbiased square root filters (Livings et al., 2008) if $p < n$.

Remark 2.2.3. Theorems 2.2.1 and 2.2.2 concern the effects caused by a large t, $\mathbf{S}$ and $\mathbf{R}$ separately, by means of sending one quantity to infinity while keeping others fixed. In fact, these quantities do interact in the optimal and EnKF updates, which will affect the comparison in a much more complicated manner. Although in this more interesting case analytical results seem hard to derive, we do think these theorems provide some qualitative view of the comparison as we will see in the numerical experiments.

2.2.4 Connection to bias comparison

The robustness tells us about the stability of the filters when the data distribution is nearly ideal. However, as mentioned earlier, a more direct comparison would be to just look at the bias, that is, the difference between the limiting distribution of the updated ensemble ($f_{\mathrm{s},r}$ and $f_{\mathrm{d},r}$), and the optimal conditional distribution ($f_{\mathrm{o},r}$). A first observation is that when r is small, then $f_{\mathrm{o},r} \approx f_{\mathrm{o},0}$, i.e., $f_{\mathrm{o},r}$ would mostly be as if there is no contamination at all, as long as $\mathbf{y}$ is not too far from 0 or not too close to t, which is often the case when $||t||_2 \gg 0$ and $\mathbf{y}^{\mathrm{o}}$ is randomly drawn from f_r. This can be seen from the fact that

$$F_{\mathrm{o},r} = (1-\pi(r))N(\mu^{\mathrm{a}}_{\mathrm{o},1}, \mathbf{P}^{\mathrm{a}}_{\mathrm{o},1}) + \pi(r)N(\mu^{\mathrm{a}}_{\mathrm{o},2}, \mathbf{P}^{\mathrm{a}}_{\mathrm{o},2}), \tag{2.22}$$

where, letting $\phi(\mathbf{x}; \mu, \mathbf{P})$ be the density of $N(\mu, \mathbf{P})$ at $\mathbf{x}$,

$$\pi(r) = \frac{r\phi(\mathbf{y}; t, \mathbf{S}+\mathbf{R})}{r\phi(\mathbf{y}; t, \mathbf{S}+\mathbf{R}) + (1-r)\phi(\mathbf{y}; 0, \mathbf{I}+\mathbf{R})},$$

and, for $j = 1, 2$, with the convention that $\mu^{\mathrm{f}}_1 = 0$, $\mathbf{P}^{\mathrm{f}}_1 = \mathbf{P}^{\mathrm{f}}$, $\mu^{\mathrm{f}}_2 = t$, and $\mathbf{P}^{\mathrm{f}}_2 = \mathbf{S}$,

$$\mathbf{K}_j = \mathbf{P}^{\mathrm{f}}_j(\mathbf{P}^{\mathrm{f}}_j + \mathbf{R})^{-1}, \quad \mu^{\mathrm{a}}_j = (\mathbf{I}-\mathbf{K}_j)\mu^{\mathrm{f}}_j + \mathbf{K}_j\mathbf{y}, \quad \mathbf{P}^{\mathrm{a}}_j = (\mathbf{I}-\mathbf{K}_j)\mathbf{P}^{\mathrm{f}}_j.$$

For the proof of (2.22), we refer the reader to Bengtsson et al. (2003) and references therein. As a result, when $||t||_2 \gg 0$, and $\mathbf{y}$ not far from 0, we have $\pi(r)/r \approx 0$.

As a result, we have $f_{\mathrm{o},r} \approx f_{\mathrm{o},0}$, for large t. Note further that $f_{\mathrm{s},0} = f_{\mathrm{d},0} = f_{\mathrm{o},0}$, which means that $f_{\mathrm{s},r} - f_{\mathrm{o},r} \approx f_{\mathrm{s},r} - f_{\mathrm{o},0} = f_{\mathrm{s},r} - f_{\mathrm{s},0}$. That is, robustness actually indicates small bias. In Section 2.3 we present simulation results to verify this idea.

A limitation of our analysis to this point is that the L_2 distance provides only partial information about the deviation of the analysis distribution from the optimal (Bayes)

update. In fact, data assimilation is best evaluated by 1) the distance between the analysis center and the true posterior center and 2) the size of the analysis covariance which needs to be large enough to have the analysis ensemble cover a substantial proportion of the true posterior distribution including its center. These two criteria are labeled in Sacher & Bartello (2009) as "accuracy" and "stability" respectively (recall that in this chapter the notion of "stability" is different). In the context of large ensemble behavior, the analysis center is almost the same for the stochastic filter and the deterministic filter. Therefore they should perform similarly in this aspect given they are starting from the same forecast ensemble. On the other hand, although both filters have the same second order statistics, the updated ensemble is distributed differently for a non-Gaussian prior. This difference will affect the future forecast ensemble and hence the filter performance in sequential applications, which needs to be explored further.

Another class of criteria are higher order moments since in a non-Gaussian distribution the higher moments contain much information about the distribution. In the next subsection we consider the third moment as another measure of performance to support our previous results.

2.2.5 The third moment

The third moment is an indication of the skewness of the distribution. Therefore it seems a natural criterion beyond the first two moments to evaluate the updated ensemble. Lawson & Hansen (2004) also considered the ensemble skewness in their experiments. Here for presentation simplicity we consider the one dimensional model given by (2.5).

Assuming model (2.5), let $M_{\mathrm{s}}(\mathbf{y}) = \int \mathbf{x}^3 f_{\mathrm{s}}(\mathbf{x}|\mathbf{y})d\mathbf{x}$ be the third moment of the limiting updated distribution given by the stochastic filter and similarly define $M_{\mathrm{d}}(\mathbf{y})$ for the deterministic filter. Then we have the following theorem:

Theorem 2.2.4. *Under model (2.5), if both $X_{\mathrm{f}} \in \mathbb{R}^1$ and $Y \in \mathbb{R}^1$, then*

(i) For all $\mathbf{S}$ *and* $\mathbf{y}$

$$\lim_{|t|\to\infty} |\nu(G, F; M_{\mathrm{s}}(\mathbf{y}))| = \infty, \quad \textit{and} \lim_{|t|\to\infty} \frac{|\nu(G, F; M_{\mathrm{s}}(\mathbf{y}))|}{|\nu(G, F; M_{\mathrm{d}}(\mathbf{y}))|} < 1. \tag{2.23}$$

(ii) For all t *and* $\mathbf{y}$,

$$\lim_{|\mathbf{S}|\to\infty} |\nu(G, F; M_{\mathrm{s}}(\mathbf{y}))| = \infty, \quad \textit{and} \lim_{|\mathbf{S}|\to\infty} \frac{|\nu(G, F; M_{\mathrm{s}}(\mathbf{y}))|}{|\nu(G, F; M_{\mathrm{d}}(\mathbf{y}))|} < 1. \tag{2.24}$$

Proof. See Section 2.6. □

These results are similar to those in the previous theorems, except that the third moment is a scalar which allows us to derive results for each value of $\mathbf{y}$. The intuition behind Theorem 2.2.4 can be seen from Figure 2.2, where the deterministic filter tends to produce two components which are less spread and further away from each other than in the stochastic filter. As a result, the deterministic filter puts a little more density in the region which are likely outliers (the bump near =7 on the bottom right plot). Despite maintaining the right mean and covariance, these outliers will have a substantial impact on the higher moments as shown in Theorem 2.2.4. The empirical comparison of the bias of the third moments is provided in Section 2.3.

2.3 Simulation results

In this section we present simulation results comparing the performance of the two versions of ensemble Kalman filters. As we will see later, the simulations do support the analytical results and intuitive discussion in Section 2.2.3 and 2.2.4.

2.3.1 The 1-dimensional case

In the 1-dimensional case, n random samples are drawn from $F_r = (1-r)F + rG$ as described in model (2.5), under different combinations of model parameters $(r, \mathbf{R}, t, \mathbf{S})$ as defined in Section 2.2.1. Both versions of EnKF are applied to the same random sample and observation from which the optimal conditional distribution is calculated. We first check the expected square of L_2 distance as a measure of bias as a direct verification of Theorem 2.2.1 and 2.2.2, then we look at the third moment to further confirm our results.

The expected square of L_2 distance Once all the parameters in Model (2.5) are specified, for any value of $\mathbf{y}$, the functions $f_{\mathrm{s},r}(\mathbf{x})$, $f_{\mathrm{d},r}(\mathbf{x})$ and $f_{\mathrm{o},r}(\mathbf{x})$ can be calculated analytically. The expected square of L_2 distances

$$E_{\mathbf{y}} \int \left(f_{\mathrm{s},r}(\mathbf{x}) - f_{\mathrm{o},r}(\mathbf{x})\right)^2 d\mathbf{x}, \quad \text{and} \quad E_{\mathbf{y}} \int \left(f_{\mathrm{d},r}(\mathbf{x}) - f_{\mathrm{o},r}(\mathbf{x})\right)^2 d\mathbf{x} \tag{2.25}$$

are calculated numerically. That is, $\mathbf{y}$ is simulated many times, and for each simulated value of $\mathbf{y}$ the above integrals are calculated numerically and averaged. In Table 2.1, we set $t = 8$, $\mathbf{S} = 1$, the same setup as in Figure 2.2. Actually the simulation is quantifying the difference between the density curves shown in Figure 2.2, except that it takes further expectation over all possible values of $\mathbf{y}$. Three different values of $\mathbf{R}$ are chosen according

Table 2.1: Mean square L_2 distance to the true conditional distribution in 1-D, with $t = 8$, $\mathbf{S} = 1$, averaged over 1000 realizations of $\mathbf{y}$. Numbers in parentheses indicate standard deviations for each result. Recall that the contamination distribution $G = N(t, \mathbf{R})$

		$\mathbf{R} = 0.25\mathbf{P}_r^{\mathrm{f}}$	$\mathbf{R} = \mathbf{P}_r^{\mathrm{f}}$	$\mathbf{R} = 4\mathbf{P}_r^{\mathrm{f}}$
r=0.05	Sto.	0.369(0.195)	0.435(0.105)	0.112(0.037)
	Det.	0.409(0.405)	0.586(0.137)	0.150(0.051)
r=0.1	Sto.	0.255(0.112)	0.286(0.094)	0.099(0.029)
	Det.	0.356(0.350)	0.464(0.161)	0.150(0.054)
r=0.5	Sto.	0.117(0.034)	0.124(0.006)	0.055(0.005)
	Det.	0.240(0.156)	0.199(0.064)	0.050(0.018)

to its relative size with $\mathbf{P}_r^{\mathrm{f}} = \mathrm{var}(\mathbf{x}|F_r)$. This result supports the analysis in Section 2.2.3 and the intuition in Section 2.2.4: when r is small, $f_{\mathrm{s},r}$ is closer to $f_{\mathrm{o},r}$. Moreover, it seems that the asymptotic statement can be extended to much larger value of r, e.g., $r = 0.5$ as shown in Table 2.1. The expectation over $\mathbf{y}$ is approximated by averaging over 1000 simulated values of $\mathbf{y}$ (standard deviations are shown in the parentheses).

The third moment The EnKF forces the updated ensemble to have the correct first and second moment, therefore the third moment becomes a natural criterion of comparison. The empirical third moments of the two updated ensembles are compared with the optimal third moment which is calculated analytically.

Here, instead of taking expectation over $\mathbf{y}$, we investigate the impact of the value $\mathbf{y}$ on the comparison. That is, we look at all $\mathbf{y} \in \left[-2(1+\mathbf{R})^{\frac{1}{2}}, 2(1+\mathbf{R})^{\frac{1}{2}}\right]$, which covers a majority of probability mass in F_r. In the experiment, $(\mathbf{R}, \mathbf{S}) \in \{1/4, 1, 4\}^2$, and $t \in \{1, 10, 30, 50, 100\}$. We choose $(n, r) = (500, 0.05)$. Several representative pictures are displayed in Figure 2.3. We see that for small t, both filters give very small bias for almost the whole range of $\mathbf{y}$, and when t gets bigger, the stochastic filter gives smaller bias

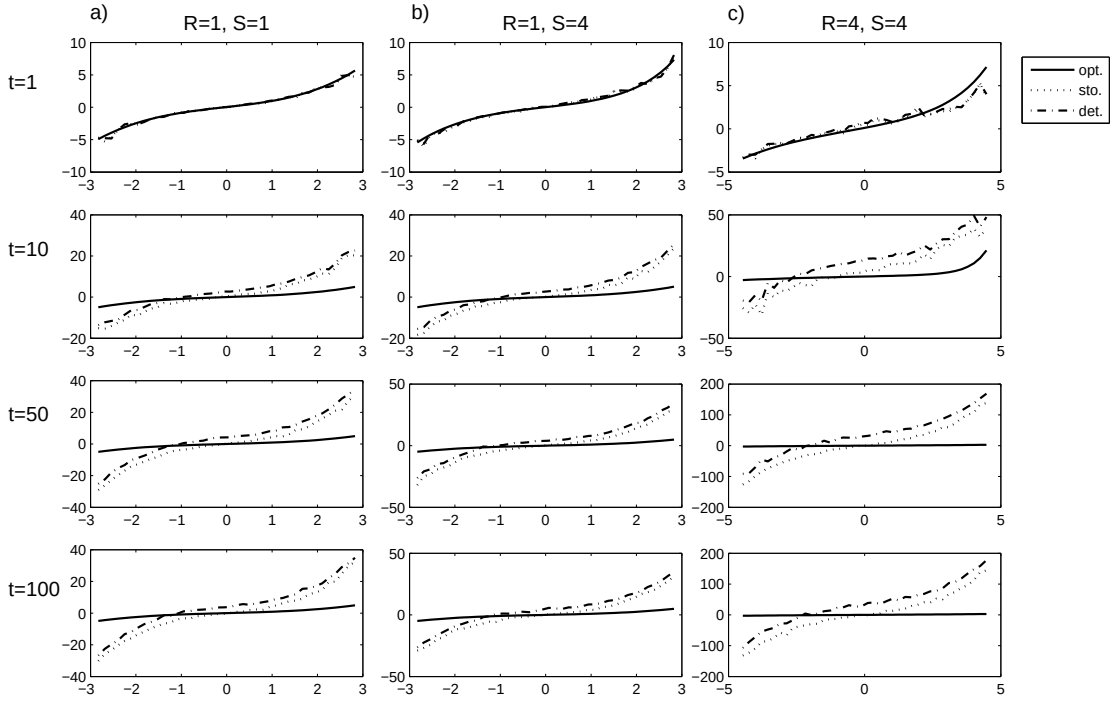

Figure 2.3: The conditional third moments. Horizontal coordinate: the observation $\mathbf{y}$; vertical coordinate: $E_{F_{\mathrm{o},r}}\mathbf{x}^3$ (solid), $E_{\hat{F}_{\mathrm{s},r}}\mathbf{x}^3$ (dotted) and $E_{\hat{F}_{\mathrm{d},r}}\mathbf{x}^3$ (dash-dotted). Parameters: $t = 1$ (top row), $t = 10$ (second row), $t = 50$ (third row), $t = 100$ (bottom row); $\mathbf{R} = \mathbf{S} = 1$ (left column), $\mathbf{R} = 1, \mathbf{S} = 4$ (middle column), $\mathbf{R} = \mathbf{S} = 4$ (right column).

for a wide range of $\mathbf{y}$, which covers the majority of probability mass of its distribution.

Moreover, the difference is enhanced by larger values of $\mathbf{R}$ and $\mathbf{S}$.

2.3.2 2-dimensional case

In the 2-dimensional case, our theory claims that it is the magnitude (i.e., the smallest eigenvalue) of the covariance matrices that determine the amount of deviation. However, in the finite sample simulation, it seems necessary to consider not only the magnitude, but also different orientations of the matrices. We consider two instances:

- *Orientation 1*: $\mathbf{P} = \mathbf{I}_2$, $\mathbf{R} = c_1\mathbf{R}_0$ and $\mathbf{S} = c_2\mathbf{S}_0$, where $(c_1, c_2) \in \{1/4, 1, 4\}^2$ tunes the magnitude of $\mathbf{R}$ and $\mathbf{S}$, where $\mathbf{R}_0 = \mathrm{diag}(1.5, 1)$, and $\mathbf{S}_0$ is a simulated 2 by 2 Wishart matrix:
$$\mathbf{S}_0 = \begin{pmatrix} 1.15 & 0.14 \\ 0.14 & 0.70 \end{pmatrix}.$$

- *Orientation 2*: In this case we consider a contamination distribution G with very different shape from F, i.e., $\mathbf{S}_0$ that has very different orientation from $\mathbf{P} = \mathrm{cov}_F(\mathbf{x})$. Here we choose $\mathbf{P}$ to be, up to a scaling constant, the covariance matrix of the stationary distribution of the first two coordinates in the Lorenz 63 system, and $\mathbf{S}_0$ is obtained by switching the eigenvalues of P.
$$\mathbf{P} = \begin{pmatrix} 1.06 & 1.05 \\ 1.05 & 1.35 \end{pmatrix}, \quad \mathbf{S}_0 = \begin{pmatrix} 1.35 & -1.05 \\ -1.05 & 1.06 \end{pmatrix}.$$
Here $\mathbf{S}_0$ has the same eigenvectors as Σ, but with the eigenvalues switched. That is,
$$\mathbf{P} = \mathbf{Q}\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}\mathbf{Q}^{\mathrm{T}}, \quad \mathbf{S}_0 = \mathbf{Q}\begin{pmatrix} d_2 & 0 \\ 0 & d_1 \end{pmatrix}\mathbf{Q}^{\mathrm{T}},$$
where $\mathbf{Q}$ is a orthogonal matrix and $d_1 = 0.15$, $d_2 = 2.27$ are eigenvalues of $\mathbf{P}$ and $\mathbf{S}_0$. The other settings are the same as above expect that $\mathbf{R}_0 = \mathbf{I}_2$.

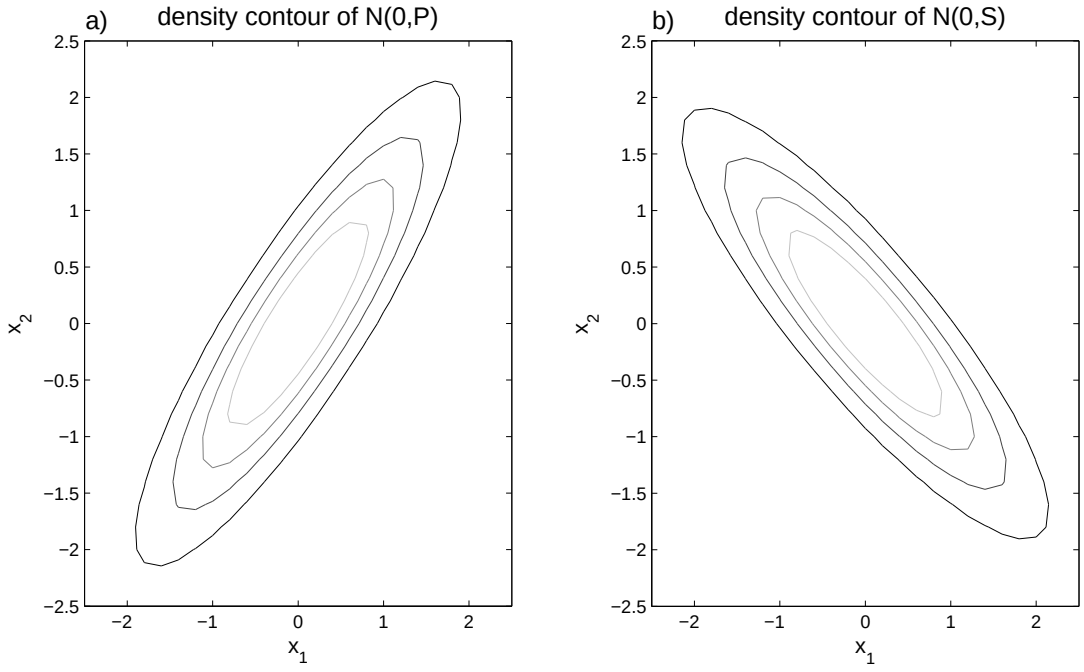

Figure 2.4: The contour of the densities of the two components in Orientation 2 (up to shift). Left: $N(0, \mathbf{P})$; right: $N(0, \mathbf{S})$. The levels are (from inner to outer): 0.2, 0.15, 0.1, 0.05.

The contour of the two Gaussian densities are plotted in Figure 2.4.

In the deterministic algorithm we try two choices of $\mathbf{U}$ in (2.3). The first is simply to choose $\mathbf{U} = \mathbf{I}$. The second choice is based on the "ensemble adjustment Kalman filter" (EAKF) proposed by Anderson (2001). Similar to the 1-dimensional case, the expectation over $\mathbf{y}$ is approximated by averaging over 120 simulated $\mathbf{y}$. Standard deviations are shown in the parentheses. Some representative results are summarized in Table 2.2, where $r = 0.05$, $c_2 = 1$, and $t = (10, 10)$ (other values make no qualitative difference).

Recall that c_1 indicates the size of $\mathbf{R}$. We can see that for small c_1, the stochastic filter is remarkably less biased, agreeing with the experiments in Lawson & Hansen (2004).

Also note that in Model (5) both the forecast and the analysis distribution are a mixture of two Gaussian components, where the major component (i.e., the one with a weight close to 1) contains mostly "normal" ensemble members whereas the minor component (the one whose weight is close to 0) contains mostly ensemble members that are likely outliers. When the observation is accurate, the optimal filter puts more weight on the major Gaussian component. On the other hand neither of the two EnKFs adjusts the component weights in the analysis. The two components in the analysis distribution given by the deterministic filter are less spread than those given by the stochastic filter. In order to have the same covariance, the less spread components have to be moved further away from each other. As a result, the outliers tends to be even more outlying in the deterministic update. An instance of this intuition can be seen in the right panel of Figure 2 where the deterministic filter always produces a small bump in the right tail, especially for small observation errors.

Another interesting observation is the comparison of the choices of the rotation matrix $\mathbf{U}$. For small observation noise, the difference is negligible. One can imagine that when the observation is accurate, the optimal analysis distribution tends to be closer to a Gaussian, whose distribution is determined by the first two moments, therefore the rotation does not make too much difference. While when c_1 gets bigger, the analysis ensemble becomes much less Gaussian and the choice $\mathbf{U} = \mathbf{I}$ shows significant advantage as compared with the EAKF, agreeing with Theorem 3. This basically says that when the observation is very uninformative, there is no need to change, and hence no need to rotate, the ensemble.

Table 2.2: Mean square L_2 distance to the true conditional distribution in 2-D, with $t = (10, 10)$, $r = 0.05$, $c_2 = 1$, averaged over 120 realizations of $\mathbf{y}$. Numbers in parentheses indicate standard deviations for each result.

		$c_1 = 1/4$	$c_1 = 1$	$c_1 = 4$	$c_1 = 16$
Orient. 1	Sto.	.035(.040)	.041(.039)	.043(.031)	.040(.027)
	Det. ($\mathbf{U} = \mathbf{I}$)	.462(.114)	.183(.093)	.100(.071)	.065(.055)
	Det. (EAKF)	.454(.111)	.183(.094)	.105(.075)	.086(.058)
Orient. 2	Sto.	.066(.142)	.047(.071)	.049(.056)	.050(.051)
	Det. ($\mathbf{U} = \mathbf{I}$)	.492(.224)	.204(.119)	.114(.098)	.077(.079)
	Det. (EAKF)	.500(.207)	.208(.118)	.128(.101)	.103(.085)

Moreover, the results shown in Table 2.2 also confirm the theory in that only the magnitude of the contamination matters since similar behavior is observed for two very different shapes of contamination distribution.

2.4 Conclusion

We have studied the large-ensemble performance of ensemble Kalman filters using the robustness approach. In the contaminated Gaussian model, the updated distribution is another mixture with two components, where the stochastic filter is more stable against small model violation due to the fact that its main component in the updated distribution is closer to that of the optimal filter. Our theoretical results are supported by intensive simulation over a wide range of the model parameters, agreeing with the empirical findings in Lawson & Hansen (2004), where the intuitive argument says that deterministic shifting and re-scaling exaggerates the dispersion of some ensemble members.

Although our study focuses on large-ensemble behavior under a classical model, our method can be extended in at least two directions. First, the influence function theory

enables one to study other shapes of contamination, rather than Gaussian. Second, in geophysical studies the model deviation might come from the observation, instead of the state variable. In other words, the modeling error could come from the mis-specification of the distribution of the observation error. The approach developed in this chapter is applicable to analysis of situations where the observation error is not exactly Gaussian.

The choice of the orthogonal matrix $\mathbf{U}$ in the deterministic filter is an unsettled issue in data assimilation literature. Our L_2-based stability criterion gives an answer to this question which is intuitively reasonable: you do almost nothing when the observation is uninformative.

In practice, there are many factors determining which filtering method to use, such as the computational constraints, the modeling error, the particular prediction task, and the specific shapes of the forecasting distribution and error distribution, etc. But this cannot be done before we fully understand the properties of all the candidates. We hope our study contributes to that understanding.

2.5 Large-ensemble behavior of EnKFs

Following the discussion in Section 2.1.3, we have:

Proposition 2.5.1. *As $n \to \infty$, we have*

$$\hat{F}_{\mathrm{s}} \Rightarrow F_{\mathrm{s}}, \quad \hat{F}_{\mathrm{d}} \Rightarrow F_{\mathrm{d}},$$

where F_{s} and F_{d} are the distribution functions of $(\mathbf{I}-\mathbf{KH})\mathbf{x}^{\mathrm{f}}+\mathbf{K}(\mathbf{y}+\epsilon)$ and $\mu^{\mathrm{a}}+\mathbf{A}(\mathbf{x}^{\mathrm{f}}-\mu^{\mathrm{f}})$, respectively.

Our theoretical result on comparing the stochastic and deterministic filters are based on F_{s} and F_{d}.

Proof. We show the weak convergence of $\hat{F}_{\mathrm{s}}$. The proof for $\hat{F}_{\mathrm{d}}$ is similar.

Let J be a random index uniformly drawn from $\{1, \ldots, n\}$. Let $\hat{Z}_n = (\mathbf{I} - \hat{\mathbf{K}}\mathbf{H})\mathbf{x}^{\mathrm{f}(J)} + \hat{\mathbf{K}}(y + \epsilon^{(J)})$ and $Z_n = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{x}^{\mathrm{f}(J)} + \mathbf{K}(\mathbf{y} + \epsilon^{(J)})$. Then $\hat{Z}_n \sim \hat{F}_{\mathrm{s}}$, and $Z_n \sim F_{\mathrm{s}}$, so it is enough to show that $\hat{Z}_n - Z_n \xrightarrow{P} 0$.

Consider the random variable $W = \mathbf{H}\mathbf{x}^{\mathrm{f}} - \mathbf{y} - \epsilon$. For any $\xi > 0$, $\delta > 0$, one can find an M large enough such that $P(||W||_2 \geq M\xi) \leq \delta/2$. On the other hand, since $\hat{\mathbf{K}} - \mathbf{K} \xrightarrow{P} 0$, one can find $N_{\xi,\delta}$ such that $P(||\hat{\mathbf{K}} - \mathbf{K}||_2 \geq 1/M) \leq \delta/2$ whenever $n \geq N_{\xi,\delta}$. Then for all $n \geq N_{\xi,\delta}$, we have

$$\begin{aligned}
P\left(||\hat{Z}_n - Z_n||_2 \geq \xi\right) &= P\left(||(\hat{\mathbf{K}} - \mathbf{K})(\mathbf{H}\mathbf{x}^{\mathrm{f}(J)} - \mathbf{y} - \epsilon^{(J)})|| \geq \xi\right) \\
&\leq P\left(||\hat{\mathbf{K}} - \mathbf{K}||_2 \geq 1/M\right) + P\left(||\mathbf{H}\mathbf{x}^{\mathrm{f}(J)} - \mathbf{y} - \epsilon^{(J)}||_2 \geq M\xi\right) \\
&= P\left(||\hat{\mathbf{K}} - \mathbf{K}||_2 \geq 1/M\right) + P\left(||\mathbf{H}\mathbf{x} - \mathbf{y} - \epsilon||_2 \geq M\xi\right) \\
&\leq \delta/2 + \delta/2 \\
&= \delta.
\end{aligned}$$

□

Remark 2.5.2. In Proposition 2.5.1, there is nothing special about Gaussianity, so the result holds for any random variable $\mathbf{x}^{\mathrm{f}}$ such that $E\mathbf{x}^{\mathrm{f}} = \mu^{\mathrm{f}}, \mathrm{Var}(\mathbf{x}^{\mathrm{f}}) = \mathbf{P}^{\mathrm{f}}$.

2.6 Proofs of the main theorems

Proof of Theorem 2.2.1

We give a sketchy proof for part (i), the argument applies similarly to other parts.

We first consider the simpler case: $t = \rho t_0$, where $||t_0||_2 = 1$.

Letting $\mathbf{K} = (\mathbf{I}+\mathbf{R})^{-1}$, $\mathbf{B} = \mathbf{I}-\mathbf{K}$, $\Gamma = tt^{\mathrm{T}} + \mathbf{S} - \mathbf{I}$, $\mathbf{A} = \mathbf{A}(0) = \mathbf{B}^{\frac{1}{2}}\mathbf{U}$ for some orthogonal $\mathbf{U}$, and $V_{\mathrm{s}} = \mathbf{B}\Gamma\mathbf{B}^{\mathrm{T}} - \mathbf{A}\Gamma\mathbf{A}^{\mathrm{T}}$, then, in the deterministic filter, we have

$$\begin{aligned}
&\frac{\partial}{\partial r} f_{\mathrm{d},r}(\mathbf{x})\bigg|_{r=0} \\
&= \bigg[-\frac{1}{2}\mathrm{tr}\left(\mathbf{B}^{-1}V_{\mathrm{s}}\right) + (\Gamma\mathbf{K}\mathbf{y} + \mathbf{B}^{-1}(\mathbf{B}-\mathbf{A})t)^{\mathrm{T}}(\mathbf{x}-\mathbf{K}\mathbf{y}) \\
&\quad + \frac{1}{2}(\mathbf{x}-\mathbf{K}\mathbf{y})^{\mathrm{T}}\mathbf{B}^{-1}V_s\mathbf{B}^{-1}(\mathbf{x}-\mathbf{K}\mathbf{y}) - 1\bigg]\phi(\mathbf{x};\mathbf{K}\mathbf{y},\mathbf{B}) + \phi(\mathbf{x};\mathbf{K}\mathbf{y}+\mathbf{A}t,\mathbf{A}\mathbf{S}\mathbf{A}^{\mathrm{T}}).
\end{aligned}$$

Then it can be shown, via some algebra, that

$$E_{\mathbf{y}}\int\left(\frac{\partial}{\partial r} f_{\mathrm{d},r}(\mathbf{x})\bigg|_{r=0}\right)^2 d\mathbf{x} = C\cdot a_{\mathrm{d}}(t_0)\rho^4 + P_{\mathrm{d}}(\rho) + e^{-\kappa_{\mathrm{d}}\rho^2}Q_{\mathrm{d}}(\rho), \tag{2.26}$$

where $P_{\mathrm{d}}(\rho)$ and $Q_{\mathrm{d}}(\rho)$ are polynomials of degree 3; $C>0$ is a constant depending only on $\mathbf{B}$; $\kappa_{\mathrm{d}}>0$ is a constant; and

$$a_{\mathrm{d}}(t_0) = \frac{1}{2}\mathrm{tr}(t_0t_0^{\mathrm{T}}\mathbf{K}t_0t_0^{\mathrm{T}}\mathbf{B}) + \frac{1}{16}E\left(z^{\mathrm{T}}\left(\mathbf{B}^{\frac{1}{2}}t_0t_0^{\mathrm{T}}\mathbf{B}^{\frac{1}{2}} - \mathbf{U}t_0t_0^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\right)z\right)^2. \tag{2.27}$$

On the other hand, in the stochastic filter,

$$\begin{aligned}
&\frac{\partial}{\partial r} f_{\mathrm{s},r}(\mathbf{x})\bigg|_{r=0} \\
&= \left[(\Gamma\mathbf{K}\mathbf{y})^{\mathrm{T}}(\mathbf{x}-\mathbf{K}\mathbf{y}) - 1\right]\phi(\mathbf{x};\mathbf{K}\mathbf{y},\mathbf{B}) + \phi(\mathbf{x};\mathbf{K}\mathbf{y}+\mathbf{B}t,\mathbf{B}\mathbf{S}\mathbf{B}^{\mathrm{T}}+\mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}}).
\end{aligned}$$

Similarly,

$$E_{\mathbf{y}}\int\left(\frac{\partial}{\partial r} f_{\mathrm{s},r}(\mathbf{x})\bigg|_{r=0}\right)^2 d\mathbf{x} = C\cdot a_{\mathrm{s}}(t_0)\rho^4 + P_{\mathrm{s}}(\rho) + e^{-\kappa_{\mathrm{d}}\rho^2}Q_{\mathrm{s}}(\rho), \tag{2.28}$$

where $P_{\mathrm{s}}(\rho)$ and $Q_{\mathrm{s}}(\rho)$ are polynomials of degree 3; C is the same constant as in (2.26); $\kappa_{\mathrm{s}} > 0$ is a constant; and

$$a_{\mathrm{s}}(t_0) = \frac{1}{2}\mathrm{tr}(t_0 t_0^{\mathrm{T}} \mathbf{K} t_0 t_0^{\mathrm{T}} \mathbf{B}). \tag{2.29}$$

Note that $||\mathbf{B}^{\frac{1}{2}} t_0||_2 < ||\mathbf{U} t_0||_2$, for all $t_0 \neq 0$. Therefore,

$$\lim_{\rho\to\infty} E_{\mathbf{y}} \int \nu^2 \left(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y})\right) d\mathbf{x} = \infty, \quad \text{and} \quad \lim_{\rho\to\infty} \frac{E_{\mathbf{y}} \int \nu^2 \left(G, F; f_{\mathrm{s}}(\mathbf{x}|\mathbf{y})\right) d\mathbf{x}}{E_{\mathbf{y}} \int \nu^2 \left(G, F; f_{\mathrm{d}}(\mathbf{x}|\mathbf{y})\right) d\mathbf{x}} = \frac{a_{\mathrm{s}}(t_0)}{a_{\mathrm{d}}(t_0)} < 1.$$

The statement of Theorem 2.2.1 (i) follows easily via a standard argument using the compactness of the set $\{t_0 \in \mathbb{R}^p : ||t_0||_2 = 1\}$.

The proofs for part (ii) and (iii) are simply repeating the argument above on $\mathbf{S}$ and $\mathbf{R}$, respectively.

2.6.1 Proof of Theorem 2.2.2

The argument is essentially the same as in the proof of Theorem 2.2.1. Starting from the easy facts:

$$\lim_{||\mathbf{R}||_2\to\infty} \mathbf{K} = 0, \quad \lim_{||\mathbf{R}||_2\to\infty} \mathbf{B} = \mathbf{I}, \quad \text{and} \lim_{||\mathbf{R}||_2\to\infty} \mathbf{A} = \mathbf{U},$$

then

$$\begin{aligned}
&\lim_{||\mathbf{R}||\to\infty} \frac{\partial}{\partial r} f_{\mathrm{d},r}(\mathbf{x})\bigg|_{r=0} \\
&= \left[((\mathbf{I}-\mathbf{U})t)^{\mathrm{T}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathrm{T}}(\Gamma - \mathbf{U}\Gamma\mathbf{U}^{\mathrm{T}})\mathbf{x} - 1\right] \phi(\mathbf{x}; 0, \mathbf{I}) + \phi(\mathbf{x}; \mathbf{U}t, \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}),
\end{aligned}$$

and

$$\lim_{||\mathbf{R}||\to\infty} \frac{\partial}{\partial r} f_{\mathrm{s},r}(\mathbf{x})\bigg|_{r=0} = -\phi(\mathbf{x}; 0, \mathbf{I}) + \phi(\mathbf{x}; t, \mathbf{S}).$$

The rest of the proof is simply repeating the argument for the proof of Theorem 2.2.1 (i) and (ii).

2.6.2 Proof of Theorem 2.2.4

The result is straight forward if one realizes that $F_{\mathrm{s},r}$ and $F_{\mathrm{d},r}$ are both Gaussian mixtures with two components. One can calculate analytically the parameters of each component. Then straight calculus gives:

$$\begin{aligned}&\nu(G,F;M_{\mathrm{s}}(\mathbf{y}))\\ =&\beta^3t^3+\left(3\alpha^3\beta+6\alpha\beta^2\right)t^2+\left(3\alpha^2\beta+3\beta^2+3\beta^3(\mathbf{S}-1)\right)t+(\mathbf{S}-1)(3\alpha^3\beta+6\alpha\beta^2),\end{aligned} \tag{2.30}$$

and

$$\begin{aligned}&\nu(G,F;M_{\mathrm{d}}(\mathbf{y}))\\ =&\beta^{\frac{3}{2}}t^3+\left(3\alpha^3\beta+6\alpha\beta^2\right)t^2+\left(3\alpha^2\beta+3\beta^2-3\beta^{\frac{3}{2}}+3\beta^{\frac{1}{2}}\mathbf{S}\right)t+(\mathbf{S}-1)(3\alpha^3\beta+6\alpha\beta^2),\end{aligned} \tag{2.31}$$

where

$$\alpha=\frac{\mathbf{y}}{1+\mathbf{R}},\quad \beta=\frac{\mathbf{R}}{1+\mathbf{R}}.$$

Then the results in Theorem 2.2.4 follows immediately because $0<\beta<1$ for all $\mathbf{R}$.

Chapter 3

Improving the EnKF: the NLEAF Algorithm

In this chapter we consider state space models (SSM) with state sequence $\{X_t : t \geq 1\}$ and observation sequence $\{Y_t : t \geq 1\}$ with the following form:

$$\begin{aligned} X_{t+1} &= f_t(X_t, U_t), \quad f_t(\cdot,\cdot) : \mathbb{R}^p \times [0,1] \mapsto \mathbb{R}^p, \\ (Y_t | X_t = x) &\sim g(\cdot; x), \quad g(\cdot;\cdot) : \mathbb{R}^q \times \mathbb{R}^p \mapsto \mathbb{R}^+. \end{aligned} \tag{3.1}$$

where U_t is a random variable independent of everything else with uniform distribution on $[0,1]$ and $g(\cdot; x)$ is a density function for each x. The state variable X_t evolving according to the dynamics $f_t(\cdot, U_t)$ is usually of interest but never directly observed. Instead it can only be learned indirectly through the observations Y_t. SSM have been widely used in sciences and engineering including signal processing, public health, ecology, economics and geophysics. For a comprehensive summary, please see Fan & Yao (2003); Künsch (2001). A central problem in SSM is the filtering problem: assume that $f(\cdot,\cdot)$ and $g(\cdot;\cdot)$ are known,

how can one approximate the distribution of X_t given the observations $Y_1^t := (Y_1, \ldots, Y_t)$ and the initial distribution of X_0, for every $t \geq 1$? A related problem of much practical interest is the tracking problem: for a realization of the SSM, how can one locate the current hidden state X_t based on the past observations Y_1^t? Usually the filtered expectation $E(X_t|Y_1^t = y_1^t)$ can be used as the best guess for X_t.

A closed form solution to the filtering problem is available only for a few special cases such as the Gaussian linear model (Kalman filter). The Kalman filter variants for non-linear dynamics include the extended Kalman filter (EKF), the unscented Kalman filter (UKF, Julier & Uhlmann (1997)) and the Ensemble Kalman filter (EnKF, Evensen (2003)). The ensemble Kalman filter (EnKF), a combination of sequential Monte Carlo (SMC, see below) and the Kalman filter, mostly used in geophysical data assimilation, has performed successfully in high dimensional models (Evensen, 2007).

Despite the ease of implementation of the Kalman filter variants, they might still be seriously biased because the accuracy of the Kalman filter update relies on the linearity of the observation function and the Gaussianity of the distribution of X_t given y_1^{t-1}, both of which are likely to fail in reality. Another class of ensemble filtering technique is sequential Monte Carlo (SMC (Liu & Chen, 1998)) method, or the particle filter (PF, Gordon et al. (1993)), which feature a fully non-parametric update and are less biased under general non-linear models.

The basic idea of the PF (also the EnKF) is using a discrete set of n weighted particles to represent the distribution of X_t, where the distribution is updated at each time by changing the particle weights according to their likelihoods. It can be shown

that the PF is consistent under certain conditions, e.g., when the hidden Markov chain $\{X_t : t \geq 1\}$ is ergodic and the state space is compact (Künsch, 2005), whereas the EnKF in general is not (Le Gland et al., 2009; Lei et al., 2009).

A major challenge arises when p, q are very large in model (3.1) while n is relatively small. In typical climate models p can be a few thousands with n being only a few tens or hundreds. Even Kalman filter variants cannot work on the whole state variable because it is hard to estimate very large covariance matrices. It is also known that the particle filter suffers from the "curse of dimensionality" due to its nonparametric nature Bengtsson et al. (2008) even for moderately large p. As a result, dimension reduction must be employed in the filtering procedure. For example, a widely employed technique in geophysics is "localization": the whole state vector and observation vector are decomposed into many overlapping local patches according to their physical location. Filtering is performed on each local patch and the local updates are pieced back to get the update of the whole state vector. Such a scheme works for the EnKF but not for the PF because the former keeps track of each particle whereas the PF involves a reweighting/resampling step in the update of each local patch and there is no straightforward way of reconstructing the whole vector since the correlation among the patches is lost in the local reweighting/resampling step.

To sum up it is desirable to have an non-linear filtering method that is easily localizable like the EnKF and adaptive to non-linearity and non-Gaussianity like the PF. In this chapter we propose a nonlinear filter that combines the advantages of both the EnKF and the PF. This is a filter that keeps track of each particle and uses direct particle

transformation like the EnKF while using importance sampling as the PF to avoid serious bias. The new filter, which we call the Non-Linear Ensemble Adjustment Filter (NLEAF), is indeed a further combination of the EnKF and the PF in that it uses a moment-matching idea to update the particles while using importance sampling to estimate the posterior moments. It is conceptually and practically simple and performs competitively in simulations. Single step consistency can be shown for certain Gaussian linear models.

In Section 3.1 we describe EnKF and PF with emphasis on the issue of dimension reduction. The NLEAF method is described and the consistency issue is discussed in Section 3.2. In Section 3.3 we present the simulation results on two synthesized chaotic systems which are common testbeds used in numerical weather forecasting.

3.1 Ensemble filtering at a single time step

Since the filtering methods considered in this chapter are all recursive, from now on we focus on a single time step and drop the time index t whenever there is no confusion. Let X_f denote the variable $(X_t|y_1^{t-1})$ where the subindex f stands for "forecast", and Y denote Y_t. Let X_u denote the conditional random variable $(X_t|y_1^t)$.

Suppose the forecast ensemble $\{x_f^{(i)}\}_{i=1}^n$ is a random sample from X_f, and the observation $Y = y$ is also available. There are two inference tasks in the filtering/tracking procedure:

(a) Estimate $E(X_u)$ to locate the current state.

(b) Generate the updated ensemble $\{x_u^{(i)}\}_{i=1}^n$, i.e., a random sample from X_u, which

will be used to generate the forecast ensemble at next time.

3.1.1 The ensemble Kalman filter

We first revise the Kalman filter in a one-step context. Assuming a Gaussian forecast distribution and a linear observation model

$$\begin{aligned} X_f &\sim N(\mu_f, \Sigma_f), \\ Y &= HX_f + \epsilon, \quad \epsilon \sim N(0, R), \end{aligned} \tag{3.2}$$

then $X_u = (X_f|y)$ is still Gaussian:

$$X_u \sim N(\mu_u, \Sigma_u),$$

where

$$\mu_u = \mu_f + K(y - H\mu_f), \quad \Sigma_u = (I - KH)\Sigma_f, \tag{3.3}$$

and

$$K = \Sigma_f H^T (H\Sigma_f H^T + R)^{-1} \tag{3.4}$$

is the *Kalman gain*.

The EnKF (Evensen, 1994, 2003, 2007) approximates the forecast distribution by a Gaussian with the empirical mean and covariance, then updates the parameters using the Kalman filter formula. Recall the two inference tasks listed in the beginning of this section. The estimation of $E(X_u)$ is straightforward using the Kalman filter formula. To generate the updated ensemble, a naïve (and necessary if in the Gaussian case) idea is to sample directly from the updated Gaussian distribution. This will, as verified widely

in practice, lose much information in the forecast ensemble, such as skewness, kurtosis, clustering, etc. Instead, in the EnKF update, the updated ensemble is obtained by shifting and re-scaling the forecast ensemble. A brief EnKF algorithm is described as below:

The EnKF procedure

1. Estimate $\hat{\mu}_f$, $\hat{\Sigma}_f$.
2. Let $\hat{K} = \hat{\Sigma}_f H^T (H\hat{\Sigma}_f H^T + R)^{-1}$.
3. $\hat{\mu}_u = (I - \hat{K}H)\hat{\mu}_f + \hat{K}y$.
4. $x_u^{(i)} = x_f^{(i)} + \hat{K}(y - Hx_f^{(i)} - \epsilon^{(i)})$, with $\epsilon^{(i)} \overset{iid}{\sim} N(0, R)$.[1]
5. The next forecast ensemble is obtained by plugging each particle into the dynamics: $x_{t+1,f}^{(i)} = f_t(x_u^{(i)}, u_i)$, $i = 1, \ldots, n$.

Under model (3.2) the updated ensemble is approximately a random sample from X_u and that $\hat{\mu}_u \to \mu_u$ as $n \to \infty$. The method is biased if the model (3.2) does not hold (Furrer & Bengtsson, 2007). Large sample asymptotic results can be found in Le Gland et al. (2009), where the first two moments of the EnKF are shown to be consistent under the Gaussian linear model, see also Lei et al. (2009).

3.1.2 The particle filter

The particle filter (Gordon et al., 1993; Liu & Chen, 1998) also approximates the distribution of X_f by a set of particles. It differs from the EnKF in that instead

[1] In step 4 there is another update scheme which does not use the random perturbations $\epsilon^{(i)}$. This deterministic update, also known as the Kalman square-root filter, is usually used to avoid sampling error when the ensemble size is very small (Anderson, 2001; Bishop et al., 2001; Whitaker & Hamill, 2002; Tippett et al., 2003; Lei et al., 2009).

of assuming a Gaussian and linear model, it reweights the particles according to their likelihood. Formally, one simple version of the PF acts as the following:

A simple version of the particle filter

1. Compute weight $W_i = \frac{g(y;x_f^{(i)})}{\sum_{j=1}^n g(y;x_f^{(i)})}$ for $i = 1, \ldots, n$.
2. The updated mean $\hat{\mu}_u = \frac{\sum_{i=1}^n x_f^{(i)} g(y;x_f^{(i)})}{\sum_{i=1}^n g(y;x_f^{(i)})}$.
3. Generate n random samples $x_u^{(1)}, \ldots, x_u^{(n)}$ i.i.d from $\{x_f^{(i)}\}_{i=1}^n$ with probability $P(X_u^{(1)} = x_f^{(i)}) = W_i$ for $i = 1, \ldots, n$.

It can be shown (Künsch, 2005) that under strong conditions such as compactness of the state space and mixing conditions of the dynamics, the particle approximation of the forecast distribution is consistent in L_1 norm uniformly for all $1 \leq t \leq T_n$, for $T_n \to \infty$ subexponentially in n. However, it is well-known that the PF has a tendency to collapse (also known as sample degeneracy) especially in high-dimensional situations, see Liu (2001), and rigorous results in Bengtsson et al. (2008). It is suggested that the ensemble size n needs to be at least exponential in p to avoid collapse.

Another fundamental difference between the PF and the EnKF is that in the PF, $x_u^{(i)}$ is generally not directly related to the $x_f^{(i)}$ because of reweighting/resampling. Recall that in the EnKF update, each particle is updated explicitly and $x_u^{(i)}$ does correspond to $x_f^{(i)}$. This difference materializes in the dimension reduction as discussed below.

3.1.3 Dimension reduction via localization

Dimension reduction becomes necessary for both the EnKF and PF when X and Y are high dimensional, e.g., in numerical weather forecasting X and Y represents the

underlying and observed weather condition. It is usually the case that the coordinates of the state vector X and observation Y are physical quantities measured at different grid points in the physical space. Therefore it is reasonable to assume that two points far away in the physical space have little correlation, and the corresponding coordinates of the state vector can be updated independently using only the "relevant" data (Houtekamer & Mitchell, 1998; Bengtsson et al., 2003; Ott et al., 2004; Anderson, 2007). Formally, let $X = (X(1), ..., X(p))^T$. One can decompose the index set $\{1, \ldots, p\}$ into L (possibly overlapping) local windows $N_1, \ldots, N_L$ such that $|N_l| \ll p$ and $\bigcup_l N_l = \{1, \ldots, p\}$, and correspondingly decompse $\{1, \ldots, q\}$ into $\{N'_1, \ldots, N'_L\}$ such that $|N'_l| \ll q$ and $\bigcup_l N'_l = \{1, \ldots, q\}$. Let $X_f(N_l)$ denote the subvector of X_f consisting of the coordinates in N_l, and similarly define $Y(N'_l)$. $Y(N'_l)$ is usually chosen as the *local observation* of local state vector $X_f(N_l)$.

The localization of the EnKF is straightforward: For each local window N_l and its corresponding local observation window N'_l, one can apply the EnKF on $\{x_f^{(i)}(N_l)\}_{i=1}^n$ and $y(N'_l)$ with local observation matrix $H(N'_l, N_l)$, which is the corresponding submatrix of H. In the L local EnKF updates, each coordinate of X might be updated in multiple local windows. The final update is a convex combination of these multiple updates. Such a localized EnKF has been successfully implemented in the Lorenz 96 system (a 40 dimensional chaotic system, see Section 3.3) with the sample (ensemble) size being only 10 (Ott et al., 2004). The localization idea will be further explained in Section 3.2.1. To be clear, we summarize the localized EnKF as simply L parallel runs of EnKF plus a piecing step:

The localized EnKF

1. For $l = 1, \ldots, L$, run the EnKF on $\{x_f^{(i)}(N_l)\}_{i=1}^n$ and $y(N_l')$, with local observation matrix $H(N_l', N_l)$. Store the results: $\hat{\mu}_u(N_l)$ and $\{x_u^{(i)}(N_l)\}_{i=1}^n$.

2. For each $j = 1, \ldots, p$, let $\hat{\mu}_u(j) = \sum_{l:j\in N_l} w_{j,l}\hat{\mu}_u(N_l; j)$, and $x_u^{(i)}(j) = \sum_{l:j\in N_l} w_{j,l}x_u^{(i)}(N_l; j)$, where $X(N_l; j)$ is the coordinate of $X(N_l)$ that corresponds to $X(j)$, and $w_{j,l} \geq 0$, $\sum_{l:j\in N_l} w_{j,l} = 1$.

The choices of local windows N_l, N_l' and combination coefficients $w_{j,l}$ can be pre-determined since in many applications there are simple and natural choices. They can also be chosen in a data-driven fashion. For example, as we will explain later, the Kalman filter is essentially a linear regression of X on Y. Therefore for each coordinate of X one can use sparse regression techniques to select the most relevant coordinates in Y. Similarly the choice of $w_{j,l}$ in the algorithm can be viewed as a problem of combining the predictions from multiple regression models and can be calculated from the data (Breiman, 1996; Yang, 2001; Bunea et al., 2007). We will return to this issue in Section 3.2.1.

On the other hand, such a dimension reduction scheme is not applicable to the PF because each particle is reweighted differently in different local windows. In words, the reweighting breaks the strong connection of a single particle update in different local windows and it is not clear how to combine the updated particle across the local windows. This can be viewed as a form of sample degeneracy: in high dimension situations, a particle might be plausible in some coordinates but absurd in other coordinates.

So far, the properties of the EnKF and the PF can be summarized as in Table

Table 3.1: A quick comparison of the EnKF and the PF.

	consistent	stable	localizable
EnKF	✗	✓	✓
PF	✓	✗	?

3.1, where the only check mark for the PF is higher accuracy. A natural idea to reduce the bias of EnKF is to update the mean of X using importance sampling as in the PF. Meanwhile, a possible improvement of the PF is avoiding the reweighting/resampling step. One possibility is generating an ensemble using direct transformations on each particle as in the EnKF. In the next section we present what we call the "nonlinear ensemble adjustment filter" (NLEAF) as a combination of the EnKF and the PF. Some relevant works (Bengtsson et al., 2003; Chorin & Tu, 2009) also have the flavor of combining the EnKF and the PF, but both involve some form of resampling, which destroys the spatial smoothness of the forecast ensemble.

3.2 The NonLinear Ensemble Adjustment Filter (NLEAF)

3.2.1 A regression perspective on the EnKF and two sources of bias

In equation (3.4), the Kalman gain K^T is simply the linear regression coefficient of X_f on Y. In fact, from Model (3.2) we have $\mathrm{Cov}(X, Y) = \Sigma_f H^T$ and $\mathrm{Var}(Y) = H\Sigma_f H^T + R$, therefore $K^T = \mathrm{Var}(Y)^{-1}\mathrm{Cov}(Y, X_f)$. The conditional expectation of X_f given y is

$$\mu_f + K(y - H\mu_f) := m_1(y).$$

Let $y^{(i)} = Hx_f^{(i)} + \epsilon^{(i)}$ be an observation given $X_f = x_f^{(i)}$ then $(x_f^{(i)}, y^{(i)})$ is a

random sample from the joint distribution of (X_f, Y). $\hat{m}_1(\cdot) = \hat{\mu}_f + \hat{K}(\cdot - H\hat{\mu}_f)$ is an estimator of $m_1(\cdot)$. The update step of the EnKF can be written as

$$x_u^{(i)} = \hat{m}_1(y) + x_f^{(i)} - \hat{m}_1(y^{(i)}). \tag{3.5}$$

Under Model (3.2) we have that $(X_f - m_1(y)|Y = y) \sim N(0, \Sigma_u)$ where Σ_u does not depend on y. Note further that $\left(x_f^{(i)}, y^{(i)}\right) \sim (X_f, Y)$, so $x_f^{(i)} - m_1(y^{(i)})$ is a random draw from $N(0, \Sigma_u)$. Therefore $x_u^{(i)} = m_1(y) + x_f^{(i)} - m_1(y^{(i)})$ is a random draw from $N(\mu_u, \Sigma_u)$ by noting that $m_1(y) = \mu_u$, which validates the update formula (3.5).

The procedure described above is an abstraction of the EnKF which can be viewed as a solution to the sampling problem of generating a random sample of $(X_f|Y = y)$ given a sample of X_f. Classical approaches to this problem includes rejective sampling and importance sampling (with possibly a resampling step). However, the approach described above uses direct transformations on the particles $x_f^{(i)}$, with randomness involved only in generating $y^{(i)}$. This procedure is effective in the sense that each particle in the forecast ensemble correspond to exactly one particle in the updated ensemble, without sample degeneracy.

Based on the discussion above, an effective way of updating the ensemble is directly transforming each particle so that the transformed particles have the desired distribution. In a Gaussian linear model, it suffices to adjust the mean by a simple shift as in equation (3.5) and the posterior variance is implicitly obtained by generating the random number $x_f^{(i)} - \hat{m}_1(y^{(i)})$. For general models where the likelihood function $g(y; x)$ and the forecast distribution are not Gaussian, the EnKF introduces bias from two sources.

The first source of bias is estimating $E(X_f|Y = y)$ using a linear function of y.

If the model is non-Gaussian non-linear, $E(X-\hat{m}_1(y))$ would be non-zero even asymptotically. As a result, the updated ensemble $\{x_u^{(i)}\}_{i=1}^n$ no longer has the desired mean $m_1(y)$. We call this bias the first order bias which is due to using the wrong estimator $\hat{m}_1(\cdot)$.

On the other hand, under non-Gaussian non-linear models, the two variables $(X_f|Y=y)-m_1(y)$ and $(X_f|Y=y')-m_1(y')$ will also have different shape, because higher order moments might depend on Y as well. Although this will not cause any problem in the current step if one is only interested in the ensemble average, it might cause problems in the future since such a bias in shape will be propagated by the dynamics. We call this bias the higher order bias.

3.2.2 Reducing the first order bias: the NLEAF algorithm

According to the previous discussion, reducing the first order bias amounts to finding a better estimator for $m_1(\cdot)$, which can be written as the following,

$$m_1(y) = E(X_f|Y=y) = \frac{\int x p_f(x) g(y;x) dx}{\int p_f(x) g(y;x) dx},$$

where $p_f(\cdot)$ is the density of X_f. Given a random sample $\{x_f^{(i)}\}_{i=1}^n$ from X_f, a well-known non-parametric estimator is using importance sampling (Hammersley & Handscomb, 1965):

$$\hat{m}_1(y) = \frac{\sum_{i=1}^n x_f^{(i)} g(y;x_f^{(i)})}{\sum_{i=1}^n g(y;x_f^{(i)})}.$$

Based on this idea, we propose an alternative to the EnKF update: the first order NLEAF algorithm. It is a direct generalization of the EnKF by using importance sampling to estimate $m_1(\cdot)$:

The first order NLEAF

1. Generate $y^{(i)} \sim g(\cdot; x_f^{(i)})$, for $i = 1, \ldots, n$.
2. Estimate $m_1(\cdot)$ by $\hat{m}_1(\cdot) = \frac{\sum_{i=1}^n x_f^{(i)} g(\cdot; x_f^{(i)})}{\sum_{i=1}^n g(\cdot; x_f^{(i)})}$.
3. Updated mean $\hat{\mu}_u = \hat{m}_1(y)$. Updated particle $x_u^{(i)} = \hat{m}_1(y) + x_f^{(i)} - \hat{m}_1(y^{(i)})$.

This approach is valid if $\mathcal{L}\left(X_f - m_1(y^{(i)})|y^{(i)}\right) \approx \mathcal{L}\left(X_f - m_1(y)|y\right)$, where $\mathcal{L}(X)$ denotes the distribution of the random variable X. That is, $\mathcal{L}(X_f|y)$ depends on y mostly in terms of the mean. A simple example is the Gaussian linear model, where only the posterior mean depends on y. One can also expect such a situation when the likelihood $g(y; x)$ has a lighter tail than the forecast distribution X_f. To formalize, let

$$\eta = \sup_{y', y} \mathrm{TV}\left(\mathcal{L}(X_f - m_1(y')|y'), \mathcal{L}(X_f - m_1(y)|y)\right),$$

where $\mathrm{TV}(\mathcal{L}_1, \mathcal{L}_2) = \sup_A |P_{\mathcal{L}_1}(A) - P_{\mathcal{L}_2}(A)|$ denotes the total variation distance between two distributions $\mathcal{L}_1$ and $\mathcal{L}_2$. Then the smaller η is, the better is the approximation given by the first order NLEAF. To state a rigorous result, we need the following technical conditions on the likelihood function $g(x; y)$ which make the argument simple.

(A0) X_f has density function $f(\cdot) > 0$.

(A1) $0 < g(x; y) \leq M < \infty$ for all (x, y), $\sup_{x \in \mathbb{R}^p, y \in K} |xg(x; y)| \leq M_K < \infty$ for all compact $K \subset \mathbb{R}^q$.

(A2) For any compact set $K \subseteq \mathbb{R}^q$, there exists a measurable function $v_K(x)$, such that $E(v_K^2(X)) < \infty$ and for any $y_1, y_2 \in K$,

$$\max\left(|xg(x; y_1) - xg(x; y_2)|, |g(x; y_1) - g(x; y_2)|\right) \leq v_K(x)|y_1 - y_2|.$$

The conditions A1 and A2 are standard conditions for the maximal inequalities in empirical processes. They imply that the likelihood function $x \mapsto g(x;y)$ ($x \mapsto xg(x;y)$) depends on y continuously, which controls the complexity of the class of functions $x \mapsto g(x;y)$ ($x \mapsto xg(x;y)$) indexed by y and enables the use of the classical results of empirical processes. They also imply that the observation Y provides information for the whole vector of X, which precludes the degenerate situations such as $X = (X_1, X_2)^T$ and $Y = h(X_1)$. These conditions are reasonably general, including models like $g(x;y) \propto \phi(|x-y|)$ with $\phi(\cdot)$ decaying fast enough, e.g., for the Gaussian density function one can find the $v_K(x)$ is bounded by a constant. We have the following theorem whose proof is in Section 3.5:

Theorem 3.2.1. *Suppose* $\left(x^{(i)}, y^{(i)}\right)$, $i = 1, \ldots, n$ *is an i.i.d sample from the joint distribution of* (X_f, Y). *Let* $x_u^{(i)}$, $i = 1, \ldots, n$, *be the updated particles given by the first order NLEAF algorithm. For any* y, *consider the empirical distribution*

$$\hat{F}_u(A|y) = \frac{1}{n}\delta_{x_u^{(i)}}(A), \quad \forall A,$$

where

$$x_u^{(i)} = \hat{m}_1(y) + x_f^{(i)} - \hat{m}_1(y^{(i)}).$$

Also let $F_u(A|y) = P(X_f \in A|y)$ *be the true conditional measure. Then, under (A0-A2) for Borel set* A *with* $\lambda(\partial A) = 0$, *we have*

$$\limsup_{n\to\infty} |\hat{F}_u(A|y) - F_u(A|y)| \leq \eta, \quad a.s.,$$

where $\lambda(\cdot)$ *is the Lebesgue measure and* $\partial A := \bar{A} \backslash A^\circ$ *is the boundary of* A, *with* $\bar{A}$ *and* A° *being the compact closure and interior of* A, *respectively.*

A by-product of the proof of Theorem 3.2.1 is the consistency of mean update:

Corollary 3.2.2. *Under (A0-A2), we have for any y,*

$$\hat{m}_1(y) \to m_1(y), \quad a.s., \quad n \to \infty.$$

Under the Gaussian linear model we have $\eta = 0$. The above results indicate the consistency of the NLEAF of order one:

Corollary 3.2.3. *Under Model (3.2), for any y,*

$$\hat{F}_u \xrightarrow{d} F_u, \quad a.s., \quad n \to \infty.$$

3.2.3 Second order correction

The basic idea of NLEAF algorithm can be easily generalized to correct the second order moment. Note that the conditional variance can be estimated using importance sampling as following:

$$\hat{m}_1(y) = \frac{\sum_{i=1}^n g(y; x_f^{(i)}) x_f^{(i)}}{\sum_{i=1}^n g(y; x_f^{(i)})}, \tag{3.6}$$

$$\hat{m}_2(y) = \frac{\sum_{i=1}^n g(y|x_f^{(i)})(x_f^{(i)} - \hat{m}_1(y))(x_f^{(i)} - \hat{m}_1(y))^T}{\sum_{i=1}^n g(y; x_f^{(i)})}. \tag{3.7}$$

If the likelihood $g(\cdot;\cdot)$ is not known explicitly (eg, y is generated by a black-box function), one may use regression methods to estimate the conditional moments. For example, the EnKF uses a linear regression of X_f on Y to find $\hat{m}_1(y)$. However, under general models, one might need more general methods, such as polynomial regressions, to avoid serious bias. This idea is further explained in Section 3.3.2.

Based on the estimated conditional variance in (3.7), one can easily develop a second order NLEAF algorithm. Now the update is naturally chosen as

$$x_u^{(i)} = \hat{m}_1(y) + (\hat{m}_2(y))^{\frac{1}{2}} \left(\hat{m}_2\left(y^{(i)}\right)\right)^{-\frac{1}{2}} \left(x_f^{(i)} - \hat{m}_1\left(y^{(i)}\right)\right). \tag{3.8}$$

The update formula is intuitively reasonable: Suppose $x, y \in \mathbb{R}^1$, then a large $m_2(y^{(i)})$ means that the region where $x_f^{(i)}$ lies in is highly uncertain which is possibly due to the irregular behavior of the dynamics in that region. Such a particle $x_f^{(i)}$ can provide little information on the true hidden state, therefore it is down-weighted in the transformation ξ_i, which tends to drag $x_f^{(i)}$ towards $\hat{\mu}_u = \hat{m}_1(y)$ in the updated ensemble.

It should be noted that the update $x_u^{(i)}$ is apparently not unique. For example, for any orthogonal matrix U, one can define $x_u^{(i)} = \xi_i(x_f^{(i)}; U)$ through the function $\xi_i(x; U)$:

$$\xi_i(x; U) = \hat{m}_1(y) + (\hat{m}_2(y))^{\frac{1}{2}} U \left(\hat{m}_2\left(y^{(i)}\right)\right)^{-\frac{1}{2}} \left(x - \hat{m}_1\left(y^{(i)}\right)\right).$$

It is easily seen that the choice of U does not change the first two moments of $\mathcal{L}\left(\xi_i(X_f; U)|y^{(i)}\right)$. The choice $U = I$ is natural in the sense that under Model (3.2) with $\Sigma_u = \sigma^2 I$, if $U = I$ then the second order NLEAF is asymptotically equivalent to the first order NLEAF, which is proved to be consistent. This is analogous to the issue of choosing the scaling matrix in the Kalman square-root filter (Lei et al., 2009). In the rest of this chapter, we will focus on the natural choice $U = I$.

3.2.4 Localization for the NLEAF algorithm

As seen above, the NLEAF algorithm is similar to the EnKF in that it updates each particle explicitly instead of resampling. As a result, one may expect a similar

localization procedure as described in Section 3.1.3 applicable to the NLEAF algorithm. Recall that the EnKF localization involves three major steps:

a) Decompose the state vector X_f into local windows $X_f(N_l)$, $l = 1, \ldots, L$, find the corresponding local observation vector $Y(N_l')$, and the local likelihood function $g_l(y(N_l'); x_f(N_l))$;

b) Update each localized ensemble;

c) Construct the whole updated ensemble by combining the local updated ensembles.

In step a), one can usually construct a local window for each coordinate of X_f, where $X_f(N_j)$ is the subset of coordinates most relevant to $X_f(j)$, $j = 1, \ldots, p$. One can also choose these coordinates by subject knowledge. For example, in geophysics each coordinate corresponds to a physical location, then one can choose the coordinates in a neighborhood of the physical location of $X_f(j)$. Or one can use data-driven variable selection procedures to determine the relevant neighborhood $X_f(N_j)$. The choice of N_j' is similar. In many cases the special structure of the observation model (the second equation in (3.1)) enables natural and simple solutions. For example, under the linear model $Y = HX_f + \epsilon$, if H is sparse or banded, it is possible to find a submatrix $H_j = H(N_j', N_j)$ such that $Y(N_j') \approx H_j X_f(N_j) + \epsilon(N_j')$, where N_j' is a subset of $1, \ldots, q$ such that $y_{N_j'}$ is the local observation corresponding to $X_f(N_j)$.

Once step a) is done, in step b) one only needs to apply the NLEAF algorithm as described above on each of the localized ensemble. The major issue is step c). Recall that the local windows overlap with each other, therefore each coordinate might be

updated simultaneously in multiple local patches. To be concrete, for any local window $N_j \subseteq \{1, \ldots, p\}$, let $N_j' \subseteq \{1, \ldots, q\}$ be the corresponding local observation window, and $g_j(x_{N_j}; y_{N_j'})$ be the local likelihood function. Define N_k, N_k' and $g_k(\cdot;\cdot)$ similarly for another local window N_k. Suppose $r \in N_j \cap N_k$, then $X_f(r)$ is updated in both of these two local windows. From now on we consider the first order and second order NLEAF separately.

In the first order NLEAF, we write the update formula for the mean in both local windows as in equation (3.5):

$$\hat{\mu}_u(N_j) = \hat{m}_{1,j}(y(N_j')),$$

$$\hat{\mu}_u(N_k) = \hat{m}_{1,k}(y(N_k')),$$

where $\hat{m}_{1,j}(\cdot)$ denotes the local estimation of $m_{1,j}(\cdot) := E(X_f(N_j)|y(N_j'))$. Recall that we denote $(N_j; r)$ the position of the index r in the vector N_j. Then $\hat{\mu}_u(N_j; r)$ and $\hat{\mu}_u(N_k; r)$ can be viewed as predictions of $X_f(r)$ given different sets of predictors, namely $Y(N_j')$ and $Y(N_k')$, respectively. A natural method of combining the predictions of the same variable from different models is convex combination, which is chosen either conventionally or in a data-driven manner (Breiman, 1996; Yang, 2001; Bunea et al., 2007). In our numerical experiment we follow the conventional choice described in Ott et al. (2004) where the combination is simply averaging the updates in a few spatially coherent local windows. It is straightforward that this combination procedure is also applicable to the update of each single particle for exactly the same reason. However, a theoretically justifiable method of pasting together local updates is still to be developed.

On the other hand, the above method of combining local updates does not apply directly to the second order NLEAF because in equation (3.8) the left-multiplication of the matrix $(\hat{m}_2(y))^{\frac{1}{2}} \left(\hat{m}_2(y^{(i)})\right)^{-\frac{1}{2}}$ mixes the coordinates in the local window, which makes coordinates in the left hand side no longer an estimate of the corresponding coordinate of the state variable, which invalidates the convex combination.

3.3 Numerical experiments

We present numerical experiments on two dynamical systems, both proposed by E. Lorenz in studying the predictability of chaotic systems. These systems have been widely used as test beds for atmospheric data assimilation methods (Bengtsson et al., 2003; Ott et al., 2004; Anderson, 2007).

3.3.1 Experiments on L63

The L63 system is first introduced by Lorenz (1963), as one of the earliest study of chaos. This three dimensional system is determined by an ordinary differential equation

$$\frac{dx(\tau)}{d\tau} = -\sigma x + \sigma y, \tag{3.9}$$

$$\frac{dy(\tau)}{d\tau} = -xz + rx - y, \tag{3.10}$$

$$\frac{dz(\tau)}{d\tau} = xy - bz, \tag{3.11}$$

where τ denotes the time, $(x(\tau), y(\tau), z(\tau))^T$ is the state vector and (b, σ, r) are parameters of the system. When $b = 8/3$, $r = 28$ and $\sigma = 10$, the system is chaotic and its orbit is the well-known *Butterfly Attractor*.

In the simulation the system is discretized using the fourth order Runge-Kutta method. It is clear that the linearity of the evolution of the state vector between two successive time points depends on the length of the time interval $\Delta\tau$ between t and $t+1$ which we call the *step size*: The smaller is $\Delta\tau$, the more linear is the evolution between t and $t+1$.

In the simulation, there is a hidden true orbit $\{x_t, t \geq 0\}$. The starting point, x_0, of the true orbit is randomly chosen from the attractor. At the starting time, an ensemble of state vectors $\{x_0^{(i)}\}_{i=1}^n$, surrounding x_0 is available (e.g., perturbations of x_0 with random noise or a random sample from a small neighborhood of x_0 in the attractor). For all $t > 0$ a noisy observation $y_t = x_t + \epsilon_t$ is available with

$$\epsilon_t \overset{iid}{\sim} N(0, \sigma^2 I_3). \tag{3.12}$$

At each time $t \geq 1$, The updated ensemble average is used as the best single estimate of x_t. Therefore, the data assimilation performance is evaluated by the root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{p}||\hat{\mu}_{u,t} - x_t||_2^2}. \tag{3.13}$$

We consider two time steps: 0.05 and 0.2, corresponding to the nearly linear case and the non-linear case respectively. In each case the system is propagated 2000 steps and at each time the data assimilation is performed using four different methods: the EnKF, the PF, the first order NLEAF (NLEAF1) and the second order NLEAF (NLEAF2), each with an ensemble of size 400. Also we consider three values of σ^2 in (3.12): 0.25, 1 and 4, corresponding to different levels of the observation accuracy.

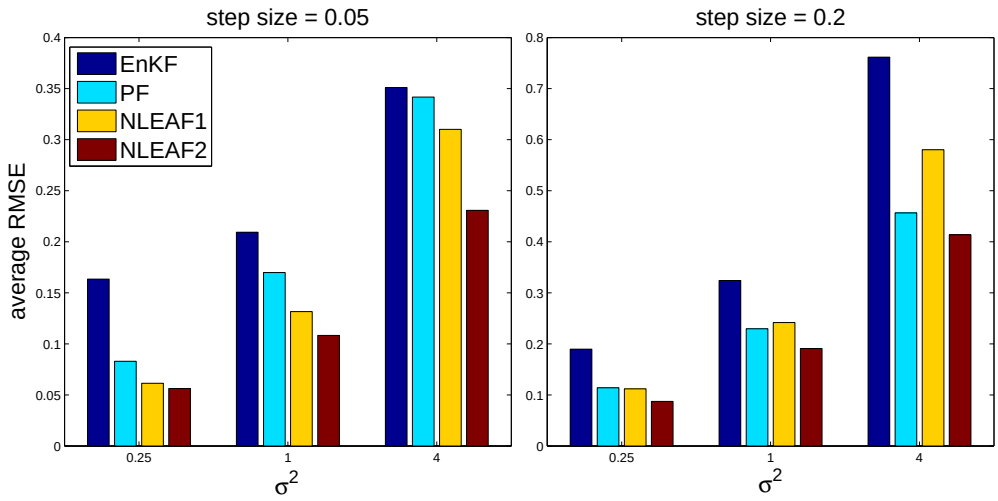

Figure 3.1: Average RMSE over 2000 cycles.

In Figure (3.1) we see that the EnKF gives the largest RMSE because of the non-linear dynamics. The NLEAF2 performs the best under all circumstances considered here. When step size is small, the system is nearly linear so that the NLEAF1 performs better than the PF. When the step size is large and the distribution is significantly non-Gaussian and non-linear, the PF shows some advantage against the NLEAF1 which ignores the higher order moments.

3.3.2 Experiments on L96

The L96 system is introduced in (Lorenz, 1996) in the study of predictability of high dimensional chaotic systems. The state vector is 40 dimensional, and the dynamics is given by an ODE as follows:

$$\frac{dx_j(t)}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 8, \quad \text{for } j = 1, \ldots, 40, \tag{3.14}$$

where $x_0 = x_{40}$, $x_{-1} = x_{39}$ and $x_{41} = x_1$. This system mimics the evolution of some meteorological quantity at 40 equally spaced grid points along a latitude circle. The system is discretized with a time step of $\Delta\tau = 0.05$, which is analogous to a 6 hour in the real world.

Although the dimensionality of the L96 system is still far from the reality, it has been challenging for many standard data assimilation methods including the Kalman filter variants. Among the vast literature, we mention only two previous works: Ott et al. (2004) considered the localized ensemble Kalman filter in an approximately linear case ($\delta\tau = 0.05$) and a complete observation, that is

$$Y_t = X_t + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} N(0, I_{40}). \tag{3.15}$$

We call this set-up the *easy case*. On the other hand, Bengtsson et al. (2003) studied a localized Gaussian mixture filter in a highly non-linear case ($\delta\tau = 0.4$) and an incomplete observation: for $j = 1, \ldots, 20$,

$$Y_t(j) = X_t(2j-1) + \epsilon_t(j), \quad \epsilon_t \overset{iid}{\sim} N(0, I_{20}/2). \tag{3.16}$$

We call this set-up the *hard case*.

The major criterion is still the RMSE defined in (3.13). Moreover, because of its dimensionality and resemblance to real atmospheric data, we do care about the computation, where the main restriction is the ensemble size.

We consider both the easy case and the hard case. The system is propagated 2000 steps from a random starting point with data assimilation performed at each step. Because of the localization, we do not use the second order NLEAF. Instead, we use a

variant of NLEAF1, namely NLEAF1q, with the letter "q" for "quadratic", in which the function $m_1(\cdot) = E(X_f|Y = \cdot)$ is estimated using a quadratic regression of X_f on Y. To be concrete, in the NLEAF1q algorithm $\hat{m}_1(\cdot)$ is the minimizer over all quadratic functions $m(\cdot)$ of the square loss:

$$\sum_{i=1}^{n} \left(m(y^{(i)}) - x_f^{(i)} \right)^2.$$

We consider the NLEAF1q algorithm because we believe sometimes $g(\cdot, \cdot)$ may not be available explicitly and the y's are generated by a black-box function of x. We emphasize that in the NLEAF1q algorithm, the function $g(\cdot, \cdot)$ is pretended to be unknown and not used.

In both NLEAF1 and NLEAF1q, the localization is as described in Section 3.1.1, which is also essentially the same as in Ott et al. (2004): Let l be a pre-chosen window size. For each $j = 1, \ldots, 40$, let $N_j = (j - l, \ldots, j, \ldots, j + l)$ be the local window centered at j. The corresponding local observation window N_j' is the local observations of $X(N_j)$. For example, if $l = 2$, then $N_1 = (39, 40, 1, 2, 3)$. In the easy case, $N_1' = (39, 40, 1, 2, 3)$ since the observation is complete (eq. (3.15)); In the hard case the observation is incomplete (eq. (3.16)) and we have $N_1' = (20, 1, 2)$.For each j, the coordinate $X(j)$ of the state variable X is updated in $2l + 1$ local windows. In the first order NLEAF algorithm, for $k \in N_j$, $X(j)$ is updated in the local window N_k using the conditional expectation given $y_{N_k'}$ (or $y_{N_k'}^{(i)}$). Similar to the scheme proposed in Ott et al. (2004), we combine the local updates of $X_f(j)$ from N_{j-1}, N_j and N_{j+1} by simply averaging them. One can also use a data-driven method at a higher computational cost (Breiman (1996); Yang (2001); Bunea et al. (2007)).

Table 3.2: Summarizing statistics of RMSE's over 2000 time steps in the hard case. Ensemble size = 400.

NLEAF			NLEAFq			EnKF			XEnsF		
mean	med	std	mean	med	std	mean	med	std	mean	med	std
0.65	0.63	0.20	0.71	0.67	0.22	0.83	0.75	0.31	0.92	0.85	0.31

The hard case

In the hard case we compare four methods: the NLEAF1; the NLEAF1q; the mixture ensemble filter (XEnsF Bengtsson et al. (2003)); the EnKF without localization. Following the set-up in Bengtsson et al. (2003), the ensemble size is fixed to be 400. The system is run for a total of 2000 time steps. At each time step three different filtering methods are applied to obtain the updates individually, resulting an RMSE value for each method. We compare the performance of NLEAF1 and NLEAF1q directly with those reported in Bengtsson et al. (2003), summarized in Table 3.2, where we see similar results as in the L63 experiment: The NLEAF1 gives much smaller RMSE than both the XEnsF and the EnKF. This is the first time the authors see the average RMSE goes below 0.7 in this set-up.

The easy case

In the easy case we compare three methods: the NLEAF1, the NLEAF1q and the local ensemble transform Kalman filter (LETKF) proposed by Ott et al. (2004), which achieves the best known performance in this set-up, with an average RMSE of about 0.2 using an ensemble as small as 10. It is reported that enlarging the ensemble size does not improve the accuracy of EnKF (LETKF) while the NLEAF is expected to work better for

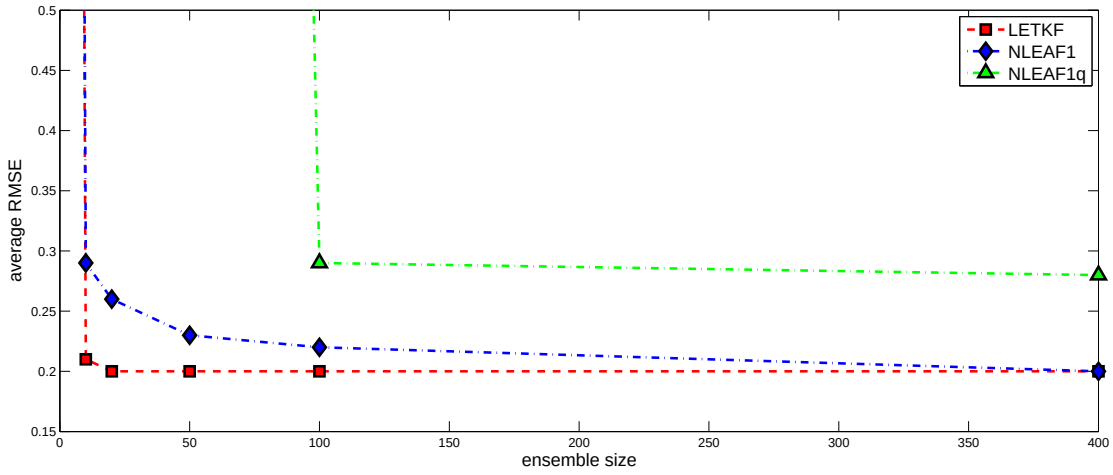

Figure 3.2: Average RMSE over 2000 cycles in the easy case of L96 system, ensemble size = 400.

larger ensembles. Here we consider different ensemble sizes ranging from 10 to 400. The result is summarized in Figure 3.2 where only the mean of the average RMSE is plotted. The median and the variance are qualitatively similar to those presented in the hard case and are omitted here. We see that the LETKF still gives the best performance especially for small ensemble sizes. The NLEAF1 becomes competitive when the ensemble size is moderately large. From the plot it is also reasonable to expect even smaller RMSE of NLEAF1 given even larger ensembles. The performance of NLEAF1q is not as good as the other two methods but we believe it is of practical interest since it requires much less *a priori* knowledge on the observation mechanism.

An intermediate case

So far both the easy and the hard cases are of practical interests: The easy case is analogous to 6-hour operational data assimilation; The hard case challenges forecast in the presence of high nonlinearity and incomplete observation which is often the case in practice. As a result, it would be interesting to consider an *intermediate case* where the time step is still short as in the easy case but the observation is incomplete as in the hard case, with a larger observation noise:

$$Y_t(j) = X_t(2j-1) + \epsilon_t(j), \quad \epsilon_t \overset{iid}{\sim} N(0, 2I_{20}). \tag{3.17}$$

Again we let the ensemble size vary from 10 to 400. The results are summarized in Figure 3.3. Now the NLEAF1 and NLEAF1q gives much better relative results than in the easy case. The NLEAF1 is competitive for a ensemble as large as 100. Here again we see the potentiality of improvement for the NLEAF1 when the ensemble gets large. The NLEAF1q algorithm does a decent job for large ensembles too.

It should be noted that the LETKF tends to lose accuracy when the ensemble size gets beyond 20. There are two possible reasons for this phenomenon: first, the method of combining updates in different local windows might not be optimal for this set-up in varying ensemble sizes; second, the the mis-specification of the linear model assumed by the ensemble Kalman filter incurs a larger bias when the ensemble size gets large.

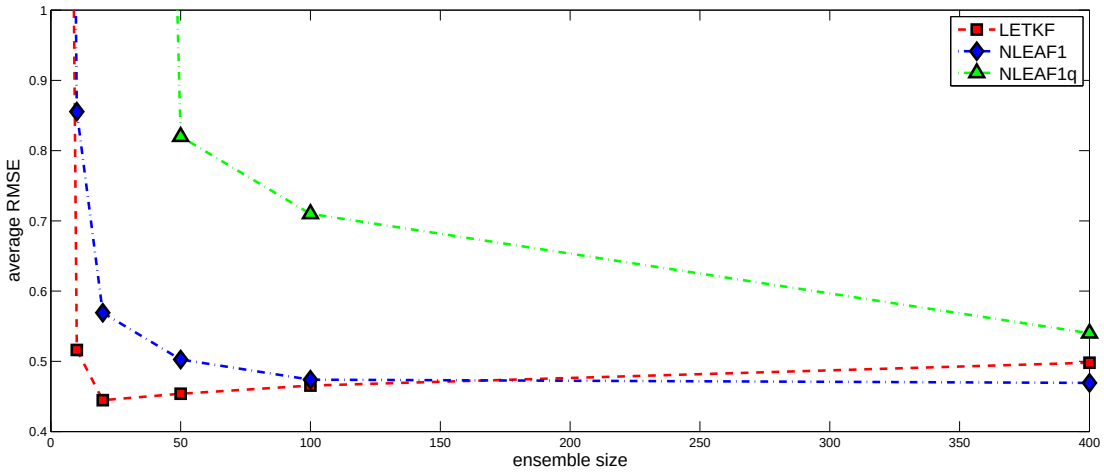

Figure 3.3: Average RMSE over 2000 cycles in the intermediate case of L96 system, ensemble size = 400.

3.4 Conclusion

As the increasing availability of both sophisticated climate models and massive sequential data, scientific applications such as numerical weather forecasting pose new challenges on statistical inference on high-dimensional nonlinear state space model. The proposed NLEAF algorithm is a combination of the traditional ensemble Kalman filter and particle filter which is adaptive to the nonlinearity of the dynamics and also easily scalable to high-dimensional situations. In two classical test beds for atmospheric data assimilation, very simple NLEAF algorithms give reasonably good performances. They outperforms the state-of-art methods in the nonlinear set-up, while still being competitive even in the linear situation where the EnKF is expected to be nearly optimal. We also observe that the NLEAF algorithm has the potential to improve its accuracy for larger ensembles, while the EnKF does not. Furthermore, the NLEAF algorithm is flexible and

allows the observation model to be unknown and estimated from the data, which makes itself more applicable for many real world problems where the observation error can hardly be specified *a priori*.

There are still issues to be addressed. For example, the localization for NLEAF of order two or higher will be useful since we observed a substantial improvement of accuracy by NLEAF2 in the L63 system. A further question is that whether the NLEAF algorithm can be used in combination with other dimension reduction methods such as manifold learning and regularization. Finally, it would be interesting to do more simulations with non-Gaussian observations which is a realistic situation in geophysical sciences. In this case, one can expect even better relative performance for the NLEAF algorithm.

3.5 Proofs

3.5.1 Proof of Theorem 3.2.1

Suppose $(x_f^{(i)}, y^{(i)})$, $i = 1, \ldots, n$ is an i.i.d sample from the joint distribution of (X_f, Y), For any y, consider the empirical distribution

$$F_u^*(A|y) = \frac{1}{n}\delta_{x_u^{*(i)}}(A), \quad \forall A,$$

with

$$x_u^{*(i)} = m_1(y) + x_f^{(i)} - m_1(y^{(i)}).$$

Note that the NLEAF update in equation (3.5) uses $\hat{m}_1(\cdot)$ instead of $m_1(\cdot)$. The rough idea is that if $\hat{m}_1(\cdot)$ approximates $m_1(\cdot)$ well enough, one might expect $x_u^{(i)} \approx x_u^{*(i)}$ and the result follows from Hoeffding's inequality. To show that $x_u^{(i)}$ does approximates

$x_u^{*(i)}$ we use the empirical process theory. The maximal inequality of the empirical process requires the majority of $y^{(i)}$ lies in a compact set, which is of high probability if the compact set is large enough.

For any $0 < \epsilon < 1$, one can find a compact set $K(\epsilon)$ such that $P(Y \in K) \geq 1 - \epsilon$. Define the set J as

$$J = \{i : y^{(i)} \in K(\epsilon)\}.$$

Consider the event

$$E_1 = \left\{ \frac{|J|}{n} \geq 1 - 2\epsilon \right\},$$

then we have, by Hoeffding's inequality,

$$P(E_1) \geq 1 - \exp\left(-2n\epsilon^2\right). \tag{3.18}$$

Let $B(\epsilon) = \inf_{y \in K(\epsilon)} \int g(y;x) f(x) dx > 0$. Consider the events

$$E_2 = \left\{ \sup_{y \in K(\epsilon)} \left| \frac{1}{n} \sum_{i=1}^{n} g(y; x_f^{(i)}) - \int g(y;x) f(x) dx \right| \leq \min\left(\frac{B(\epsilon)}{2}, \frac{B^2(\epsilon)}{8M(\epsilon)} \right) \right\},$$

where $M(\epsilon) = M_{K(\epsilon)}$ as defined in Assumption A1, and

$$E_3 = \left\{ \sup_{y \in K(\epsilon)} \left| \frac{1}{n} \sum_{i=1}^{n} x_f^{(i)} g(y; x_f^{(i)}) - \int x g(y;x) f(x) dx \right| \leq \frac{B(\epsilon)\epsilon}{8} \right\}.$$

By assumption A1 and A2 and the maximal inequality of empirical process (van der Vaart, 2001; Talagrand, 1994), there exist functions $c_i(\epsilon)$, $i = 1, 2$, such that

$$P(E_2^C) \leq c_1(\epsilon) n^{q-1} \exp\left(-n c_2(\epsilon)\right),$$

and

$$P(E_3^C) \leq c_1(\epsilon) n^{q-1} \exp\left(-n c_2(\epsilon)\right).$$

Note that on $E_2 \bigcap E_3$, we have $|\hat{m}_1(y) - m_1(y)| \leq \epsilon/2$, for all $y \in K(\epsilon)$. As a result, on $E_2 \bigcap E_3$, we have,

$$|x_u^{(i)} - x_u^{*(i)}| \leq \epsilon, \quad \forall i \in J.$$

Then we have, on $E_1 \bigcap E_2 \bigcap E_3$,

$$\begin{aligned}
\hat{F}_u(A|y) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_A(x_u^{(i)}) \geq \frac{1}{n}\sum_{i\in J} \mathbb{1}_A(x_u^{(i)}) \\
&\geq \frac{1}{n}\sum_{i\in J} \mathbb{1}_{A_\epsilon^-}(x_u^{*(i)}) \\
&\geq \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_\epsilon^-}(x_u^{*(i)}) - \frac{|J^C|}{n} \\
&\geq \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_\epsilon^-}(x_u^{*(i)}) - 2\epsilon,
\end{aligned}$$

where the set A_ϵ^- is defined as

$$A_\epsilon^- = \{x \in A : D(x, \epsilon) \subseteq A\},$$

with $D(x, \epsilon)$ being the ϵ-open ball centering at x.

Consider event E_4:

$$E_4 = \left\{ \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_\epsilon^-}(x_u^{*(i)}) \geq F_u(A_\epsilon^-|y) - \eta - \epsilon \right\}.$$

Again, note that $\mathbb{1}_{A_\epsilon^-}(x_u^{*(i)})$ are independent Bernoulli random variables with probability at least $F_u(A_\epsilon^-|y) - \eta$, by Hoeffding's inequality, we have

$$P(E_4) \geq 1 - \exp(-2n\epsilon^2).$$

Then on $\bigcap_{k=1}^{4} E_k$, we have

$$\hat{F}_u(A|y) - F(A|y) \geq F(A_\epsilon^-|y) - F(A|y) - \eta - 3\epsilon$$

$$= -\eta - \rho^{-}(\epsilon) - 3\epsilon,$$

where $\rho^{-}(\epsilon) = F(A|y) - F(A_{\epsilon}^{-}|y)$ is a continuous non-decreasing function of ϵ with $\rho^{-}(0) = 0$ because $\lambda(\partial A) = 0$. As a result, there exists functions $C_1(\epsilon) > 0$, $C_2(\epsilon) > 0$ independent of n, such that

$$P\left(\hat{F}_u(A|y) - F_u(A|y) \geq -\eta - \epsilon\right) \geq 1 - C_1(\epsilon) n^{q-1} \exp\left(-C_2(\epsilon) n\right).$$

A similar bound for the other direction can be obtained using the same argument. By the Borel-Cantelli lemma we have,

$$\left|\hat{F}_u(A|y) - F_u(A|y)\right| \leq \eta + \epsilon, \quad \text{a.s.}$$

Note that the above convergence is for any $\epsilon > 0$, therefore we have

$$\left|\hat{F}_u(A|y) - F_u(A|y)\right| \leq \eta, \quad \text{a.s.}$$

Chapter 4

Uniform Convergence of Sequential Monte Carlo Filters

In this chapter we consider state space models which consist of a Markovian state process $\{X_i \in \mathcal{X} = \mathbb{R}^p, i \geq 0\}$ with transition kernel $q(\cdot, \cdot)$:

$$(X_{i+1}|X_i = x) \sim q(x, \cdot), \quad i \geq 0,$$

and an observation sequence $\{Y_i \in \mathcal{Y} = \mathbb{R}^d, i \geq 1\}$, where Y_i's are conditionally independent given X_i's, with likelihood $g(\cdot; \cdot)$:

$$(Y_i|X_i = x) \sim g(\cdot; x), \quad i \geq 1.$$

The joint distribution of (X_i, Y_i) is determined by q, g and ϕ_0, the initial distribution of X_0. Models of this form are also known as hidden Markov model (Künsch, 2001; Cappé et al., 2005). Typical inference tasks in state space models include: 1) estimation of parameters in the dynamics $q(\cdot, \cdot)$ and/or the observation mechanism $g(\cdot; \cdot)$ (Bickel et al., 1998; Olsson

& Rydén, 2008); and 2) calculating the conditional distribution, $\phi_{i|s}$, of state variables X_i given the observations Y_1^s (Liu & Chen, 1998), where $Y_1^s = (Y_1, \ldots, Y_s)^T$. Calculating $\phi_{i|s}$ for $s = i$, $s > i$ and $s < i$ are called filtering, smoothing and predicting, respectively. State space models have found wide application in signal processing, robotics, biology, finance, and geophysics. For a thorough introduction and more related problems on state space models, we refer the reader to Liu (2001); Künsch (2001); Cappé et al. (2005).

This chapter focuses on the filtering problem which has been a classical topic in probability and statistics. The major challenge is that in general the object $\phi_{i|i}$ cannot be characterized by a finite number of parameters except in few special cases. Gordon et al. (1993) proposed a novel approach to approximate the conditional distributions in a non-parametric fashion, which is now known as particle filters (see also sequential Monte Carlo methods Liu & Chen (1998), recursive Monte Carlo filters Künsch (2005)). The basic idea is using a discrete set of sample points to represent the state space, while the distribution is updated at each time step by modifying the weights associated to each sample point, followed by an optional resampling step. Particle filters can easily be implemented in general state space models and can be proved to be consistent. Doucet et al. (2001) provides a thorough introduction to the basic theory and application of particle filters.

Despite the fast development of particle filters in both theory and applications, an important question that remains open is how to quantify the relationship between the approximation error, the sample size and the time interval length. The aim of this chapter is to study the time-uniform convergence of the particle filter approximation. For example, let $\hat{\phi}_{i|i}$ be the approximation of $\phi_{i|i}$. What can we say about the relationship between the

time-uniform approximation error

$$\sup_{1\leq i\leq t} ||\hat{\phi}_{i|i} - \phi_{i|i}||$$

and the sample size n? Here $||\cdot||$ can be any suitable function norm.

Many previous results about time-uniform convergence (Del Moral & Guionnet, 2001; Le Gland & Oudjane, 2004; Künsch, 2005) depends on mixing conditions on the state process of the form:

$$c_- a(\cdot) \leq q(x,\cdot) \leq c_+ a(\cdot), \quad \forall x \in \mathcal{X}, \tag{4.1}$$

for some density function $a(\cdot)$ and positive constants c_-, c_+. Condition (4.1) was originally introduced to show the filter stability (also known as the "forgetting" property). The filter is called stable if

$$||\phi_{i|i}[\phi_0, Y_1^i] - \phi_{i|i}[\phi_0', Y_1^i]|| \to 0$$

as $i \to \infty$ for any pair of initial distributions (ϕ_0, ϕ_0').

However, (4.1) is often too strong to hold when $\mathcal{X}$ is not compact. Many recent works on filter stability have successfully weakened the strong assumption of (4.1), extending the theory to non-compact state spaces (Douc et al., 2009b,a). In the meantime, similar extensions for particle filter theory have also appeared. Heine & Crisan (2008) developed time-uniform convergence in the weak sense for a class of truncated particle filters in autoregressive models with informative observations. van Handel (2009) proved time-average convergence in terms of the bounded Lipschitz norm under conditions that hold for certain autoregressive models.

This chapter extends the uniform convergence theory of particle filters by developing non-asymptotic upper bounds on the uniform approximation error as functions of the sample size n for models in which the function q and g have appropriate tail behavior. To be specific, our methods are particulary applicable to autoregressive models of the form:

$$X_i = a(X_{i-1}) + U_i,$$

$$Y_i = b(X_i) + V_i.$$

In the first case where U_i is heavy-tailed and V_i light-tailed, we show that

$$\sup_{i>0} E\left|\hat{\phi}_{i|i} - \phi_{i|i}\right| \leq c_0 n^{-c},$$

for some $c \in \left(0, \frac{1}{2}\right)$ depending on the tail behavior of U_i, and constant c_0 depending on the model only.

In another case, where a, b are linear and U_i, V_i are Gaussian, we show that

$$E\left(\sup_{1\leq i\leq t}\left|\hat{\phi}_{i|i} - \phi_{i|i}\right|\right) \leq c_1 t^2 \left(c_2 \bigvee \frac{\sqrt{n}}{t^{1+2c\theta}}\right)^{\nu} \exp\left(-\frac{c_3 n}{t^{2+4c\theta}}\right) + 3t^{-\frac{\theta-2}{2}} + c_4 t^{-c_5\theta},$$

for any $\theta > 0$, with constants ν, c, and c_i, $i = 1, \ldots, 5$ depending on the model only.

In the first case, the result is on the time supremum of the expected total variation norm at each time step, therefore it is stronger than the similar results in Heine & Crisan (2008) and van Handel (2009), which did not consider the total variation norm. As for the second case, a similar result is available in Künsch (2005, Theorem 2), which relies largely on (4.1) as well as the compactness of $\mathcal{X}$ and $\mathcal{Y}$.

In Section 4.1 we briefly review some prerequisites about optimal filtering in general state space models, which are useful for the later discussion. In Section 4.2 we develop the theory for models with heavy-tailed state process and light-tailed observation. In Section 4.3 we give an alternative set of conditions, paying special attention to the Gaussian linear model. The proofs are included in Section 4.4.

4.1 Preliminaries on filtering

Recall that the function $p_Z(\cdot)$ denotes the density of random variable Z. The conditional density of X_i given Y_1^s is specially written as $\phi_{i|s}(\cdot)$.

The dependence structure of a state space model can be described as the following diagram:

$$\begin{array}{ccccccccc} \ldots & \longrightarrow & X_{i-1} & \longrightarrow & X_i & \longrightarrow & X_{i+1} & \longrightarrow & \ldots \\ & & \downarrow & & \downarrow & & \downarrow & & \\ \ldots & & Y_{i-1} & & Y_i & & Y_{i+1} & & \ldots \end{array}$$

This graph representation leads to some basic recursive formulas which we state without proof (see Künsch (2001)).

4.1.1 The forward propagation and Monte Carlo approximation

Suppose at time $i \geq 1$ we have obtained $\phi_{i-1|i-1}$, then the one-step forecast distribution of X_i giving Y_1^{i-1} is obtained by applying the Markov transition kernel q on the density function $\phi_{i-1|i-1}$:

$$\phi_{i|i-1}(x_i) = \int \phi_{i-1|i-1}(x_{i-1}) q(x_{i-1}, x_i) dx_{i-1}. \tag{4.2}$$

When the new observation $Y_i = y_i$ is available, the distribution of X_i given $Y_1^i = y_1^i$ is obtained by applying the Bayes rule on the forecast density $\phi_{i|i-1}$ with likelihood function g_i :

$$\phi_{i|i}(x_i) = \frac{\phi_{i|i-1}(x_i)g_i(x_i)}{\int \phi_{i|i-1}(x)g_i(x)dx}, \tag{4.3}$$

where

$$g_i(\cdot) := g(y_i; \cdot).$$

In practice the prediction (4.2) and Bayes update (4.3) do not permit any analytical forms. Particle filters tackle this difficulty using Monte Carlo methods to approximate the conditional distributions. We consider the recursive Monte Carlo (RMC, Künsch (2005)) filter as a generic form of particle filters.

In RMC filters, the integral in (4.2) is substituted by averaging over a random sample:

$$\hat{\phi}_{i|i-1}(x_i) = \frac{1}{n}\sum_{j=1}^{n} q(x_{i-1}^j, x_i), \tag{4.4}$$

where $\{x_{i-1}^j, j = 1, \ldots, n\}$ is an i.i.d sample from $\hat{\phi}_{i-1|i-1}$. The Bayes update step is not much different:

$$\hat{\phi}_{i|i}(x_i) = \frac{\hat{\phi}_{i|i-1}(x_i)g_i(x_i)}{\int \hat{\phi}_{i|i-1}(x)g_i(x)dx}.$$

The recursion starts from $\hat{\phi}_{0|0} = \phi_{0|0} = \phi_0$. See Künsch (2005) for a detailed discussion on implementation details of RMC filters.

4.1.2 The operator notation

In later discussions we will find operator notation for the recursion introduced by Künsch (2005) quite helpful.

Define the Markov transition operator Q:

$$Q\phi(x) = \int \phi(x')q(x', x)dx',$$

for any density function $\phi(\cdot)$. The Bayes operator B is defined as

$$B(\phi, g)(x) = \frac{\phi(x)g(x)}{\int \phi(x')g(x')dx'},$$

for a pair of density $\phi(\cdot)$ and likelihood $g(\cdot)$. As a result, the forward recursion can be written as

$$\phi_{i|i} = B(Q\phi_{i-1|i-1}, g_i) := F_{i-1}\phi_{i-1|i-1}.$$

For RMC approximation, define the random Markov transition kernel $\hat{Q}$ as

$$\hat{Q}\phi(x) = \frac{1}{n}\sum_{j=1}^{n} q(z^j, x),$$

with $\{z^j, j = 1, \ldots, n\}$ an i.i.d sample from $\phi(\cdot)$. Therefore, the RMC recursion becomes

$$\hat{\phi}_{i|i} = B(\hat{Q}\hat{\phi}_{i-1|i-1}, g_i) := \hat{F}_{i-1}\hat{\phi}_{i-1|i-1}.$$

The following equations show how the desired density and its approximation is obtained from the beginning:

$$\phi_{i|i} = F_{i-1}F_{i-2}\ldots F_0\phi_{0|0}, \tag{4.5}$$

$$\hat{\phi}_{i|i} = \hat{F}_{i-1}\hat{F}_{i-2}\ldots \hat{F}_0\phi_{0|0}. \tag{4.6}$$

We wish to control

$$||\hat{\phi}_{i|i} - \phi_{i|i}||_{\text{TV}},$$

where $||\cdot||_{\text{TV}}$ refers to the total variation norm:

$$||f||_{\text{TV}} := \frac{1}{2}\int |f(x)|dx,$$

for any function f. In the rest of this chapter, we simply use $|f|$ as $||f||_{\text{TV}}$ whenever f is a function.

Loosely speaking, the Monte Carlo approximation $\hat{Q}$ of Q will introduce a sampling error of order $O_P\left(n^{-\frac{1}{2}}\right)$ at each time step. Such an error will subsequently be propagated by the Bayes operators $B(\cdot, g_i)$ which is non-linear and might be expanding (Künsch, 2001, Lemma 3.6). Therefore, propagating through multiple Bayes operator might result in an exponential growth of the sampling error. One can bypass this difficulty by looking at a different way of getting $\phi_{i|i}$ from $\phi_{0|0}$.

4.1.3 The backward recursion and the alternative filter representation

Define the backward function $\beta_{i,s}$:

$$\beta_{i,s}(x_i) = \begin{cases} p_{Y_{i+1}^s}\left(y_{i+1}^s \middle| X_i = x_i\right), & i \leq s-1. \\ 1, & i \geq s. \end{cases}$$

It is easy to check that for all $i \leq s-1$, $\beta_{i,s}$ follow a simple but very useful backward recursion:

$$\beta_{i,s} = \int q(x_i, x_{i+1}) g_{i+1}(x_{i+1}) \beta_{i+1,s}(x_{i+1}) dx_{i+1}. \tag{4.7}$$

The backward function can be used to calculate $\phi_{i|s}$ for $s > i$:

$$\phi_{i|s} = B(\phi_{i|i}, \beta_{i,s}).$$

The alternative representation relies on the following well-known lemma:

Lemma 4.1.1. *For any $s \geq 1$, the conditional chain $\{X_i, 0 \leq i \leq s\}$ given $Y_1^s = y_1^s$ is a (possibly non-homogenous) Markov chain, with transition kernel $F_{i|s} : \mathcal{X} \times \mathcal{B}_{\mathcal{X}} \mapsto [0, 1]$:*

$$F_{i|s}(x_i, A) = \frac{\int_A q(x_i, x_{i+1}) g_{i+1}(x_{i+1}) \beta_{i+1,s}(x_{i+1}) dx_{i+1}}{\beta_{i,s}(x_i)},$$

for any $x_i \in \mathcal{X}$ and measurable set A, where $\mathcal{B}_{\mathcal{X}}$ denotes the Borel σ-filed on $\mathcal{X}$.

We refer the reader to Cappé et al. (2005) for a proof of Lemma 4.1.1.

Lemma 4.1.1 suggests an alternative representation of equation (4.5):

$$\phi_{s|s} = F_{s-1|s} \dots F_{0|s} B\left(\phi_{0|0}, \beta_{0,s}\right), \tag{4.8}$$

or more generally for all $i \leq s - 1$ and any density ϕ

$$F_{s-1} \dots F_i \phi = F_{s-1|s} \dots F_{i|s} B(\phi, \beta_{i,s}). \tag{4.9}$$

Equations (4.8) and (4.9) show how to obtain $\phi_{s|s}$ with only a single Bayes operator followed by a sequence of Markov operator. This is the origin of most studies on filter stability, because the Markov operator is contracting under total variation norm: for any Markov operator F and densities f_1, f_2

$$|F f_1 - F f_2| \leq \delta_F |f_1 - f_2|, \tag{4.10}$$

for some $0 \leq \delta(F) \leq 1$. Clearly we have for any pair of Markov kernels F and F',

$$\delta(FF') \leq \delta(F)\delta(F'). \tag{4.11}$$

4.1.4 Controlling error propagation in RMC filters

To use Eq. (4.8) and (4.9) in RMC theory, we first introduce the intermediate Monte Carlo approximations of $\phi_{s|s}$. For $0 \leq i \leq s$, define

$$\phi_{s|s}^{(i)} = F_{s-1} \dots F_i \hat{F}_{i-1} \dots \hat{F}_0 \phi_{0|0} = F_{s-1} \dots F_i \hat{\phi}_{i|i}.$$

That is, the density we would get if we only apply Monte Carlo approximations up to time i. Apparently $\phi_{s|s}^{(s)} = \hat{\phi}_{s|s}$ and $\phi_{s|s}^{(0)} = \phi_{s|s}$.

Consider the following decomposition of the total approximation error for $\phi_{s|s}$:

$$\left| \hat{\phi}_{s|s} - \phi_{s|s} \right| \leq \sum_{i=1}^{s} \left| \phi_{s|s}^{(i)} - \phi_{s|s}^{(i-1)} \right|. \tag{4.12}$$

The ith term in the RHS of (4.12) is the contribution of the sampling error at time i to the total approximation error. Using Eq. (4.9) and the fact that $B(B(\phi, g), h) = B(\phi, gh)$, we have

$$\phi_{s|s}^{(i)} = F_{s-1|s} \dots F_{i|s} B\big(\phi_{i|i-1}^{(i)}, g_i \beta_{i,s}\big) = F_{s-1|s} \dots F_{i|s} \phi_{i|s}^{(i)}, \tag{4.13}$$

$$\phi_{s|s}^{(i-1)} = F_{s-1|s} \dots F_{i|s} B\big(\phi_{i|i-1}^{(i-1)}, g_i \beta_{i,s}\big) = F_{s-1|s} \dots F_{i|s} \phi_{i|s}^{(i-1)}. \tag{4.14}$$

By (4.13), (4.14) and contracting property of Markov kernels, the ith term of (4.12) is bounded by

$$\begin{aligned}
&\left| \phi_{i|s}^{(i)} - \phi_{i|s}^{(i-1)} \right| = \left| B\left(\phi_{i|i-1}^{(i)}, g_i \beta_{i,s}\right) - B\left(\phi_{i|i-1}^{(i-1)}, g_i \beta_{i,s}\right) \right| \\
=&\frac{1}{2} \int \left| \frac{\phi_{i|i-1}^{(i)}(x_i) g_i(x_i) \beta_{i,s}(x_i)}{\int \phi_{i|i-1}^{(i)}(x_i') g_i(x_i') \beta_{i,s}(x_i') dx_i'} - \frac{\phi_{i|i-1}^{(i-1)}(x_i) g_i(x_i) \beta_{i,s}(x_i)}{\int \phi_{i|i-1}^{(i-1)}(x_i') g_i(x_i') \beta_{i,s}(x_i') dx_i'} \right| dx_i \\
\leq&\frac{\int \left| \phi_{i|i-1}^{(i)}(x_i) - \phi_{i|i-1}^{(i-1)}(x_i) \right| g_i(x_i) \beta_{i,s}(x_i) dx_i}{\int \phi_{i|i-1}^{(i-1)}(x_i) g_i(x_i) \beta_{i,s}(x_i) dx_i},
\end{aligned} \tag{4.15}$$

where the last inequality follows from the argument of Lemma 3.6 in Künsch (2001).

Inequality (4.15) is the major building block of the arguments in the rest of this chapter. In the next two sections we present two arguments which lead to different bounds on (4.15), from which the uniform convergence is then established.

4.2 Light-tailed observations and heavy-tailed state processes

4.2.1 General conditions and preliminary results

Our main assumptions consist of three parts:

1. Conditions on $q(\cdot,\cdot)$: basic conditions such as bounded and Lipschitz.

2. Conditions on $g(\cdot;\cdot)$: bounded, Lipschitz and light-tailed (to be specified later).

3. Conditions on the relationship between q and g: $g(y;\cdot)$ has lighter tails than both $q(\cdot,x)$ and $q(x,\cdot)$ for all y and x.

For conditions on $q(\cdot,\cdot)$, we simply require

(**A1**) $q(\cdot,\cdot)$ is bounded and Lipschitz: $\sup_{x,x'} q(x,x') \leq M < \infty$, where M is a positive constant; and $|q(x,x') - q(x,x'')| \leq A|x' - x''|$, for all (x,x',x'') with some constant $A < \infty$.

As for conditions on $g(\cdot;\cdot)$, first look at (4.15) in the relatively simpler case $i = s$. When $i = s$, we have $\beta_{i,s} \equiv 1$, and the numerator of (4.15) becomes

$$\int \left|\phi^{(i)}_{i|i-1}(x_i) - \phi^{(i-1)}_{i|i-1}(x_i)\right| g_i(x_i)dx_i. \tag{4.16}$$

Note that

$$\begin{aligned}&\phi^{(i)}_{i|i-1}(x_i) - \phi^{(i-1)}_{i|i-1}(x_i)\\ =&\frac{1}{n}\sum_{j=1}^{n} q(x^j_{i-1}, x_i) - \int q(x_{i-1}, x_i)\phi^{(i-1)}_{i-1|i-1}(x_{i-1})dx_{i-1},\end{aligned}$$

with $\left(x^j_{i-1},\ j = 1, \ldots, n\right)$ an i.i.d sample from $\phi^{(i-1)}_{i-1|i-1}$. As a result, one would expect the above quantity is of order $O_P(n^{-\frac{1}{2}})$. Then intuitively the integral in (4.16) is of the same order provided that $g_i(x_i)$ is integrable. To give a rigorous argument, one needs to bound $\left|\phi^{(i)}_{i|i-1}(x_i) - \phi^{(i-1)}_{i|i-1}(x_i)\right|$ simultaneously for all x_i, which is a well-studied problem in empirical process theory. Unfortunately, $\sup_{x_i}\left|\phi^{(i)}_{i|i-1}(x_i) - \phi^{(i-1)}_{i|i-1}(x_i)\right|$ is hard to bound because the class of functions $\{q(\cdot, x_i) : x_i \in \mathcal{X}\}$ might be too rich. However, since in the integral (4.16), $\left|\phi^{(i)}_{i|i-1}(x_i) - \phi^{(i-1)}_{i|i-1}(x_i)\right|$ is multiplied by $g_i(x_i)$, the x_i's outside a compact set becomes negligible if g_i decays fast enough. More concretely, we wish to find functions $\tilde{g}_i$ and $\bar{g}_i$, such that

- $g_i \leq \tilde{g}_i\bar{g}_i$.
- $\sup_{x_i}\left|\phi^{(i)}_{i|i-1}(x_i) - \phi^{(i-1)}_{i|i-1}(x_i)\right|\tilde{g}_i$ is small, with high probability.
- $\int \bar{g}_i < \infty$.

The idea is that we want the function g_i to spare some of its light tail to control the sampling error simultaneously for each x_i, while the remaining part of g_i still behaves like a density. Based on this idea, we formally introduce the definition of light-tailed functions:

Definition 4.2.1 (Light-tailed functions)**.** A function $g(x)$ is light-tailed with parameter (A, M, α, γ), if it there exists non-negative functions $\bar{g}(x)$ and $\tilde{g}(x)$ such that

(S1) $0 \leq g(x) \leq \tilde{g}(x)\bar{g}(x)$.

(S2) $\int \bar{g}(x)dx \leq M$, and $\sup_x \bar{g}(x) \leq M$.

(S3) $\tilde{g}(\cdot)$ is bounded and Lipshctiz: $\sup_x \tilde{g}(x) \leq M$; $|\tilde{g}(x) - \tilde{g}(x')| \leq A|x - x'|$ for all (x, x').

(S4) For all $\delta > 0$, there exists a set $K(\delta)$ with $\text{diam}(K(\delta)) \leq \alpha\delta^{-\gamma}$, such that $\tilde{g}(x) \leq \delta$ for all $x \notin K(\delta)$.

These conditions simply require the function $g(\cdot)$ can be decomposed as products of a function $\tilde{g}$ which decays at least polynomially fast and a function $\bar{g}$ which is more or less like a density (possibly scaled). For example, if $g(y; x) = \exp\{-\alpha_1|y - x|^{\alpha_2}\}$ for some positive α_1, α_2, one can choose $\tilde{g}_y(\cdot) = \bar{g}_y(\cdot) = g^{\frac{1}{2}}(y; \cdot)$.

Remark 4.2.2. The constants A, M (and κ, introduced below) appear in multiple statements. They are chosen to be large (or small) enough to satisfy all the statements.

Remark 4.2.3. In the following arguments we let c_j, $j = 1, 2, 3, 4$, be positive constants that do not depend on anything other than the parameters $(A, M, \alpha, \gamma, \kappa)$ which determined by the model and their value might vary among difference displays. Also in the argument about a particular observation sequence y_1^t, the corresponding functions $\tilde{g}_{y_i}$, $\bar{g}_{y_i}$, and sets K_{y_i} in Definition 4.2.1 are written simply as $\tilde{g}_i$, $\bar{g}_i$ and K_i.

Light-tailed functions as defined above are of interest because of the following lemma whose proof is in Section 4.4.1:

Lemma 4.2.4. *For any $q(\cdot, \cdot)$ satisfying (A1), $\tilde{g}(\cdot)$ satisfying (S3), (S4), then for any*

$\epsilon > 0$

$$P\left(\sup_{x_i}\left|\frac{1}{n}\sum_{j=1}^{n} q(x_{i-1}^{j}, x_i) - \int \phi(x_{i-1})q(x_{i-1}, x_i)dx_i\right|\tilde{g}(x_i) \geq \epsilon\right)$$
$$\leq c_1\left(c_2 \bigvee \sqrt{n}\epsilon\right)^{p(1+\gamma)} \exp\left(-c_3 n\epsilon^2\right), \tag{4.17}$$

with positive constants c_1, c_2, c_3 depending only on the model, and γ is the constant defined in (S4) for function $\tilde{g}$, and p is the dimensionality of the state space $\mathcal{X}$.

Our conditions on $g(\cdot;\cdot)$ simply requires it to be light-tailed:

(**A2**) For every $y \in \mathcal{Y}$, the likelihood function $g(y;\cdot)$ is light-tailed with common parameter (A, M, α, γ).

Assumptions (A1) and (A2) enable one to bound the integral in (4.16). However, when $i \neq s$, we still need to control $\beta_{i,s}$, which can be completely unbounded from either above or below. However, because $\beta_{i,s}$ appears in both the denominator and numerator in (4.15), one can expect some cancelation. This is possible if one can separate X_i from the far future. In particular, we consider the following assumption:

(**A3**) For every y, there exists a set $C_y \subseteq \mathcal{X}$, such that for all x, x', y,

$$\min\left\{\frac{\int_{C_y} q(x,x')g(y;x')dx'}{\int q(x,x')g(y;x')dx'}, \frac{\int_{C_y} g(y;x)q(x,x')dx}{\int \bar{g}(y;x)q(x,x')dx}, \frac{1}{M}\int_{C_y} g(y;x)dx\right\} \geq \kappa > 0,$$

where $\bar{g}(y;\cdot)$ and $\tilde{g}(y;\cdot)$ are the corresponding functions defined in (S1-S4) for $g(y;\cdot)$. Here $\kappa \leq \int_{C_0}\phi_0(x_0)dx_0 \leq 1$ is a positive constant independent of (x, x', y). Since there is no Y_0, we conventionally set $y_0 = 0$ and the corresponding set $C_{y_0} = C_0$.

Condition (A3) is useful in showing cancelation of $\beta_{i,s}$ in (4.15), as expected (see Lemma 4.4.6 in Section 4.4.1). It is essentially requiring that the observation provides more information about current state than the previous and future state, which is satisfied when the likelihood has lighter tails than the transition kernel. Also, in the following we use C_i to denote C_{y_i} as defined in (A3).

Now we state a lemma which provides an upper bound on the conditional tail probability of the propagated single step sampling error $\left|\phi_{i|s}^{(i)} - \phi_{i|s}^{(i-1)}\right|$. Its proof can be found in Section 4.4.1:

Lemma 4.2.5. *Assuming (A1-A3), then there exist positive constants c_1, c_2, c_3 depending on the model only , such that for any $\epsilon > 0$ and $1 \leq i \leq s$,*

$$P\left\{\left|\phi_{i|s}^{(i)} - \phi_{i|s}^{(i-1)}\right| > \epsilon\xi_i \middle| Y_1^s\right\} \leq c_1\left(c_2 \bigvee \sqrt{n}\epsilon\right)^{p(1+\gamma)} \exp\left(-c_3 n\epsilon^2\right),$$

where

$$\xi_i = \sup_{C_{i-1}\times C_i} q^{-1}(x_{i-1}, x_i),$$

with p being the dimension of state space $\mathcal{X}$ and γ defined in (A2).

Lemma 4.2.5 immediately suggests the following corollary, which provides an upper bound of the propagated sampling error introduced at time i, at the expected rate $O_P(n^{-\frac{1}{2}})$.

Corollary 4.2.6. *Assuming (A1-A3), then*

$$E\left(\left|\phi_{i|s}^{(i)} - \phi_{i|s}^{(i-1)}\right| \middle| Y_1^s\right) \leq \frac{c_1\xi_i}{\sqrt{n}},$$

with some constant c_1 depending on the model only.

Further refinements By definition, the total variation distance between two density functions is always no more than 1, therefore

$$\left|\phi_{i|s}^{(i)} - \phi_{i|s}^{(i-1)}\right| \leq 1.$$

Then one can try to bound the expected propagated approximation error at time i by

$$E\left|\phi_{i|s}^{(i)} - \phi_{i|s}^{(i-1)}\right| \leq E\left(1 \bigwedge \frac{c_1}{\sqrt{n}}\xi_i\right),$$

where the rate of decay as $n \to \infty$ depends on the tail behaviors of both $q(\cdot,\cdot)$ and $g(\cdot;\cdot)$.

Another useful technique for tighter bounds is taking into account the contracting property of Markov kernels $F_{i|s}$. Note that by (4.10)

$$\left|\phi_{s|s}^{(i)} - \phi_{s|s}^{(i-1)}\right| \leq \delta\left(\prod_{i}^{s-1} F_{i|s}\right)\left|\phi_{i|s}^{(i)} - \phi_{i|s}^{(i-1)}\right|.$$

As a result, if one can show the contraction of the product kernel

$$\delta\left(\prod_{i}^{s-1} F_{i|s}\right)$$

decays exponentially as $s-i$ increases, then it is possible to get a rate of convergence that is uniform over time.

4.2.2 Case study: functional autoregressive model

Consider a non-linear non-Gaussian state space model which has been considered in Douc et al. (2009b) (see also Le Gland & Oudjane (2003)):

$$\begin{aligned} X_i &= a(X_{i-1}) + U_i, \\ Y_i &= b(X_i) + V_i, \end{aligned} \tag{4.18}$$

for $i \geq 1$ with $X_0 \sim \phi_0$. Here (U_i) and (V_i) are two independent sequences of random variables, with probability density p_U and p_V on $\mathcal{X} = \mathcal{Y} = \mathbb{R}^p$. For presentation simplicity we focus on the scalar case $p = 1$. Extensions to $p > 1$ is straightforward.

Condition (A1) and (A2) Now the transition kernel $q(\cdot, \cdot)$ becomes

$$q(x, x') = p_U(x' - a(x)).$$

Therefore condition (A1) is satisfied when p_U is bounded and Lipschitz:

$$||p_U||_\infty \leq M < \infty, \quad |a(x) - a(x')| \leq a_+|x - x'|. \tag{4.19}$$

On the other hand, the likelihood function is

$$g(y; x) = p_V(y - b(x)).$$

Then (A2) holds if $p_V(\cdot)$ satisfies the light-tailed condition (S1-S4) and $b(\cdot)$ is one-to-one differentiable with Jacobian b' bounded and bounded a way from zero:

$$b_- \leq |b'(x)| \leq b_+, \quad \forall x. \tag{4.20}$$

Too see this, it is enough to verify (S1-S4) for $g(y; x)$ with constants (A, M, α, γ) independent of y. Suppose p_V is light-tailed with constants $(A_0, M_0, \alpha_0, \gamma_0)$, let $\bar{p}_V$ and $\tilde{p}_V$ be corresponding functions satisfying (S1-S4). Let $\bar{g}(y; x) = \bar{p}_V(y - b(x))$, $\tilde{g}(y; x) = \tilde{p}_V(y - b(x))$. Then $||\bar{g}(y; x)||_\infty \leq M_0$, $||\tilde{g}(y; x)||_\infty \leq M_0$, $\sup_{x,x'} |\tilde{g}(y; x) - \tilde{g}(y; x')| \leq A_0|x - x'|$ and

$$\begin{aligned} \int \bar{g}(y; x)dx &= \int \bar{p}_V(y - b(x))dx \\ &\leq \int \bar{p}_V(y - z) \left| \frac{db^{-1}(z)}{dz} \right| dz \end{aligned}$$

$$\leq (b_-)^{-1} M_0.$$

For any $\delta > 0$, let $K_V(\delta)$ be the corresponding set for $\tilde{p}_V$ in (S4). Consider $K_y(\delta) = b^{-1}(y - K_V(\delta))$. Then $\tilde{g}(y;x) \leq \delta$ for any $x \in K_y(\delta)$. Meanwhile we have $\operatorname{diam}(K_y(\delta)) \leq (b_-)^{-1} \operatorname{diam}(K_y(\delta)) \leq (b_-)^{-1}\alpha_0\delta^{-\gamma_0}$. Therefore $g(y;x)$ is light-tailed with parameter $\left(A_0, \left(1 \vee (b_-)^{-1}\right) M_0, (b_-)^{-1}\alpha_0, \gamma_0\right)$. In the following discussion, we always assume that p_V satisfies (S1-S4) with corresponding $\tilde{p}_V$ and $\bar{p}_V$.

Condition (A3) Condition (A3) requires $\bar{g}$ to have lighter tails than the transition kernel. Here it would be enough to assume that $\bar{p}_V$ has lighter tails than p_U. Formally, (A3) holds if p_U and p_V satisfy, in addition to (4.19), (4.20),

- For any x, $p_U(x) = p_U(|x|)$, non increasing on $[0, \infty)$. Moreover, for all $w \geq 0$ and $w' \geq 0$,
$$\frac{p_U(w + w')}{p_U(w)p_U(w')} \geq r > 0. \tag{4.21}$$
- For any y, $p_V(y) = p_V(|y|)$, and p_V, $\bar{p}_V$ are non increasing on $[0, \infty)$ and satisfy
$$\int [p_U(cx)]^{-2} \bar{p}_V(x)dx < \infty, \quad \forall c > 0. \tag{4.22}$$
- The initial distribution also has lighter tail than p_U:
$$\int \left[p_U(a_+^{-1}b_-x)\right]^{-1} \phi_0(x)dx < \infty. \tag{4.23}$$

Similar conditions are also considered by Douc et al. (2009b) in the study of filter stability. (4.21) indicates a somewhat heavy tail of p_U, which is satisfied for exponential, logistic and Pareto-type tails (not for Gaussian). The condition of p_U, p_V and $\bar{p}_V$ being

non decreasing on $[0, \infty)$ can be relaxed to have them non increasing on $[L, \infty)$ and strictly positive on $[0, L]$ (Douc et al., 2009b). The case $L = 0$ is qualitatively not special but allows concise presentation.

To verify condition (A3), we first specify the sets C_y:

$$C_y := \{x : |x - b^{-1}(y)| \leq D\},$$

with a constant D to be chosen later with

$$\inf_{[0,D]} \tilde{p}_V > 0, \tag{4.24}$$

which is reasonable for choices of $\tilde{p}_V(x)$ such as $1 \bigwedge |x|^{-\gamma}$ and p_V^{λ} for some $\gamma > 0$ or $0 < \lambda < 1$.

Under these conditions, one can show the following lemma verifying (A3):

Lemma 4.2.7. *Assuming (4.19)-(4.24), then for each y the correspondingly defined C_y's satisfy*

$$\min \left\{ \frac{\int_{C_y} q(x, x')g(y; x')dx'}{\int q(x, x')g(y; x')dx'}, \frac{\int_{C_y} g(y; x)q(x, x')dx}{\int \bar{g}(y; x)q(x, x')dx}, \frac{1}{M} \int_{C_y} g(y; x)dx \right\}$$
$$\geq \kappa > 0,$$

where κ does not depend on (x, x', y).

The proof is in Section 4.4.2.

Now we have verified all the conditions necessary to apply Lemma 4.2.5. We next develop a bound for the term $E\left(1 \bigwedge \frac{C}{\sqrt{n}}\xi_i\right)$

Controlling ξ_i Under the autoregressive model (4.18), using (4.21) repeatedly, we have for $i \geq 2$

$$\begin{aligned}
\xi_i &= \sup_{C_{i-1}\times C_i} p_U^{-1}\left(x_i - a(x_{i-1})\right) \\
&= \sup_{C_{i-1}\times C_i} p_U^{-1}\left(x_i - b^{-1}(Y_i) + b^{-1}(Y_i)\right. \\
&\qquad\qquad \left. -a(b^{-1}(Y_{i-1})) + a(b^{-1}(Y_{i-1})) - a(x_{i-1})\right) \\
&\leq c p_U^{-1}(b^{-1}(Y_i) - a(b^{-1}(Y_{i-1}))) \\
&\leq c p_U^{-1}(b^{-1}(Y_i) - X_i + X_i - a(X_{i-1}) + a(X_{i-1}) - a(b^{-1}(Y_{i-1}))) \\
&\leq c p_U^{-1}(b_-^{-1}V_i) p_U^{-1}(U_i) p_U^{-1}(a_+ b_-^{-1}(V_{i-1})),
\end{aligned} \tag{4.25}$$

note again that the constant c may take different values in different displays.

Therefore, we have for any $\theta > 0$

$$\begin{aligned}
&E\left(1 \bigwedge \frac{c}{\sqrt{n}}\xi_i \middle| V_i, V_{i-1}\right) \\
\leq &\frac{c}{\sqrt{n}} p_U^{-1}(b_-^{-1}V_i) p_U^{-1}(a_+ b_-^{-1}V_{i-1}) \int_{-\theta}^{\theta} p_U^{-1}(u) p_U(u) du + \int_{[-\theta,\theta]^c} p_U(u) du \\
= &\frac{c\theta}{\sqrt{n}} p_U^{-1}(b_-^{-1}V_i) p_U^{-1}(a_+ b_-^{-1}V_{i-1}) + P(|U_1| > \theta).
\end{aligned} \tag{4.26}$$

The case $i = 1$ is similar. Actually (4.26) still holds if note that

$$\begin{aligned}
\xi_1 &= \sup_{C_0\times C_1} p_U^{-1}(x_1 - a(x_0)) \\
&\leq \sup_{C_0\times C_1} p_U^{-1}(b_-^{-1}V_1) p_U^{-1}(U_1) p_U^{-1}(a_+ b_-^{-1}|X_0|).
\end{aligned} \tag{4.27}$$

As a result, we obtain the following bound on the expected one step propagated sampling error.

Proposition 4.2.8. *Under Model (4.18), assuming (4.19)-(4.24), we have*

$$E\left(\left|\phi_{i|s}^{(i)}-\phi_{i|s}^{(i-1)}\right|\right)\leq E\left(1\bigwedge\frac{c_1}{\sqrt{n}}\xi_i\right)\leq\frac{c_1\theta}{\sqrt{n}}+P(|U_1|>\theta),$$

for some constant c_1 depending only on the model.

Making use of Markov kernels $F_{i|s}$ Another lemma from Douc et al. (2009b) will enable one to make use of the contraction of Markov kernels $F_{i|s}$. We state it without proof:

Lemma 4.2.9. *Under Model (4.18), assuming (4.19)-(4.24), then*

$$\delta_{F_{i|s}F_{i-1|s}}\leq\rho<1,\quad\forall\ 1\leq i\leq s-1,$$

for some constant ρ depending only on the model.

From Lemma 4.2.9 and (4.11) we have

$$\delta\left(\prod_i^{s-1}F_{i|s}\right)\leq\rho^{\left\lfloor\frac{s-i}{2}\right\rfloor}\leq\rho^{-\frac{1}{2}}\sqrt{\rho}^{s-i},$$

which implies that

$$\begin{aligned}E\left(\left|\phi_{s|s}^{(s)}-\phi_{s|s}\right|\right)&\leq\sum_{i=1}^{s}E\left|\phi_{s|s}^{(i)}-\phi_{s|s}^{(i-1)}\right|\\&\leq\sum_{i=1}^{s}\rho^{-\frac{1}{2}}\sqrt{\rho}^{s-i}E\left|\phi_{i|s}^{(i)}-\phi_{i|s}^{(i-1)}\right|\\&\leq\left(\frac{c_1\theta}{\sqrt{n}}+P(|U_1|>\theta)\right)\sum_{i=1}^{s}\rho^{-\frac{1}{2}}\sqrt{\rho}^{s-i}\\&\leq\frac{c_1\theta}{\sqrt{n}}+c_2P(|U_1|>\theta).\end{aligned}\tag{4.28}$$

The results obtained so far can be summarized in the following theorem:

Theorem 4.2.10. *Under Model (4.18), assuming (4.19)- (4.24), then there exists constants c_1 and c_2, such that*

$$\sup_{s\geq 0} E\left|\phi_{s|s}^{(s)} - \phi_{s|s}\right| \leq \frac{c_1\theta}{\sqrt{n}} + c_2 P(|U_1| > \theta), \quad \forall\ \theta > 0.$$

Theorem 4.2.10 indicates that the time-uniform expected approximation error is bounded by the sum of two parts: one determined by the sample size and one by the tail behavior of the state noise. In general, one can always choose a θ to optimize the rate of convergence.

Example 4.2.11. If the state noise U_1 has Pareto-type (power law) tails:

$$P(|U_1| > \theta) = O(\theta^{-c}),$$

for some $c > 0$, then one can choose $\theta = n^{\frac{1}{2+2c}}$, and Theorem 4.2.10 yields:

$$\sup_{s\geq 0} E\left|\phi_{s|s}^{(s)} - \phi_{s|s}\right| = O(n^{-\frac{c}{2+2c}}).$$

Example 4.2.12. If the state noise U_1 has exponential tails:

$$P(U_1 > \theta) = O\left(e^{-\theta^c}\right),$$

for some $0 < c \leq 1$, then one can choose $\theta = \left(\frac{1}{2}\log n\right)^{\frac{1}{c}}$ and

$$\sup_{s\geq 0} E\left|\phi_{s|s}^{(s)} - \phi_{s|s}\right| = O\left(\left((\log n)^{\frac{1}{c}} + 1\right) n^{-\frac{1}{2}}\right) = o(n^{-\frac{1}{2}+\delta}),$$

for any $\delta > 0$. That is, when U_1 has exponential tails, the rate of convergence suggested by Theorem 4.2.10 can be arbitrarily close to $n^{-\frac{1}{2}}$.

4.3 Conditions based on normalization: the Gaussian case

4.3.1 When the state process is not heavy-tailed

The conditions developed in Section 4.2 work well for models with heavy-tailed state transition kernel and light-tailed likelihood. However, Condition (A3) does not hold when the transition kernel and likelihood have similar tail behavior. One common example is the Gaussian linear model:

$$X_i = aX_{i-1} + \sigma U_i,$$

$$Y_i = X_i + \tau V_i,$$

with $X_0 \sim N(\mu_0, \sigma_0^2)$ and U_i, V_i independent standard Gaussian random variables. Again, for presentation simplicity we consider the scalar case. Extension to the vector case is straightforward provided that the matrix a is non-singular and U_i, V_i are non-degenerate. In such a model Condition (A3) usually fails because the observation Y_i no longer provides overwhelming control on X_i. However, one can nevertheless have light-tailed likelihood, which is the key to prove Lemma 4.2.4. Therefore similar results might be obtained if one can control $\beta_{i,s}$ and the denominator of (4.15) using alternative methods. Following this direction, we consider a normalized version of function $g_i\beta_{i,s}$:

$$\beta^*_{i,s} = \frac{g_i\beta_{i,s}}{\int g_i\beta_{i,s}},$$

where the integrability of $g_i\beta_{i,s}$ is guaranteed if g_i is light-tailed. The normalization tries to put $\beta^*_{i,s}$ on the same order of magnitude as $\phi_{i|i-1}$, because $|\phi_{i|i-1}| = |\beta^*_{i,s}| = \frac{1}{2}$. Then

we can substitute $g_i\beta_{i,s}$ by its normalized version in (4.15):

$$\left|\phi^{(i)}_{s|s} - \phi^{(i-1)}_{s|s}\right| \leq \frac{\int \left|\phi^{(i)}_{i|i-1}(x_i) - \phi^{(i-1)}_{i|i-1}(x_i)\right| \beta^*_{i,s}(x_i)dx_i}{\int \phi^{(i-1)}_{i|i-1}(x_i)\beta^*_{i,s}(x_i)dx_i}, \tag{4.15'}$$

Consider the following alternative to (A2)

(**A2'**) For all $1 \leq i \leq s \leq t$, $\beta^*_{i,s}$ is light-tailed with common parameters (A, M, α, γ).

Conditions (A2') is generally harder to check than (A2) since it involves the normalization of the whole backward functions $\beta_{i,s}$. However, this is possible in some structured models as illustrated in the following example.

4.3.2 Example: Gaussian autoregressive model

Consider a one dimensional model (multivariate models are qualitatively similar): for all $i \geq 1$

$$X_i = aX_{i-1} + \sigma U_i,$$

$$Y_i = X_i + \tau V_i,$$

where U_i, V_i are independent standard Gaussian variables. Assume $X_0 \sim N(\mu_0, \sigma_0^2)$ and $|a| < 1$. This model is only of theoretical interest since we can use the explicit Kalman filter in this situation.

Forward recursion: computing $\phi_{s|s}$ Suppose $1 \leq i \leq t$, then $\phi_{i|i} = \psi(\cdot; \mu_i, \sigma_i^2)$, a Gaussian density with mean μ_i and variance σ_i^2 following a recursion:

$$\sigma_i^2 = \frac{(a^2\sigma_{i-1}^2 + \sigma^2)\tau^2}{(a^2\sigma_{i-1}^2 + \sigma^2) + \tau^2}, \tag{4.29}$$

$$\mu_i = a\rho_i\mu_{i-1} + (1-\rho_i)y_i, \tag{4.30}$$

where

$$\rho_i = \frac{\tau^2}{a^2\sigma_{i-1}^2 + \sigma^2 + \tau^2}.$$

Clearly

$$\frac{\tau^2\sigma^2}{\tau^2+\sigma^2} = \sigma_-^2 \leq \sigma_i^2 \leq \sigma_+^2 = \sigma^2, \tag{4.31}$$

$$\frac{\tau^2}{a^2\sigma^2+\sigma^2+\tau^2} = \rho_- \leq \rho_i \leq \rho_+ = \frac{\tau^2(\sigma^2+\tau^2)}{a\tau^2\sigma^2+(\tau^2+\sigma^2)} \tag{4.32}$$

for all $i \geq 1$.

Backward recursion: computing $\beta_{i,s}^*$ In a Gaussian linear model the function $\beta_{i,s}$ is also proportional to a Gaussian density. For $1 \leq i \leq s$,

$$\beta_{i,s}(x) \propto \psi(x; \mu_{i,s}, \sigma_{i,s}^2),$$

where for all $1 \leq i \leq s-1$,

$$\sigma_{i,s}^2 = \frac{\left(\sigma_{i+1,s}^2 + \sigma^2\right)\tau^2}{\sigma_{i+1,s}^2 + \sigma^2 + a^2\tau^2}, \tag{4.33}$$

$$\mu_{i,s} = a^{-1}\rho_{i,s}\mu_{i+1,s} + (1-\rho_{i,s})y_i, \tag{4.34}$$

with $\mu_{s,s} = y_s$, $\sigma_{s,s}^2 = \tau^2$, and

$$\rho_{i,s} = \frac{a^2\tau^2}{\sigma_{i+1,s}^2 + \sigma^2 + a^2\tau^2}.$$

Similarly we have

$$\frac{\tau^2\sigma^2}{a^2\tau^2+\sigma^2} = (\sigma_-^*)^2 \leq \sigma_{i,s}^2 \leq (\sigma_+^*)^2 = \tau^2, \tag{4.35}$$

$$\frac{a^2\tau^2}{\tau^2+\sigma^2+a^2\tau^2} = \rho_-^* \le \rho_{i,s} \le \rho_+^* = \frac{a^2\tau^2(a^2\tau^2+\sigma^2)}{\tau^2\sigma^2+(a^2\tau^2+\sigma^2)^2}. \tag{4.36}$$

The fact that $\beta_{i,s}^* \propto \beta_{i,s}$ and $\int \beta_{i,s}^* = 1$ indicates $\beta_{i,s}^*(x) = \psi(x; \mu_{i,s}, \sigma_{i,s}^2)$. Therefore (A2') holds easily by taking $\bar{\beta}_{i,s}^* = \tilde{\beta}_{i,s}^* = \left(\beta_{i,s}^*\right)^{\frac{1}{2}}$, with constants (A, M, α, γ) depending only on (a, σ, τ), since $\sigma_{i,s}$ is bounded from up and below uniformly for all i, s. Note also that $\tilde{\beta}_{i,s}$ has Gaussian tail, so condition (S4) holds for any positive constant γ, here we just choose any arbitrary $\gamma = \gamma_0$, with the corresponding constant $\alpha = \alpha_0$.

Lower bound of the denominator From the forward recursion we know that

$$\phi_{i|i-1}(x_i) = \psi\left(x_i; \mu_{i|i-1}, \sigma_{i|i-1}^2\right),$$

with $\mu_{i|i-1} = a\mu_{i-1}$ and $\sigma_{i|i-1}a^2 = \sigma_{i-1}^2 + \sigma^2$. Then

$$\begin{aligned}
\int \phi_{i|i-1}\beta_{i,s}^* &= \frac{1}{\sqrt{2\pi(\sigma_{i|i-1}^2+\sigma_{i,s}^2)}} \exp\left(-\frac{1}{2}\frac{(\mu_{i|i-1}-\mu_{i,s})^2}{\sigma_{i|i-1}^2+\sigma_{i,s}^2}\right) \\
&\ge \frac{1}{\sqrt{2\pi(a^2\sigma_+^2+\sigma^2+(\sigma_+^*)^2)}} \exp\left(\frac{1}{2}\frac{(\mu_{i|i-1}-\mu_{i,s})^2}{a\sigma_-^2+\sigma^2+(\sigma_-^*)^2}\right) \\
&= c_1 \exp\left(-c_2(\mu_{i|i-1}-\mu_{i,s})^2\right).
\end{aligned} \tag{4.37}$$

The next step is to develop an upper bound of $|a\mu_{i-1} - \mu_{i|s}|$ uniformly for all i, s. Let $\xi_t = \max\{|U_i|, |V_i|, 0 \le i \le t\}$.

Lemma 4.3.1. *Assuming $0 < |a| < 1$, then there exists constant c_1 depending on the model only, such that for all $0 \le i \le s \le t$,*

$$\max\left\{|\mu_i|, |\mu_{i,s}|, 0 \le i \le s\right\} \le c_1\xi_t.$$

The following lemma shows that $\xi_t = O_P(\sqrt{\log t})$.

Lemma 4.3.2. *Define event*

$$E^* = \{\xi_t \leq \sqrt{\theta \log t}\},$$

for some $\theta > 0$. *Then for all* $t \geq 2$,

$$P(E^*) \geq 1 - 3t^{-\frac{\theta-2}{2}}. \tag{4.38}$$

Then we have our main result about the Gaussian linear model:

Theorem 4.3.3. *Under the Gaussian linear model, assuming* $0 < |a| < 1$, *then, there exists positive constants* c *and* $c_0, \ldots, c_3$, *depending on the model only, such that for every* $0 < \epsilon \leq c_0 t^{-c\theta}$,

$$P\left(\sup_{1\leq s\leq t} \left|\phi_{s|s}^{(s)} - \phi_{s|s}\right| \geq \epsilon\right) \leq c_1 t^2 \left(c_2 \bigvee \frac{\sqrt{n}\epsilon}{t^{1+c\theta}}\right)^{p(1+\gamma)} \exp\left(-\frac{c_3 n\epsilon^2}{t^{2+2c\theta}}\right) + 3t^{-\frac{\theta-2}{2}}, \tag{4.39}$$

where θ *is a free parameter defined as in Lemma 4.3.2, and* γ *is a constant defined as in (S4) which depends only on the model.*

The proofs of Lemma 4.3.1, 4.3.2 and Theorem 4.3.3 are in Section 4.4.3.

In Theorem 4.3.3 letting $\epsilon = c_0 t^{-c\theta}$, and using the fact that $\sup_s \left|\phi_{s|s}^{(s)} - \phi_{s|s}\right| \leq 1$, one immediately get an upper bound of the expected uniform approximation error:

Corollary 4.3.4. *Under the same conditions of Theorem 4.3.3, there exits positive constants* c, $c_1, \ldots, c_4$ *depending on the model only, such that*

$$E\left(\sup_{1\leq s\leq t} \left|\phi_{s|s}^{(s)} - \phi_{s|s}\right|\right) \leq c_1 t^2 \left(c_2 \bigvee \frac{\sqrt{n}}{t^{1+2c\theta}}\right)^{p(1+\gamma)} \exp\left(-\frac{c_3 n}{t^{2+4c\theta}}\right) + 3t^{-\frac{\theta-2}{2}} + c_4 t^{-c\theta}. \tag{4.40}$$

In Theorem 4.3.3 and Corollary 4.3.4 the approximation bound is no longer uniform in t. However, it is about the uniform approximation error up to time t instead of the expected approximation error at each time, as we considered in the previous section. These results implies that it is enough to have the sample size n to increase polynomially in t in order to control the uniform approximation error. In fact, if t is large, it is enough to have $n_t = t^{2+4c\theta+\eta}$, for any $\eta > 0$.

4.4 Proofs

4.4.1 Proofs of Section 4.2.1

We take three steps to prove Lemma 4.2.5.

Controlling the numerator

We first prove Lemma 4.2.4, which provides partial upper bound for the numerator of (4.15).

Before proving Lemma 4.2.4, we introduce some useful concept in empirical processes (van der Vaart & Wellner, 1996, Ch. 2). Here let $\mathcal{F}$ be any family of functions and $||\cdot||$ be any function norm.

Definition 4.4.1 (Bracketing numbers)**.** Given two functions f_l and f_u, the bracket $[f_l, f_u]$ is the set of all functions f with $f_l \leq f \leq f_u$. An δ-bracket is a bracket $[f_l, f_u]$ with $||f_u - f_l|| \leq \delta$. The bracketing number $N_{[]}(\delta, \mathcal{F}, ||\cdot||)$ is the minimum number of δ-brackets needed to cover $\mathcal{F}$.

A related notion is the covering number:

Definition 4.4.2 (Covering numbers)**.** The covering number $N(\delta, \mathcal{F}, ||\cdot||)$ is the minimal number of balls $\{g : ||g - f|| < \delta\}$ of radius δ needed to cover $\mathcal{F}$.

The proof of Lemma 4.2.4 requires an upper bound of the bracketing number of the function class $\mathcal{F} \equiv \{f(x, x') : x \mapsto q(\cdot, x')\tilde{g}_i(x')|x' \in \mathcal{X}\}$, which is in turn bounded by the corresponding covering number. A useful result relating the covering number and bracketing number is the following:

Lemma 4.4.3 (Theorem 2.7.11, van der Vaart & Wellner (1996))**.** *Let $\mathcal{F}$ be a class of functions $x \mapsto f(x, x')$ indexed by a parameter $x' \in K$. Suppose that*

$$|f(x, x') - f(x, x'')| \leq d(x', x'')F(x), \quad \forall x, x', x'',$$

for some metric d on the index set and function F. Then for any norm $||\cdot||$,

$$N_{[]}(\delta, \mathcal{F}, ||\cdot||) \leq N\left(\frac{\delta}{2||F||}, K, d\right).$$

Proof of Lemma 4.2.4. Let $f(x, x') : x \mapsto q(x, x')\tilde{g}(x')$ and $\mathcal{F} = \{f(\cdot, x') : x' \in \mathcal{X}\}$. We want to control the bracketing number $N_{[]}(\delta, \mathcal{F}, ||\cdot||_\infty)$.

First note that $f(x, x')$ is Lipschitz in x' under Conditions (A1) and (A2):

$$\begin{aligned}
&|f(x, x') - f(x, x'')| \\
\leq &|\tilde{g}(x')q(x, x') - \tilde{g}(x')q(x, x'')| + |\tilde{g}(x')q(x, x'') - \tilde{g}(x'')q(x, x'')| \\
\leq &2AM|x' - x''|. \qquad (4.41)
\end{aligned}$$

For any $\delta > 0$, let $K(\delta/M)$ as defined in (S4) with respect to $\tilde{g}$, we have for all $x' \in K(\delta/M)^c$,

$$0 \leq f(x, x') \leq q(x, x')\delta/M \leq \delta.$$

Let $\mathcal{F}_\delta = \{f(\cdot, x') : x' \in K(\delta/M)\}$, then the above inequality indicates:

$$\begin{aligned} N_{[]}(\delta, \mathcal{F}, ||\cdot||_\infty) &\leq N_{[]}(\delta, \mathcal{F}_\delta, ||\cdot||_\infty) + N_{[]}(\delta, \mathcal{F}\backslash\mathcal{F}_\delta, ||\cdot||_\infty) \\ &\leq N_{[]}(\delta, \mathcal{F}_\delta, ||\cdot||_\infty) + 1. \end{aligned} \tag{4.42}$$

On the other hand, let $||\cdot||_\mathcal{X}$ be the Euclidean norm on the state space, then by (S4),

$$\begin{aligned} N\left(\frac{\delta}{4AM}, K\left(\frac{\delta}{M}\right), ||\cdot||_\mathcal{X}\right) &\leq \left(\frac{4AM\text{diam}\left(K(\delta/M)\right)}{\delta} + 1\right)^p \\ &\leq \left(\frac{4\alpha AM^{1+\gamma}\delta^{-\gamma}}{\delta} + 1\right)^p. \end{aligned} \tag{4.43}$$

Therefore, using Lemma 4.4.3, (4.42), and (4.43) we have

$$\begin{aligned} N_{[]}\left(\delta, \mathcal{F}, ||\cdot||_\infty\right) \leq & N_{[]}(\delta, \mathcal{F}_\delta, ||\cdot||_\infty) + 1 \\ \leq & N\left(\frac{\delta}{4AM}, K_i\left(\frac{\delta}{M}\right), ||\cdot||_\mathcal{X}\right) + 1 \\ \leq & \left(\frac{4\alpha AM^{1+\gamma}\delta^{-\gamma}}{\delta} + 1\right)^p + 1 \\ \leq & 2\left(\frac{(4\alpha AM^{1+\gamma} + M^{2(1+\gamma)})\delta^{-\gamma}}{\delta}\right)^p \\ = & \left(\frac{c(A, M, \alpha, \gamma)}{\delta}\right)^{p(1+\gamma)}, \end{aligned} \tag{4.44}$$

where the last inequality uses the fact that $\delta \leq M^2$ since $K(M)$ can be chosen as empty if $\delta > M^2$.

By a classical result in empirical processes (Talagrand, 1994, Theorem 1.1), there exists positive constants c_1, c_2 such that for any $\epsilon > 0$,

$$\begin{aligned} & P\left(\sup_{x_i}\left|\frac{1}{n}\sum_{j=1}^{n} q(x_{i-1}^{j}, x_i) - \int \phi(x_{i-1})q(x_{i-1}, x_i)\right|\tilde{g}(x_i) > \epsilon\right) \\ \leq & c_1\left(c_2 \bigvee \sqrt{n}\epsilon\right)^{p(1+\gamma)} \exp\left(-\frac{2n\epsilon^2}{M^4}\right). \end{aligned}$$

□

Note that the result of Lemma 4.2.4 is uniform with respect to the underlying measure ϕ. Substitute ϕ by $\phi_{i-1|i-1}^{(i-1)}$ we have

Corollary 4.4.4. *Under (A1) and (A2), there exists constants c_1, c_2 depending on the model only, such that for any $\epsilon > 0$, $1 \leq i \leq s$,*

$$\begin{aligned} & P\left(\sup_{x_i}\left|\phi_{i|i-1}^{(i)}(x_i) - \phi_{i|i-1}(x_i)\right|\tilde{g}(x_i) > \epsilon\right) \\ & \leq c_1\left(c_2 \bigvee \sqrt{n}\epsilon\right)^{p(1+\gamma)} \exp\left(-\frac{2n\epsilon^2}{M^4}\right). \end{aligned}$$

Controlling the denominator

We have the following lemma providing a partial lower bound for the denominator of (4.15):

Lemma 4.4.5. *Under (A1-A3), we have, for any $1 \leq i \leq s-1$, conditioning on Y_1^s,*

$$\inf_{x_i \in C_i} \phi_{i|i-1}^{(i-1)}(x_i) \geq \kappa \xi_i^{-1}.$$

proof of Lemma 4.4.5. for $i \geq 2$, by assumption (A3), for all $x_i \in C_i$

$$\phi_{i|i-1}^{(i-1)}(x_i)$$

$$
\begin{aligned}
&\geq \int \phi_{i-1|i-1}^{(i-1)}(x_{i-1}) q(x_{i-1}, x_i) dx_{i-1} \\
&\geq \int_{C_{i-1}} \phi_{i-1|i-1}^{(i-1)}(x_{i-1}) q(x_{i-1}, x_i) dx_{i-1} \\
&\geq \xi_i^{-1} \int_{C_{i-1}} \phi_{i-1|i-1}^{(i-1)}(x_{i-1}) dx_{i-1} \\
&= \xi_i^{-1} \frac{\int_{C_{i-1}} \int \phi_{i-2|i-2}^{(i-1)}(x_{i-2}) q(x_{i-2}, x_{i-1}) dx_{i-2} g_{i-1}(x_{i-1}) dx_{i-1}}{\int \int \phi_{i-2|i-2}^{(i-1)}(x_{i-2}) q(x_{i-2}, x_{i-1}) dx_{i-2} g_{i-1}(x_{i-1}) dx_{i-1}} \\
&= \xi_i^{-1} \frac{\int \int_{C_{i-1}} q(x_{i-2}, x_{i-1}) g_{i-1}(x_{i-1}) dx_{i-1} \phi_{i-2|i-2}^{(i-1)}(x_{i-2}) dx_{i-2}}{\int \int q(x_{i-2}, x_{i-1}) g_{i-1}(x_{i-1}) dx_{i-1} \phi_{i-2|i-2}^{(i-1)}(x_{i-2}) dx_{i-2}} \\
&\geq \xi_i^{-1} \kappa \frac{\int \int q(x_{i-2}, x_{i-1}) g_{i-1}(x_{i-1}) dx_{i-1} \phi_{i-2|i-2}^{(i-1)}(x_{i-2}) dx_{i-2}}{\int \int q(x_{i-2}, x_{i-1}) g_{i-1}(x_{i-1}) dx_{i-1} \phi_{i-2|i-2}^{(i-1)}(x_{i-2}) dx_{i-2}} \\
&\geq \xi_i^{-1} \kappa.
\end{aligned}
$$

For $i = 1$, we have according to (A3), for all $x_1 \in C_1$

$$
\begin{aligned}
&\phi_{1|0}(x_1) \\
&= \int \phi_0(x_0) q(x_0, x_1) dx_0 \\
&\geq \kappa \int \int_{C_0} \phi_0(x_0) q(x_0, x_1) dx_0 \\
&\geq \xi_1 \int_{C_0} \phi_{0|0}(x_0) dx_0 \\
&\geq \xi_1 \kappa.
\end{aligned}
$$

□

Controlling the effect of $\beta_{i,s}$

Lemma 4.4.6. *Under (A2-A3), we have for all $1 \leq i \leq s$, and all y_1^s,*

$$
\frac{\int_{C_i} g_i(x_i) \beta_{i,s}(x_i) dx_i}{\int \bar{g}_i(x_i) \beta_{i,s}(x_i) dx_i} \geq \kappa^2. \tag{4.45}
$$

proof of Lemma 4.4.6. When $i = s$, we have $\beta_{i,s} \equiv 1$, then the result follows easily from (A3) and the fact that $\kappa \leq 1$.

When $i \leq s-1$, by (A3),

$$\begin{aligned}
&\int_{C_i} g_i(x_i)\beta_{i,s}(x_i)dx_i \\
\geq &\inf_{C_i} \tilde{g}_i \int_{C_i} \bar{g}_i(x_i)\beta_{i,s}(x_i)dx_i \\
\geq &\kappa \int\int_{C_i} \bar{g}_i(x_i)q(x_i, x_{i+1})dx_i g_{i+1}(x_{i+1})\beta_{i+1,s}(x_{i+1})dx_{i+1} \\
\geq &\kappa^2 \int\int \bar{g}_i(x_i)q(x_i, x_{i+1})dx_i g_{i+1}(x_{i+1})\beta_{i+1,s}(x_{i+1})dx_{i+1} \\
= &\kappa^2 \int \bar{g}_i(x_i)\beta_{i,s}(x_i)dx_i.
\end{aligned}$$

□

Putting things together

Now we can put Corollary 4.4.4, Lemma 4.4.5 and 4.4.6 together to control the RHS of (4.15), thereby proving Lemma 4.2.5.

Proof of Lemma 4.2.5. From Equation (4.15) and Lemma 4.4.5 and 4.4.6, we have

$$\begin{aligned}
&\left|\phi_{i|s}^{(i)} - \phi_{i|s}^{(i-1)}\right| \\
\leq &\frac{\int \left|\phi_{i|i-1}^{(i)}(x_i) - \phi_{i|i-1}^{(i-1)}(x_i)\right| g_i(x_i)\beta_{i,s}(x_i)dx_i}{\int \phi_{i|i-1}^{(i-1)}(x_i)g_i(x_i)\beta_{i,s}(x_i)dx_i} \\
\leq &\frac{\sup_{x_i}\left[\left|\phi_{i|i-1}^{(i)}(x_i) - \phi_{i|i-1}^{(i-1)}(x_i)\right| \tilde{g}_i(x_i)\right] \int \bar{g}_i(x_i)\beta_{i,s}(x_i)dx_i}{\inf_{C_i} \phi_{i|i-1}^{(i-1)}(x_i) \int_{C_i} g_i(x_i)\beta_{i,s}(x_i)dx_i} \\
\leq &\frac{\xi_i}{\kappa^3} \sup_{x_i}\left[\left|\phi_{i|i-1}^{(i)}(x_i) - \phi_{i|i-1}^{(i-1)}(x_i)\right| \tilde{g}_i(x_i)\right]. \qquad (4.46)
\end{aligned}$$

Therefore,

$$\begin{aligned} & P\left(\left|\phi_{i|s}^{(i)}-\phi_{i|s}^{(i-1)}\right| \geq \epsilon\xi_i \middle| Y_1^s\right) \\ \leq & P\left(\sup_{x_i}\left[\left|\phi_{i|i-1}^{(i)}(x_i)-\phi_{i|i-1}^{(i-1)}(x_i)\right| \tilde{g}_i(x_i)\right] \geq \kappa^3\epsilon\right) \\ \leq & c_1\left(c_2 \bigvee \sqrt{n}\epsilon\right)^{p(1+\gamma)} \exp\left(-c_3 n\epsilon^2\right), \end{aligned} \tag{4.47}$$

for constants c_1, c_2 and c_3 depending only on the model. □

4.4.2 Proofs of Section 4.2.2

The proofs follow largely from Douc et al. (2009b, Lemma 10,11 and 12). First we have the following lemma:

Lemma 4.4.7 (Lemma 10 of Douc et al. (2009b))**.** *Assume diam$(C) < \infty$. Then for all $x \in C$ and $x' \in \mathcal{X}$,*

$$\rho(C)h_C(x') \leq q(x,x') \leq \rho^{-1}(C)h_C(x'), \tag{4.48}$$

with

$$\rho(C) = rp_U(\mathrm{diam}(C)) \wedge \inf_{|u|\leq \mathrm{diam}(C)} p_U \wedge \left(\sup_{|u|\leq \mathrm{diam}(C)} p_U\right)^{-1},$$

$$h_C(x') = \mathbb{1}\left(x' \in a(C)\right) + \mathbb{1}\left(x' \notin a(C)\right) p_U(|x' - a(z_0)|),$$

where r is defined in (4.21) and z_0 is an arbitrary element of C. In addition, for all $x \in \mathcal{X}$ and $x' \in C$,

$$\nu(C)k_C(x) \leq q(x,x'), \tag{4.49}$$

with

$$\nu(C) = \inf_{|u|\leq \mathrm{diam}(C)} p_U,$$

$$k_C(x) = \mathbb{1}\left(a(x) \in C\right) + r\mathbb{1}\left(a(x) \notin C\right) p_U(|z' - a(x)|),$$

where z' is an arbitrary element in C.

Proof of Lemma 4.2.7. Choose $C_y = \{x : |x - b^{-1}(y)| \leq D\}$, for some $D > 0$.

We first show

$$\inf_y \frac{\int_{C_y} g(y;x)dx}{\int g(y;x)dx} > 0.$$

In fact

$$\int_{C_y^c} p_V(y - b(x))dx \leq \int_{C_y^c} p_V\left(b_-|b^{-1}(y) - x|\right)dx \leq \int_{|x|\geq D} p_V(b_-|x|)dx,$$

which is independent of y. Also note that $\int p_V(y - b(x))dx$ is bounded from below uniformly in y by change of variables and the assumption that the Jacobian of b is bounded. Now we can choose D large enough so that $\int_{|x|\geq D} p_V(b_-|x|)dx < \inf_y \int p_V(y - b(x))dx$. Then we have

$$\inf_y \int_{C_y} g(y;x)dx > 0.$$

Then we are going to show

$$\inf_{y,x} \frac{\int_{C_y} q(x,x')g(y;x')dx'}{\int q(x,x')g(y;x')dx'} > 0, \tag{4.50}$$

which is equivalent to

$$\inf_{y,x} \frac{\int_{C_y} q(x,x')g(y;x')dx'}{\int_{C_y^c} q(x,x')g(y;x')dx'} > 0.$$

Note that in Lemma 4.4.7 the constants $\rho(C_y)$ and $\nu(C_y)$ depends on C_y only through its diameter. That is, $\rho(C_y)$ and $\nu(C_y)$ depends only on D which is independent of y. As a result, in the following arguments we will drop the dependence on y when using these notations.

Consider two cases:

1. $a(x) \in C_y$.

   In this case $k_{C_y}(x) \equiv 1$ as define in Lemma 4.4.7. We have

   $$\frac{\int_{C_y} q(x,x')g(y;x')dx'}{\int q(x,x')g(y;x')dx'} \geq \frac{\nu}{M}\frac{\int_{C_y} g(y;x')dx'}{\int g(y;x')dx'} \geq \frac{\nu}{M}\inf_y \frac{\int_{C_y} g(y;x')dx'}{\int g(y;x')dx'} > 0, \quad (4.51)$$

   where we used the fact $q(x,x') \leq ||p_U||_\infty \leq M$.

2. $a(x) \notin C_y$.

   In this case $k_{C_y}(x) = rp_U(|b^{-1}(y) - a(x)|)$ as defined in Lemma 4.4.7, where z' is chosen as $b^{-1}(y)$. In (4.21) let $w = x' - a(x)$, $w' = b^{-1}(y) - x'$, then by monotonicity of p_U (4.21) we have

   $$p_U(|w + w'|) \geq p_U(|w| + |w'|) \geq rp_U(|w|)p_U(|w'|),$$

   which implies

   $$\frac{p_U(|x' - a(x)|)}{p_U(|b^{-1}(y) - a(x)|)} \leq r^{-1}p_U^{-1}(|b^{-1}(y) - x'|).$$

   Therefore, using (4.22)

   $$\begin{aligned}
   \frac{\int_{C_y} q(x,x')g(y;x')dx'}{\int_{C_y^c} q(x,x')g(y;x')dx'} &\geq \frac{r\nu p_U(|b^{-1}(y) - a(x)|)\int_{C_y} g(y;x')dx'}{\int_{C_y^c} p_U(|x' - a(x)|)p_V(y - b(x'))dx'} \\
   &\geq \frac{r^2\nu \int_{C_y} g(y;x')dx'}{\int_{C_y^c} p_U^{-1}(|b^{-1}(y) - x'|)p_V(y - b(x'))dx'} \\
   &\geq \frac{r^2\nu \int_{C_y} g(y;x')dx'}{\int_{C_y^c} p_U^{-1}(|b^{-1}(y) - x'|)p_V(b_-|b^{-1}(y) - x'|)dx'} \\
   &\geq \frac{r^2\nu \int_{C_y} g(y;x')dx'}{\int_{|z|\geq D} p_U^{-1}(|z|)p_V(b_-|z|)dz} \\
   &\geq \frac{r^2\nu \inf_y \int_{C_y} g(y;x')dx'}{\int_{|z|\geq D} p_U^{-1}(|z|)p_V(b_-|z|)dz} > 0, \quad (4.52)
   \end{aligned}$$

where the last inequality is based on (4.22) and the fact that $p_V = \tilde{p}_V \bar{p}_V \leq M\bar{p}_V$. Note that the bounds of both (4.51) and (4.52) are independent of y and x. As a result (4.50) is true.

It remains to show

$$\inf_{y,x'} \frac{\int_{C_y} g(y;x)q(x,x')dx}{\int \bar{g}(y;x)q(x,x')dx} > 0. \tag{4.53}$$

The argument is very similar to those of (4.50). Again, consider two cases:

1. $x' \in a(C_y)$.

   In this case $h_{C_y}(x') \equiv 1$, and we have

   $$\frac{\int_{C_y} g(y;x)q(x,x')dx}{\int g(y;x)q(x,x')dx} \geq \frac{\rho \int_{C_y} g(y;x)dx}{M \int \bar{g}(y;x)dx} \geq \frac{\rho}{M^2} \inf_y \int_{C_y} g(y;x)dx > 0. \tag{4.54}$$

2. $x' \notin a(C_y)$.

   In this case $h_{C_y}(x') = p_U(|x' - a(b^{-1}(y))|)$ choosing $z_0 = b^{-1}(y)$ in Lemma 4.4.7, and

   $$\frac{\int_{C_y} g(y;x)q(x,x')dx}{\int \bar{g}(y;x)q(x,x')dx} \geq \frac{\rho p_U(|x' - a(b^{-1}(y))|) \int_{C_y} g(y;x)dx}{\int_{C_y^c} \bar{p}_V(y - b(x))p_U(x' - a(x))dx}. \tag{4.55}$$

   It suffices to show

   $$\int_{C_y^c} p_U^{-1}(|x' - a(b^{-1}(y))|)p_U(x' - a(x))\bar{p}_V(y - b(x))dx$$

   is bounded uniformly for all y and x'.

   Again, let $w = x' - a(x)$, $w' = a(x) - a(b^{-1}(y))$. Then $|w + w'| = |x' - a(b^{-1}(y))| > L$. Therefore

   $$p_U(|w + w'|) \geq p_U(|w| + |w'|) \geq r p_U(|w|)p_U(|w'|),$$

which indicates

$$p_U^{-1}(|x' - a(b^{-1}(y))|)p_U(x' - a(x)) \leq r^{-1}p_U^{-1}(|a(x) - a(b^{-1}(y))|).$$

Also note that for all $z, z' \in \mathcal{X}$,

$$p_U^{-1}(|a(z) - a(z')|) \leq p_U^{-1}(a_+|z - z'|).$$

As a result,

$$\begin{aligned}
&\int_{C_y^c} p_U^{-1}(|x' - a(b^{-1}(y))|)p_U(x' - a(x))\bar{p}_V(y - b(x))dx \qquad (4.56)\\
\leq &\int_{C_y^c} r^{-1}p_U^{-1}(|a(x) - a(b^{-1}(y))|)\bar{p}_V(y - b(x))dx\\
\leq &r^{-1}\int_{C_y^c} p_U^{-1}(a_+|z - z'|)\bar{p}_V(y - b(x))dx\\
\leq &r^{-1}\int_{|x|>D} p_U(a_+|x|)\bar{p}_V(b_-x)dx\\
< &\infty,
\end{aligned}$$

where the last inequality uses (4.22). Therefore (4.53) is true because the bounds in (4.54) and (4.56) do not depend on y or x'.

□

4.4.3 Proofs of Section 4.3

Proof of Lemma 4.3.1. We will simply show

$$\sup_{1\leq i\leq t} |\mu_{i-1}| \leq c_1\xi_t, \qquad \sup_{1\leq i\leq s\leq t} |\mu_{i,s}| \leq c_2\xi_t,$$

for some constant c_1, c_2. The proof for $|\mu_{i,s}|$ is essentially the same. For the first inequality, the recursive formula (4.30) implies

$$\begin{aligned} |\mu_i| &= |a\rho_i\mu_{i-1} + (1-\rho_i)Y_i| \\ &= \left|\sum_{j=0}^{i-1} a^j \rho_i \dots \rho_{i-j+1}(1-\rho_{i-j})Y_{i-j}\right| \\ &= \left[\sum_{j=0}^{i-1}(a\rho_+)^j\right](1-\rho_-)\max_{1\le j\le i}|Y_j| \\ &= \frac{1-\rho_-}{1-a\rho_+}\max_{1\le j\le i}|Y_j|, \end{aligned} \tag{4.57}$$

where ρ_-, ρ_+ are defined in (4.32). Therefore it suffices to show $\max_{1\le i\le t}|Y_i| \le c_1\xi_t$ for some $c_1 > 0$:

$$\begin{aligned} |Y_i| &= |X_i + \tau V_i| = |aX_{i-1} + \sigma U_i + V_i| \\ &= \left|\tau V_i + \sum_{j=0}^{i} a^{i-j}\sigma U_j\right| \\ &\le \left(\tau + \frac{\sigma}{1-a}\right)\xi_t. \end{aligned} \tag{4.58}$$

On the other hand, let $h = \frac{\sigma^2}{\tau^2}$, and $h_{i,s} = \frac{\sigma_{i,s}^2}{\tau^2}$, then (4.33) becomes

$$h_{i,s} = \frac{h_{i+1,s} + h}{h_{i+1,s} + h + a^2}.$$

Let $h_0 \in (0,1)$ be the fixed point of the above recursion:

$$h_0 = \frac{1 - h - a^2 + \sqrt{(1-h-a^2)^2 + 4h}}{2}.$$

Note also that $h_{s,s} = 1$, then the recursion formula for $h_{i,s}$ indicates:

$$|h_{i,s} - h_0| \le \left(\frac{1-h_0}{h+a^2}\right)^{s-i}|1-h_0|.$$

It is easy to check that $0 < \frac{1-h_0}{h+a^2} < 1$. Therefore $h_{i,s} \to h_0$ exponentially fast as $s-i$ increases. Furthermore, there exists constant k, such that for all $s-i \geq k$,

$$h_{i,s} \geq \frac{1-a^2-h}{2}.$$

As a result, we have for any $s-i \geq k$,

$$\left|\frac{\rho_{i,s}}{a}\right| \leq \frac{|a|}{a^2+h+\frac{1-a^2-h}{2}} \leq \frac{2|a|}{2|a|+h} < 1.$$

Then

$$\begin{aligned}\mu_{i,s} &= \frac{\rho_{i,s}}{a}\mu_{i+1,s} + (1-\rho_{i,s})y_i \\ &= \sum_{j=i}^{s-1}\left(\prod_{l=i}^{j-1}\frac{\rho_{l,s}}{a}\right)(1-\rho_{j,s})y_j + \left(\prod_{j=i}^{s-1}\frac{\rho_{j,s}}{a}\right)y_s \end{aligned} \tag{4.59}$$

The desired inequality follows using the fact that

$$\left|\prod_{l=i}^{j-1}\frac{\rho_{l,s}}{a}\right| \leq \left(\frac{2|a|}{2|a|+h}\right)^{j-i}\left(\frac{2|a|+h}{2|a|}\frac{\rho_+^*}{|a|}\right)^k.$$

□

Proof of Lemma 4.3.2. For a standard Gaussian random variable U,

$$\begin{aligned}P(|U|>m) &= \frac{2}{\sqrt{2\pi}}\int_m^\infty e^{-\frac{1}{2}u^2}du \leq \frac{2}{\sqrt{2\pi}}\int_m^\infty \frac{u}{m}e^{-\frac{1}{2}u^2}du \\ &= \frac{2}{\sqrt{2\pi}m}e^{-\frac{m^2}{2}}. \end{aligned} \tag{4.60}$$

Let $m = \sqrt{\theta \log t}$, we have

$$P(|U| > \sqrt{\theta \log t}) \leq \frac{2}{\sqrt{2\pi\theta \log t}}t^{-\frac{\theta}{2}}.$$

Therefore, for $t \geq 2$.

$$P(\xi_t > \sqrt{\theta \log t}) \leq (2t+2)\frac{2}{\sqrt{2\pi\theta \log t}}t^{-\frac{\theta}{2}} \leq 3t^{-\frac{\theta-2}{2}}.$$

□

Proof of Theorem 4.3.3. On E^*, by (4.37), and Lemma 4.3.1, there exists constants κ and c depending on the model only, such that

$$\int \phi_{i|i-1}(x_i)\beta^*_{i,s}(x_i)dx_i \geq \kappa t^{-c\theta}. \tag{4.61}$$

Then one would expect that $\int \phi^{(i-1)}_{i|i-1}\beta^*_{i,s}$ is lower bounded too if $\phi^{(i-1)}_{i-1|i-1}$ is close to $\phi_{i-1|i-1}$. For $0 < \epsilon \leq \frac{\kappa}{2Mt^{c\theta}}$, where $M > 0$ is defined in (S1-S4) for function $\beta^*_{i,s}$, define the events:

$$E_i = \left\{ \left| \phi^{(i)}_{i|i} - \phi_{i|i} \right| \leq \epsilon \right\}.$$

Then on $E_{i-1} \bigcap E^*$, the denominator of (4.15') is lower bounded by

$$\int \phi^{(i-1)}_{i|i-1}\beta^*_{i,s} \geq \frac{\kappa}{2t^{c\theta}}.$$

Consider events $E_{i,s}$:

$$E_{i,s} = \left\{ \sup_{x_i} \left| \phi^{(i)}_{i|i-1}(x_i) - \phi^{(i-1)}_{i|i-1}(x_i) \right| \tilde{\beta}^*_{i,s}(x_i) \leq \frac{\kappa\epsilon}{2Mt^{1+c\theta}} \right\}.$$

We have on $E_{i-1} \bigcap E_{i,s} \bigcap E^*$,

$$\left| \phi^{(i)}_{s|s} - \phi^{(i-1)}_{s|s} \right| \leq \left| \phi^{(i)}_{i|s} - \phi^{(i-1)}_{i|s} \right| \leq \frac{\epsilon}{t}.$$

Therefore

$$E_s \supseteq \bigcap_{i=1}^{s} \left(E_{i-1} \bigcap E_{i,s} \bigcap E^* \right).$$

Note also that E_0 is simply the whole probability space, therefore it is easy to show

$$E_s \supseteq E^* \bigcap \left(\bigcap_{i=1}^{s} \bigcap_{j=1}^{i} E_{j,i} \right),$$

and hence

$$\bigcap_{s=1}^{t} E_s \supseteq E^* \bigcap \left(\bigcap_{s=1}^{t} \bigcap_{i=1}^{s} E_{i,s} \right).$$

By Lemma 4.2.4, there exists constants c_1, c_2, c_3 depending on the model only, such that

$$P\left(E_{i,s}\right) \geq 1 - c_1 \left(c_2 \bigvee \frac{\sqrt{n}\epsilon}{t^{1+c\theta}} \right)^{p(1+\gamma)} \exp\left(-c_3 \frac{n\epsilon^2}{t^{2+2c\theta}} \right),$$

where c is the same as in (4.61).

Therefore,

$$P\left(\bigcap_{s=1}^{t} E_s \right) \geq 1 - c_1 t^2 \left(c_2 \bigvee \frac{\sqrt{n}\epsilon}{t^{1+c\theta}} \right)^{p(1+\gamma)} \exp\left(-c_3 \frac{n\epsilon^2}{t^{2+2c\theta}} \right) - 3t^{-\frac{\theta-2}{2}}.$$

In other words, we have for all $0 < \epsilon \leq \frac{\kappa}{2Mt^{c\theta}}$,

$$\begin{aligned} & P\left(\sup_{1 \leq s \leq t} \left| \phi_{s|s}^{(s)} - \phi_{s|s} \right| \geq \epsilon \right) \\ & \qquad \leq c_1 t^2 \left(c_2 \bigvee \frac{\sqrt{n}\epsilon}{t^{1+c\theta}} \right)^{p(1+\gamma)} \exp\left(-\frac{c_3 n\epsilon^2}{t^{2+2c\theta}} \right) + 3t^{-\frac{\theta-2}{2}}, \end{aligned} \tag{4.62}$$

with positive constants c_1, c_2, c_3 depending only on (a, τ, σ) and c is the constant in (4.61).

□

Bibliography

Anderson, J. (2001). An ensemble adjustment kalman filter for data assimilation. *Monthly Weather Review*, *129*, 2884–2903.

Anderson, J. L. (2003). A local least squares framework for ensemble filtering. *Monthly Weather Review*, *131*, 634–642.

Anderson, J. L. (2007). Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, *230*, 99–111.

Anderson, J. L., & Anderson, S. L. (1999). A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, *127*, 2741–2758.

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, *37*(6), 1554–1563.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, *41*(1), 164–171.

Bengtsson, T., Bickel, P., & Li, B. (2008). Curse of dimensionality revisited: the collapse of importance sampling in very large scale systems. *IMS Collections: Probability and Statistics*, *2*, 316–334.

Bengtsson, T., Snyder, C., & Nychka, D. (2003). Toward a nonlinear ensemble filter for high-dimensional systems : Application of recent advances in space-time statistics to atmospheric data. *J. Geophys. Res.*, *108(D24)*, STS2.1–STS2.10.

Bickel, P., Ritov, Y., & Rydén, T. (1998). Asymptotic normality of the maximum likelihood estimator for general hidden Markov models. *The Annals of Statistics*, *26*(4), 1614–1635.

Bickel, P. J., & Doksum, K. A. (to appear). *Mathematical statistics, basic ideas and selected topics. Volume II*. Prentice Hall.

Bishop, C. H., Etherton, B., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transformation kalman filter. part i: theoretical aspects. *Monthly Weather Review*, *129*, 420–436.

Breiman, L. (1996). Stacked regressions. *Machine Learning*, *24*, 49–64.

Bunea, F., Tsybakov, A. B., & Wegkamp, M. H. (2007). Aggregation for gaussian regression. *The Annals of Statistics*, *35*(4), 1674–1697.

Cappé, O., Moulines, E., & Rydén, T. (2005). *Inference in hidden Markov models*. Springer.

Chorin, A., & Tu, X. (2009). Non-Bayesian particle filters.

Del Moral, P., & Guionnet, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, *37*, 155–194.

Douc, R., Fort, G., Moulines, E., & Priouret, P. (2009a). Forgetting the initial distribution for hidden Markov models. *Stochastic Processes and their Applications*, *119*, 1235–1256.

Douc, R., Moulines, E., & Ritov, Y. (2009b). Forgetting of the initial condition for the filter in general state-space hidden Markov chain: a coupling approach. *The Electronic Journal of Probability*, *14*(2), 27–49.

Doucet, A., de Freitas, N., & Gordon, N. (Eds.) (2001). *Sequential Monte Carlo in Practice*. Springer-Verlag.

Evensen, G. (1994). Sequential data assimilation with a non-linear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.*, *99(C5)*, 10143–10162.

Evensen, G. (2003). The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, *53*, 343–367.

Evensen, G. (2007). *Data assimilation: the ensemble Kalman filter*. Springer.

Fan, J., & Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*. Springer.

Furrer, R., & Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior

covariance matrices in kalman filter variants. *Journal of Multivariate Analysis*, *98*, 227–255.

Gordon, N., Salmon, D., & Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F*, *140*, 107–113.

Hammersley, J. M., & Handscomb, D. C. (1965). *Monte Carlo Methods*. Methuen & Co.

Hampel, F., Ronchetti, E., Rousseeuw, P., & Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley.

Heine, K., & Crisan, D. (2008). Uniform approximations of discrete-time filters. *Advances in Applied Probability*, *40*, 979–1001.

Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, *126*, 796–811.

Julier, S. J., & Uhlmann, J. K. (1997). A new extension of the Kalman filter to nonlinear systems. In *Proc. of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, vol. Multi Sensor Fusion, Tracking and Resource Management II. Orlando, Florida.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, *82*(Series D), 35–45.

Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Transactions of the ASME. Series D, Journal of Basic Engineering*, *83*, 95–107.

Künsch, H. R. (2001). State space and hidden Markov models. In O. E. Barndorff-Nielsen, D. R. Cox, & C. Klüppelberg (Eds.) *Complex Stochastic Systems*, (pp. 109–173). Chapman and Hall.

Künsch, H. R. (2005). Recursive Monte Carlo filters: algorithms and theoretical analysis. *The Annals of Statistics*, *33*, 1983–2021.

Lawson, G. W., & Hansen, J. A. (2004). Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Monthly Weather Review*, *132*, 1966–1981.

Le Gland, F., Monbet, V., & Tran, V. (2009). Large sample asymptotics for the ensemble Kalman filter.

Le Gland, F., & Oudjane, N. (2003). A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. *Stochastic Processes and Their Applications*, *106*, 279–316.

Le Gland, F., & Oudjane, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *The Annals of Applied Probability*, *14*, 144–187.

Lei, J., Bickel, P., & Snyder, C. (2009). Comparison of ensemble Kalman filters under non-Gaussianity. *submitted to Monthly Weather Review*.

Liu, J. (2001). *Monte Carlo strategies in scientific computing*. Springer.

Liu, J., & Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, *93*(443), 1032–1044.

Livings, D. M., Dance, S. L., & Nicols, N. K. (2008). Unbiased ensemble square root filters. *Physica D*, *237*, 1021–1028.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*, 130–141.

Lorenz, E. N. (1996). Predictability: a problem partly solved. In *Proc. Seminar on Predictability*, vol. 1. Shinfield Park, Reading, Berkshire, United Kingdom: European Centre for Medium-Range Weather Forecast.

Olsson, J., & Rydén, T. (2008). Asymptotic properties of particle filter-based maximum likelihood estimators for state space models. *Stochastic Processes and Their Applications*, *118*, 649–680.

Ott, E., Hunt, B., Szunyogh, I., Zimin, A., Kostelich, E., Corazza, M., Kalnay, E., Patil, D., & Yorke, J. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, *56A*, 415–428.

Sacher, W., & Bartello, P. (2008). Sampling errors in ensemble Kalman filtering. part i: theory. *Monthly Weather Review*, *136*, 3035–3049.

Sacher, W., & Bartello, P. (2009). Sampling errors in ensemble Kalman filtering. part ii: application to a barotropic model. *Monthly Weather Review*, *137*, 1640–1654.

Sakov, P., & Oke, P. R. (2007). Implications of the form of the ensemble transformation in the ensemble square root filters. *Monthly Weather Review*, *136*, 1042–1053.

Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *The Annals of Probability*, *22*, 28–76.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., & Whitaker, J. S. (2003). Ensemble square root filters. *Monthly Weather Review*, *131*, 1485–1490.

van der Vaart, A. W. (2001). *Asymptotic Statistics*, chap. 19. Cambridge University Press.

van der Vaart, A. W., & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

van Handel, R. (2009). Uniform time average consistency of Monte Carlo particle filters. *Stochastic Processes and their Applications*, *119*, 3835–3861.

Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, *130*, 1913–1924.

Yang, Y. (2001). Adaptive regression by mixing. *Journal of the American Statistical Association*, *96*(454), 574–588.