Unbiased estimators applied to the Ensemble Kalman-Bucy Filter

Thesis by

Miguel Angel Alvarez Ballesteros

In Partial Fulfillment of the Requirements

For the Degree of

Masters of Science

King Abdullah University of Science and Technology

Thuwal, Kingdom of Saudi Arabia

April, 2022

EXAMINATION COMMITTEE PAGE

The thesis of Miguel Angel Alvarez Ballesteros is approved by the examination committee

Committee Chairperson: Ajay Jasra
Committee Members: Raul F. Tempone, Raphael Huser

ABSTRACT

Unbiased estimators applied to the Ensemble Kalman-Bucy Filter

Miguel Angel Alvarez Ballesteros

Recent debiasing techniques are incorporated into the Ensemble Kalman-Bucy Filter (EnKBF). Specifically, a novel double randomization is applied. The EnKBF is a Monte Carlo (MC) method that approximates the Kalman-Bucy Filter (KBF), which in turn can be seen as the continuous-time version of the celebrated discrete-time Kalman Filter (KF). The KF is a method that combines sequential observations with an underlying dynamics model to predict the state of the quantity of interest. Our interest in the EnKBF comes from its relevance in high dimensions, where it overcomes the curse of dimensionality and outperforms other standard methods like the Particle Filter. We will consider debiasing techniques (also termed unbiased estimators) in order to improve the error-to-cost rate. Unbiased estimators are variance reduction techniques that produce unbiased and finite variance estimators. Applications of the EnKBF are numerous, from atmospheric sciences, numerical weather prediction, finance, machine learning, among others. Thus, improving the EnKBF is of interest. Numerical tests are done in order to evaluate the cost and the error-to-cost rate of the algorithm, where we consider Ornstein-Uhlenbeck processes. Specifically, a numerical comparison with the Multilevel Ensemble Kalman-Bucy Filter (MLEnKBF) is made using two different unbiased estimators, the coupled sum and the single term estimators. Additionally, we test two variants of the EnKBF, the Vanilla EnKBF, and the Deterministic EnKBF. We find that the error-to-cost rate is virtually the same, although the cost of the unbiased EnKBF is much higher.

ACKNOWLEDGEMENTS

I want to thank Prof. Ajay and Neil for their council and promptness. I also want to thank Carolina for lending me a helping hand.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

Filtering problems, also termed data assimilation [1, 2, 3], are predictive models that incorporate stochastic observations of an underlying stochastic phenomenon. We introduce them theoretically in the following. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t\geqslant 0}$ and let two $\mathcal{F}$-measurable processes $X = \{X_t\}_{t\geqslant 0}$ and $Y = \{Y_t\}_{t\geqslant 0}$ be d-dimensional stochastic process representing the quantity of interest (signal process) and the observation process, respectively. In this thesis we consider diffusion processes as the underlying model and observations, respectively, in the form

$$\begin{aligned} dX_t =& f(X_t)dt + \sigma(X_t)dW_t, \quad t \geqslant 0, \\ dY_t =& h(X_t)dt + dV_t, \end{aligned}$$

where $f : \mathbb{R}^d \to \mathbb{R}^d$, $h : \mathbb{R}^d \to \mathbb{R}^d$ are potentially nonlinear functions, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is the diffusion coefficient and $W = \{W_t\}_{t\geqslant 0}$, $V = \{V_t\}_{t\geqslant 0}$ are two independent $d-$dimensional Brownian motions. The filtering problem is concerned with the computation of the quantity $\mathbb{E}(\varphi(X_t)|\mathcal{Y}_t)$, where $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ and $\mathcal{Y} = \{\mathcal{Y}_t\}_{t\geqslant 0}$ is the filtration generated by the observation process. Practical implementation of the filtering problem can be divided into two steps; the first one is the understanding of the theoretical properties and the construct of a framework that allows rigorous treat of the filter; this topic is already well understood, being developed by prominent figures as Stratonovich, Kushner, Zakai, Kalman, Bucy, among others (for a historical review see [4], for an extensive treatment of the theory see [1, 2, 3, 5]). The second step is to obtain the solution to the specific problem at hand, numerical or exact. Analytic solutions are challenging to find except for some specific configurations. Thus we recur to numerical approximations. Two of the most important algorithms to approximate the filter are the Particle Filter (PF) and approximate Gaussian filters.

The filtering problem is important because it has multiple applications in diverse fields, including tracking problems [6, 7], numerical weather prediction [8], ecology [9], mathematical finance [10], language processing [11], geophysical sciences [12], we remark that this list is not exhaustive. As stated before, one of the more popular methods to approximate filters is the PF [13, 14, 15], a Monte Carlo (MC) method that propagates a number $N \in \mathbb{Z}^+$ of particles and associated weights such that the approximated filtering distribution converges to the exact as the number of particles goes to infinite, thus its appealing. Unfortunately, PFs do not perform well in high dimensions, where stability requires an exponential increment of the particles depending on the dimension [16]. As an alternative to PFs, the Ensemble Kalman-Bucy Filter (EnKBF) is another MC method derived by assuming Gaussianity on the signal process, and by propagating an ensemble of particles that update the mean of covariance of the filter. Due to the assumption of Gaussianity, the EnKBF convergence to the exact distribution has only been proven for Gaussian signaling processes. Nevertheless, the EnKBF takes less computational resources compared, for example, to PFs; additionally, it avoids the curse of dimensionality. For these reasons, we will consider the EnKBF in this thesis. The EnKBF has found multiple applications such as oil reservoir simulations [17], inference problems in ocean and atmosphere sciences [18, 19, 20], weather forecasting [21, 22, 23], ecological statistics [24], and others. A common characteristic in these application is that the number of interacting variables is high. For a recent review that shows different variants of the EnKBF along a framework to analyze its errors, see [25].

Time discretization of the EnKBF equations is necessary due to its continuous-time nature. As might be expected, this discretization induces unwanted consequences, among them a bias, and instability, in certain cases. The most straightforward way of dealing with the bias is to make the discretization finer, increasing the cost of the computation. More sophisticated methodologies have been proposed; among the most important are debiasing techniques [26] and Multilevel Monte Carlo (MLMC) methods [27]. Both techniques improve the cost-to-error rate and are closely related [28]. Unbiased techniques have been used in different contexts [29, 30, 31, 32]. However, these techniques have not been applied to Kalman-based filtering, providing us motivation for this thesis.

In particular, in this work, we implement the unbiased techniques to the EnKBF based on [26, 33]. We will apply two of the estimators developed in [26], the single term estimator and the coupled sum estimator. The main contribution of this thesis is the development of an unbiased EnKBF method, and its comparison with the Multilevel Ensemble Kalman-Bucy Filter (MLEnKBF) [34].

The MLEnKBF is an application of MLMC method to the EnKBF, and it is interest because it reduces the error-to-cost rate with a relative low cost. We remark that the comparison of the EnKBF and the MLEnKBF has already been made in [34], providing us with a reference on the EnKBF performance. The numerical examples will be performed on an Ornstein-Uhlenbeck process with different final times. Our goal is to observe the error-to-cost rates discuss the viability of the unbiased EnKBF and its practical implementation.

This thesis is outlined as follows, in chapter 2 we make a literature review including the following topics, the filtering problem, PFs, EnKBF, debiasing techniques and MLMC. In chapter 3 we introduce the unbiased EnKBF, review the MLEnKBF, and at the end of the chapter, numerical tests are presented and discussed. Finally, in the conclusion chapter, 4, we summarize our work, conclude, and layout future work.

1.1 Objectives and Contributions

The contributions of this thesis folds in the following streams:

- Compare the numerical performance of the unbiased EnKBF with respect to the MLEnKBF by computing the error-to-cost rates in certain configurations.
- Postulate unbiased methods with double randomization applied to the EnKBF.
- Compare the numerical performance of the unbiased EnKBF for the single term estimator and the couple sum estimators. Additionally, compare two variants of the EnKBF, the Vanilla EnKBF and the Deterministic EnKBF.

Chapter 2

Literature Review

2.1 Introduction

This chapter includes the relevant framework needed in this thesis. Including the filtering equations, numerical implementations and variance reduction techniques.

In section (2.2) we consider the filtering problem and its equations in the nonlinear and linear-Gaussian settings. In section 2.3, two computational methods relevant to current numerical filtering are reviewed, the Particle Filter (PF) and the Ensemble Kalman-Bucy Filter (EnKBF). Finally, in section 2.4 we review two multilevel techniques that are applicable to the EnKBF and other filtering schemes: Multilevel Monte Carlo (MLMC) and debiasing techniques or unbiased estimators.

2.2 Filtering Problem

2.2.1 Signal and Observation Processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t\geqslant 0}$. Let also $X = \{X_t\}_{t\geqslant 0}$ be the quantity of interest and a $\mathcal{F}_t$-adapted process. Let $\mathbb{S}$ be a complete separable metric space with $\mathcal{B}(\mathbb{S})$ as the associated Borel $\sigma-$algebra. $B(\mathbb{S})$ is defined to be the space of bounded $\mathcal{B}(\mathbb{S})$-measurable functions. Let $\mathcal{A}$ be defined as an operator $\mathcal{A} : B(\mathbb{S}) \to B(\mathbb{S})$ and $\mathcal{D}(\mathcal{A})$ be the domain of $\mathcal{A}$.

X is commonly referred to as the *signal process*, in practice, we do not have access to the process perse, what we do have access to is to the $m-$dimensional *observation* *process* $Y = \{Y_t\}_{t \geqslant 0}$, which is defined by the Stochastic Differential Equation (SDE)

$$dY_t = h(X_t)dt + dV_t, \quad t \geqslant 0, \tag{2.1}$$

with initial condition $Y_0 = 0$ and where $h : \mathbb{S} \to \mathbb{R}^m$ is a m-dimensional vector that satisfies $\mathbb{P}\left(\int_0^t \|h(X_s)\| \, ds < \infty\right) = 1, \ \forall \ t \geqslant 0$, and V_t is a $\mathcal{F}_t-$measurable m-dimensional Wiener process. Lastly, let the information of the observation process be in the augmented process $\mathcal{Y} = \{\mathcal{Y}_t\}_{t \geqslant 0}$, where $\mathcal{Y}_t = \sigma(Y_s, s \in [0, t])$.

2.2.2 Diffusion Process

As a specific subset of the broader classes of processes X, we have the diffusion processes, which are defined by the d-dimensional SDE

$$dX_t = f(X_t)dt + \sigma(X_t)dW_t, \quad t \geqslant 0, \tag{2.2}$$

with the process X_0 as initial condition, where $W = \{W_t\}_{t \geqslant 0}$ is a $p-$dimensional $\mathcal{F}_t$-adapted Wiener process, and where the d-dimensional function $f : \mathbb{R}^d \to \mathbb{R}^d$ and the $d \times p$ matrix $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times p}$ are globally Lipschitz functions, i.e., there exists positive constants K and C such that

$$\|f(X) - f(Y)\| \leqslant K\|X - Y\|, \quad \|\sigma(X) - \sigma(Y)\| \leqslant C\|X - Y\|,$$

where $\| \cdot \|$ is the Euclidean norm for vectors and the Hilbert-Schmidt or Frobenius norm for matrices. With this conditions the existence and uniqueness of a solution X_t for (2.2) is guaranteed.

There exists a generator $\mathcal{A}$ of the diffusion process

$$\mathcal{A} = f^\top \nabla + \mathbb{T}\mathrm{r}\left[a \nabla \nabla^\top\right],$$

where the superscript $^\top$ represents the transpose, ∇ the gradient, $\mathbb{T}\mathrm{r}(\cdot)$ represents the trace, and a is defined as $a = \frac{1}{2}\sigma\sigma^\top$.

2.2.3 Filtering distribution

The information that we know about the signal is encoded in the SDE (2.2) and the observation process. Since this is the only information that we have, it is not possible to know exactly the state of X_t, thus the new quantity of interest is the expected value of the signal conditioned on $\mathcal{Y}_t$. This quantity is denoted as

$$\eta_t(\varphi) \equiv \mathbb{E}(\varphi(X_t)|\mathcal{Y}_t), \quad t \geqslant 0, \tag{2.3}$$

where the expectation is with respect to the measure $\mathbb{P}$ and $\varphi \in B(\mathbb{S})$. Since $Y_0 = 0$, the conditioned expectation at time zero is the expectation of $\varphi(X_0)$.

The filtering problem refers to the set of equations describing the evolution of the filter $\eta_t(\varphi)$.

2.2.4 Filtering equations

One of the ways to deduce the filtering equations is using Girsanov's theorem and performing a change of measure such that the observation process becomes a Wiener process under the new measure. We define

$$Z_t = \exp\left(-\int_0^t h(X_s)^\top dV_s - \frac{1}{2}\int_0^t h(X_s)^\top h(X_s)ds\right), \quad t \geqslant 0,$$

where $Z = \{Z_t\}_{t\geqslant 0}$ is a martingale under the following condition

$$\mathbb{E}\left(\exp\left(\frac{1}{2}\int_0^t h^\top h ds\right)\right) < \infty,$$

since this condition can be challenging to prove, two other conditions can replace it

$$\mathbb{E}\left(\int_0^t \|h(X_s)\|^2 ds\right) < \infty, \qquad \mathbb{E}\left(\int_0^t Z_s\|h(X_s)\|^2 ds\right) < \infty. \tag{2.4}$$

Let $\tilde{\mathbb{P}}^t$ on $\mathcal{F}_t$ be a measure defined with the Radon-Nikodym derivatives such that

$$\left.\frac{d\tilde{\mathbb{P}}^t}{d\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t, \quad t \geqslant 0,$$

additional, let $\tilde{Z}_t$ be defined as

$$\left.\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}^t}\right|_{\mathcal{F}_t} = \tilde{Z}_t, \quad t \geqslant 0.$$

where $\tilde{\mathbb{P}}^t$ is precisely the measure in which Y_t is an independent Brownian motion.

The Kallianpur-Streibel formula is essential because it offers a way of deducing the filtering equations; Kallianpur makes use of the measure $\tilde{\mathbb{P}}$ and $\tilde{Z}_t$ defined previously.

Assume that condition (2.4) holds, then $\forall\varphi \in B(\mathbb{S})$ and for fixed $t \in [0, \infty)$ the filter can be expressed as

$$\eta_t(\varphi) = \frac{\tilde{\mathbb{E}}(\tilde{Z}_t\varphi(X_t)|\mathcal{Y})}{\tilde{\mathbb{E}}(\tilde{Z}_t|\mathcal{Y})}, \quad \mathbb{P}^t \quad \text{and} \quad \tilde{\mathbb{P}} \quad a.s., \tag{2.5}$$

where the expectation is with respect to the measure $\tilde{\mathbb{P}}$.

Let the measure $\rho = \{\rho_t\}_{t\geqslant 0}$ be the unnormalized filter such that

$$\rho_t(\varphi) \equiv \eta_t(\varphi)\tilde{\mathbb{E}}(\tilde{Z}_t|\mathcal{Y}),$$

then we can identify $\rho_t(\varphi) = \tilde{\mathbb{E}}(\tilde{Z}_t\varphi(X_t)|\mathcal{Y})$, which follows the Zakai equation

$$d\rho_t(\varphi) = \rho_t(\mathcal{A}\varphi)dt + \rho_t(\varphi h^\top)dY_t, \quad \tilde{\mathbb{P}} \quad a.s. \quad \forall t \geqslant 0,$$

with initial condition $\rho_0(\varphi)$ and for all $\varphi \in \mathcal{D}(\mathcal{A})$. Where the condition

$$\tilde{\mathbb{P}}\left(\int_0^t \rho_s(\|h\|)^2 ds < \infty\right) = 1, \tag{2.6}$$

must be assumed. Note that this equation is not the equation for the filtering distribution since $\rho_t(\varphi)$ is unnormalized. Thus, using again the Kallianpur-Streibel formula it can be deduced

$$d\eta_t(\varphi) = \eta_t(\mathcal{A}\varphi)dt + (\eta_t(\varphi h^\top) - \eta_t(h^\top)\eta_t(\varphi))\left(dY_t - \eta_t(h)dt\right), \tag{2.7}$$

with initial condition $\eta_0(\varphi)$ and for all $\varphi \in \mathcal{D}(\mathcal{A})$, and where both condition (2.6) and (2.4) must be assumed, and additionally, $\mathbb{P}\left(\int_0^t \|\eta_s(h)\|^2 ds < \infty\right) = 1$ needs to be satisfied.

2.2.5 Kalman-Bucy Filter

The Kalman-Bucy Filter (KBF) is the continuous-time analogous of the discrete-time Kalman Filter (KF). The KBF is a linear and Gaussian filter, originated by a linear signal process X_t

$$dX_t = (A_t X_t + f_t)dt + R_t^{\frac{1}{2}} dW_t, \quad \forall t \geqslant 0, \tag{2.8}$$

which is a specific case of diffusion process, where A_t, is a $d \times d$ matrix, f_t is a $d-$dimensional vector and R_t is a $d \times p$ vector. The initial value of the process is $X_0 \sim \mathcal{N}(m_0, \mathcal{C}_0)$, where $\mathcal{N}(m, C)$ represents a multivariate Gaussian random distribution with mean m and covariance matrix $\mathcal{C}$. The observation process is defined as

$$dY_t = (H_t X_t + h_t)dt + R_1^{\frac{1}{2}} dV_t, \quad \forall t \geqslant 0, \tag{2.9}$$

where H_t is defined as a $m \times d$ matrix, R_1 is positive definite $m \times m$ matrix and h_t is a m-dimensional vector. A_t, f_t, H_t, h_t and R_t depend only on time. It can be proven that the system (X, Y) is Gaussian, then it can be shown that X conditioned on $\mathcal{Y}_t$ is also normal. One of the properties of the Gaussian distributions is that can be characterized by its mean and covariance matrix. Let $\xi_t = \eta_t(e)$ where $e(x) = x$, and $\mathcal{P}_t = \eta_t(\tilde{\varphi}) - \xi_t \xi_t^\top$ where $\tilde{\varphi}(\xi) = \xi\xi^\top$, these quantities represent the mean and covariance of X conditioned on $\mathcal{Y}_t$, respectively.

The Kushner-Stratonovich equation takes the form

$$d\xi_t = (A_t\xi_t + f_t)dt + \mathcal{P}_t H_t^\top R_1^{-1}(dY_t - (H_t\xi_t + h_t)dt), \quad \forall t \geqslant 0, \tag{2.10}$$

$$\frac{d\mathcal{P}_t}{dt} = \mathrm{Ricc}(\mathcal{P}_t), \tag{2.11}$$

where the Riccati drift is defined as

$$\mathrm{Ricc}(Q) = A_t Q + Q A_t^\top - QSQ + R, \quad \text{with } R = R_t^{\frac{1}{2}} R_t^{\frac{1}{2}\top} \quad \text{and } S \equiv H_t^\top R_1^{-1} H_t,$$

with initial conditions $(\xi_0, \mathcal{P}_0) = (m_0, \mathcal{C}_0)$. The equation of the filter covariance is a Riccati-type differential equation. One important remark about this equation is that is deterministic meaning that it does not depend on the observations. Additionally, the covariance of the filter can be shown to converge exponentially fast to a constant matrix, which is precisely the solution of a stationary Riccati equation [35, 25]. This behavior provides insight into the filter in the long term time.

2.3 Computational Methods

Current challenges in filtering computations include high-dimensional and nonlinear systems, datasets of observations in discrete time for continuous systems of particles, among others. In order to address those problems we present two numerical filtering methods capable of dealing with those problems to some extent.

2.3.1 Particle Filter

Particle Filters (PFs) are a class of algorithms that estimate sequentially filtering distributions and its expectations, furthermore, PF are part of a broader class of methods called Sequential Monte Carlo (SMC) methods. This kind of method is particularly appealing due to its applicability to approximate both Gaussians and nonlinear filters, among other qualities. PFs sample from auxiliary probability kernels in an importance sampling fashion, additionally, they use a technique most commonly known as *resampling*.

Let $\{\eta_k\}_{k\in\mathbb{Z}^+}$ be a sequence of target measures. Similarly to the previous session, these measures represent the distribution of a signal process conditioned on observations. The measures are such that lie in the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and are dominated by the Lebesgue measure with Radon-Nikodym derivative $\rho_k : \mathbb{R}^d \to \mathbb{R}^+$. Let $\rho : \mathbb{R}^{d+m} \to \mathbb{R}^+$ be the density distribution or both the signal and the observations of a discrete-time state space model, and $p : \mathbb{R}^m \to \mathbb{R}^+$ be the marginal density distribution of the observations, defined as

$$\rho(x_{1:k}, y_{1:k}) = \frac{\prod_{i=1}^{k} h\left(y_i \mid x_i\right) g\left(x_i \mid x_{i-1}\right)}{\int_{\mathbb{R}^{(d+m)k}} \prod_{j=1}^{k} h\left(v_j \mid u_j\right) g\left(u_j \mid u_{j-1}\right) du_{1:k} dv_{1:k}}, \tag{2.12}$$

$$p(y_{1:k}) = \int_{\mathbb{R}^{d\times k}} \prod_{i=1}^{k} h\left(y_i \mid x_i\right) g\left(x_i \mid x_{i-1}\right) dx_{1:k}, \quad \forall k \in \mathbb{Z}^+, \tag{2.13}$$

where here we assume $g(x_1|x_0) = g(x_1)$. The sequence of densities $\{\rho_k\}_{k\in\mathbb{Z}^+}$ is such that

$$\rho_k(x_k) = \rho(x_k|y_{1:k}) \propto \int_{\mathbb{R}^{(k-1)d}} \left(\prod_{i=1}^{k} h\left(y_i \mid x_i\right) g\left(x_i \mid x_{i-1}\right) \right) dx_{1:k-1}, \quad \forall k \in \mathbb{Z}^+.$$

It is assumed that the functions $h(y_i|x_i)g(x_i|x_{i-1}) > 0$ for $i \in \mathbb{Z}^+$ and $(x_i, x_{i-1}, y_i) \in$

$\mathbb{R}^{2d+m}$.

We assume that we have a sequence of auxiliary probability measures w.r.t. the Lebesgue measure $\{q_k(x_k|x_{k-1})\}_{k\in\mathbb{Z}^+}$, with strictly positive function $q_1(x_1|x_0) = q_1(x_1) > 0$, $q_1(x_1) : \mathbb{R}^d \to \mathbb{R}^+$; and such that $q_k(x_k|x_{k-1}) : \mathbb{R}^{2d} \to \mathbb{R}^+$ are also strictly positive for $k \in \mathbb{Z}^+$ and $(x_i, x_{i-1}) \in \mathbb{R}^{2d}$.

Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{B}(\mathbb{R}^d)$-measurable function, this function can be identified with the function in Eq. (2.3). One of the quantities we are most interested in is the filter, this time conditioned in given observations $\{y_k\}_{k\in\mathbb{Z}^+}$, i.e.

$$\eta_k(\varphi) = \int_{\mathbb{R}^{kd}} \varphi(x_k)\rho_k(x_k)dx_k,$$

where the PF, defined in algorithm 1 converges in the infinite particles limit [2] $(N \to \infty)$ to,

$$\sum_{i=1}^{N} \omega_k^i \varphi(x_k^i) \xrightarrow{a.s.} \eta_k(\varphi), \quad \forall k \in \mathbb{Z}^+, \tag{2.14}$$

additionally, the PF estimates the density distribution using Dirac-delta $\delta(x)$ distributions as

$$\rho_k^N(x) = \sum_{i=1}^{N} \omega_k^i \delta(x - x_k^i), \quad \forall k \in \mathbb{Z}^+, \tag{2.15}$$

and the marginal distribution (2.13) can also be unbiasedly approximated as

$$p^N(y_{1:k}) \equiv \prod_{i=1}^{k} \left(\frac{1}{N} \sum_{j=1}^{N} \frac{h\left(y_i \mid x_i^j\right) g\left(x_i^j \mid \hat{x}_{i-1}^j\right)}{q_i\left(x_i^j \mid \hat{x}_{i-1}^j\right)} \right), \quad \forall k \in \mathbb{Z}^+. \tag{2.16}$$

Algorithm 1 can be considered the backbone of the PF; many other specialized variants have been proposed, some of them can be found in [36]. It is noted that although the algorithm is simple, it is general in the choice of the resampling method and the auxiliary densities.

Algorithm 1 Particle Filter

1. **Initialize:** Sample x_1 from q_1 N times, i.e. obtain N i.i.d. samples $x_1^i \sim q_1$. Compute the weights using the following formula

$$\omega_1^i = \frac{h(y_1^i|x_1^i)g(x_1^i)}{q(x_1^i)} \left(\sum_{j=1}^{N} \frac{h(y_1^j|x_1^j)g(x_1^j)}{q(x_1^j)} \right)^{-1}.$$

   Set $k = 1$

2. **Iterate:** Set $k = k + 1$.
   - Resampling: Using the weights $\{\omega_{k-1}^i\}_{i=1}^N$ resample $\{\hat{x}_{k-1}^i\}_{i=1}^N$.
   - Using the auxiliar densities sample $x_k^i \sim q_k(\cdot|\hat{x}_{k-1}^i)$ and compute

$$\omega_k^i = \frac{h(y_k^i|x_k^i)g(x_k^i|\hat{x}_{k-1}^i)}{q(x_k^i|\hat{x}_{k-1}^i)} \left(\sum_{j=1}^{N} \frac{h(y_k^j|x_k^j)g(x_k^j|\hat{x}_{k-1}^j)}{q(x_k^j|\hat{x}_{k-1}^j)} \right)^{-1}.$$

   Go back to the second step until reaching the target k.

Each iteration of the algorithm is divided into two parts: resampling and importance sampling. Importance sampling is a straightforward way to compute expectations, with the perk of the possibility of reducing the variance, say, for example, of the left part Eq. (2.14). It would be natural to ask what happens if we remove the resampling step; in this case, a problem known as weight-degeneracy appears, in which the variance of the weights ω_k^i grows algorithmically with the time parameter k, the introduction of resampling is motivated by this problem.

A known issue of the PF is the computational complexity in terms of the state space dimension d. As the dimension grows, the PF becomes unstable [16] and the particles converge to one weight, which is, of course, undesired. To solve this problem, the number of particles can be increased exponentially on the dimension, i.e. $N = \mathcal{O}(\beta^d)$ [15], for $\beta > 0$, where $\mathcal{O}(\cdot)$ is the big O notation. There are cases in which the system has certain properties that alleviate this curse of dimensionality, but the general PF suffers from it.

Among others, some of the qualities of the PF are the convergence of Eq. (2.14) in several norms [2] as the L_p norm, with canonical error on the number of particles with certain regularity on the function φ. Many of these results do not depend on the time parameter k, making it appealing for large-time computations. Its applicability spans linear, nonlinear, and Gaussian filters, making it versatile.

2.3.2 Ensemble Kaman-Bucy Filter

The Ensemble Kalman-Bucy Filter (EnKBF) is a numerical method concerned with the filtering problem, it is part of the so-called approximate Gaussian filters, and its authorship can be traced to Evensen [37, 17].

EnKBF is mainly used in high dimensions filtering problems as oil reservoir simulations [17], inference problems in ocean and atmosphere sciences [18, 19, 20], weather forecasting [21, 22, 23], ecological statistics [24], among others. In practice, the EnKBF is used as an alternative to the PF when dealing with relatively high dimensions; a list of these implementations can be found in [17, 38, 25]. The approximations in high-dimensions are used in both Gaussian-linear and non-Gaussians settings; in the former, the approximation can be shown to converge in the infinite particle limit. In the latter, the EnKBF is most regarded as a state estimator where we cannot say much about statistics in a rigorous fashion except for the long time stability that the method offers [39, 40, 41].

The method works by propagating particles that mimic the evolution of the linear-Gaussian filter, where the ensemble mean and covariance are used in the propagation equations instead of the exact mean and covariance. Two ways of propagating the particles are considered here.

In contrast to the computation of the KBF, the EnKBF does not use the Riccati equation to compute the covariance matrix, instead it computes an MC-like approximation. There exists an analytical solution of the Riccati equation but its computations cost becomes unfeasible for high dimensions. Likewise, a numerical computation of the covariance using the Riccati differential is costly and its hard to store for high dimensions.

Vanilla EnKBF and Deterministic EnKBF

Let us consider the linear diffusion processes (2.8), (2.9) with $f_t = 0$, $h_t = 0$ and where A_t, H_t, and R_t do not depend on time and will be denoted as A, H, and R respectively. For the dimension of the Wiener process W we choose the same as the signal process, i.e. $p = d$. The resulting system is:

$$\begin{aligned} dX_t =& AX_t dt + R^{\frac{1}{2}} dW_t, \quad \forall t \geqslant 0, \\ dY_t =& HX_t dt + R_1^{\frac{1}{2}} dV_t, \end{aligned}$$

where initial value of the signal process X is $X_0 \sim \mathcal{N}(m_0, \mathcal{C}_0)$, and the initial value of the observation process Y is $Y_0 = 0$. The equation for the linear-Gaussian filtering problem was presented in Eq. (2.10). For this specific setting, the filtering equation reduces to

$$d\xi_t =A\xi_t dt + \mathcal{P}_t H^\top R_1^{-1}(dY_t - H\xi_t dt), \quad \forall t \geqslant 0, \tag{2.17}$$

this is a conditional McKean-Vlasov diffusion process (see for example [42]), with initial condition $\xi_0 = m_0$. Now we introduce two other contiditional McKean-Vlasov processes

$$\begin{aligned} &\text{(F1)} \quad d\hat{\xi}_t = A\hat{\xi}_t dt + R^{\frac{1}{2}} d\overline{W}_t + \mathcal{P}_{\widehat{\eta}_t} H^\top R_1^{-1} \left(dY_t - \left[H\xi_t dt + R_1^{\frac{1}{2}} d\bar{V}_t \right] \right), \quad \forall t \geqslant 0, \\ &\text{(F2)} \quad d\hat{\xi}_t = A\hat{\xi}_t dt + R^{\frac{1}{2}} d\overline{W}_t + \mathcal{P}_{\widehat{\eta}_t} H^\top R_1^{-1} \left(dY_t - \frac{1}{2} H[\xi_t + \widehat{\eta}_t(e)] dt \right), \end{aligned}$$

where $(\hat{\xi}_0, \overline{W}, \bar{V})$ are independent copies of (ξ_0, W, V), $e(\xi) = \xi$, $\widehat{\eta}_t(\cdot)$ is similarly defined as in (2.3), and the covariance is defined as

$$\begin{aligned}\mathcal{P}_{\widehat{\eta}_t} =& \widehat{\eta}_t \left([e - \widehat{\eta}_t(e)] \, [e - \widehat{\eta}_t(e)]^\top \right), \quad t \geqslant 0, \\ \widehat{\eta}_t(\varphi) \equiv& \mathbb{E}(\varphi(\hat{\xi}_t) | \mathcal{Y}_t).\end{aligned}$$

for $\varphi : \mathbb{R}^d \to \mathbb{R}$. It can be shown [25] that both distributions $\widehat{\eta}_t(\varphi)$ for (F1) and (F2) are Gaussian and equivalent. Moreover,

$$\widehat{\eta}_t(e) = \mathbb{E}(\hat{\xi}_t | \mathcal{Y}_t) = \mathbb{E}(X_t | \mathcal{Y}_t) = \eta_t(e), \quad t \geqslant 0,$$

and then the covariances $\mathcal{P}_{\widehat{\eta}_t} = \mathcal{P}_t$. Since $\mathcal{P}_t$ is independent on the filtration $\mathcal{Y}$ so it is $\mathcal{P}_{\widehat{\eta}_t}$. The randomness introduced by the processes $\overline{W}$ and $\bar{V}$ are key for the EnKBF formulation since this allows the propagation of random particles $\hat{\xi}^i$. MC-type estimators can be formulated to approximate both the filter covariance $\mathcal{P}_{\widehat{\eta}_t}$ and its mean $\hat{\eta}_t(e)$, these estimators are in turn used in the equations to propagate the particles.

Let $N \in \mathbb{Z}^+$, $N \geqslant 2$ be the number of particles in the ensemble; the following estimators associated with (F1) and (F2) are termed Vanilla EnKBF (VEnKBF) and Deterministic EnKBF (DEnKBF) respectively

$$\text{(F1)} \quad d\hat{\xi}_t^i = A\hat{\xi}_t^i dt + R^{\frac{1}{2}} d\overline{W}_t^i + P_t^N H^\top R_1^{-1} \left(dY_t - \left[H\xi_t^i dt + R_1^{\frac{1}{2}} d\overline{V^i}_t \right] \right), \quad \forall t \geqslant 0, \tag{2.18}$$

$$\text{(F2)} \quad d\hat{\xi}_t^i = A\hat{\xi}_t^i dt + R^{\frac{1}{2}} d\overline{W}_t + P_t^N H^\top R_1^{-1} \left(dY_t - \frac{1}{2} H[\xi_t^i dt + \eta_t^N(e)] \right), \tag{2.19}$$

for $i \in \{j\}_{j=1}^N$, where $\left\{\left(\overline{W}^i, \bar{V}^i\right)\right\}_{i=1}^N$ are N i.i.d. copies of $(\overline{W}, \bar{V})$ and

$$\begin{aligned} P_t^N =& \frac{1}{N-1}\sum_{i=1}^N (\xi_t^i - \eta_t^N(e))(\xi_t^i - \eta_t^N(e))^\top, \\ \eta_t^N(e) =& \frac{1}{N}\sum_{i=1}^N \xi_t^i. \end{aligned}$$

In both DEnKBF and VEnKBF we have consistency in the infinite particles limit [25] for the mean and the covariances, i.e. $\eta_t^N(e) \to \eta_t(e)$ and $P_t^N \to \mathcal{P}_t$ as $N \to \infty$. The DEnKBF is denominated Deterministic because, in comparison to the VEnKBF, it does not have the randomness associated to $\bar{V}$. Its postulation can be found in [43]. We will denote

$$\eta_t^N(\varphi) = \frac{1}{N}\sum_{i=1}^N \varphi\left(\xi_t^i\right). \tag{2.20}$$

as the EnKBF approximation of the filter.

Nonlinear Processes

In practice, the EnKBF is applied to nonlinear signal and observations processes. This is useful when the dimension of the signal process is relatively high, so other options like the PF are not feasible. Let the signal system be a diffusion process as in (2.2) and the observations (2.1), i.e.

$$\begin{aligned} dX_t =& f(X_t)dt + \sigma(X_t)dW_t, \quad t \geqslant 0, \\ dY_t =& h(X_t)dt + dV_t, \end{aligned}$$

where the necessary regularity conditions have been considered in section 2.2. In this case, the propagation of the Ensemble is given by

$$
\begin{aligned}
\text{(F1)} \quad d\hat{\xi}_t^i =& f(\hat{\xi}_t^i)dt + R^{\frac{1}{2}} d\overline{W}_t^i + \widehat{P}_t^h R_1^{-1} \left(dY_t - \left[h(\xi_t^i)dt + R_1^{\frac{1}{2}} d\overline{V}^i{}_t \right] \right), \quad \forall t \geqslant 0, \\
\text{(F2)} \quad d\hat{\xi}_t^i =& f(\hat{\xi}_t^i)dt + R^{\frac{1}{2}} d\overline{W}_t + \widehat{P}_t^h R_1^{-1} \left(dY_t - \frac{1}{2}[h(\xi_t^i) + \widehat{h}_t]dt \right),
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{h}_t \equiv& \frac{1}{N} \sum_{i=1}^{N} h(\hat{\xi}_t^i), \quad t \geqslant 0, \\
\widehat{P}_t^h \equiv& \frac{1}{N-1} \sum_{i=1}^{N} (\hat{\xi}_t^i - \eta_t^N(e))(h(\hat{\xi}_t^i) - \widehat{h}_t)^\top.
\end{aligned}
$$

The limiting behavior of this system is not appropriately studied yet. In general, we have that $\widehat{\eta}_t(\varphi) \neq \eta_t(\varphi)$, meaning that the particles limit of the nonlinear EnKBF does not provide an exact source of information. This kind of approximation is usually termed as state estimator. The information that provides is relatively reliable, given the stability of the EnKBF. Such stability has been studied in nonlinear and Linear Gaussian settings in [39, 40, 41].

Discretization

In practice we cannot solve the system of SDES (2.18), (2.19) and we must resort to numerical solutions such as the Euler-Maruyama discretization, which is precisely the one we will consider here. Let the time step Δ_l be parametrized by the *level* $l \in \mathbb{N}_0$ and let the discretization at level l be $\{k\Delta_l\}_{k\in\mathbb{N}_0}$. The Euler-Maruyama method at level l, applied to Eqs. (2.18) and (2.19), respectively, reads

$$
\begin{aligned}
\hat{\xi}_{(k+1)\Delta_l}^{i,l} =& \hat{\xi}_{k\Delta_l}^{i,l} + A\hat{\xi}_{k\Delta_l}^{i,l}\Delta_l + R^{\frac{1}{2}}\Delta\overline{W}_{k\Delta_l}^i \\
&+ P_{k\Delta_l}^{N,l} H^\top R_1^{-1} \left([Y_{(k+1)\Delta_l} - Y_{k\Delta_l}] - \left[H\hat{\xi}_{k\Delta_l}^{i,l}\Delta_l + R_1^{\frac{1}{2}}\Delta\bar{V}_{k\Delta_l}^i \right] \right),
\end{aligned}
$$

$$\hat{\xi}^{i,l}_{(k+1)\Delta_l} = \hat{\xi}^{i,l}_{k\Delta_l} + A\hat{\xi}^{i,l}_{k\Delta_l}\Delta_l + R^{\frac{1}{2}}\Delta\overline{W}^i_{k\Delta_l}$$
$$+ P^{N,l}_{k\Delta_l}H^\top R_1^{-1}\left([Y_{(k+1)\Delta_l} - Y_{k\Delta_l}] - \frac{1}{2}H[\hat{\xi}^{i,l}_{k\Delta_l}\Delta_l + \eta^{N,l}_{k\Delta_l}(e)]\right),$$

for a system of equations spanned by all $(k, i) \in \mathbb{N}_0 \times \{j\}_{j=1}^N$, where $\Delta\overline{W}^i_{k\Delta_l} = \overline{W}^i_{(k+1)\Delta_l} - \overline{W}^i_{k\Delta_l}$ and $\Delta\bar{V}^i_{k\Delta_l} = \bar{V}^i_{(k+1)\Delta_l} - \bar{V}^i_{k\Delta_l}$, and where the initial conditions $\hat{\xi}^{i,l}_0$ are i.i.d. copies of $\hat{\xi}_0 \sim \mathcal{N}(m_0, C_0)$. The covariance and the mean are

$$P^{N,l}_{k\Delta_l} = \frac{1}{N-1}\sum_{i=1}^N(\hat{\xi}^{i,l}_{k\Delta_l} - \eta^{N,l}_{k\Delta_l}(e))(\hat{\xi}^{i,l}_{k\Delta_l} - \eta^{N,l}_{k\Delta_l}(e))^\top, \tag{2.21}$$

$$\eta^{N,l}_{k\Delta_l}(e) = \frac{1}{N}\sum_{i=1}^N \hat{\xi}^{i,l}_{k\Delta_l}. \tag{2.22}$$

The time is restricted to the discretization, i.e. $t \in \{k\Delta_l\}_{k\in\mathbb{N}_0}$. Denote

$$\eta^{N,l}_t(\varphi) = \frac{1}{N}\sum_{i=1}^N \varphi\left(\xi^{i,l}_t\right). \tag{2.23}$$

as the approximation of (2.20) when we discretize.

2.4 Unbiased Estimators and Multilevel Monte Carlo

Multilevel Monte Carlo (MLMC) and unbiased estimators are Monte Carlo (MC) methods designed to reduce the Mean Square Error (MSE)

$$\text{MSE}(\Xi) = \text{Variance}(\Xi) + (\text{bias}(\Xi))^2, \quad \text{bias}(\Xi) = \mathbb{E}(\Xi - \hat{\Xi}),$$

where Ξ is an estimator of $\hat{\Xi}$. MLMC techniques are not applicable to all MC estimators, in order to apply the method and see and improvement we need a consistent sequence of estimators with *increasing cost*.

MLMC is a variance reduction method developed by Giles in 2008 [27] and it can be seen as a generalization of the control variates technique. Debiasing methods are techniques initially developed by Rhee and Glynn in 2012 [26]. Debiasing methods produce unbiased estimator and that seek to improve the error-to-cost rate. In the following we present a common structure that is used in both unbiased estimators and MLMC.

Let η be a probability measure on the space $(\Omega, \mathcal{F})$ and for a η-integrable function $\varphi : \Omega \to \mathbb{R}$ and assume that we have a sequence of increasing in *cost* measures parametrized by the *level* l, $\{\eta^l\}_{l=0}^{\infty}$, defined on $(\Omega, \mathcal{F})$, and such that $\lim_{l\to\infty} \mathbb{E}_{\eta^l}(\varphi) = \lim_{l\to\infty} \eta^l(\varphi) = \eta(\varphi)$. Where the target value is $\eta(\varphi)$.

Assume that we have access to a mean square integrable sequence of estimators $\{\overline{\eta}^l(\varphi)\}_{l\in\mathbb{N}_0}$, and that with it we can approximate the estimator $\overline{\eta}(\varphi)$, such that $\mathbb{E}(\overline{\eta}(\varphi)) = \eta(\varphi)$. An example of these kind of estimator can be constructed with the Euler-Maruyama discretization of an SDE, let say X from eq. (2.2), where in this case $\overline{\eta}^l(\varphi)$ can be the evaluation of φ at the discretized version of X of *level* l.

The mean of the estimators is such that $\mathbb{E}(\overline{\eta}^l(\varphi)) = \eta^l(\varphi)$. Its mean square relation with the limiting estimator is $\lim_{l\to\infty} \mathbb{E}[(\overline{\eta}^l(\varphi) - \overline{\eta}(\varphi))^2] = 0$. Then we can construct a sequence $\{\Xi_l\}_{l\in\mathbb{N}_0}$ such that $\mathbb{E}(\Xi_l) = \eta^l(\varphi) - \eta^{l-1}(\varphi)$ for $l \in \mathbb{Z}^+$, and $\mathbb{E}(\Xi_0) = \eta^0(\varphi)$, therefore having the telescoping property, i.e.

$$\mathbb{E}\left(\sum_{l=0}^{L} \Xi_l\right) = \sum_{l=0}^{L} \mathbb{E}(\Xi_l) = \sum_{l=0}^{L} \left(\eta^l(\varphi) - \eta^{l-1}(\varphi)\right) = \eta^L(\varphi), \tag{2.24}$$

where we denote $\eta_{-1}(\varphi) = 0$. As metioned before, this structure can be used in different ways to either construct MLMC estimators or different types of unbiased estimators.

2.4.1 Unbiased estimators

Unbiased estimators [26] or debiasing techniques are schemes that use the multilevel strategy along with an additional randomization to produce unbiased estimators. This technique is appealing given how easy is parallelizable. In this section, we will review some of the most popular unbiased estimators: the *single term* and the *coupled sum* estimators.

Let the measure of Ξ_l be denoted as $[\eta^l - \eta^{l-1}]$ for $l \in \mathbb{Z}^+$. The main idea behind debiasing techniques is to use a suitable p.m.f. $\mathbb{P}_L : \mathbb{N}_0 \to \mathbb{R}^+$ to generate random samples of $L \sim \mathbb{P}_L$, and subsequently generate samples from Ξ_L. The randomness of the level L allows the MC computation of quantities proportional to $[\eta^l - \eta^{l-1}](\varphi)$ for $l \in \mathbb{Z}^+$ and $\eta^0(\varphi)$, in the mean sense. Therefore allowing the use of the telescoping property of (2.24) that makes the estimator unbiased.

It is noted that choice of the p.m.f. $\mathbb{P}_L$ is of utmost importance and can influence variance and its relation with the cost.

Single term estimators

The single term estimator is defined as

$$\widehat{\eta(\varphi)_S} = \frac{\Xi_L}{\mathbb{P}_L},$$

where $\mathbb{P}_L(l) > 0$ for $l \in \mathbb{N}_0$, and $L \sim \mathbb{P}_L$. Its mean can be easily computed and it is

$$\begin{aligned}
\mathbb{E}\left(\widehat{\eta(\varphi)_S}\right) =& \mathbb{E}\left(\frac{\Xi_L}{P_L}\right), \\
=& \sum_{l \in \mathbb{N}_0} \mathbb{P}_L(l) \mathbb{E}\left[\frac{\Xi_l}{\mathbb{P}_L(l)}\right], \\
=& \sum_{l \in \mathbb{N}_0} \mathbb{E}\left[\Xi_l\right], \\
=& \eta^0(\varphi) + \sum_{l \in \mathbb{N}} \left[\eta^l - \eta^{l-1}\right](\varphi), \\
=& \lim_{l \to \infty} \eta^l(\varphi),
\end{aligned}$$

$$=\eta(\varphi).$$

Since the mean of the estimator is finite, we can compute just the second moment to obtain the variance

$$\begin{aligned}\mathbb{E}\left[\widehat{\eta(\varphi)_S}^2\right] &= \sum_{l\in\mathbb{N}_0} \mathbb{P}_L(l)\mathbb{E}\left[\frac{\Xi_l^2}{\mathbb{P}_L(l)^2}\right], \\ &= \sum_{l\in\mathbb{N}_0} \frac{\mathbb{E}\left[\Xi_l^2\right]}{P_L(l)},\end{aligned}$$

where the measure P_L has to be chosen so the second moment is finite.

With finite variance, an MC-like estimator can be defined as

$$\widehat{\eta(\varphi)_{S_M}} = \frac{1}{M}\sum_{i=1}^{M} \widehat{\eta(\varphi)_S}^{(i)} = \frac{1}{M}\sum_{i=1}^{M} \frac{\Xi_{L^{(i)}}}{\mathbb{P}_{L^{(i)}}}, \tag{2.25}$$

where $L^{(i)} \sim \mathbb{P}_L$ are M i.i.d. distributed, and $\Xi_{L^{(i)}}$ and $\Xi_{L^{(j)}}$ are independent for $j \neq i$.

The Monte Carlo canonical rate refers to the relation between the computational cost and the MSE for unbiased estimators with finite cost and variance, this relation is Cost $= \frac{1}{\text{MSE}} = \frac{1}{\text{Var}}$. The number of samples M of the unbiased estimator (2.25) allow us to control the variance as in the usual MC, i.e. the variance is inversely proportional to the number of samples M. We remark that this does not mean that we obtain the canonical rate since the expected cost of the estimator might be infinite.

Coupled sum estimator

The couple sum estimator is defined as

$$\widehat{\eta(\varphi)_C} = \sum_{l=0}^{L} \frac{\Xi_l}{\overline{\mathbb{P}}_L(l)},$$

where $\overline{\mathbb{P}}_L(k) \equiv \sum_{l=k}^{\infty} \mathbb{P}_L(l) > 0$ for each $k \in \mathbb{N}_0$ and $L \sim \mathbb{P}_L$. Notice that we must not only compute Ξ_L but also for Ξ_l in $0 \leqslant l < L$. The sampling of the estimators $\{\Xi_L, \Xi_{L-1}, \ldots, \Xi_0\}$ is allowed to be coupled, thus the name of the estimator (there is another version of this estimator in which the samples are independent, this estimator is called the Independent sum estimator). We give a sketch of the proof of the mean of the estimator:

$$\begin{aligned}
\mathbb{E}\left[\widehat{\eta(\varphi)_C}\right] &= \sum_{l \in \mathbb{N}_0} \mathbb{P}_L(l) \sum_{k=0}^{l} \frac{\mathbb{E}\left[\Xi_l\right]}{\overline{\mathbb{P}}_L(k)}, \\
&= \sum_{l \in \mathbb{N}_0} \mathbb{P}_L(l) \sum_{k=0}^{l} \frac{\left[\eta^k - \eta^{k-1}\right](\varphi)}{\overline{\mathbb{P}}_L(k)}, \\
&= \sum_{k \in \mathbb{N}_0} \frac{\left[\eta^k - \eta^{k-1}\right](\varphi)}{\overline{\mathbb{P}}_L(k)} \sum_{l=k}^{\infty} \mathbb{P}_L(k), \\
&= \sum_{k \in \mathbb{N}_0} \left[\eta^k - \eta^{k-1}\right](\varphi), \\
&= \lim_{l \to \infty} \eta^l(\varphi), \\
&= \eta(\varphi).
\end{aligned}$$

From Theorem 1 in [26] the second moment of the coupled estimator is

$$\mathbb{E}\left(\widehat{\eta(\varphi)_C}^2\right) = \sum_{l=0}^{\infty} \frac{\nu_l}{\overline{\mathbb{P}}_L(l)}, \tag{2.26}$$

where $\nu_l = \mathbb{E}[(\overline{\eta}^{l-1} - \overline{\eta})^2] - \mathbb{E}[(\overline{\eta}^l - \overline{\eta})^2]$. The second moment converges if

$$\sum_{l=0}^{\infty} \frac{\mathbb{E}[(\overline{\eta}^{l-1})]}{\overline{\mathbb{P}}_L(l)} < \infty. \tag{2.27}$$

We remark that in practice, the coupled sum estimator has a smaller variance than the single-term estimator. This difference in the variances has a counterpart in the complexity of the algorithm, given that the realizations of correlated

$\{\Xi_L, \Xi_{L-1}, \ldots, \Xi_0\}$ are harder to compute and store numerically.

One might think that the unbiased property of the estimators makes it better compared to other methods such as the MLMC, in general, this is not the case if we use the MSE as a criterion, given that the variance plays an important role and it is typically more significant for the unbiased estimators. Nonetheless, unbiased estimators are easily parallelizable, we can see some examples in [44, 45, 32]. The parallelization property is useful when we have multiple computer processors available, this allows to reduce the net computing time of the algorithm for a given variance.

2.4.2 Multilevel Monte Carlo

Given that we have access to i.i.d. samples from the random variables Ξ_l, $l \in \mathbb{N}_0$, it is possible to construct

$$\eta_{ML}^L = \sum_{l=0}^{L} \sum_{i=0}^{N_l} \frac{\Xi_l^{(i)}}{N_l},$$

which has mean $\mathbb{E}(\eta_{ML}^L) = \eta^L(\varphi)$, and where $\Xi_l^{(i)} \sim [\eta^l - \eta^{l-1}]$ and $\Xi_0^{(i)} \sim \eta^0$ are mutually independent for $(i, l) \in \mathbb{Z}^+ \times \mathbb{N}_0$. We also have that $\{N_l\}_{l=0}^L$, $N_l \in \mathbb{Z}^+$ is a sequence that represent the number of samples at each level. Under some regularity conditions on the variance of each estimator Ξ_l, the variance of the ML estimator is finite for all L.

The main goal of the MLMC is to reduce the cost of computing the estimation for a given MSE. To this end, an equilibrium has to be achieved between the bias and the variance of the estimator; the bias and the variance can be controlled by the level L and the sequence of samples $\{N_0, N_1, \ldots, N_L\}$, respectively. We are interested in the relation of the MSE with the computational cost of the estimator. In the context of this thesis, we are interested in the MLMC applied to SDEs; the following theorem (based on Theorem 3.1 in [27]) summarizes the pertinent error-to-cost rates.

Theorem 1. Let $\eta \in (\Omega, \mathcal{F})$ be the measure of the solution of an SDE at time t. Let $\{\eta^l\}_{l=0}^{\infty}$ for $\eta^l \in (\Omega, \mathcal{F})$ be the correspondent approximations of the measure with a numerical discretization with time step $\Delta_l = 2^{-l}$ and let $\varphi : \Omega \to \mathbb{R}$ be a η-integrable function.

If there exists independent estimators Ξ_l and positive constants $\alpha \geqslant \frac{1}{2}$, β, c_1, c_2, c_3 such that

- $|\eta^l(\varphi) - \eta(\varphi)| \leqslant c_1 \Delta_l^{\alpha}$,
- $\mathbb{E}(\Xi_l) = \begin{cases} \eta^0(\varphi) & \text{if} \quad l = 0, \\ \eta^l(\varphi) - \eta^{l-1}(\varphi) & \text{if} \quad l = 0. \end{cases}$
- $\mathrm{Var}(\Xi_l) \leqslant c_2 \Delta_l^{\beta}$,
- K_l, the computational cost of Ξ_l is bounded by $K_l \leqslant c_3 \Delta_l^{-1}$.

then there exists a positive constant c_4 such that for any $\epsilon < e^{-1}$ (where e is the Euler constant) there are values L and $\{N_l\}_{l=0}^{L}$ for which the multilevel estimator η_{ML}^{L} has a MSE bound

$$\mathrm{MSE} \leqslant \epsilon^2$$

with a computational complexity C bounded by

$$K \leqslant \begin{cases} c_4 \epsilon^{-2}, & \beta > 1, \\ c_4 \epsilon^{-2} (\log \epsilon)^2, & \beta = 1, \\ c_4 \epsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

In the SDE context, the most common way to construct ML estimators is to discretize the differential equation with time steps $\Delta_l = 2^{-l}$ and to use subsequent numbers l with the same underlying Brownian motion $\{W_t\}_{t\in\mathbb{R}^+\cup\{0\}}$. The Wiener processes have their respective discretizations $\{W_{\Delta_l k}\}_{k=0}^{T/\Delta_l}$ and $\{W_{\Delta_{l-1} k}\}_{k=0}^{T/\Delta_{l-1}}$ where the fist discretization includes all elements of the second one. A similar approach can also be used to construct unbiased estimators.

Chapter 3

Multilevel techniques applied to the EnKBF

In this chapter we discuss two numerical methods for the EnKBF, which are the Multilevel Ensemble Kalman-Bucy Filter (MLEnKBF) and the unbiased EnKBF. Additionally, tests and comparison of the methods are made, with more focus on the unbiased EnKBF, which is the main contribution of the thesis. In section 3.1 we present the MLEnKBF, its theoretical results, and the practical correspondent implementation. In sections 3.2 the unbiased estimators are presented along with its algorithmic representation. In section 3.3 we implement and compare the unbiased estimators and the MLEnKBF numerically.

3.1 Multilevel Ensemble Kalman-Bucy Filter

Using the main ideas of the MLMC and the EnKBF estimator presented in the past chapter (section 2.4.2 and 2.4.1 respectively), we have the necessary background material to introduce the MLEnKBF, which is postulated in [34]. The goal of the MLEnKBF is to approximate the KBF using the discretized EnKBF (2.23), where the levels in the MLMC scheme are identified with the discretization levels of the EnKBF. The number of samples per level is taken to be the size of the ensemble, N. Monte Carlo methods are used in order to approximate estimators up to a desired error. Specifically, for MSE=$\mathcal{O}(\epsilon^2)$ the Monte Carlo application of the EnKBF needs a computational cost of $\mathcal{O}(\epsilon^{-3})$. MLEnKBF is a MLMC methodology designed to improve this rates.

Let the discretized EnKBF equations at levels l and $l-1$, $l \in \mathbb{Z}_+$ for the Vanilla EnKBF (VEnKBF) be

$$(\mathbf{F1}) \begin{cases} \xi^{i,l}_{(k+1)\Delta_l} = \xi^{i,l}_{k\Delta_l} + A\xi^{i,l}_{k\Delta_l}\Delta_l + R^{\frac{1}{2}}\Delta\overline{W}^i_{k\Delta_l} \\ \qquad + P^{N,l}_{k\Delta_l}H^\top R_1^{-1}\left([Y_{(k+1)\Delta_l} - Y_{k\Delta_l}] - \left[H\xi^{i,l}_{k\Delta_l}\Delta_l + R_1^{\frac{1}{2}}\Delta\bar{V}^i_{k\Delta_l}\right]\right), \\ \xi^{i,l-1}_{(k+1)\Delta_{l-1}} = \xi^{i,l-1}_{k\Delta_{l-1}} + A\xi^{i,l-1}_{k\Delta_{l-1}}\Delta_{l-1} + R^{\frac{1}{2}}\Delta\overline{W}^i_{k\Delta_{l-1}} \\ \qquad + P^{N,l-1}_{k\Delta_{l-1}}H^\top R_1^{-1}\left([Y_{(k+1)\Delta_{l-1}} - Y_{k\Delta_{l-1}}] - \left[H\xi^{i,l-1}_{k\Delta_{l-1}}\Delta_{l-1} + R_1^{\frac{1}{2}}\Delta\bar{V}^i_{k\Delta_{l-1}}\right]\right), \end{cases} \tag{3.1}$$

for all $(k,i) \in \mathbb{N}_0 \times \{i\}_{i=1}^N$, where the discretization length is $\Delta_l = 2^{-l}$, and the initial conditions $\xi_0^{i,l} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(m_0, C_0)$, $\xi_0^{i,l-1} = \xi_0^{i,l}$. The approximations of the mean and covariance are

$$\begin{aligned} P^{N,l}_{k\Delta_l} &= \frac{1}{N-1}\sum_{i=1}^N \left(\xi^{i,l}_{k\Delta_l} - \eta^{N,l}_{k\Delta_l}(e)\right)\left(\xi^{i,l}_{k\Delta_l} - \eta^{N,l}_{k\Delta_l}(e)\right)^\top, \\ \eta^{N,l}_{k\Delta_l}(e) &= \frac{1}{N}\sum_{i=1}^N \xi^{i,l}_{k\Delta_l}, \\ P^{N,l-1}_{k\Delta_{l-1}} &= \frac{1}{N-1}\sum_{i=1}^N \left(\xi^{i,l-1}_{k\Delta_{l-1}} - \eta^{N,l-1}_{k\Delta_{l-1}}(e)\right)\left(\xi^{i,l-1}_{k\Delta_{l-1}} - \eta^{N,l-1}_{k\Delta_{l-1}}(e)\right)^\top, \\ \eta^{N,l-1}_{k\Delta_{l-1}}(e) &= \frac{1}{N}\sum_{i=1}^N \xi^{i,l-1}_{k\Delta_{l-1}}, \end{aligned}$$

and for the Deterministic EnKBF (DEnKBF)

$$(\mathbf{F2}) \begin{cases} \xi^{i,l}_{(k+1)\Delta_l} = \xi^{i,l}_{k\Delta_l} + A\xi^{i,l}_{k\Delta_l}\Delta_l + R^{\frac{1}{2}}\Delta\overline{W}^i_{k\Delta_l} \\ \qquad + P^{N,l}_{k\Delta_l}H^\top R_1^{-1}\left([Y_{(k+1)\Delta_l} - Y_{k\Delta_l}] - \frac{1}{2}H[\xi^{i,l}_{k\Delta_l}\Delta_l + \eta^{N,l}_{k\Delta_l}(e)]\right), \\ \xi^{i,l-1}_{(k+1)\Delta_{l-1}} = \xi^{i,l-1}_{k\Delta_{l-1}} + A\xi^{i,l-1}_{k\Delta_{l-1}}\Delta_{l-1} + R^{\frac{1}{2}}\Delta\overline{W}^i_{k\Delta_{l-1}} \\ \qquad + P^{N,l-1}_{k\Delta_{l-1}}H^\top R_1^{-1}\left([Y_{(k+1)\Delta_{l-1}} - Y_{k\Delta_{l-1}}] - \frac{1}{2}H[\xi^{i,l-1}_{k\Delta_{l-1}}\Delta_{l-1} + \eta^{N,l-1}_{k\Delta_{l-1}}(e)]\right), \end{cases} \tag{3.2}$$

for all $(k, i) \in \mathbb{N}_0 \times \{i\}_{i=1}^N$ and the same initial conditions of the vanilla variant. For $\varphi : \mathbb{R}^d \to \mathbb{R}^d$, the discretized EnKBF approximation of $\eta_t^l(\varphi) - \eta_t^{l-1}(\varphi) = [\eta_t^l - \eta_t^{l-1}](\varphi)$ is a Monte Carlo-type approximation:

$$[\eta_t^{N,l} - \eta_t^{N,l-1}](\varphi) = \frac{1}{N}\sum_{i=0}^{N}[\varphi(\xi_t^{i,l}) - \varphi(\xi_t^{i,l-1})], \tag{3.3}$$

and similarly for $\eta_t^0(\varphi)$ we have

$$\eta_t^{N,0}(\varphi) = \frac{1}{N}\sum_{i=0}^{N}\varphi(\xi_t^{i,0}), \tag{3.4}$$

for a time t in the discretization $\{k\Delta_{l-1}\}_{k=1}^N$. The MLEnKBF estimator is defined as

$$\eta_t^{ML}(\varphi) \equiv \eta_t^{N_0,0}(\varphi) + \sum_{l=1}^{L}[\eta_t^{N_l,l} - \eta_t^{N_l,l-1}](\varphi).$$

for $L \in \mathbb{Z}^+$ and $N_l \in \{2, 3, \dots\}$, $l \in \{j\}_{j=0}^L$, and where each term in the sum is computed independently. Note that this estimator is not an exact application of the MLMC presented in 2.4.2 having a slight difference on the estimation of (3.3) because the samples ξ^i are correlated as opposed to the independent samples in MLMC. The detailed application of the MLEnKBF can be found in algorithm 2.

3.1.1 Error-to-cost rate of MLEnKBF

In [34] we are presented with a ML analysis for the vanilla variant **(F1)**, with a slight modification. The theory derived is presented for a simpler ideal i.i.d. coupled particle system for $(i, k) \in \{i\}_{i=1}^N \times \mathbb{N}_0$:

$$\begin{aligned}
\zeta_{(k+1)\Delta_l}^{i,l} &= \zeta_{k\Delta_l}^{i,l} + A\zeta_{k\Delta_l}^{i,l}\Delta_l + R^{1/2}\Delta\overline{W}_{k\Delta_l}^i + U_{k\Delta_l}^l\Big([Y_{(k+1)\Delta_l}^i - Y_{k\Delta_l}^i] \\
&\quad - \Big[H\zeta_{k\Delta_l}^{i,l}\Delta_l + R_1^{1/2}\Delta\overline{V}_{k\Delta_l}^i\Big]\Big), \\
\zeta_{(k+1)\Delta_{l-1}}^{i,l-1} &= \zeta_{k\Delta_{l-1}}^{i,l-1} + A\zeta_{k\Delta_{l-1}}^{i,l-1}\Delta_{l-1} + R^{1/2}\Delta\overline{W}_{k\Delta_{l-1}}^i
\end{aligned}$$

$$+ U^{l-1}_{k\Delta_{l-1}}\Big([Y^i_{(k+1)\Delta_{l-1}} - Y^i_{k\Delta_{l-1}}] - \Big[H\zeta^{i,l-1}_{k\Delta_{l-1}}\Delta_{l-1} + R_1^{1/2}\Delta\overline{V}^i_{k\Delta_{l-1}}\Big]\Big),$$

where $U^s_{k\Delta_s} = P^s_{k\Delta_s}H^\top R_1^{-1}$, $s \in \{l-1, l\}$, and also $\xi_0^{i,l} = \xi_0^{i,l-1} = \zeta_0^{i,l} = \zeta_0^{i,l-1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(m_0, C_0)$, with $\zeta^{i,l}_{(k+1)\Delta_l}|\mathcal{Y}_{(k+1)\Delta_l} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(m^l_{(k+1)\Delta_l}, P^l_{(k+1)\Delta_l})$, where

$$m^l_{(k+1)\Delta_l} = m^l_{k\Delta_l} + Am^l_{k\Delta_l}\Delta_l + U^l_{k\Delta_l}\Big([Y_{(k+1)\Delta_l} - Y_{k\Delta_l}] - Cm^l_{k\Delta_l}\Delta_l\Big),$$
$$P^l_{(k+1)\Delta_l} = P^l_{k\Delta_l} + \text{Ricc}(P^l_{k\Delta_l})\Delta_l + (A - P^l_{k\Delta_l}S)P^l_{k\Delta_l}(A^\top - SP^l_{k\Delta_l})\Delta_l^2,$$

and similarly for level $l-1$. Where in this case the Riccati drift is defined as

$$\text{Ricc}(Q) = AQ + QA^\top - QSQ + R, \quad \text{with} \quad \text{and } S \equiv H^\top R_1^{-1}H.$$

Then from this i.i.d. coupled system, with a slight change of notation, we set the measure of interest as,

$$\hat{\eta}^{ML}_t(\varphi) \equiv \hat{\eta}^{N_0,0}_t(\varphi) + \sum_{l-1}^{L}[\hat{\eta}^{N_l,l}_t - \hat{\eta}^{N_l,l-1}_t](\varphi),$$

where $[\hat{\eta}^{N_l,l}_t - \hat{\eta}^{N_l,l-1}_t](\varphi) = \frac{1}{N_l}\sum_{i=1}^{N_l}[\varphi(\zeta^{i,l}_t) - \varphi(\zeta^{i,l-1}_t)]$.

The motivation behind doing so was that the i.i.d. coupled system is simpler to work with, and through a limit analysis, one can show that the $\hat{\eta}_t$ coincides with η_t as $N \to \infty$. In particular, it has been shown in [34] that they coincide with high probability, i.e., for any $\varepsilon > 0$ and $q > 0$

$$\mathbb{P}\left(\left|\left[\eta^{ML}_t - \hat{\eta}^{ML}_t\right](e)\right| > \varepsilon\right) \leqslant \frac{C}{\varepsilon^{2q}}\left(\sum_{l=0}^{L}\frac{1}{N_l^{q/2}}\right)$$

where C is a constant that can depend on (q, L, t), but not on $\{N_l\}_{l=0}^L$. Thus for a large number of particles the multilevel estimator are similar. (We leave out the rest of the technicalities for this thesis as it is not our primary goal, but refer to the reader to [34]). Now provided this new ML estimator, we recall the main theorem from [34] which is an error-to-cost rate of the MSE w.r.t. to the MLEnKBF.

Theorem 2. For any $T \in \mathbb{N}$ fixed and $t \in [0, T-1]$ there exists a $C < +\infty$ such that for any $(L, N0, N_1, \ldots, N_L) \in \mathbb{N} \times \{2, 3, \ldots\}^{L+1}$,

$$\mathbb{E}\left[\left\|[\hat{\eta}_t^{ML} - \eta_t](e)\right\|_2^2\right] \leqslant C\left(\sum_{l=0}^{L} \frac{\Delta_l}{N_l} + \sum_{l=1}^{L} \sum_{q=1, q\neq l}^{L} \frac{\Delta_l \Delta_q}{N_l N_q} + \Delta_L^2\right).$$

Proof. The proof of the above theorem for **(F1)** can be found in [34]. □

Algorithm 2 (**MLEnKBF**) Multilevel Estimation of the Ensemble Kalman–Bucy Filter

---

1. **Input:**
   - Target level $L \in \mathbb{N}$
   - Start level $l_* \in \mathbb{N}$ such that $l_* < L$
   - Number of particles on each level $\{N_l\}_{l=l_*}^{L}$
   - The time parameter $T \in \mathbb{N}$
   - Initial independent ensembles $\left\{\{\tilde{\xi}_0^{i,l_*}\}_{i=1}^{N_{l_*}}, \cdots, \{\tilde{\xi}_0^{i,L}\}_{i=1}^{N_L}\right\}$
2. **Initialize:** Set $l = l_*$. For $(i, k) \in \{1, \cdots, N_l\} \times \{0, \cdots, T\Delta_l^{-1} - 1\}$, set $\{\xi_0^{i,l}\}_{i=1}^{N_l} = \{\tilde{\xi}_0^{i,l}\}_{i=1}^{N_l}$ .Then using (2.23), return $\eta_T^{N_l,l}(\varphi)$.
3. **Iterate:** For $l \in \{l_* + 1, \cdots, L\}$ and $(i, k) \in \{1, \cdots, N_l\} \times \{0, \cdots, T\Delta_l^{-1} - 1\}$, set $\{\xi_0^{i,l-1}\}_{i=1}^{N_l} = \{\xi_0^{i,l}\}_{i=1}^{N_l} = \{\tilde{\xi}_0^{i,l}\}_{i=1}^{N_l}$ (the one which corresponds to the case used in Step 2.). Then using (2.23), return $\eta_T^{N_l,l-1}(\varphi)$ & $\eta_T^{N_l,l}(\varphi)$.
4. **Output:** Return the multilevel estimation of the EnKBF:

$$\eta_T^{ML}(\varphi) = \eta_T^{N_{l_*},l_*}(\varphi) + \sum_{l=l_*+1}^{L} \{\eta_T^{N_l,l}(\varphi) - \eta_T^{N_l,l-1}(\varphi)\}. \tag{3.5}$$

The above theorem can be understood as that, in order to attain an MSE of order $\mathcal{O}(\epsilon^2)$, the cost associated with this is of order $\mathcal{O}(\epsilon^{-2}\log(\epsilon)^2)$. So by considering the MLEnKBF this provides a reduction in cost compared to the EnKBF, which, to attain an MSE of the same magnitude, the associated cost is more expensive, i.e.,

$\mathcal{O}(\epsilon^{-3})$. This is relevant to mention because, as we will see in the numerics, we will use this error-to-cost rate to compare the unbiased estimators and the MLEnKBF.

We emphasis that the result from Theorem 2 is specific to the vanilla variant of the MLEnKBF, i.e. **(F1)**, which is for the following modified ideal i.i.d. coupled system

$$\zeta^{i,l}_{(k+1)\Delta_l} = \zeta^{i,l}_{k\Delta_l} + A\zeta^{i,l}_{k\Delta_l}\Delta_l + R_1^{1/2}[\overline{W}^i_{(k+1)\Delta_l} - \overline{W}^i_{k\Delta_l}] + U^l_{k\Delta_l}\Big([Y^i_{(k+1)\Delta_l} - Y^i_{k\Delta_l}] - \frac{1}{2}\Big[C\zeta^{i,l}_{k\Delta_l}\Delta_{l-1} + Cm^l_{k\Delta_l}\Delta_l\Big]\Big), \tag{3.6}$$

$$\zeta^{i,l-1}_{(k+1)\Delta_{l-1}} = \xi^{i,l-1}_{k\Delta_{l-1}} + A\zeta^{i,l-1}_{k\Delta_{l-1}}\Delta_{l-1} + R_1^{1/2}[\overline{W}^i_{(k+1)\Delta_{l-1}} - \overline{W}^i_{k\Delta_{l-1}}] + U^{l-1}_{k\Delta_{l-1}}\Big([Y^i_{(k+1)\Delta_{l-1}} - Y^i_{k\Delta_{l-1}}] - \frac{1}{2}\Big[C\zeta^{i,l-1}_{k\Delta_{l-1}}\Delta_{l-1} + Cm^{l-1}_{k\Delta_{l-1}}\Delta_{l-1}\Big]\Big). \tag{3.7}$$

For **(F2)** the new coupled system is given by the equations (3.6) - (3.7).

Remark. As conducted in [34, 46], the numerical experiments were tested using the originally stated ensemble Kalman–Bucy filters, while the theory was derived for the i.i.d. coupled particle systems. As stated in those previous works, our reason for this is that the recursion of the MLEnKBF make it difficult to derive exact multilevel rates, so instead, one can consider a simpler system and use a limiting argument, which is to say as $N \to \infty$ these systems behave similarly and coincide. This can be found in [Prop. 2.1., [34]].

3.2 Unbiased Estimators

In this section, we present the method and algorithm of the main topic of this thesis: unbiased techniques applied to the EnKBF. Analogously as the MLEnKBF is not an exact application of the MLMC, the unbiased EnKBF is not an exact application of the debiasing schemes to the EnKBF. The main reason for this is that our debiasing scheme applies a *double randomization*, similar to the one proposed and implemented in [32]. Double randomization is possible for the EnKBF because there are two kinds of levels of *discretization* or two *degrees of freedom*. The first kind is the time discretization level, parametrized by l. The second kind of level depends on the number of particles of the ensemble N_p. We consider it a level of discretization given that regardless on the number of particles, as long as it is finite, the time discretized EnKBF (or even in the continuous EnKBF) is biased, this changes when we take the number of particles limit to infinity. The estimator proposed here can be understood as the application of two steps, each of them associated with a different randomization. In the first step, we randomize on the number of particles, obtaining unbiased estimators time discretized EnKBF, which are in turn used to secure unbiased estimators of the EnKBF.

Let $\{N_p\}_{p\in\mathbb{N}_0}$ be an increasing sequence of positive integers with $N_0 > 1$ such that $\lim_{p\to\infty} N_p = \infty$. Furthermore, let $\eta_t^{N_p,0}(e)$ (3.4) be a EnKBF of ensemble size N_p, and also be an estimator of $\eta_t^0(e) = m_t^0$. Similarly $[\eta_t^{N_p,l} - \eta_t^{N_p,l-1}](e)$ (3.3) is a EnKBF estimator of $[\eta_t^l - \eta_t^{l-1}](e) = m_t^l - m_t^{l-1}$. We recall that $e(\xi) = \xi$. Note that here we are trying to estimate the time discretized levels, and that the time discretized EnKBF estimators are biased. In order to obtain unbiased estimators of the time continuous EnKBF, we resort to using an additional randomization associated to the number of the ensemble. This randomization is different and independent on the randomization applied to the levels of the time discretization.

The following proposition (Proposition (2.1) from [34]) will help us to establish the consistency of the estimators

Proposition 1. For any $(l, t, k_1) \in \mathbb{N}_0 \times \mathbb{R}^+ \times \{0, 1, \ldots, \triangle_l^{-1}\}$, almost surely

$$\lim_{N\to\infty} [\eta_{t+k_1\Delta_l}^{N,l} - \eta_{t+k_1\Delta_l}](e) = 0.$$

This result can be rearranged to obtain consistent estimators, i.e.

$$\lim_{p\to\infty}\eta_t^{N_p,0}(e)=\eta_t^0(e),\quad \lim_{p\to\infty}\left\{\left[\eta_t^{N_p,l}-\eta_t^{N_p,l-1}\right](e)\right\}=\left[\eta_t^l-\eta_t^{l-1}\right](e),$$
$$\forall\,(l,t,k_1)\in\mathbb{Z}^+\times\mathbb{R}^+\left\{0,1,\ldots,\triangle_l^{-1}\right\}.$$

almost surely.

Let $\mathbb{P}_P(p):\mathbb{N}_0\to(0,1)$, and $\mathbb{P}_L(l):\mathbb{N}_0\to(0,1)$ be strictly positive probability measures. Now we define the random quantity that includes both degrees of freedom p and l:

$$\Xi_{l,p}\equiv\begin{cases}\frac{1}{\mathbb{P}_P(p)\mathbb{P}_L(l)}\left[\eta_t^{N_{0:p},0}-\eta_t^{N_{0:p-1},0}\right](\varphi), & \text{if } l=0,\\ \frac{1}{\mathbb{P}_P(p)\mathbb{P}_L(l)}\left(\left[\eta_t^{N_{0:p},l}-\eta_t^{N_{0:p},l-1}\right](\varphi)-\left[\eta_t^{N_{0:p-1},l}-\eta_t^{N_{0:p-1},l-1}\right](\varphi)\right), & \text{otherwise.}\end{cases}$$

for all $(l,p)\in\mathbb{N}_0\times\mathbb{N}_0$, with the convention $N_{-1}=0$ and $\eta_t^{N_{0:-1},l}(\varphi)=0$, where the Monte Carlo estimators $\eta_t^{N_{0:p},0}(\varphi)$ of $\eta_t^0(\varphi)$, and $[\eta_t^{N_{0:p},l}-\eta_t^{N_{0:p},l-1}](\varphi)$ of $\left[\eta_t^l-\eta_t^{l-1}\right](\varphi)$ will be discussed below. This quantity has telescoping properties in degrees of freedom p and l.

Let us set $\Xi_l=\Xi_{l,P}$, where P is a random variable with p.m.f. $\mathbb{P}_P(p)$. From section 2.4.1 we get

$$\mathbb{P}_L(l)\mathbb{E}\left[\Xi_l\right]=\begin{cases}\eta_t^0(\varphi) & \text{if } l=0\\ \left[\eta_t^l-\eta_t^{l-1}\right](\varphi) & \text{otherwise}\end{cases},$$

which is a single term estimator in the level l, with finite variance provided that the second moment is finite, i.e.

$$\mathbb{E}\left[\Xi_l^2\right]=\sum_{p\in\mathbb{N}_0}\mathbb{P}_P(p)\mathbb{E}\left[\Xi_{l,p}^2\right]<+\infty.$$

Let us define $\Xi \equiv \Xi_L(l)$ as a single term-type estimator in both degrees of freedom, where L is a random variable with p.m.f. $\mathbb{P}_L(l)$. Additionally, it attains a finite variance if the second moment is finite

$$\mathbb{E}\left(\Xi^2\right) = \sum_{(p,l)\in\mathbb{N}_0\times\mathbb{N}_0} \mathbb{P}_P(p)\mathbb{P}_L(l)\mathbb{E}\left[\Xi_{l,p}^2\right] < +\infty.$$

Now we consider the computation of the estimators $\eta_t^{N_{p:0},0}(\varphi)$ and $[\eta_t^{N_{p:0},l} - \eta_t^{N_{p:0},l-1}](\varphi)$, which is analogous to the one presented in [32]. Let us start with $\eta^{N_{p:0},0}(\varphi)$. To form our approximation with N_0 samples we run the EnKBF with N_0 samples. Next we run the EnKBF independently of the first EnKBF with $N_1 - N_0$ samples, and continue this process with $N_2 - N_1, \ldots, N_p - N_{p-1}$ samples, for any $p \geqslant 2$. Now we can define

$$\eta_t^{N_{0:p},0}(\varphi) \equiv \sum_{q=0}^{p} \left(\frac{N_q - N_{q-1}}{N_p}\right) \eta_t^{N_q - N_{q-1},0}(\varphi),$$
$$\eta_t^{N_{0:p-1},0}(\varphi) \equiv \sum_{q=0}^{p-1} \left(\frac{N_q - N_{q-1}}{N_{p-1}}\right) \eta_t^{N_q - N_{q-1},0}(\varphi),$$

where, as defined before

$$\eta_t^{N_q - N_{q-1},0}(\varphi) = \frac{1}{N_q - N_{q-1}} \sum_{i=N_{q-1}+1}^{N_q} \varphi\left(\xi_t^{i,0}\right),$$

where the values $\{\xi_t^{n,0}\}_{n=1}^{N_0}$ are generated from the first EnKBF, $\{\xi_t^{n,0}\}_{n=N_0+1}^{N_1}$ secondly and so on. A similar procedure is performed to obtain the estimator $[\eta_t^{N_{p:0},l} - \eta_t^{N_{p:0},l-1}](\varphi)$, we generate $p+1$ mutually independent EnKBFs with number of particles $N_0, N_1 - N_0, \ldots, N_p - N_{p-1}$. For each EnKBF we use equations (3.1) and (3.2) to compute the levels $\{l, l-1\}$ of the vanilla and deterministic variants, respectively; building them correlated. Where $\{(\xi_t^{n,l}, \xi_t^{n,l-1})\}_{n=1}^{N_0}$ are the N_0 particles of the first ensemble, $\{(\xi_t^{n,l}, \xi_t^{n,l-1})\}_{n=N_0+1}^{N_1}$ of the second and so on. Now we construct

$$[\eta_t^{N_{0:p},l} - \eta_t^{N_{0:p},l-1}](\varphi) \equiv \sum_{q=0}^{p} \left(\frac{N_q - N_{q-1}}{N_p}\right) [\eta_t^{N_q-N_{q-1},l} - \eta_t^{N_q-N_{q-1},l-1}](\varphi), \quad (3.8)$$

$$[\eta_t^{N_{0:p-1},l} - \eta_t^{N_{0:p-1},l-1}](\varphi) \equiv \sum_{q=0}^{p-1} \left(\frac{N_q - N_{q-1}}{N_{p-1}}\right) [\eta_t^{N_q-N_{q-1},l} - \eta_t^{N_q-N_{q-1},l-1}](\varphi), \quad (3.9)$$

where the estimates

$$[\eta_t^{N_q-N_{q-1},l} - \eta_t^{N_q-N_{q-1},l-1}](\varphi) = \frac{1}{N_q - N_{q-1}} \sum_{i=N_{q-1}+1}^{N_q} \left(\varphi\left(\xi_t^{i,l}\right) - \varphi\left(\xi_t^{i,l-1}\right)\right),$$

are sampled only once and used for (3.8) and (3.9). We will give a detailed procedure on how to compute estimator in algorithm 3. Furthermore, we will consider an additional debiasing scheme, where we use a coupled sum type of unbiased estimator in the number of particles. The details of its implementation will also be provided in algorithm 3. As mentioned on 2.4.1, the choice of the probability measures $\mathbb{P}_L(l)$ affects the error-to-cost rate, in the case of the double randomization the measure $\mathbb{P}_P(p)$ plays a similar role. The choice of the measures and the practical implementation of the algorithm will discussed in the following section.

Remark. With the above discussion on the unbiased estimator, we note that we do not prove that our estimator is unbiased and has finite variance. This is for two reasons, firstly because of the difficulty, and secondly, the multilevel analysis only derived in [34] holds only the i.i.d. filter, not the EnKBF described in 3.1.1 Therefore, we leave this for future work.

3.3 Numerical Results

In this section, we will describe our algorithm and its practical implementation along with the numerical results obtained. A brief discussion over the choice of the parameters of the algorithm is also presented. The computation of the expected cost of the

Algorithm 3 Unbiased Estimate of the Ensemble Kalman-Bucy Filter

1: **Input:** Two positive probability mass functions $\mathbb{P}_L$ and $\mathbb{P}_P$ on $\mathbb{N}_0$. For $i = 1, \cdots, M$, run Step 2. independently:

2: **Iterate:** Sample $(l_i, p_i) \in \mathbb{N}_0^2$ from $\mathbb{P}_L \otimes \mathbb{P}_P$.

I. If $l_i = 0$. For $s \in \{0, \dots, p_i\}$, independently generate an EnKBF with $N_s - N_{s-1}$ samples where $0 = N_{-1} < N_0 < \cdots < N_{p_i}$ are given(e.g. $N_{p_i} = N_0\, 2^{p_i}$, $N_0 \in \{2, 3, \dots\}$ fixed). Return

$$\Xi_{l_i,p_i} = \frac{1}{\mathbb{P}_L(l_i)\mathbb{P}_P(p_i)} \Big(\eta_t^{N_{0:p_i},0}(\varphi) - \eta_t^{N_{0:p_i-1},0}(\varphi)\Big),$$

II. Otherwise for $s \in \{0, \dots, p_i\}$, independently generate coupled EnKBFs with $N_s - N_{s-1}$ samples where $0 = N_{-1} < N_0 < \cdots < N_{p_i}$ are given (e.g. $N_{p_i} = N_0\, 2^{p_i}$, $N_0 \in \{2, 3, \dots\}$ fixed). Return

$$\Xi_{l_i,p_i} = \frac{1}{\mathbb{P}_L(l_i)\mathbb{P}_P(p_i)} \Big\{ \Big(\eta_t^{N_{0:p_i},l_i}(\varphi) - \eta_t^{N_{0:p_i},l_i-1}(\varphi)\Big) - \Big(\eta_t^{N_{0:p_i-1},l_i}(\varphi) - \eta_t^{N_{0:p_i-1},l_i-1}(\varphi)\Big) \Big\},$$

For the *coupled sum estimator* we instead return

$$\Xi_{l_i,p_i} = \sum_{s=0}^{p_i} \frac{1}{\sum_{q=s}^{\infty} \mathbb{P}_L(l_i)\mathbb{P}_P(q)} \Big\{ \Big(\eta_t^{N_{0:s},l_i}(\varphi) - \eta_t^{N_{0:s},l_i-1}(\varphi)\Big) - \Big(\eta_t^{N_{0:s-1},l_i}(\varphi) - \eta_t^{N_{0:s-1},l_i-1}(\varphi)\Big) \Big\},$$

3: **Output:** Return the unbiased estimator

$$\widehat{\eta_t(\varphi)} = \frac{1}{M} \sum_{i=1}^{M} \Xi_{l^i,p^i}. \tag{3.10}$$

unbiased estimator Ξ is

$$\text{cost} = \sum_{(l,p) \in \mathbb{N}_0 \times \mathbb{N}_0} \mathbb{P}_L(l)\mathbb{P}_P(p)\, \text{cost}_{l,p} = \sum_{(l,p) \in \mathbb{N}_0 \times \mathbb{N}_0} \mathbb{P}_L(l)\mathbb{P}_P(p)\mathcal{O}(t2^{p+l}).$$

In most of the practical cases (see, for example, [32, 26]) is extremely difficult or impossible to obtain measures $\mathbb{P}_L$ or $\mathbb{P}_P$ such that the variance and the expected cost are finite. In such cases, it is opted by setting a measure that leaves the variance finite. In [32] the authors find one configuration with finite variance in which for a MSE= $\mathcal{O}(\epsilon^2)$, the cost is $\mathcal{O}(\epsilon^{-2}\log(\epsilon)^{2+\delta})$ for any $\delta > 0$, and even though we have this rate, the expected cost is still infinite. In this case an error-to-cost rate is established, nevertheless, a bound for the maximum cost is desired. This leads to the truncation of the level measures $\mathbb{P}$ in practical computations of the unbiased estimators.

Let $(p_{max}, l_{max}) \in \mathbb{N}_0 \times \mathbb{N}_0$ be the truncation level and $\widetilde{\mathbb{P}}_P(p) : \{0, 1, \ldots, p_{max}\} \to (0, 1)$ and $\widetilde{\mathbb{P}}_L(l) : \{0, 1, \ldots, l_{max}\} \to (0, 1)$ be the truncated probability measures. The (truncated) unbiased estimator is defined as $\widetilde{\Xi} \equiv \Xi_{\widetilde{L},\widetilde{P}}$, with $\widetilde{L} \sim \widetilde{\mathbb{P}}_L$ and $\widetilde{P} \sim \widetilde{\mathbb{P}}_P$. Its computation follows exactly algorithm 3 replacing the measures $\mathbb{P}_P(p)$ and $\mathbb{P}_L(l)$ by $\widetilde{\mathbb{P}}_P(p)$ and $\widetilde{\mathbb{P}}_L(l)$, respectively. There are some implications of the truncation in the cost, bias and variance of the estimator, the fist one, is that none of them diverges, the maximum cost of computing, say, a fixed $M < \infty$ number of samples is bounded, as intended. Unfortunately, the estimator is no longer unbiased. Nevertheless, given the telescoping property of the estimator, the bias can be controlled with the truncation levels p_{max} and l_{max}.

It is known that the unbiased estimators and MLMC schemes are closely related [28]. In the truncated setting, this similarity is more recognizable, given that we have to control the bias and the variance.

In the following we show the results of the numerical tests and discuss the choice of the different parameters of the model such as the initial number of particles N_0 for the first level $p = 0$, the initial time discretization l^*, and the truncated measures. The tests were carried out with a $d-$dimensional Ornstein-Uhlenbeck signal process $X = \{X_t\}_{t\geqslant 0}$ with $d-$dimensional linear observations $Y = \{Y_t\}_{t\geqslant 0}$ and additive Gaussian noise, i.e. the state space model is

$$
\begin{aligned}
dX_t =& AX_t dt + R^{\frac{1}{2}} dW_t, \quad \forall t \geqslant 0, \\
dY_t =& HX_t dt + R_1^{\frac{1}{2}} dV_t,
\end{aligned}
$$

with initial conditions $X_0 \sim \mathcal{N}(6, I_d)$, where I_d is the $d-$dimensional identity matrix, and where the matrices A, R, H and R_1 are generated randomly. We simulate exactly from the process X in the discretization $\{\Delta_l k\}_{k=0}^{T/q}$, certain regularity conditions on the matrices are necessary so we can simulate exactly [47]. In this section we consider the function $\varphi(\xi) = e(\xi) = \xi$. In fig. 3.1 we can see a realization of the signal process, its correspondent KBF and some EnKBF samples. The observations (not displayed) that generated the dynamics for the figure will be used in two of the tests below ($d = 2$, $T = 30$ and $T = 80$), additionally, the plot gives us an idea of the variation of the EnKBF with respect to the KBF.

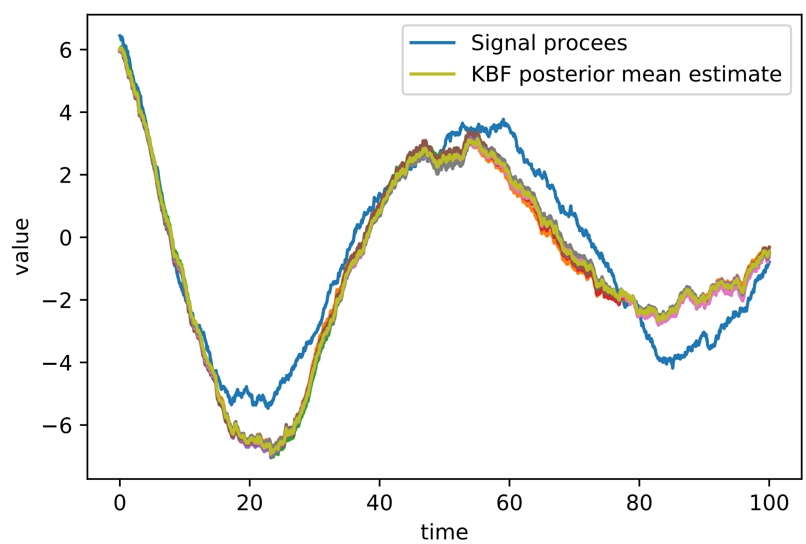

Figure 3.1: On display is the first dimension of a two dimensional signal process (blue) along a span of time $t \in (0, 100)$. The corresponding KBF (light green) and realizations of the EnKBF are also showed (the remaining colors).

The telescoping property of the estimator ensures that $\mathbb{E}(\widetilde{\Xi}) = \mathbb{E}\left(\eta_t^{N_{p_{max}}, l_{max}}(e)\right)$, thus, for a given MSE$= \epsilon^2$ we control the bias in the following way

$$
\begin{aligned}
|\,\text{bias}\,| = \left\| \mathbb{E}\left(\eta_t^{N,l}(e) - \eta_t(e)\right) \right\| &= \left\| \mathbb{E}\left(\eta_t^{N,l}(e) - \eta_t^l(e)\right) + \mathbb{E}\left(\eta_t^l(e) - \eta_t(e)\right) \right\|, \\
&\leqslant k_1 \Delta_l + k_2 \left(\frac{1}{N}\right) = \frac{\epsilon}{\sqrt{2}},
\end{aligned}
$$

where $\| \cdot \|$ is the euclidean norm and $k_1 \in \mathbb{R}^+$ $k_2 \in \mathbb{R}^+$ are constants that do not depend on l or N, respectively.

The results $\left\|\mathbb{E}\left(\eta_t^{N,l}(e) - \eta_t^l(e)\right)\right\| \leqslant \frac{k_2}{N}$ and $\left\|\mathbb{E}\left(\eta_t^l(e) - \eta_t(e)\right)\right\| \leqslant k_1\Delta_l$ are assumed, and numerically checked in order to obtain the constants k_1 and k_2. A figure with the numerical bounds can be seen in 3.2. For a specific setting, we can observe that the numerical rates coincide with the rates of the assumed relations, i.e. the slopes of the figures are approximately one. We obtain these results via Richardson extrapolation. Analogously to the MLEnKBF procedure, we choose the truncation levels (p_{max}, l_{max}) so bias$^2 \leqslant \epsilon^2/2$.

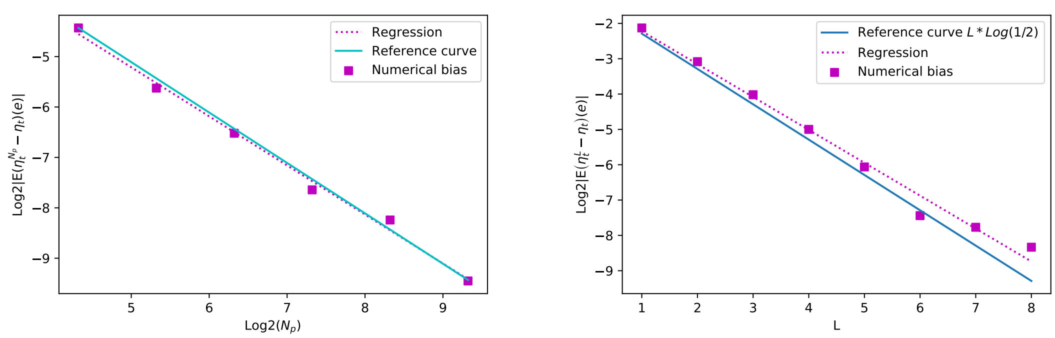

Figure 3.2: Bias in terms of the number of particles (left) and the discretization level (right). The dimension considered is $d = 2$ with final time $T = 30$.

The variance of the estimator $\widehat{\eta_t(\varphi)}$ (3.10) is proportional to M^{-1}, we can control it by estimating the variance of $\tilde{\Xi}$, which can be computed in parallel with $\widehat{\eta_t(\varphi)}$, i.e., the samples used to compute the sample variance can be used for the estimator. The parallel calculation of the variance along the estimator can be seen as a advantage of the unbiased technique with respect to the MLEnKBF. This because we need to estimate some constants related to the variance of the MLEnKBF, such constants have to be estimated before the implementation of the algorithm.

Two different setting were considered to compute the relation of the cost and the error, $d = 2$ with $T = 30$ and $d = 2$ with $T = 80$, we can see them in fig. 3.3 and 3.4 respectively. In the figure we can see four types of lines, each of them with its respective rate m. Three of the curves correspond to different algorithms and one of them is plotted as a reference (blue, dashed) that represents the error-to-cost relation targeted by the MLEnKBF, i.e. MSE= $\mathcal{O}(\epsilon^2)$, cost= $\mathcal{O}(\epsilon^{-2}\log(\epsilon)^2)$.

The first observation we make is that the unbiased estimators, coupled sum and single term, have similar rates and costs, such rates change across the different settings, but we cannot say that there is a substantial difference. Similar rates are expected for the unbiased estimators since, in theory, both algorithms have the same rates (for more details on the rates, we refer the reader to [26]). Similar costs are also expected since for these settings, the maximum level of truncation was relatively small, $p_{max} = 3$. On the other hand, there is a significant difference in the cost of the MLEnKBF compared to the unbiased methods; the unbiased EnKBF are approximately 60 times more expensive. This can be explained by the variance introduced by the randomization, which is proportional to the cost for a given truncation. Regarding the error-to-costs rates of the MLEnKBF compared to the unbiased EnKBF, we observe that they are similar across the different configurations. We notice that there is a slight improvement in the rates of fig. 3.4 of the DEnKBF compared to VEnKBF. In summary, the only consistent difference across the settings, of the MLEnKBF with respect to the unbiased EnKBF, is the cost, which is much higher for the unbiased EnKBF. We also note that the rates are similar to the rate of the reference curve.

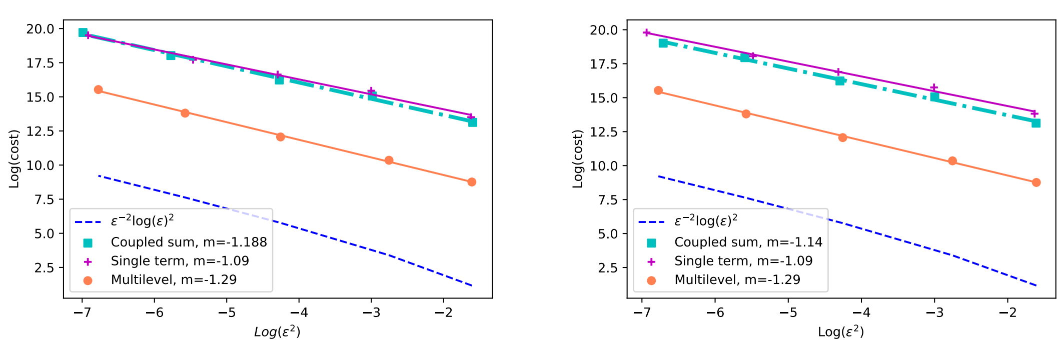

Figure 3.3: Error-to-cost rate for the Vanilla (left) and Deterministic (right) EnKBF. The dimension considered is $d = 2$ with final time $T = 30$.

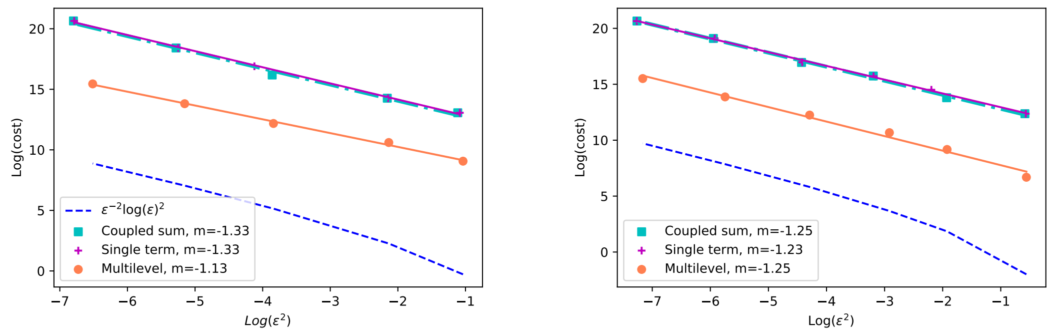

Figure 3.4: Error-to-cost rate for the Vanilla (left) and Deterministic (right) EnKBF. The dimension considered is $d = 2$ with final time $T = 80$.

Now we discuss the choice of the parameters of the algorithm. We choose a geometrical sequence for the number of particles in the ensemble $N_p = N_0 2^p$ following the results in [32]. In this thesis we considered truncated measures in the form $\widetilde{\mathbb{P}}_P(p) \propto 2^{-\alpha p}$ and $\widetilde{\mathbb{P}}_L(l) \propto 2^{-\alpha l}$ with $\alpha \in \{3/2, 2\}$. We chose specifically the geometric probability distributions following the resemblance of the unbiased estimators with the MLMC methods.

For the computations that generated 3.3 and 3.4 we used $\alpha = 3/2$, furthermore we computed the variance and the estimator with the same samples. Numerically, these two values $\{3/2, 2\}$ performed better than the other options, it is noted that a optimization of this value is not possible for the moment because we ignore the structure of the variance of $\Xi_{l,p}$, this result can be pursued in future work.

The second moment of the truncate unbiased estimator is

$$\mathbb{E}(\widetilde{\Xi}^2) = \sum_{(p,l)=(0,0)}^{p_{max},l_{max}} \mathbb{P}_{\widetilde{P}}(p)\mathbb{P}_{\widetilde{L}}(l)\mathbb{E}\left[\Xi_{l,p}^2\right], \tag{3.11}$$

Relatively high values of α, i.e. $\alpha \geqslant 2$, make the sampling of levels close to (p_{max}, l_{max}) highly unlikely, and since $\Xi_{l,p} \propto (\mathbb{P}_{\widetilde{P}}(p)\mathbb{P}_{\widetilde{L}}(l))^{-1}$, this terms are significant in the computation of the second moment. For this reason, the sample variance (computed along the estimator) might not be a good estimator of the variance. In order to avoid this scenario, we compute the variance of 100 i.i.d. realizations of $\widehat{\eta_t(\varphi)}$ for $\alpha = 2$ and the single term unbiased DEnKBF configuration. In fig. 3.5 we show the cost-to-rate relation of the previously described estimator. We have three curves in the figure, two of them with its respective rates m. The upper line represents the unbiased estimator and the line in the middle represents the MLEnKBF, the comments for this figure a similar to the ones for figs. 3.3 and 3.4. We remark that the difference of the costs of the methods with $\alpha = 2$ is smaller. The unbiased EnKBF cost is about 45 times the one of the MLEnKBF, making this estimator slightly better, with the caveat that we might not have a good estimation of the variance in individual samplings.

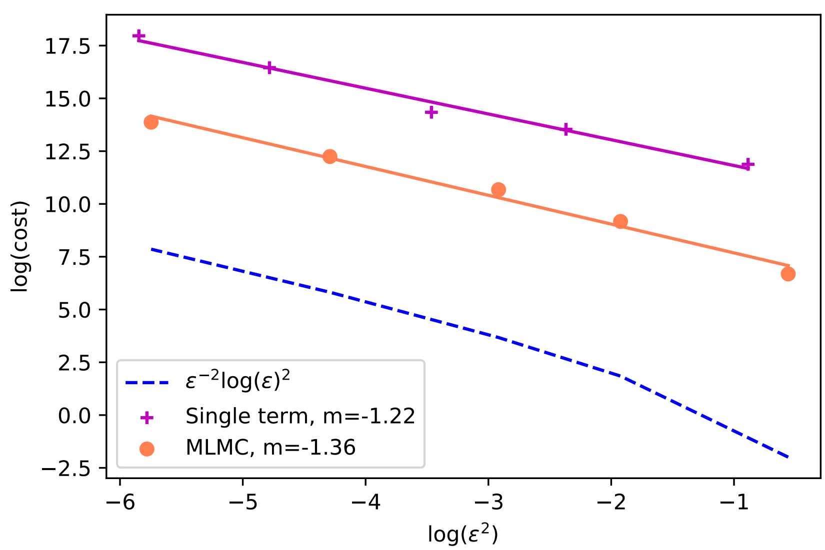

Figure 3.5: Error-to-cost rate of the Deterministic EnKBF with $\alpha = 2$. The dimension considered is $d = 2$ with final time $T = 30$.

The choice of N_0 is quite important since it defines the base number of particles, i.e., the number of particles in level zero. A smaller number N_0 reduces the cost for a given variance, thus it must be taken as smallest as possible so that the EnKBF converges. A similar comment can be made for the base level of the time discretization l^*, where again, the EnKBF can numerically diverge for low l^*. This phenomenon is widely documented and it is termed catastrophic divergence [48, 49, 50].

In conclusion, in this chapter we presented the unbiased methods for different variants of the EnKBF, the VEnKBF and the DEnKBF. Numerical tests were made to compare the performance of the unbiased EnKBF relative to the MLEnKBF (the comparison of them MLEnKBF with the EnKBF is already inverstigated and it can be found in [34]). It is found that the main difference between these two methods is the complexity cost, being much greater for the unbiased methods. Nevertheless, in the unbiased EnKBF case, the estimation of the parameters related to the variance can be done with the same samples that will form the main estimator $\widehat{\eta}_t$, thus sparing additional computations. The error-to-cost rates of the unbiased EnKBF and the MLEnKBF change overall depending on the setting (e.g. final time, EnKBF variant), but for each specific configuration are similar and comparable with the subcanonical relation MSE= $\mathcal{O}(\epsilon^2)$ and cost=$\mathcal{O}(\epsilon^{-2}\log(\epsilon)^2)$.

Chapter 4

Concluding Remarks

Ensemble Kalman Filters (EnKBF) are essential methods in the computation of filters due to their low computational cost and stability in high dimensions. The success of the EnKF is followed by its continuous-time analogous, the Ensemble Kalman-Bucy Filter, which is helpful in linear signaling processes or, more generally, in the context of the continuous-time observations. Debiasing techniques are relatively recent techniques that aim to reduce the error-to-cost rate of Monte Carlo estimators by building unbiased estimators. In this work, we integrate the unbiased methodology with the EnKBF and propose an algorithm to (in practice) obtain unbiased estimators-like with arbitrary bias. We observe that the error-to-cost rate of this algorithm is similar to the rate of a previously studied method (for a MSE= $\mathcal{O}(\epsilon^2)$ a cost=$\mathcal{O}(\epsilon^{-2}\log(\epsilon)^2)$ is necessary), the MLEnKBF, where the latter shows improvement overall with respect the plain EnKBF (to observe the rates and costs you can go to [34] in the conclusions section). We applied the debiasing methodology to two variants of the EnKBF, the Vanilla and Deterministic EnKBF. There is no strong sign in the numerical test to differentiate these two variants, given that both produced similar results.

In general, the complexity cost to attain a certain MSE= $\mathcal{O}(\epsilon^2)$ is much higher for the unbiased EnKBF compared to the MLEnKBF. This disadvantage might be alleviated by the parallelazing property of the Unbiased estimators, given that additional processors are available. Estimating the error for this kind of estimator usually takes several samples of associated quantities (e.g., the variance of the EnKBF at certain levels MLEnKBF); such quantities might not be implemented in the computations of the estimators. On the other hand, the unbiased EnKBF methodology allows computing quantities related to the variance that play a role in the stopping criteria of the algorithm, in this way sparing additional computations.

4.1 Future Research Work

There are two main ways in which we can expand our work. The first one is theoretical, in which several results are needed, the most important being a demonstration of the unbiasedness of the estimator and the computation of bounds that ensure the convergence of the variance. Specifically, the variance of each specific level is desired in order to optimize the double randomization.

The second way in which we can expand is in the structure o the estimators, e.g., the specific classes of the unbiased algorithm, the probability measures considered, the sampling method for $\eta_{k\Delta_l}^{N_{0:p},l}$, among others. In general, there are many details to tweak that might improve the algorithm's performance. In order to reduce the variance, the coupled sum and independent sum estimators can be implemented in the time discretization degree of freedom. These kinds of algorithms are known to have a smaller variance compared to the single term estimator but are harder to implement and need additional memory to store the Brownian motions of each particle. Finally, another direction that we can pursue is related to the parallelization of the algorithm and implementation in a supercomputer with several processors available.

REFERENCES

[1] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation: A Mathematical Introduction.* Cham: Springer International Publishing, 2015.

[2] A. Bain and D. Crisan, *Fundamentals of stochastic filtering.* New York: Springer, 2009.

[3] S. Reich and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation.* Cambridge: Cambridge University Press, 2015.

[4] D. Crisan, "The stochastic filtering problem: a brief historical account," *Journal of Applied Probability*, vol. 51, no. A, p. 13–22, 2014.

[5] A. Jazwinski, *Stochastic Processes and Filtering Theory.* New York: Dover Publications, 2007.

[6] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 2, p. 107, 1993.

[7] L. D. Stone, R. L. Streit, T. L. Corwin, and K. L. Bell, *Bayesian multiple target tracking*, 2nd ed. New York: Artech House, 2013.

[8] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," no. 2, pp. 47–55, 2015.

[9] W. H. Aeberhard, J. Mills Flemming, and A. Nielsen, "Review of state-space models for fisheries science," *Annual Review of Statistics and Its Application*, vol. 5, no. 1, pp. 215–235, 2018.

[10] P. Date and K. Ponomareva, "Linear and non-linear filtering in mathematical finance: a review," *IMA Journal of Management Mathematics*, vol. 22, no. 3, pp. 195–211, 2011.

[11] D. Suleiman, A. Awajan, and W. Al Etaiwi, "The use of hidden markov model in natural arabic language processing: a survey," *Procedia Computer Science*, vol. 113, no. 1877-0509, pp. 240–247, 2017.

[12] A. Majda and X. Wang, *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows.* Cambridge: Cambridge University Press, 2006.

[13] D. Crisan and T. Lyons, "A particle approximation of the solution of the kushner–stratonovitch equation," *Probability Theory and Related Fields*, vol. 115, no. 4, pp. 549–578, 1998.

[14] A. Beskos, A. Jasra, N. Kantas, and A. Thiery, "On the convergence of adaptive sequential Monte Carlo methods," *The Annals of Applied Probability*, vol. 26, no. 2, pp. 1111 – 1146, 2016.

[15] A. Beskos, D. Crisan, A. Jasra, K. Kamatani, and Y. Zhou, "A stable particle filter for a class of high-dimensional state-space models," *Advances in Applied Probability*, vol. 49, no. 1, p. 24–48, 2017.

[16] P. Bickel, B. Li, and T. Bengtsson, "Sharp failure rates for the bootstrap particle filter in high dimensions," in *Pushing the limits of contemporary statistics: Contributions in honor of Jayanta K. Ghosh.* Beachwood: Institute of Mathematical Statistics, 2008, pp. 318–329.

[17] G. Evensen, *Data assimilation: the ensemble Kalman filter.* Berlin: Springer, 2009.

[18] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge: Cambridge Univ. Press, 2003.

[19] A. J. Majda and J. Harlim, *Filtering Complex Turbulent Systems.* Cambridge: Cambridge University Press, 2012.

[20] K. A. Lisæter, J. Rosanova, and G. Evensen, "Assimilation of ice concentration in a coupled ice–ocean model, using the ensemble kalman filter," *Ocean Dynamics*, vol. 53, no. 4, pp. 368–388, 2003.

[21] G. Evensen, J. Hove, H. Meisingset, E. Reiso, K. S. Seim, and Espelid, "Using the EnKF for Assisted History Matching of a North Sea Reservoir Model," ser. SPE Reservoir Simulation Conference, vol. All Days, 2007.

[22] A. Seiler, G. Evensen, J.-A. Skjervheim, J. Hove, and J. Vabø, "Using the ensemble kalman filter for history matching and uncertainty quantification of complex reservoir models," *Large-Scale Inverse Problems and Quantification of Uncertainty*, 2010.

[23] G. Nævdal, L. Johnsen, S. Aanonsen, and E. Vefring, "Reservoir monitoring and continuous model updating using ensemble kalman filter," *SPE Journal*, vol. 10, no. 1, pp. 66–74, 2005.

[24] J. I. Allen, M. Eknes, and G. Evensen, "An ensemble kalman filter with a complex marine ecosystem model: hindcasting phytoplankton in the cretan sea," *Annales Geophysicae*, vol. 21, no. 1, pp. 399–411, 2003.

[25] A. N. Bishop and P. D. Moral, "On the mathematical theory of ensemble (linear-gaussian) kalman-bucy filtering," *arXiv e-prints*, 2020.

[26] C.-H. Rhee and P. W. Glynn, "Unbiased estimation with square root convergence for sde models," *Operations Research*, vol. 63, no. 5, pp. 1026–1043, 2015.

[27] M. B. Giles, "Multilevel monte carlo path simulation," *Oper. Res.*, vol. 56, no. 3, p. 607–617, may 2008.

[28] M. Vihola, "Unbiased estimators and multilevel monte carlo," *Operations Research*, vol. 66, no. 2, pp. 448–462, 2018.

[29] N. K. Chada, J. Franks, A. Jasra, K. J. Law, and M. Vihola, "Unbiased inference for discretely observed hidden markov model diffusions," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 9, no. 2, pp. 763–787, 2021.

[30] J. Heng, J. Houssineau, and A. Jasra, "On Unbiased Score Estimation for Partially Observed Diffusions," *arXiv e-prints*, 2021.

[31] J. Heng, A. Jasra, K. J. H. Law, and A. Tarakanov, "On Unbiased Estimation for Discretized Models," *arXiv e-prints*, 2021.

[32] A. Jasra, K. Law, and F. Yu, "Unbiased Filtering of a Class of Partially Observed Diffusions," *arXiv e-prints*, 2020.

[33] P. W. Glynn and C.-H. Rhee, "Exact estimation for markov chain equilibrium expectations," *Journal of Applied Probability*, vol. 51, no. A, p. 377–389, 2014.

[34] N. K. Chada, A. Jasra, and F. Yu, "Multilevel Ensemble Kalman-Bucy Filters," *arXiv e-prints*, 2020.

[35] R. Johnson and C. Nunez, "The Kalman-Bucy filter revisited," *Discrete and Continuous Dynamical Systems*, vol. 34, no. 10, pp. 4139–4153, 2014.

[36] N. Chopin and O. Papaspiliopoulos, *An Introduction to Sequential Monte Carlo*. Gewerbestrasse: Springer International Publishing, 2020.

[37] G. Burgers, P. J. van Leeuwen, and G. Evensen, "Analysis scheme in the ensemble kalman filter," *Monthly Weather Review*, vol. 126, no. 6, p. 1719–1724, 1998.

[38] G. Evensen, "The ensemble kalman filter: Theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, 2003.

[39] X. T. Tong, A. J. Majda, and D. Kelly, "Nonlinear stability and ergodicity of ensemble based Kalman filters," *Nonlinearity*, vol. 29, no. 2, p. 657, 2016.

[40] P. Del Moral, A. Kurtzmann, and J. Tugaut, "On the stability and the uniform propagation of chaos of a class of extended ensemble kalman–bucy filters," *SIAM Journal on Control and Optimization*, vol. 55, no. 1, pp. 119–155, 2017.

[41] D. T. B. Kelly, K. J. H. Law, and A. M. Stuart, "Well-posedness and accuracy of the ensemble kalman filter in discrete and continuous time," *Nonlinearity*, vol. 27, no. 10, pp. 2579–2603, 2014.

[42] P. Del Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems With Applications*, 05 2004, vol. 100.

[43] P. Sakov and P. R. Oke, "A deterministic formulation of the ensemble kalman filter: an alternative to ensemble square root filters," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 60, no. 2, pp. 361–371, 2008.

[44] J. H. Blanchet, P. W. Glynn, and Y. Pei, "Unbiased Multilevel Monte Carlo: Stochastic Optimization, Steady-state Simulation, Quantiles, and Other Applications," *arXiv e-prints*, 2019.

[45] A. Jasra, K. J. H. Law, and D. Lu, "Unbiased Estimation of the Gradient of the Log-Likelihood in Inverse Problems," *arXiv e-prints*, 2020.

[46] H. Ruzayqat, N. K. Chada, and A. Jasra, "Multilevel Estimation of Normalization Constants Using the Ensemble Kalman-Bucy Filter," *arXiv e-prints*, 2021.

[47] D. T. Gillespie, "Exact numerical simulation of the ornstein-uhlenbeck process and its integral," *Phys. Rev. E*, vol. 54, no. 2, pp. 2084–2091, 1996.

[48] J. Harlim and A. Majda, "Catastrophic filter divergence in filtering nonlinear dissipative systems," *Communications in Mathematical Sciences*, vol. 8, no. 1, pp. 27–43, 2010.

[49] G. A. Gottwald and A. J. Majda, "A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks," *Nonlinear Processes in Geophysics*, vol. 20, no. 5, pp. 705–712, 2013.

[50] D. Kelly, A. J. Majda, and X. T. Tong, "Concrete ensemble kalman filters with rigorous catastrophic filter divergence," *Proceedings of the National Academy of Sciences*, vol. 112, no. 34, pp. 10 589–10 594, 2015.

5 Papers Submitted and Under Preparation

- Miguel Alvarez, Neil K. Chada, Ajay Jasra, "Unbiased Estimation of the Vanilla and Deterministic Ensemble Kalman-Bucy Filters", *In preparation.*