Asymptotic Properties of Non-linear Filters

A thesis

submitted to the
Tata Institute of Fundamental Research, Mumbai
for the degree of
Doctor of Philosophy

in

Physics

by

Anugu Sumith Reddy

International Centre for Theoretical Sciences,
Tata Institute of Fundamental Research,
Bengaluru, Karnataka.

June, 2020
Final version submitted in October, 2020

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

The work was done under the guidance of Professor Amit Apte, at the International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bengaluru.

Anugu Sumith Reddy

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Amit Apte
Date: 28 October 2020

"Confusion is a word we have invented for
an order which is not understood."

-HENRY MILLER

Abstract

The theory of nonlinear filtering concerns with the estimation of underlying signal (which is Markovian) given that it has been indirectly and partially observed *via.* an observational model that is noisy. These estimates can be understood by studying the conditional distribution of the signal at any instant given observations made upto that instant (referred to as filter). The filter as a function of time is given by a recursive algorithm for discrete time and is a measure valued solution to a stochastic partial differential equation in continuous time. The filter necessarily depends on the initial condition of the signal. In practice, if the signal cannot be observed directly, it is highly unlikely that we have the information about the initial condition of the signal. Therefore, it is desirable for the filter to behave nearly independently with respect to initial conditions after large times. This thesis studies the long time behavior of the filter in terms of sensitivity of the filter to the initial conditions, in the framework of deterministic signal. A good understanding of the long behavior of the filter with respect to initial conditions is important for practical applications of nonlinear filters.

The first topic of my thesis is the long time behavior of the filter with respect to the initial condition in the case of linear deterministic dynamics and linear observation model. This is a much studied problem in the context of noisy linear non-autonomous dynamics with linear non-autonomous observation model. Here, we study filter stability in the case of deterministic signal using the notion of uniform complete observability. We showed that the filter (in this case) is stable in the almost sure sense. We have also showed that for a given set of observations, true conditional distribution (coming from true signal dynamics with small stochastic perturbations) and distribution given by the incorrect filter (assuming that the underlying signal is the true dynamics which is deterministic) are close to each other for all times. This problem is again important in cases where the simulation of the filter is efficient when the signal dynamics is deterministic.

The second and final topic of my thesis is the long time behavior of the filter with respect to initial conditions in the case of non-linear deterministic dynamics and non-linear observation model. Here, we study the stability of the filter by studying the proximity of conditional distribution of initial condition given observations and the Dirac measure at the initial condition. The main result is that conditional distribution of initial condition given observations converges to the Dirac measure at the initial condition, if the dynamics is such that the orbits are neither converging towards each other nor going far away from each other and the observations are rich enough. It is observed through numerical computations that for sufficiently rich observations, very common chaotic systems like Lorenz 63 and Lorenz 96 models satisfy the assumptions of the result.

Acknowledgements

First and foremost, I would like to thank my advisor, Prof. Amit Apte, who has introduced me to filtering theory and was very supportive throughout this entire endeavour. In addition to his guidance on my doctoral work, he also shared his invaluable knowledge on general academic life after graduation. He made me give series of lectures on the topics that I was learning. This helped me tremendously in improving my teaching skills and also understand the finer details of those topics.

I am very thankful to Prof. Vishal Vasan for numerous discussions on the topic of functional analysis. I have obtained significant part of my knowledge and a better perspective on this topic through these discussions. He was always available for clarifying any questions I had in functional analysis. I would also like to thank Prof. Sreekar Vadlamani for the interaction we had during our collaboration and for his advice on general academic life, like Prof. Amit Apte.

I am indebted to Prof. Amarjit Budhiraja for letting me audit his course on stochastic analysis at the University of North Carolina, Chapel Hill. Not only was it a fantastic course in its own right but it was also a very important experience in the context of my doctoral work. I got much better clarity on the subject of filtering theory (and probability, in general) by the end of the course. Numerous discussions I had with him, had a significant bearing on the output of my doctoral work.

I am grateful to Prof. Eric Van Vleck for the discussions we had on Lyapunov exponents and non-autonomous linear systems. I am also thankful to Prof. Chris Jones for hosting me during my visit to the University of North Carolina, Chapel Hill and The Statistical and Applied Mathematical Sciences Institute which was funded by Infosys Foundation Excellence Program of ICTS. This was a very interesting visit during which I had plenty of fruitful interactions.

Finally, I would like to express my gratitude to my parents, without whose love and support, this journey would not have started and to my friends, without whom this journey would not have been so exciting.

List of publications relevant to the thesis

- Anugu Sumith Reddy, Amit Apte, Sreekar Vadlamani. "Asymptotic properties of linear filter for deterministic processes ", In: *Systems & Control Letters 139 (2020), 104676*
- Anugu Sumith Reddy and Amit Apte. "Stability of Non-linear Filter for Deterministic Dynamics". In: *arXiv preprint arXiv:1910.14348 (2019)*

Contents

List of Figures

Nomenclature

$(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$	Underlying filtered probability space
$\mathcal{B}(\cdot)$	Borel σ-algebra
$\mathcal{D}(\mathcal{A})$	Domain of operator $\mathcal{A}$
$\mathcal{G} \vee \mathcal{H}$	σ- algebra generated by $\mathcal{G} \cup \mathcal{H}$
$\mathcal{L}(U)$	Law of a random variable U
$\mathcal{N}$	Set of $\mathbb{P}$- null sets
$\mathcal{N}(M, P)$	n-dimensional Gaussian distribution with mean M and covariance P
$\mathcal{S}$	σ- algebra of S
$\mathcal{S}^o$	σ- algebra of $\mathcal{S}^o$
$\frac{d\mathbb{P}}{d\mathbb{Q}}$	Radon-Nikodym derivative of $\mathbb{P}$ w.r.t $\mathbb{Q}$
$\hat{\pi}_t^\mu$	Conditional distribution of X_0 given $\mathcal{Y}_t$ with $\mathcal{L}(X_0) = \mu$
$\hat{X}_t$	Conditional expectation of X_t given $\mathcal{Y}_t$
$\langle A \rangle_t$	Quadratic variation of local martingale A_t
$\left(\hat{X}_t^{M,P}, P_t^P\right)$	Solutions of Kalman-Bucy filter equations with initial condition (M, P)
$\mathbb{1}_A$	Indicator function of set A
$\mathbb{I}_d$	$d \times d$- Identity matrix
$\mathbb{P}_W^q$	Wiener measure on $C([0,\infty), \mathbb{R}^q)$
μ	True initial condition

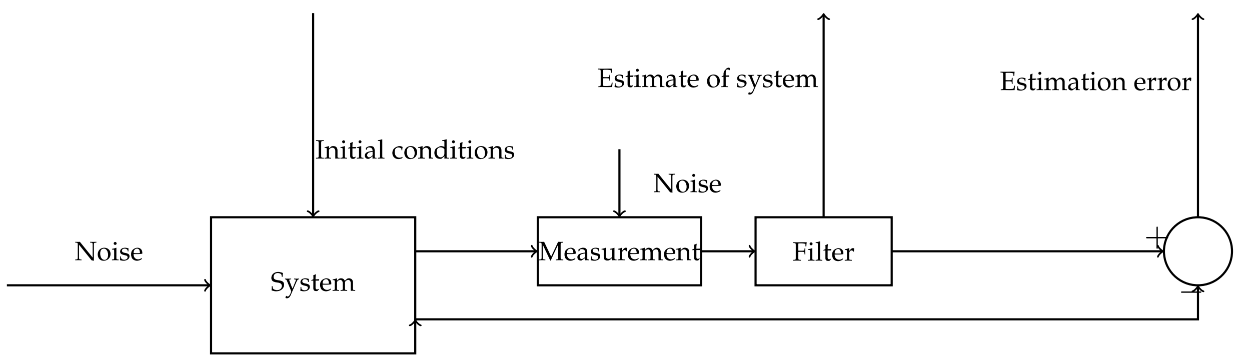
ν	Incorrect initial condition
π_t^ρ	Solution of Kushner Stratanovich equation with initial condition ρ
$\mathbb{R}^+$	Positive real numbers
$\mathbb{R}^p$	p- dimensional Euclidean space
$\sigma(A_s : 0 \leq s \leq t)$	σ- algebra generated by $A_s : 0 \leq s \leq t$
σ_t	Unnormalised conditional distribution
$\{\mathcal{Y}_t^o\}_t \geq 0$	Unaugmented observation filtration
$B_r(x)$	Closed ball of radius r around x
B_t	Signal Brownian motion
$C([0,\infty), S)$	Set of S-valued continuous functions on $[0,\infty)$
$C_b^2(\mathbb{R}^m)$	Set of twice differentiable bounded functions on $\mathbb{R}^m$
I_t	Innovations process
$L^1(\Omega, \mathcal{F}, \mathbb{P})$	Set of integrable $\mathcal{F}$- measurable functions
P_t	Conditional covariance of X_t given $\mathcal{Y}_t$
S	Signal state space
S^o	Observation state space
W_t	Observation Brownian motion
X_t	Signal process
Y_t	Observation process
$\mathcal{Y}_t$	Augmented observation filtration
m(g)	Integral of g w.r.t. measure m

To my mother and father…

Chapter 1

Introduction

Filtering theory deals with estimating the state of an underlying system from the observations that are made. Systems in engineering and nature are evolving with time according to certain rules. These rules are modelled by dynamical systems. Very often, systems in practice interact with the environment whose rules of evolution are unknown. It may happen that we only know some "averaged" version of the rules and the interaction with environment is treated as random with appropriate "averaged" properties. Effectively, these systems are modelled as noisy dynamical systems. To study these systems, we need to make measurements/observations using a measuring device. Since measurements can also interact with the environment, it is also modelled as a noisy dynamical system (with measurement noise). Now, one question of practical interest is how to deduce the state of the underlying system using the observations made on the system. We thereby need a systematic way to filter out the measurement noise to deduce useful information about the state of the system. In particular, we would like estimate the state of system (using observations made on the system) by running an algorithm (referred to as filter, for now and will be made precise below). The following diagram illustrates this idea.



In our context, the best estimate of the state given observations is considered to be the one that optimizes in the L^2 sense. In the following, we briefly introduce filtering theory and the notion of filter stability. We defer a slightly detailed introduction and relevant literature to the later chapters.

1.1 Filtering Theory

To make the question of finding the best estimate of underlying system given the observations mathematically rigorous in a general context, the language of probability theory and stochastic processes is used. We as usual, start with an abstract probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. We denote signal state space by S and observation state space by S^o. $X_t \in S$ denotes the underlying signal process (system) and $Y_t \in S^o$ denotes the observation process. Until mentioned otherwise, we only consider the case of $t \in \mathbb{R}^+$. From probability theory, it is known that best estimate (in L^2 sense) of a random variable given some information can be expressed using conditional expectation with respect to that information. Mathematically, sigma algebra plays the role of information. Therefore, the question of best estimate (in L^2 sense) of the state X_t given observations $Y_{[0,t]}$ is equivalent to finding $\mathbb{E}[X_t|\mathcal{Y}_t^o]$, where $\mathcal{Y}_t^o$ is the σ- algebra generated by $Y_{[0,t]}$. Very often, one may be interested in estimating a function of X_t given the information obtained. This requires us to compute the conditional distribution (or optimal filter or simply, filter), π_t. In general, the observation Y_t and signal X_t do not necessarily have a simple relation. Formally, if one considers the following relation between the observation Y_t and the signal X_t[1]:

$$Y_t = h_t(X_t) + N_t,$$

where, h_t is a deterministic function and N_t is the measurement noise. Note that Y_t is considered to depend only on X_t, N_t and t, but not on any X_s $(0 \leq s < t)$. For $t \in \mathbb{R}^+$, very often, the measurement noise is modelled by white noise and the above equation is not amenable to analysis. Therefore, to work with a mathematically tractable object, we have to work with redefined form of the observational model.

$$Y_t = \int_0^t h_s(X_s)ds + B_t,$$

where, Y_t is now the redefined observation process and B_t is Brownian motion (It can be intuitively thought of as integral of white noise). Note that knowing the integral of function of t from 0 to t for all $t \in \mathbb{R}^+$ is equivalent to knowing the function for any $t \in \mathbb{R}^+$. Therefore, we did not lose any information in redefining Y_t and in return, we ended up with a tractable object B_t. Given the newly defined observational model and assuming that X_t is a Markov process, the evolution of π_t

[1]In discrete time, a more general well defined model like $Y_t = f_t(X_t, N_t)$ (for a deterministic function f_t) can be considered.

is given by the following equation: for a suitable function $g : S \to \mathbb{R}$

$$\pi_t(g) = \pi_0(g) + \int_0^t \pi_s(\mathcal{A}g)\, ds + \int_0^t \left(\pi_s(gh^T) - \pi_s(g)\pi_s(h^T)\right)(dY_s - \pi_s(h)ds), \tag{1.1}$$

where, $\pi_t(g) \doteq \int_S g(x)\pi_t(dx)$, $\pi_0 = \mathcal{L}(X_0)$ and $\mathcal{A}$ is the infinitesimal generator of X_t. π_t can also written in the following form, for appropriate $g : S \to \mathbb{R}$

$$\pi_t(g) = \frac{\int_{C([0,\infty),S)} g(x_t) \exp\left(\int_0^t h(x_s)^T dY_s - \frac{1}{2}\int_0^t \|h(x_s)\|^2 ds\right) dP^X(x)}{\int_{C([0,\infty),S)} \exp\left(\int_0^t h(x_s)^T dY_s - \frac{1}{2}\int_0^t \|h(x_s)\|^2 ds\right) dP^X(x)}, \mathbb{P}\ a.s., \tag{1.2}$$

where, P^X is the law of the process. Even though (1.2) is simple looking and (1.1) is complicated stochastic partial differential equation (at least when X is a diffusion process), (1.2) has limited practical use compared to (1.1). The reason is that (1.2) requires the entire set of observations until t, whereas (1.1) requires incremental observations. That being said, one should note that π_t lies in an infinite dimensional space. And for almost all of the systems, this is genuinely an infinite dimensional object, except for linear case, Beneš filter etc. See next chapter for more details.

1.2 Applications of Filtering

Now that we know what filtering is (in a very informal way, at least), we give some of the applications of filtering theory below. We also emphasize that this list is in no way, exhaustive. In practice, the approximations of π_t are used to compute the conditional expectations numerically. Out of all different kinds of approximation schemes, a Markov chain Monte Carlo algorithm known as particle filter [69] is till now, best suited for approximating π_t, at least in low dimensional S. It fails very badly to approximate π_t in case of very high dimensional S (a phenomenon referred to as curse of dimensionality). In high dimensional setting, many ensemble methods are known to perform well [96].

- **Fault detection or changepoint detection**

  Filtering is used to detect sudden changes in the observations. When analysing seismic data or behavior of a device using the noisy observational data or biomedical signal like EEG, it is important to know that if a sudden change in data is significant or not [77, 141, 12].

- **Positioning and Tracking**

Filtering is used in positioning of underwater vessels, surface ships, cars and aircrafts using the geographical information systems containing information of the surroundings [69]. In addition, it is used in target tracking and car collision avoidance [70].

- **Active vision**

  To visually track an object and in the field of active vision, tracking deforming objects via estimating global motion and local deformation of object, filtering techniques are used [124, 123, 103].

- **Robotics**

  In robotics, filtering methods are applied to the problem of SLAM (Simultaneous Localization And Mapping), where the aim is to build a map of an environment using a moving robot making landscape measurements. The robot has to simultaneously determine its location in the map and also figure out the locations of the landmarks in the map. The solution to this problem is considered to be the pre-requisite for making autonomous robots[110, 13, 60].

- **Aerospace industry**

  One of the first applications of filtering theory came in guidance and navigation systems in the aerospace industry[34, 23]. It is also being used in GPS/INS integration which is used in navigating a moving object that has a time-dependent access to satellite. This compensates the short terms accuracy of INS (Inertial Navigation System) due to dead reckoning with long term accuracy of GPS (Global Positioning System) and also combines with robust nature of INS[40].

- **Finance**

  Filtering techniques are used in estimating coefficients of models of financial market models [128]. They are also used in the finding the optimum portfolio (such that terminal wealth is maximized) by observing only the stock prices [91, 56].

- **Space industry**

  Filtering is also used in space navigation [34], satellite orbit determination, satellite re-entry [79] and satellite launching [27].

- **Seismology**

In extracting the seismic events from the noisy seismic data [11], filtering is used. And also, it is used in estimating the structural damage due to seismic activity [131].

- **Other applications**

  Filtering is used in systems biology [145], speech recognition [65, 111, 122], audio enhancements [63], weather prediction [4, 5] and software reliability [42].

1.3 Stability of the filter

From Equation (1.1), it is evident that the evolution of π_t depends on the law of X_0. In practice, many systems of interest are accessed indirectly through a measurement device (which is in itself noisy) and therefore, it is very unlikely to have complete information of the initial condition of the system when it is not even possible to measure it directly. For this reason, it is desirable to have π_t asymptotically independent of law of X_0 and depend only on the observation process Y_t, if the filter is to be of any use. The filter is said to stable if it is asymptotically independent of the law of X_0. The problem of filter stability is studied extensively in the literature. We defer the extensive literature survey of filter stability to Chapter 3.

In this thesis, we study the asymptotic properties of the filter where X_t is deterministic process. In the following, we motivate the need to study the deterministic case: In practice, filtering for deterministic systems is quite commonly used in the context of geosciences where the problem is known as data assimilation (cf. [39, 62, 6, 127, 96, 129, 96]). The signal being observed is the ocean and/or the atmosphere through observations from sensors like satellites or those that are installed on-site. The dynamical models used in these applications such as weather predictions are derived from nonlinear partial differential equations. These models are deterministic (even though, stochasticity should be considered for better modelling) since both the theoretical and numerical developments for stochastic PDE in this context is still a fairly new research area [120, 100, and ref. therein]. Some of the important characteristics of these systems are that:

1. They are high dimensional system. This is by far the biggest hurdle in achieving efficient numerical algorithms. As mentioned earlier, particle filters (which are shown to be genuine approximations to the filter [15]) perform very badly and we have to resort to ad hoc algorithms.

2. The observations are sparse and noisy.

3. The dynamical models are very commonly deterministic.

4. They are chaotic. Even though it gives rise to a dynamically very complex system, this complexity provides some advantage in the dealing with high dimensionality. This is because significant part of the chaotic dynamics usually takes place in a much lower dimensional attractor. Many numerical algorithms that exploit this behavior are being developed. In the applications, this asymptotic low dimensional behavior and stability of the filter covariance (but not of the filter mean) for discrete time Kalman filter has been studied recently [68, 28]. These results about the rank deficiency of the filter covariance provide a mathematical foundation for commonly used data assimilation algorithms collectively called as "assimilation in unstable subspace" (AUS) [38, 119, 137].

Although many numerical algorithms that focus on one or more of these characteristics are being developed, only a few theoretical results related to filtering for deterministic, chaotic signal dynamics have been established so far. Even though the stability of filter with noisy signal was studied extensively, the case of deterministic signal is mostly unexplored in the literature (again, the relevant literature survey for both is deferred to the later chapters). The main contribution of this thesis is precisely to provide filter stability results that fill this gap.

Following is the outline of this thesis: In Chapter 2, we give briefly the fundamentals of the stochastic filtering theory. We define the filtering problem in the very general setup and later focus on a specific setup that we deal with in the rest of the thesis. We briefly describe the technical issues that come up when setting up the general version of the problem in continuous time. In Chapter 3, we define the notion of the filter stability and give a detailed survey of the literature on filter stability. In Chapter 4, we introduce the linear filter, derive the Kalman-Bucy equations and state the already existing stability results in the case of stochastic signal. We also establish filter stability in the case of linear filter with deterministic signal under the assumption of uniform complete observability and study the small signal noise asymptotics of the filter. We conclude the thesis by studying filter stability for the non-linear filter (with deterministic signal) is established in Chapter 5. We also study the support of the filtering distribution after long times.

Chapter 2

Fundamentals of Filtering Theory

In this chapter, we briefly describe the general framework of the filtering theory. We describe a general framework in the Section (2.1) and give two common approaches to filtering in continuous time namely reference probability method in (2.1.1) and innovations process approach in (2.1.2). Sections (2.2) and (2.3) contain the explicit construction of signal and observation processes and brief description of finite dimensional filters, respectively. We hope that describing the framework makes this thesis as self-contained as possible. Following textbooks are used as the references: [82, 15, 99, 148]. As already mentioned, we only work with continuous time case until stated otherwise.

Norbert Wiener [144] was the first to study the problem of estimation. He studied the problem of estimating a stationary process f_t by observing $Z_t = f_t + W_t$, where W_t is a noise, not necessarily independent of f_t. Wiener showed, using Fourier transform techniques (therefore, in frequency domain) that best estimate $\hat{X}_t$, in the sense of mean square is given by

$$\hat{X}_t = \int_{-\infty}^{t} h(t-s) Z_s ds,$$

where, $h(s)$ is called a transfer function that depends on the auto-correlation and cross-correlation coefficients of f_t and W_t. This estimator goes by name of Wiener filter. The discrete version of Wiener filter was independently derived by A. Kolmogorov [90].

The next significant progress in filtering is due to Kalman and Bucy [85] (in continuous time) and Kalman [84] (in discrete time) in their seminal work on the linear filter where they considered the estimation problem in time domain. This allowed for non-stationary process, in contrast to the independent works of Wiener and Kolmogorov. In this work, they studied the estimation of signal X_t which satisfies a linear stochastic differential equation with Brownian motion as the driving noise and X_t is indirectly observed through a linear observational model with Brownian motion as the measurement noise (more details will be given in Chapter (4)). We refer the reader to Bain and Crisan [15, Section 1.3] and Crisan [51] for a nice historical account.

2.1 General framework

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space that satisfies usual conditions, *i.e.*, it is a complete probability space with $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$ and $\mathcal{F}_0$ contains all $\mathbb{P}$- null sets. We work with this abstract probability space until we define the stability of the filter.

Before we start, let us define $(S, \mathcal{S})$ and $(S^o, \mathcal{S}^o)$ as the signal state and observation state measurable spaces, respectively. As mentioned earlier, the main question of interest in filtering theory is find the best estimate of some function of signal $X_t \in S$ given that indirect noisy observations are made up to time t, *i.e.*, we have $Y_s \in S^o$, $0 \leq s \leq t$. Intuitively, we expect to compute the best estimate if we have the knowledge of $Y_s(0 \leq s \leq t)$. Mathematically, this means that the best estimate is measurable with respect to sigma algebra $\mathcal{Y}_t^o = \sigma\{Y_s : 0 \leq s \leq t\}$ (see Theorem (A.0.2)). And also, the best estimate that we consider is in the sense of L^2 (or mean square). It is a well known result in probability [148, Lemma 5.1] that the best estimate of $f(X_t)$ given $\mathcal{Y}_t^o$ in the sense of L^2 is given by the conditional expectation of $f(X_t)$ with respect to $\mathcal{Y}_t^o$, $\mathbb{E}[f(X_t)|\mathcal{Y}_t^o]$, for a bounded measurable f. In general, we are interested in computing $\mathbb{E}[f(X_t)|\mathcal{Y}_t^o]$, for all bounded measurable $f : S \to \mathbb{R}$. Equivalently, we are interested in computing $\pi_t(A) \doteq \mathbb{E}[\mathbb{1}_{X_t\in A}|\mathcal{Y}_t^o]$, for all $A \in \mathcal{S}$.

To have a sensible notion of conditional distribution, we should be able to define $\pi_t(A)(\omega)$ on all $A \in \mathcal{S}$ for all $\omega \in \Omega$. But from Remark (A.0.5), we know that the conditional expectation $\mathbb{E}[\mathbb{1}_{X_t\in A}|\mathcal{Y}_t^o](\omega)$ can be defined uniquely for all ω not lying in $\mathbb{P}$-null set $\mathcal{N}_A$. Therefore, in general, $\mathbb{E}[\mathbb{1}_{X_t\in A}|\mathcal{Y}_t^o](\omega)$ is defined for all $\omega \in \mathcal{N}^o \doteq (\cup_{A\in\mathcal{S}}\mathcal{N}_A)^c$. But, since $\mathcal{N}^o$ is an uncountable union of N_A, $\mathcal{N}^o$ is not guaranteed to be a $\mathbb{P}$-null set or even measurable set. So, $\mathbb{E}[\mathbb{1}_{X_t\in A}|\mathcal{Y}_t^o](\omega)$ is not a well defined probability measure on $(S, \mathcal{S})$ for ω in a non-measurable set.

To overcome this problem, the notion of regular conditional distribution (see Definition (A.0.7)) is introduced. Assuming $(S, \mathcal{S})$ to be such that there is an injective map $g : S \to \mathbb{R}$ such that g is $\mathcal{S}$- measurable and g^{-1} is $\mathcal{B}(\mathbb{R})$- measurable, it is known that regular conditional distribution exists [29, Theorem 4.34].

If we only want to compute the π_t for a fixed time, the above statement gives the existence of a well behaved conditional distribution. But, we are interested in the evolution of π_t with $t \in \mathbb{R}^+$. This requires us to treat π_t as a continous time process and the regularity and the measurability properties of π_t become very important.

In rest of the section, we closely follow the Chapter 2 of [15]. We assume that X_t and Y_t are continous in time and are $\mathcal{F}_t$- adapted. Define

$$\mathcal{Y}_t \doteq \sigma\left(\sigma\{Y_s : 0 \leq s \leq t\} \cup \mathcal{N}\right),$$

where, $\mathcal{N}$ is the set of all $\mathbb{P}$-null sets. Note that in general, the way $\mathcal{Y}_t$ is defined, it does not necessarily contain all of the $\mathbb{P}$- null sets. These are important because the conditional expectation that we deal with later on, are only uniquely defined upto a $\mathbb{P}$- null set and without including these $\mathbb{P}$- null sets we run into measurability issues. So we work with $\mathcal{Y}_t$ to avoid the technicalities. From Equation (1.1), it is clear that we shall define a stochastic integral of $\pi_t(g) \doteq \int_S g(x) d\pi_t(x)$, for some appropriate $g : S \to \mathbb{R}$. And also, from stochastic calculus, we know that π_t has to be progressively measurable (see Definition (A.0.9)) with respect to $\mathcal{Y}_t$, for the stochastic integral to make sense. Another issue that arises due to continuous time is that it is non-trivial to construct a process from X_t that is $\mathcal{Y}_t$- adapted (in general, X_t is not $\mathcal{Y}_t$-measurable or else the filtering problem would become trivial). We cannot naively construct the adapted process by defining it to be $\mathbb{E}[X_t|\mathcal{Y}_t]$, since this process is defined outside a $\mathbb{P}$- null set for every $t \geq 0$. Therefore, we may end up defining that process outside an uncountable union of $\mathbb{P}$- null sets (because $t \in \mathbb{R}^+$), which can have non-zero measure or even be non-measurable.

Note that until now, no apparent relation between X_t and Y_t is assumed. The following theorem gives the partial answer to the above issue.

Theorem 2.1.1. *[15, Theorem 2.24] Let $(S, \mathcal{B}(S))$ be a Polish space. Then there exists a $\mathcal{P}(S)$- valued process denoted by π_t such that*

1. *π_t is $\mathcal{Y}_{t+}$- optional process[1] ($\mathcal{Y}_{t+} \doteq \cap_{s>t}\mathcal{Y}_s$)*
2. *For any bounded measurable $g : S \to \mathbb{R}$, the process $\pi_t(g) \doteq \int_S g(x) d\pi_t(x)$ is indistinguishable from the process $\mathbb{E}[f(X_t)|\mathcal{Y}_{t+}]$.*

In the following, we assume that the observation process is such that $Y_t : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and

$$Y_t = \int_0^t h(X_s) ds + W_t, \tag{2.1}$$

where, W_t is n- dimensional $\mathcal{F}_t$- Brownian motion independent of X_t. We also assume that

$$\mathbb{E}[\int_0^t \|h(X_s)\|^2 ds] < \infty$$

[1] See appendix for the definition

to make integrals (that we come across) finite. Theorem (2.1.1) and the following theorem gives existence of a $\mathcal{Y}_t$- adapted process that matches with regular conditional distribution, for each $t \geq 0$ outside a $\mathbb{P}$- null set (independent of t).

Theorem 2.1.2. *[15, Theorem 2.35] Suppose the observation process is of the form above and*

$$\mathbb{E}[\int_0^t \|h(X_s)\|^2 ds] < \infty.$$

Then $\mathcal{Y}_t = \mathcal{Y}_{t+}$. In other words, observation filtration is right- continuous.

Proof. See the Appendix for the proof. ∎

As we established that filtering problem is well posed, we move onto the next problem of computing $\pi_t(g)$. There are two ways to compute this quantity. One is using Kallianpur-Striebel formula [81] that does not impose too many restrictions on g and uses entire observation process $Y_s : (0 \leq s \leq t)$ to compute $\pi_t(g)$. The other way of computing is through Kusher-Stratanovich equation [94, 149, 83, 80, 59], the evolution equation for $\pi_t(g)$ that puts slightly stronger restrictions on g. In practice, it is useful to be able to compute $\pi_{t+\delta}(g)$ from $\pi_t(g)$ and the observations from Y_t to $Y_{t+\delta}$. An evolution equation will satisfy this requirement (in our case, it will be a stochastic equation).

In literature, there are two approaches to deriving the evolution equation. First approach is known as the martingale method which is based on martingale representation theorems and using properties of the innovations process. This approach uses the Fujisaki-Kallianpur-Kunita formula [80] and was inspired from the work of T. Kailath and P. Frost [64]. Detailed expositions of this approach can be found in [82, 99, 15]. The second approach is known as the reference probability method which uses the fact that filtering problems reduces to a simpler problem under the change of probability measure. One derives the evolution equations under the new probability measure and changes back to old probability measure to get Kushner-Stratanovich equation. Detailed exposition of this approach can be found in [15, 22, 148].

Until now, no explicit type of property of X_t, other than continuity is assumed. Without further assumptions, it is not possible to use the tools of stochastic calculus (especially, the Ito's integral). To that end, we make the following assumption.

Assumption 2.1.3. *Let X_t be a solution of a martingale problem [86, Section 5.4] for $(\mathcal{A}, \mu)$,* i.e., *for $\mathcal{L}(X_0) = \mu$, operator $\mathcal{A} : B(S) \to B(S)$ domain $\mathcal{D}(\mathcal{A})$ and for any*

$\phi \in \mathcal{D}(\mathcal{A})$,

$$M_t^\phi := \phi(X_t) - \phi(X_0) - \int_0^t \mathcal{A}\phi(X_s)ds$$

is a $\mathcal{F}_t$- martingale

From the definition of semi-martingale, $\phi(X_t)$ is a $\mathcal{F}_t$- semi-martingale. Therefore, Ito's formula can be applied to $\phi(X_t)$.

2.1.1 Reference probability method

We now use the reference probability method to obtain the Kushner-Stratanovich equation and the Kallianpur-Striebel formula. As mentioned already, we try to find the new probability measure (referred to as reference probability) under which filtering problem becomes simpler. We make the required computations under the new probability measure and change back to the old probability measure to interpret the results.

Before proceeding any further, it is useful to remind ourselves of a simple form of Bayes' rule.

Lemma 2.1.4 (Bayes' rule)**.** *Let $\mathbb{P}$ and $\mathbb{Q}$ be two equivalent probability measures on $(\Omega, \mathcal{F})$ and let $H \in L^1(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ be a sub σ- algebra. Then*

$$\mathbb{E}^{\mathbb{P}}[H|\mathcal{G}] = \frac{\mathbb{E}^{\mathbb{Q}}[H\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{G}]}{\mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{G}]}, \quad \mathbb{P} - \text{ a.s. and } \mathbb{Q} - \text{ a.s.,}$$

where, $\frac{d\mathbb{P}}{d\mathbb{Q}}$ is the Radon-Nikodym derivative (see Theorem (A.0.3)).

Proof. Let $A \in \mathcal{G}$. Consider

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}[\mathbb{1}_A \frac{\mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}H|\mathcal{G}]}{\mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{G}]}] &= \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_A \frac{d\mathbb{P}}{d\mathbb{Q}} \frac{\mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}H|\mathcal{G}]}{\mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{G}]}] \\
&= \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_A \mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}} \frac{\mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}H|\mathcal{G}]}{\mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{G}]}|\mathcal{G}]] \\
&= \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_A \mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{G}] \frac{\mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}H|\mathcal{G}]}{\mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{G}]}] \\
&= \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_A \mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}H|\mathcal{G}]] \\
&= \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_A \frac{d\mathbb{P}}{d\mathbb{Q}}H] \\
&= \mathbb{E}^{\mathbb{P}}[\mathbb{1}_A H]
\end{aligned}$$

$$= \mathbb{E}^{\mathbb{P}}[\mathbb{1}_A \mathbb{E}^{\mathbb{P}}[H|\mathcal{G}]]$$

From uniqueness of conditional expectation upto a $\mathbb{P}$- null set (equivalently, a $\mathbb{Q}$-null set) and arbitrariness of $A \in \mathcal{G}$, we have the result. ■

In our case, $\mathcal{G} = \mathcal{Y}_t$, $H = g(X_t)$, for nice enough g. Note that this lemma can only be used for a fixed time t. In other words, every time we do computations, we have to evaluate the integrals all over again. In principle, the above lemma allows for any measure equivalent to $\mathbb{P}$. But, this lemma is only useful in practice, if one can compute both $\mathbb{E}^{\mathbb{Q}}[H\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{G}]$ and $\mathbb{E}^{\mathbb{Q}}[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{G}]$ in an simple way. The power of reference probability method lies in choosing a particular $\mathbb{Q}$ such that computations are simple.

To that end, define the measure $\mathbb{Q}$ by the following Radon-Nikodym derivative:

$$\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]}) \doteq \frac{d\mathbb{P}}{d\mathbb{Q}} \exp\left(-\int_0^t h(X_s)^T dY_s - \frac{1}{2}\int_0^t \|h(X_s)\|^2 ds\right).$$

Suppose $\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})$ is $\mathcal{F}_t$- martingale[2]. Now, consider a measure $\mathbb{Q}_t$ on $(\Omega, \mathcal{F}_t)$ which is given by the following Radon-Nikodym derivative.

$$\frac{d\mathbb{Q}_t}{d\mathbb{P}} = \tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})$$

The martingale property ensures that $\{\mathbb{Q}_t\}_{t\geq 0}$ is a consistent family of measures on $\mathcal{F}_t$ in the following sense: Suppose $0 \leq s < t$ and $A \in \mathcal{F}_s$. Now,

$$\begin{aligned}
\mathbb{Q}_t(A) &= \mathbb{E}^{\mathbb{P}}[\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})\mathbb{1}_A] \\
&= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})\mathbb{1}_A|\mathcal{F}_s]] \\
&= \mathbb{E}^{\mathbb{P}}[\mathbb{1}_A \mathbb{E}^{\mathbb{P}}[\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})|\mathcal{F}_s]] \\
&= \mathbb{E}^{\mathbb{P}}[\mathbb{1}_A \tilde{Z}_s(X_{[0,s]}, Y_{[0,s]})] \\
&= \mathbb{Q}_s(A),
\end{aligned}$$

where, we used the fact that $A \in \mathcal{F}_s$ and the martingale property of $\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})$. In other words, $\mathbb{Q}_t$ put same measure on any $A \in \mathcal{F}_s$ as that of $\mathbb{Q}_s$, for any $0 \leq s < t$. Even though we can define a new probability measure $\mathbb{Q}_t$ on $(\Omega, \mathcal{F}_t)$, unless a strong integrability condition (uniform integrability) on $\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})$ holds, we cannot define a measure on $(\Omega, \mathcal{F}_\infty)$ with

$$\mathcal{F}_\infty \doteq \sigma(\cup_{t\geq 0}\mathcal{F}_t)$$

[2] $\tilde{Z}_t^{-1}(X_{[0,t]}, Y_{[0,t]})$ is a martingale under $\mathbb{Q}$ when $\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})$ is a martingale under $\mathbb{P}$. And also, sufficiency conditions on $\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})$ are given in [86, Section 3.5.*D*]

We refer to [86, Pg. 192] for detailed discussion. The following example from [86, Pg. 193] shows that if $\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})$ is not uniformly integrable, then $\bar{\mathbb{Q}}$ defined by $\bar{\mathbb{Q}}(A) \doteq \mathbb{E}[\mathbb{1}_A \tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})]$, for $A \in \mathcal{F}_\infty$ is not equivalent to $\mathbb{P}$.

Example 2.1.5. *Consider a process, $Y_t = \mu t + W_t$ with $\mu \neq 0$ and W_t being a Brownian motion under the probability measure $\mathbb{P}$. From Girsanov's theorem (see Theorem (A.0.22)), Y_t is Brownian motion under probability measure $\bar{\mathbb{Q}}$. Note that the corresponding Radon-Nikodym derivative is not uniformly integrable in $\mathbb{P}$. Defining,*

$$A \doteq \{\omega \in \Omega : \frac{W_t(\omega)}{t} = 0\},$$

we have

$$\mathbb{P}(A) = 1 \text{ and } \bar{\mathbb{Q}}(A) = 0$$

In the above conclusion, we used the strong law of large numbers for both probability measures. Therefore, $\mathbb{P}$ and $\mathbb{Q}$ are not equivalent.

Define, $\mathbb{Q}(A) \doteq \mathbb{E}[\mathbb{1}_A \tilde{Z}_t]$, for $A \in \cup_{t \geq 0} \mathcal{F}_t$. To summarize, $\mathbb{Q}$ can be defined on $\cup_{t \geq 0} \mathcal{F}_t$ and not on $\mathcal{F}_\infty$. But, this will not be an issue in our case. The reason is that we use the change of measure only for finite time, make the computations under new probability measure and transform back to the old probability measure using Bayes' rule. This is the main reason for not studying filter stability problem under the new measure (see next chapter).

We make the following assumption that ensures that $\tilde{Z}_t(X_{[0,t]}, Y_{[0,t]})$ is an $\mathcal{F}_t$-martingale [86, Corollary 3.5.16].

Assumption 2.1.6. $S = \mathbb{R}^m$

Assumption 2.1.7. *$h : S \to \mathbb{R}^n$ is such that $\|h(x)\| \leq K(1 + \|x\|)$, for any $x \in S$ and for some constant $K > 0$.*

Now returning to the reference probability method, from Girsanov's theorem (see Theorem (A.0.22)), Y_t is $\mathcal{F}_t$- Brownian motion under $\mathbb{Q}$. Another consequence of Girsanov's theorem is captured by the following proposition [15, Proposition 3.13].

Proposition 2.1.8. *Under $\mathbb{Q}$ and with Assumptions (2.1.6) and (2.1.7), processes X[3] and Y are independent and law of X is same under $\mathbb{Q}$ and $\mathbb{P}$.*

[3] We denote any process $\{A_t\}_{t \geq 0}$ by A

Using the above proposition and Bayes' rule, we have

$$\pi_t(g) = \frac{\int_{C([0,\infty),S)} g(x_t) \exp\left(\int_0^t h(x_s)^T dY_s - \frac{1}{2}\int_0^t \|h(x_s)\|^2 ds\right) dP^X(x)}{\int_{C([0,\infty),S)} \exp\left(\int_0^t h(x_s)^T dY_s - \frac{1}{2}\int_0^t \|h(x_s)\|^2 ds\right) dP^X(x)}, \mathbb{P}\ a.s. \quad (2.2)$$

where, $g \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, P^X is the law of X_t and $\mathbb{E}$ stands for $\mathbb{E}^{\mathbb{P}}$ (we use this notation from now on). The above equation is known as Kallianpur-Striebel formula. Note that the above formula is not incremental, *i.e.*, to compute $\pi_t(g)$, we have to use the entire observation process (upto t) $Y_t : 0 \leq s \leq t$.

In the following, we use the reference probability method to derive the evolution equations for $\sigma_t(g) \doteq \mathbb{E}^{\mathbb{Q}}[g(X_t)|\mathcal{Y}_t]$ and $\pi_t(g)$. We first derive the equation for $\sigma_t(g)$ and then use Kallianpur-Striebel formula along with Ito's rule to get the equation for $\pi_t(g)$.

To simplify the derivations, we focus only on a particular case of Assumption (2.1.3). Let X_t be a strong solution of the following stochastic differential equation:

$$dX_t = f(X_t)dt + N(X_t)dB_t, \quad (2.3)$$

where, $f : \mathbb{R}^m \to \mathbb{R}^m$, $N : \mathbb{R}^m \to \mathbb{R}^{m \times p}$ and B_t is $\mathcal{F}_t$- Brownian motion. In addition, let us also assume the following: NN^T is a strictly positive matrix and for some constant $K > 0$,

$$\|f(x) - f(y)\| + \|N(x) - N(y)\| \leq K\|x - y\|, \ \forall x,\ y \in \mathbb{R}^m$$
$$\|f(x)\|^2 + \|N(x)\|^2 \leq K(1 + \|x\|^2)$$

It is well known from Ito's stochastic calculus [86, Theorem 5.2.9] that X_t exists. Define,

$$\mathcal{A} \doteq \frac{1}{2}\sum_{i,j=1}^{m} (N(x)N(x)^T)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{m} f_i(x) \frac{\partial}{\partial x_i}$$

We choose the domain $\mathcal{D}((A) = C_b^2(\mathbb{R}^m)$, space of twice differentiable bounded continuous functions on $\mathbb{R}^m$. From Ito's formula (see Theorem (A.0.17)) and properties of Ito's Integral, we have

$$M_t^{\phi} := \phi(X_t) - \phi(X_0) - \int_0^t \mathcal{A}\phi(X_s)ds, \forall \phi \in \mathcal{D}(\mathcal{A})$$

is a $\mathcal{F}_t$- martingale. Clearly, this is special case of Assumption (2.1.3).

Remark 2.1.9. *It is important to note that the evolution equations that will be derived* *below hold for X_t belonging to a much larger class. We refer to [15] for further details on dealing with larger class of signal processes X_t.*

Let us define for any $g : \mathbb{R}^m \to \mathbb{R} \in C_b^2(\mathbb{R}^m)$

$$\sigma_t(g) \doteq \int_{C([0,\infty),\mathbb{R}^m)} g(x_t) Z_t(x_{[0,t]}, Y_{[0,t]}) P^X(dx)$$

Recall that

$$Z_t(x_{[0,t]}, y_{[0,t]}) = \exp\left(\int_0^t h(x_s)^T dy_s - \frac{1}{2}\int_0^t \|h(x_s)\|^2 ds \right)$$

Now, Kallianpur-Striebel formula can be written as

$$\pi_t(g) = \frac{\sigma_t(g)}{\sigma_t(1)}, \mathbb{P} - \ a.s.$$

Theorem 2.1.10. *Under $\mathbb{P}$, suppose $\mathcal{L}(X_0) = \mu$. Assume the following conditions:*

1. $\mathbb{E}[\int_0^t \|Z_s \mathcal{A} g(X_s)\| ds] < \infty$.
2. $\mathbb{E}[(M_s^g)^2] < \infty$.
3. $\mathbb{E}^{\mathbb{Q}}[\int_0^t \|g(X_s) Z_s h(X_s)\|^2 ds] < \infty$.

Then the process σ_t follows the evolution equation, called the Zakai equation given below

$$\sigma_t(g) = \mu(g) + \int_0^t \sigma_s(\mathcal{A}g) ds + \int_0^t \sigma_s(gh^T) dY_s, \ \ \mathbb{Q} - \ a.s. \ \forall t \geq 0, \tag{2.4}$$

where, $g \in \mathcal{D}(\mathcal{A})$

Proof. We know that, from Ito's formula, Z_t satisfies the following satisfies the following equation:

$$Z_t(X_{[0,t]}, Y_{[0,t]}) = 1 + \int_0^t Z_s(x_{[0,s]}, y_{[0,s]}) h(X_s)^T dY_s. \tag{2.5}$$

And also, recall that

$$g(X_t) = g(X_0) + \int_0^t \mathcal{A} g(X_s) ds + M_t^g \tag{2.6}$$

To keep the expressions small, we suppress the arguments in $Z_t(x_{[0,t]}, y_{[0,t]})$. Applying the Ito's formula to $g(X_t)Z_t$, we have

$$g(X_t) Z_t = g(X_0) Z_0 + \int_0^t Z_s dg(X_s) + \int_0^t g(X_s) dZ_s + \frac{1}{2}\langle g(X), Z \rangle_t$$

$$
\begin{aligned}
&= g(X_0) + \int_0^t Z_s \mathcal{A} g(X_s) ds + \int_0^t Z_s dM_s^g + \int_0^t g(X_s) Z_s h(X_s)^T dY_s \\
&+ \frac{1}{2} \int_0^t Z_s h(X_s)^T \langle Y, M^g \rangle_s \\
&= g(X_0) + \int_0^t Z_s \mathcal{A} g(X_s) ds + \int_0^t Z_s dM_s^g + \int_0^t g(X_s) Z_s h(X_s)^T dY_s
\end{aligned}
$$

In the above, $\langle Y, M^g \rangle_t = 0$ from independence of Y_t and M_t^g under the measure $\mathbb{Q}$ (see Proposition (2.1.8)). Note that X_t and Y_t are independent under $\mathbb{Q}$, from Proposition (2.1.8). Therefore, $\mathbb{E}^{\mathbb{Q}}[\cdot|\mathcal{Y}_t]$ is same as considering marginal on X. Applying $\mathbb{E}^{\mathbb{Q}}[\cdot|\mathcal{Y}_t]$ on both sides, we have

$$
\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[g(X_t)Z_t|\mathcal{Y}_t] &= \mathbb{E}^{\mathbb{Q}}[g(X_0)|\mathcal{Y}_t] + \mathbb{E}^{\mathbb{Q}}[\int_0^t Z_s \mathcal{A} g(X_s) ds|\mathcal{Y}_t] \\
&+ \mathbb{E}^{\mathbb{Q}}[\int_0^t Z_s dM_s^g|\mathcal{Y}_t] + \mathbb{E}^{\mathbb{Q}}[\int_0^t g(X_s) Z_s h(X_s)^T dY_s|\mathcal{Y}_t] \\
\sigma_t(g) &= \mu(g) + \int_0^t \mathbb{E}^{\mathbb{Q}}[Z_s \mathcal{A} g(X_s)|\mathcal{Y}_t] ds \\
&+ \mathbb{E}^{\mathbb{Q}}[\int_0^t Z_s dM_s^g|\mathcal{Y}_t] + \int_0^t \mathbb{E}^{\mathbb{Q}}[g(X_s) Z_s h(X_s)^T|\mathcal{Y}_t] dY_s \\
\sigma_t(g) &= \mu(g) + \int_0^t \mathbb{E}^{\mathbb{Q}}[Z_s \mathcal{A} g(X_s)|\mathcal{Y}_t] ds + \int_0^t \mathbb{E}^{\mathbb{Q}}[g(X_s) Z_s h(X_s)^T|\mathcal{Y}_t] dY_s \\
\sigma_t(g) &= \mu(g) + \int_0^t \sigma_s(\mathcal{A}g)\, ds + \int_0^t \sigma_s\left(gh^T\right) dY_s
\end{aligned}
$$

In the above, we used the fact that, under $\mathbb{Q}$, X and Y are independent and X has the same law under both $\mathbb{Q}$ and $\mathbb{P}$. Therefore, we have

$$
\mathbb{E}^{\mathbb{Q}}[g(X_0)|\mathcal{Y}_t] = \mathbb{E}^{\mathbb{Q}}[g(X_0)] = \mathbb{E}[g(X_0)] = \mu(g)
$$

From Fubini's theorem, we have

$$
\mathbb{E}^{\mathbb{Q}}[\int_0^t Z_s \mathcal{A} g(X_s) ds|\mathcal{Y}_t] = \int_0^t \mathbb{E}^{\mathbb{Q}}[Z_s \mathcal{A} g(X_s)|\mathcal{Y}_t] ds
$$

From Fubini's theorem for stochastic integral [99, Theorem 5.15 and 5.14],

$$
\mathbb{E}^{\mathbb{Q}}[\int_0^t g(X_s) Z_s h(X_s)^T dY_s|\mathcal{Y}_t] = \int_0^t \mathbb{E}^{\mathbb{Q}}[g(X_s) Z_s h(X_s)^T|\mathcal{Y}_t] dY_s
$$

From the independence of X and Y, we have

$$
\mathbb{E}^{\mathbb{Q}}[\int_0^t Z_s dM_s^g|\mathcal{Y}_t] = 0
$$

■

Note that

$$\sigma_t(1) = 1 + \int_0^t \sigma_s(h^T)dY_s \tag{2.7}$$

We now use the following fact to derive the evolution equation for π_t:

$$\pi_t(g) = \frac{\sigma_t(g)}{\sigma_t(1)}$$

Theorem 2.1.11. *Under the conditions of Theorem (2.1.10), π_t satisfies the following evolution equation, known as Kushner-Stratanovich equation.*

$$\pi_t(g) = \mu(g) + \int_0^t \pi_s(\mathcal{A}g)\, ds + \int_0^t \left(\pi_s(gh^T) - \pi_s(g)\pi_s(h^T)\right)(dY_s - \pi_s(h)\, ds)\,, \tag{2.8}$$

where, $g \in \mathcal{D}(\mathcal{A})$.

Proof. Applying Ito's formula (see Theorem (A.0.17)) to $\frac{\sigma_t(g)}{\sigma_t(1)}$, we have

$$\begin{aligned}
\frac{\sigma_t(g)}{\sigma_t(1)} &= \frac{\sigma_0(g)}{\sigma_0(1)} + \int_0^t \frac{1}{\sigma_s(1)} d\sigma_s(g) - \int_0^t \frac{\sigma_s(g)}{\sigma_s^2(1)} d\sigma_s(1) \\
&\quad + \frac{1}{2}\left(-\int_0^t \frac{1}{\sigma_s^2(1)} d\langle\sigma(g), \sigma(1)\rangle_s + 2\int_0^t \frac{1}{\sigma_s^3(1)} d\langle\sigma(1)\rangle_s\right) \\
\frac{\sigma_t(g)}{\sigma_t(1)} &= \mu(g) + \int_0^t \frac{1}{\sigma_s(1)}\left(\sigma_s(\mathcal{A}g)ds + \sigma_s(gh^T)dY_s\right) - \int_0^t \frac{\sigma_s(g)}{\sigma_s^2(1)}\sigma_s(h^T)dY_s \\
&\quad + \left(-\int_0^t \frac{1}{\sigma_s^2(1)}\sigma_s(gh^T)\sigma_s(h)ds + \int_0^t \frac{1}{\sigma_s^3(1)}\sigma_s(h^T)\sigma_s(h)ds\right) \\
\pi_t(g) &= \mu(g) + \int_0^t \pi_s(\mathcal{A}g)ds + \int_0^t \left(\pi_s(gh^T) - \pi_s(g)\pi_s(h^T)\right)(dY_s - \pi_s(h)ds)
\end{aligned}$$

■

Till now, we have derived the evolution equation for π_t (known as Kushner-Stratanovich equation) using the evolution equation of σ_t (known as Zakai equation). We now briefly describe the other approach to deriving the Kushner-Stratanovich equation.

2.1.2 Innovation process approach

In this approach, we directly derive the Kushner-Stratanovich equation using a representation theorem. And also, the assumption that processes X and W are independent can be easily relaxed. Here, the main object of use is the innovations process $I_t \doteq Y_t - \int_0^t \pi_s(h)ds$. Before proceeding further, we need to make following assumption on the system:

Assumption 2.1.12. $\mathbb{E}\left[\int_0^t \|h(X_s)\|^2 ds\right] < \infty, \forall t \geq 0$

Let us mention the following important representation theorem that is the basis for the entire approach.

Theorem 2.1.13. *[80, Theorem 3.1][Fujisaki-Kallianpur-Kunita formula] Let M_t be a right continuous square integrable $\mathcal{Y}_t$- martingale. Then there exists a representation such that*

$$M_t = \mathbb{E}[M_0] + \int_0^t \nu_s^T dI_s, \tag{2.9}$$

where, ν_t is a progressively measurable process that is $\mathcal{Y}_t$- adapted such that

$$\mathbb{E}\left[\int_0^T \|\nu_s\|^2 ds\right] < \infty.$$

And also, M_t is continuous.

Remark 2.1.14. *At first glance, it may seem that the above theorem is a special case of the usual martingale representation theorem [86, Theorem 3.4.15]. But, we apriori do not know if $\mathcal{Y}_t = \sigma(I_s : 0 \leq s \leq t) \vee \mathcal{N}$. So, this is still a non-trivial result which states that any right continuous square integrable $\mathcal{Y}_t$- martingale can be written as a stochastic integral with respect to I_t. In general, it is known that (from[20]) $\sigma(I_s : 0 \leq s \leq t) \vee \mathcal{N} \subsetneq \mathcal{Y}_t$. [82, Theorem 11.4.1] and [2] (due to D. F. Allinger and S. K. Mitter) give sufficient conditions for the equality to hold.*

It is clear that the process I_t plays a very important role in innovations process approach and below result gives its most important property.

Proposition 2.1.15. *The process I_t is $\mathcal{Y}_t$- Brownian motion.*

Proof. Note that $Y_t = \int_0^t h(X_s)ds + W_t$. We now prove that I_t is a continuous $\mathcal{Y}_t$-martingale and apply Levy's characterization theorem (see Theorem (A.0.21)) to get the result. To that end, since Y_t and $\int_0^t \pi_t(h) = \int_0^t \mathbb{E}[h(X_t)|\mathcal{Y}_t]$ are $\mathcal{Y}_t$- adapted, continuity of I_t follows from the continuity of both Y_t and $\int_0^t \pi_s(h)ds$.

For any $0 \leq s \leq t$, Consider

$$\begin{aligned}
\mathbb{E}[I_t - I_s|\mathcal{Y}_s] &= \mathbb{E}[Y_t - Y_s - \int_0^t \pi_u(h)du + \int_0^s \pi_u(h)du|\mathcal{Y}_s] \\
&= \mathbb{E}[W_t - W_s + \int_0^t h(X_u)du - \int_0^s h(X_u)du - \int_s^t \pi_u(h)du|\mathcal{Y}_s] \\
&= \mathbb{E}[W_t - W_s + \int_s^t h(X_u)du - \int_s^t \pi_u(h)du|\mathcal{Y}_s] \\
&= \mathbb{E}[W_t - W_s|\mathcal{Y}_s] + \mathbb{E}[\int_s^t h(X_u)du - \int_s^t \pi_u(h)du|\mathcal{Y}_s]
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}[W_t - W_s|\mathcal{Y}_s] + \int_s^t \mathbb{E}[h(X_u) - \pi_u(h)|\mathcal{Y}_s]du \\
&= \mathbb{E}[W_t - W_s|\mathcal{Y}_s] + \int_s^t \mathbb{E}[h(X_u) - \mathbb{E}[h(X_u)|\mathcal{Y}_u]|\mathcal{Y}_s]du \\
&= \mathbb{E}[W_t - W_s|\mathcal{Y}_s] + \int_s^t \mathbb{E}[h(X_u)|\mathcal{Y}_s] - \mathbb{E}[h(X_u)|\mathcal{Y}_s]du \\
&= \mathbb{E}[\mathbb{E}[W_t - W_s|\mathcal{F}_s]|\mathcal{Y}_s] = 0
\end{aligned}$$

In the last two equalities, we used the fact $\mathcal{Y}_s \subset \mathcal{Y}_u$, $\mathcal{Y}_s \subset \mathcal{F}_s$ (see properties of conditional expectation in Appendix (A)) and W_t is a $\mathcal{F}_t$- Brownian motion. Therefore, I_t is a $\mathcal{Y}_t$- martingale.

Since $I_t = \int_0^t h(X_s)ds + W_t - \int_0^t \pi_s(h)ds$, the quadratic variation $\langle I \rangle_t$ is same as that of the martingale part W_t, *i.e.*,

$$\langle I \rangle_t = \langle W \rangle_t = t$$

Therefore, from Levy's characterization theorem, I_t is $\mathcal{Y}_t$- Brownian motion. ■

Using Theorem (2.1.13), we shall derive the Kushner Stratanovich equation.

Theorem 2.1.16. *Suppose $\mathcal{L}(X_0) = \mu$. Under the Assumption* (2.1.12)*, π_t satisfies Equation* (2.1.11), i.e,

$$\pi_t(g) = \mu(g) + \int_0^t \pi_s(\mathcal{A}g)\, ds + \int_0^t \left(\pi_s(gh^T) - \pi_s(g)\pi_s(h^T)\right)(dY_s - \pi_s(h)\, ds),$$

where, $g \in \mathcal{D}(\mathcal{A})$.

Proof. We follow the proof of [15, Theorem 3.35]. Choose $g \in \mathcal{D}(\mathcal{A}) = C_b^2(\mathbb{R}^m)$. Firstly, we show that $U_t \doteq \pi_t(g) - \int_0^t \pi_s(\mathcal{A}g)ds$ is a square integrable $\mathcal{Y}_t$- martingale. Since g and $\mathcal{A}g$ are bounded, U_t is square integrable. And also, from the property of π_t, U_t is $\mathcal{Y}_t$- adapted. It only remains to show that U_t is a martingale.

To that end, for $0 \leq s \leq t$, consider

$$\mathbb{E}[U_t - U_s|\mathcal{Y}_s] = \mathbb{E}[\pi_t(g) - \pi_s(g) - \int_0^t \pi_u(\mathcal{A}g)du + \int_0^s \pi_u(\mathcal{A}g)du|\mathcal{Y}_s] \tag{2.10}$$
$$= \mathbb{E}[\mathbb{E}[g(X_t)|\mathcal{Y}_t] - \mathbb{E}[g(X_s)|\mathcal{Y}_s] - \int_s^t \mathbb{E}[\mathcal{A}g(X_t)|\mathcal{Y}_t]du|\mathcal{Y}_s] \tag{2.11}$$
$$= \mathbb{E}[\mathbb{E}[g(X_t)|\mathcal{Y}_t] - \mathbb{E}[g(X_s)|\mathcal{Y}_s]\mathbb{E}[-\int_s^t \mathcal{A}g(X_u)du|\mathcal{Y}_u]|\mathcal{Y}_s] \tag{2.12}$$
$$= \mathbb{E}[g(X_t) - g(X_s) - \int_s^t \mathcal{A}g(X_u)du|\mathcal{Y}_s] \tag{2.13}$$
$$= 0, \tag{2.14}$$

where, we used Fubini's theorem for the last term of Equation (2.11) to get Equation (2.12). We then used the property of conditional expectation in Equation (2.12) and the Dynkin's formula for Equation (2.14). This concludes that U_t is a square integrable $\mathcal{Y}_t$- martingale.

Therefore, we can apply Theorem (2.1.13) to U_t and we know that there exists a progressively measurable process ν_t with respect to $\mathcal{Y}_t$ such that

$$U_t = U_0 + \int_0^T \nu_s^T dI_s$$

If we can show that $\nu_s = \pi_s(gh^T) - \pi_s(g)\pi_s(h^T)$, then we are done. To that end, expanding above equation, we have

$$\pi_t = \mu(g) + \int_0^t \pi_s(\mathcal{A}g)ds + \int_0^t \nu_s^T dI_s. \tag{2.15}$$

From properties of X, we also have

$$g(X_t) = g(X_0) + \int_0^t \mathcal{A}g(X_s)ds + M_t^g \tag{2.16}$$

Consider a class of functions,

$$S_t \doteq \left\{\gamma_t : \Omega \to \mathbb{R} : d\gamma_t = i\gamma_t r_t^T dY_t \textit{ with } \gamma_0 = 1, \textit{ where, } r_\cdot \in L^\infty([0,t], \mathbb{R}^m)\right\}$$

Fix $r_\cdot \in L^\infty([0,t], \mathbb{R}^m)$. Applying Ito's formula to $\pi_t(g)\gamma_t$ and $g(X_t)\gamma_t$ with

$$d\gamma_t = ir_t^T dY_t, \; \gamma_0 = 1,$$

we get

$$\begin{aligned}\pi_t(g)\gamma_t &= \mu(g) + \int_0^t \gamma_s d\pi_s(g) + \int_0^t \pi_s(g)d\gamma_s + \langle \gamma, \pi(g)\rangle_t \\ &= \mu(g) + \int_0^t \gamma_s \pi_s(\mathcal{A}g)ds + \int_0^t \gamma_s \nu_s^T dI_s + \int_0^t \pi_s(g)\gamma_s r_s^T (dI_s + \pi_s(h)ds) \\ &+ \int_0^t \gamma_s (r_s^T \nu_s) ds\end{aligned}$$

$$\begin{aligned}g(X_t)\gamma_t &= \mu(g) + \int_0^t \gamma_s dg(X_s) + \int_0^t g(X_s)d\gamma_s + \langle \gamma, g\rangle_t \\ &= \mu(g) + \int_0^t \gamma_s \mathcal{A}g(X_s)ds + \int_0^t \gamma_s dM_s^g + \int_0^t g(X_s)\gamma_s r_s^T dY_s + \langle Y, M^g\rangle_t\end{aligned}$$

Since W and X are independent, $\langle Y, M^g\rangle_t = \langle W, M^g\rangle_t = 0$. Subtracting and taking expectation, we have

$$\begin{aligned}\mathbb{E}[(\pi_t(g) - g(X_t))\,\gamma_t] &= \mathbb{E}[\int_0^t \gamma_s\,(\pi_s(\mathcal{A}g) - \mathcal{A}g(X_s))\,ds] \\ &+ \mathbb{E}[\int_0^t \gamma_s r_s^T \left(\pi_s(g)\pi_s(h^T) - g(X_s)h^T(X_s)\right) ds] \\ &+ \mathbb{E}[\int_0^t \gamma_s (r_s^T \nu_s) ds]\end{aligned}$$

We used the fact that expectation of martingale terms is zero. We also note that ν_t is $\mathcal{Y}_t$-measurable and clearly, so does γ_t.

$$\begin{aligned}\mathbb{E}[\mathbb{E}[g(X_t)\gamma_t|\mathcal{Y}_t] - g(X_t)\gamma_t] &= \int_0^t \mathbb{E}[\mathbb{E}[\gamma_s \mathcal{A}g(X_s)|\mathcal{Y}_s] - \gamma_s \mathcal{A}g(X_s)]ds \\ &+ \int_0^t \mathbb{E}[\gamma_s r_s^T\,(\pi_s(g)\pi_s(h) - g(X_s)h(X_s))]ds \\ &+ \int_0^t \mathbb{E}[\gamma_s (r_s^T \nu_s)]ds \\ 0 &= \int_0^t \mathbb{E}[\gamma_s r_s^T\,(\pi_s(g)\pi_s(h) - \mathbb{E}[g(X_s)h(X_s)|\mathcal{Y}_s] + \nu_s)]ds\end{aligned}$$

In the above, we used Fubini's theorem in exchanging the integrals and also used the properties of conditional expectation (see Appendix (A)). We finally ended up

$$\int_0^t r_s^T \mathbb{E}[\gamma_s\,(\pi_s(g)\pi_s(h) - \mathbb{E}[g(X_s)h(X_s)|\mathcal{Y}_s] + \nu_s)]ds = 0, \;\forall t \geq 0.$$

From above, we have

$$\mathbb{E}[\gamma_s\left(\pi_s(g)\pi_s(h) - \mathbb{E}[g(X_s)h(X_s)|\mathcal{Y}_s] + \nu_s\right)] = 0$$

From [22, Lemma 4.1.4], we conclude that

$$\nu_s = \pi_s(gh) - \pi_s(g)\pi_s(h),\ \mathbb{P} - \ a.s.$$

■

Remark 2.1.17. *In the proof of above theorem, independence of processes X and W is used in a crucial way. But this approach can also be used to derive corresponding evolution equation for correlated noise case (where, X and W are correlated). For more details, we refer to [15, Section 3.8].*

Remark 2.1.18. *Till now, we have dealt with deriving equations for π_t and σ_t. Analysis of these equations lack mathematical rigor unless the existence and uniqueness results of these equations are established. We refer to [15, 148] for positive results in this direction. In discrete time case, the conditional expectations given in a recusive form and the conditions for existence of densities for conditional distribution is very simple. In continuous time, as we have seen already, the conditional expectations are given in an implicit way through Kushner-Stratanovich equations. In this case, the question of existence of densities for conditional distribution becomes very non-trivial. We refer to [15, Theorem 7.8, Corollary 7.18] for the results on existence and smoothness of conditional density. When the conditional density exists, the evolution of the unnormalised conditional density (p_t) is given by*

$$p_t(x) = p_0(x) + \int_0^t \mathcal{A}^* p_s(x)ds + \int_0^t h^T p_s(x)dY_s$$

One can also arrive at an evolution equation without any stochastic integral term by defining $\tilde{p}_t(x) \doteq \exp\left(-Y_t^T h(x) + \frac{1}{2}\|h(x)\|^2 t\right)$. $\tilde{p}_t(x)$ satisfies the following equation

$$d\tilde{p}_t(x) = \exp\left(-Y_t^T h(x) + \frac{1}{2}\|h(x)\|^2 t\right)\mathcal{A}^*\left(\exp\left(Y_t^T h(x) - \frac{1}{2}\|h(x)\|^2 t\right)\tilde{p}_t(x)\right)dt$$

with $\tilde{p}_0(x) = p_0(x)$. See [15, Section 7.3] for more details.

Remark 2.1.19. *Often in practice, the observations are smoother than a typical continuous path as the background noise in the observations is never truly white noise. Due to this, it is desirable for the us to have filter be continuous with respect to the observation path. In the framework that is studied until now, this is not possible. This is because the filtering equation and the Kallianpur-Striebel formula involve a stochastic integral over the observation* *path which can only be guaranteed to be measurable and can only be defined upto a measure one set. This motivated the rise of a new approach to filtering known as the pathwise filtering introduced by J. M. C. Clark [49] and subsequently studied by M. H. A. Davis [53] and other authors. In this approach, filter is a continuous functional on the observation path space. See [15, Chapter 5] for more details.*

To recap, in an abstract filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, we have the following model:

$$dX_t = f(X_t)dt + N(X_t)dB_t, \; \mathcal{L}(X_0) = \mu \tag{2.17}$$

$$dY_t = h(X_t)dt + dW_t, \tag{2.18}$$

where $W, \; B, \; X_0$ are mutually independent. $W, \; B$ are n and m- dimensional $\mathcal{F}_t$-Brownian motions, respectively. $X \in \mathbb{R}^m, \; Y \in \mathbb{R}^n$ are referred to as signal and observation process, respectively. Our goal is compute and study the following object:

$$\mathbb{E}[g(X_t)|\mathcal{Y}_t],$$

where, g belongs to an appropriate class of functions on $\mathbb{R}^m$. Since, in this entire endeavour, there are only two processes that really matter, *viz.*, X and Y, it is useful to work with the explicit construction (instead of abstract setup) of these processes for computations. We describe the explicit construction below.

2.2 Explicit construction of the X and Y

Since we desire for X and Y to be $\mathbb{R}^m$ and $\mathbb{R}^n$-valued continuous processes, we consider the space to be

$$\Omega = C\left([0,\infty), \mathbb{R}^m\right) \times C\left([0,\infty), \mathbb{R}^n\right),$$

where, $C\left([0,\infty), \mathbb{R}^q\right)$ is space of $\mathbb{R}^q$- valued continuous functions on $[0,\infty)$ with the underlying topology as the uniform topology. The σ- algebra is choosen as

$$\mathcal{F} = \mathcal{B}(\Omega),$$

where, $\mathcal{B}(\cdot)$ is the Borel σ- algebra. On this measurable space, we define the canonical coordinate process $\omega_t \doteq (x_t, y_t)$ such that

$$\omega_{\cdot} : \Omega \to \mathbb{R}^m \times \mathbb{R}^n$$

Using this canonical coordinate process, we define the processes X and Y as follows.

$$X_t : \Omega \to \mathbb{R}^m, \; X_t(\omega_t) = x_t$$
$$Y_t : \Omega \to \mathbb{R}^n, \; Y_t(\omega_t) = y_t$$

Define,

$$\mathcal{H}_t \doteq \sigma(\omega_s : 0 \leq s \leq t), \; \mathcal{Y}_t \doteq \sigma(Y_s : 0 \leq s \leq t),$$

We choose the probability measure $\mathbb{P}$ such that the following holds:

1. $W_t \doteq Y_t - \int_0^t h(X_s)ds$ is $\mathcal{H}_t$- Brownian motion under $\mathbb{P}$
2. X and W are independent.
3. Marginal of $\mathbb{P}$ on X is P^X, the law of the process given by Equation (2.17).

Consider the following as the filtration:

$$\mathcal{F}_t \doteq \mathcal{H}_t \vee \mathcal{N},$$

where, $\mathcal{N}$ is the set of $\mathbb{P}$- null sets. This filtration is right continuous [86, Section 2.7]. Therefore, the constructed probability space can be used for filtering problems.

2.3 Finite dimensional filters

From Equation (2.2), it is clear that, in general, to completely determine the conditional probability, it may not be sufficient to compute $\pi_t(g)$ for finitely many g. But there are few exceptions to this general rule. The filters that can be completely determined by finitely many g's are referred to as finite dimensional filters. The following are some concrete examples of finite dimensional filters:

1. (Kalman-Bucy Filter) This filter was given by R.E. Kalman and R.S. Bucy [85]. The signal and the observation processes satisfy linear stochastic differential equation and the initial condition of signal X_0 is Gaussian. Under this setup, the conditional probability is also Gaussian. Therefore, we can completely determine $\pi_t(g)$ if we evaluate for $g(x) = x_i$ and $g(x) = x_i x_j$, where $x = (x_1, x_2, x_3, \dots, x_m)^T$. More details will be given in Chapter (4). For non-Gaussian initial conditions, it is shown that filter is finite dimensional using different methods in [104, 21].

2. (Beněs Filter) Given the success of Kalman-Bucy filter [79, 34], V.E. Beněs showed that certain class of diffusions with non-linear drift and linear observational model correspond to finite dimensional filter [19]. The model he considered is

$$dX_t = f(X_t) + dB_t,$$
$$dY_t = X_t dt + dW_t,$$

   where, f satisfies the following equation: $f'(x) + f^2(x) = ax^2 + bx + c$ with $a \geq -1$. The linear model (signal and observation process satisfy linear SDE) belongs to the above class of models.

3. A general class of filtering models that includes both the above filters is studied in [52].

The systematic study of finite dimensionality of the filter was inspired by the works of Brockett, Mitter and Clark [109, 31, 32]. They considered a Lie algebra that is generated by operators that appear in Zakai equation (written in terms of unnormalised density). The finite dimensionality of the filter is very closely related to the properties of this Lie algebra [14, Chapter 2] [115]. For more details in this regard, see [15, Section 1.3] and the references therein. It is now believed that it is very common for a non-linear filter to be genuinely infinite dimensional [76, 75].

To summarize, in this chapter, we developed the general framework of filtering. We studied two known approaches to filtering, *viz.,* reference probability approach and innovations process approach. Even though in a general framework we worked with an abstract probability space, we explicitly constructed the signal and observation process and setup the filtering problem in an explicit probability space. We finally discussed the finite dimensionality of the filter.

Chapter 3

Stability of the Filter

In this chapter, we introduce the notion of filter stability, briefly describe the tools and give the relevant literature.

In Chapter (2), we have already seen that π_t satisfies the following equation:

$$\pi_t(g) = \mu(g) + \int_0^t \pi_s(\mathcal{A}g)\,ds + \int_0^t \Big(\pi_s(gh^T) - \pi_s(g)\pi_s(h^T)\Big)\left(dY_s - \pi_s(h)\,ds\right), \tag{3.1}$$

where, $g \in \mathcal{D}(\mathcal{A})$. Clearly, the conditional distribution depends on μ, the law of the initial condition X_0. In practice, we rarely have the knowledge of the initial condition. This can happen because initialising the signal process may not be in the control of the observer. Since the signal process is being indirectly observed, we lack the knowledge of the initial condition. The question of resolving this issue gives rise to the notion of filter stability.

3.1 Filter Stability

Because of the reasons mentioned above and the fact that Equation (3.1) depends on μ, the law of X_0, it is desirable for the filter to be independent μ. Of course, the independence we desire is asymptotic in time (that is, after large times). In other words, we can initialise Equation (3.1) with other initial conditions (referred to as incorrect initial conditions) and after large times, still be close to the actual filter. This brings us to definition below. From now on, we denote the solution of Equation (3.1) (initialised by β) by π_t^{β}. Filter π_t is now denoted by π_t^{μ}, where, μ is the law of X_0.

Definition 3.1.1. *The filter is said to be stable if $\|\pi_t^{\mu} - \pi_t^{\nu}\| \to 0$ as $t \to \infty$, in an appropriate metric, for a class of distribution ν.*

The above definition is not precise as the metric varies from case to case and so does the class of distribution.

Remark 3.1.2. *At the first glance, it may seem like a usual stability problem for dynamical systems, that is, the long time effects of initialising the system with different initial conditions. This is incorrect. In the filter stability, we do not initialise the system (which is the signal process X) with different initial conditions, but rather just the Equation* (3.1) *is initialised with different initial conditions. The main difference between these two ways of studying the problem is that in the former, even the the observation process Y is different for different initial conditions, whereas the observation process remains the same in the latter. Therefore, it is useful to think of filter stability as the stability of Equation* (3.1).

Remark 3.1.3. *Another important point to note is that unlike* π_t^μ, $\pi_t^\nu(g)$ *cannot be written as the conditional expectation of* $g(X_t)$ *with respect to* $\mathcal{Y}_t$.

The filter stability (referred to as stability, from now on) was first studied for the case of linear signal and linear observational model in the seminal paper of R.E. Kalman and R.S. Bucy [85]. [95] contains detailed exposition of discrete and continuous version of linear filter. The authors of [85, 33] establish stability under assumptions of uniform complete observability and uniform complete controllability. In [3], stability was established under weaker notion of controllability. Then after, the asymptotic properties of the filter in the non-linear case were not studied until the famous work of H. Kunita [92]. In this work, author showed that if the signal is Feller-Markov and the observational model is of the form (2.1), then π_t^μ is also Feller-Markov on the space of probability measures. Even though the result was established for compact S, it was later extended to the locally compact case [93, 136] and to non-compact spaces [24]. Using the results of [92], authors of [116] showed that if the X_t is uniquely ergodic then the filter is stable (without any rate of convergence)[1]. Since then, the relation between the ergodicity of the signal and the stability of filter was studied extensively.

In contrast to the non-linear case, as already mentioned, stability for the linear case (which is non-ergodic) was established under the assumptions that observations are rich enough. And also, when S is finite, it is observed that ergodicity and good enough observations ensured stability (we will elaborate on this later). This suggests that in order to guarantee stability, system has to obey either of the following conditions (in an appropriate sense):

[1]Unfortunately, it turned out that the proof of Theorem 3.3 in [92] is incorrect (there was an exchange of intersection and supremum operations of σ- algebras without any proof). We refer the reader to [18] for more details on this issue. Budhiraja (in [35]) showed that under appropriate conditions, unique ergodicity of the filter is equivalent to its stability. Van Handel (in [71]) showed that one can construct a nondeterministic example of a filtering problem where the exchange of the aforementioned operations is not allowed

1. (Mixing) The process X is sufficiently ergodic *i.e.*, law of X_t is asymptotically independent of law of X_0. This in turn, makes the process Y to forget about μ, the law of X_0.

2. (Observability) The observations made are rich enough that observing long enough gives the entire information about the law of X_0. In other words, even though we initialised Equation (3.1) with different initial condition ν (referred to as incorrect initial condition, from now on), π_t^ν comes close to π_t^μ due to common driving term that contains the information of X_0.

Note 3.1.4. *In this thesis, we focus on the latter one to establish the stability. We indeed show that if X is deterministic, then good enough observations will guarantee stability.*

In light of above evidence, we give the literature review by categorizing the existing results into two groups *viz.,*

1. (Ergodicity assumptions) Results which assume some kind of ergodicity of X.

2. (Observability assumptions) Results which assume that observations are rich enough.

Results are further categorized based on compactness of S.

We now give the literature review on stability in the case of both continuous and discrete times.

3.2 Literature review

In this section, for the sake of reviewing the results, we consider the following continuous time observational model (instead of (2.1)):

$$Y_t = \int_0^t h(X_s)ds + \eta W_t \tag{3.2}$$

3.2.1 Ergodicity assumption

S is compact

After the earlier work of [85, 3, 92], the stability results along with the rates of convergence were obtained in the case of finite S (for both continuous and discrete time). The continuous time filter with finite signal state space S with observation model of the form (3.2) is referred to as Wonham's filter[147].

Since S is finite (with cardinality say, d), π_t^μ and π_t^ν are coordinates of points on a d- dimensional simplex and therefore, all the norms are equivalent. Under the ergodicity of X and sufficiently high signal to noise ratio in observation, authors of [55] established that

$$\gamma_\eta \doteq \lim_{t\to\infty} \frac{1}{t} \log \|\pi_t^\mu - \pi_t^\nu\| \leq -\delta_\eta, \tag{3.3}$$

where, η denotes the strength of the observation noise and $\delta_\eta > 0$ is some constant depending on η. The Lyapunov exponent techniques were introduced for the first time in filtering theory context.

In [10], Wonham filter and discrete version of Wonham filter is considered. Applying the techniques of multiplicative ergodic theory to the unnormalised conditional density (which evolves linearly, both in discrete and continuous time), authors showed that (3.3) holds and moreover, γ_η is related to the gap of top two Lyapunov exponents of the evolution of unnormalised conditional density. In addition, dependence of γ_η is studied in the limit $\eta \to 0$ with the conclusion that nice enough observations make $\|\gamma_\eta\| \to \infty$ as $\eta \to 0$.

In [9], it is shown that if X is a diffusion with strictly elliptic generator, then results analogous to those in previous paragraph hold.

In [133], geometric ergodicity of the filter for discrete time filtering model is established along with bounds on the rates of convergence.

In [18], authors considered the continuous time Wonham filter and showed that ergodicity alone can imply (3.3) along with non-asymptotic estimates. And also, even under the relaxed condition of non-ergodicity of the signal and nice observations, filter is stable in the L^1 sense.

In [44], ergodic properties for the Wonham filter is studied. In particular, it was shown that if X is ergodic, then the Feller-Markov pair process (X, π) is uniquely ergodic. Stability was proved using the subsequent limit theorem for pair process (X, π) along with new bounds on the rates.

In [46], discrete time filter with the following observation model is studied:

$$Y_n = \sum_{i=1}^{d} \mathbb{1}_{X_n=i} \xi_n(i),$$

where, $\{\xi_n\}_{n\geq 1}$ is an i.i.d. sequence of d- dimensional random vectors. Markov chain X is such that transition probabilities to different states is small (represented by a small parameter). Assuming appropriate regularity of of the densities of ξ_0 and ergodicity of X, it is shown that discrete analog of (3.3) holds and rate of convergence is lower bounded in terms of relative entropies of $\xi_0(i)$ and $\xi_0(j)$, for $i \neq j$. Moreover, asymptotics of rate of convergence (as the parameter goes to zero) is considered.

Detailed review of techniques used in above results can be found in [45, 7].

In [30], authors show that (3.3) holds for a discrete time filtering model, where the process X is generated from iterating random i.i.d maps which are uniformly expanding.

S is non-compact

The stability question of filter in the case of non-compact S (in continuous time) is first studied by D. Ocone and E. Pardoux in [116], inspired by the work of H. Kunita [92, 93] and Ł. Stettner [136]. They showed that using results of [92], if S is Polish and X is Feller-Markov with unique invariant measure, then filter is stable (This turned out be incorrect, in the light of the error in [92]). This work paved the way for further research on stability of filter.

In [8], authors studied the case of one dimensional diffusion with constant diffusion coefficient that is being observed through a linear observation model with small observation noise. They establish that, under appropriate conditions on the filtering model, distance between conditional densities corresponding to different initial conditions converges to zero exponentially for sufficiently small observation noise. The proof of their result relies on the tight upper and lower bounds of unnormalised density obtained in [150, Theorem 1].

Authors of [9] establish that filter is stable exponentially in the case of continuous and discrete time case using Hilbert projective metric and Birkoff's contraction inequality. They also establish exponentially stability result for signal in countable state space.

In [37], authors studied the one-dimensional discrete time filtering model with bounded observation noise. They established that (3.3) holds in sense of total variation, using Hilbert's projective metric and Birkoff's contraction inequality.

In [36], (3.3) is shown to hold for conditional densities of a discrete time filter, where the signal X is Markov and signal noise satisfies certain non-degeneracy conditions.

In [113], the Beneš filter is shown to hold (3.3) by obtaining the bounds on relative entropy of incorrectly initialised filter and optimal filter.

Under a very general setup (just assuming that (X, Y) is Markovian on a Polish space), relative entropy of incorrectly initialised filter with respect to optimal filter is shown to be a non-negative supermartingale in [50]. They could not derive the sufficiency conditions for convergence of relative entropy to zero. Moreover, it is shown that if Y is of the form (3.2), then under fairly general conditions, $\mathbb{E}[\int_0^\infty \|\pi_s^\mu(h) - \pi_s^\nu(h)\|^2 ds] < \infty$. In other words, $\lim_{t\to\infty} \mathbb{E}[\|\pi_t^\mu(h) - \pi_t^\nu(h)\|^2] = 0$. In [114], assuming that X is diffusion, observation is of the form (3.2) and suitable regularity conditions, the aforementioned relative entropy is decomposed into sum of an explicit decreasing process and a local martingale process.

In [35], author studied the relation among the properties like the stability, uniqueness of invariant measure for pair process, memory of the filter, uniqueness of invariant measure for the filter etc., and showed that under suitable ergodicity assumptions, the aforementioned properties are all equivalent.

In [48], (3.3) is shown to hold for a filter with Markovian X under a relaxed condition which allowed for the Markov transition density to be zero.

In [97], a general discrete time filtering model is considered. It is shown, using Hilbert's projective metric, that (3.3) holds under suitable kind of mixing condition on the kernels that arise in the filtering evolution.

Filtering model with the signal X being an inhomogenous diffusion and the linear time varying observational model is studied in [134, 135]. Author shown that (3.3) holds, using a variational approach on pathwise filter (introduced by the same author in his earlier work) in [134] and using gradient estimates of solutions solutions to heat equations.

In [121], stability of discrete time filtering model along with unknown parameter is considered. It is shown that under certain identifiability condition and mixing conditions, filter is stable . Under appropriate ergodic assumptions, stability of filter (in both discrete and continuous time) is established in [142, 89].

In [72], unique erogodicity of filter (and thereby filter stability) is established under the assumptions of unique ergodicity of the signal and the non-degeneracy of the observation noise.

In [57, 58], in discrete time setup, upper bounds on the total variation distance between the incorrectly initialised filter and optimal filter are derived and particular examples are studied where these bounds are decaying exponentially in time.

3.2.2 Observability assumption

We now give a survey of existing results that assume that observations are sufficiently rich and encode complete information of the observations.

S is compact

A rigorous notion of rich observations (for both discrete and continuous time) is given by Van Handel [139] for the case of compact S. Informally, a filtering model is said to be observable if same law of observation process Y for two initial conditions of X implies that the initial conditions are identical in distribution. In other words, two initial conditions with different distributions can never give rise to same law of

Y, if the system is observable. Under this notion, author showed that filter is stable, without any rates of convergence.

S is non-compact

Even though the notion of rich observations was known for Kalman-Bucy filters [116, 85], for non-linear case, this notion was not studied until the work of Baxendale *et al.*[47]. For a general discrete time filtering model with $Y \in \mathbb{R}$, it is shown that if an integrable function f is such that

$$f(x) = \int_{\mathbb{R}} g(y)\lambda(x,y)R(dy),$$

for some bounded measurable function g (where $\lambda(x,y)$[2] is transition density for Y with respect to a reference measure R), then

$$\lim_{n\to\infty} \mathbb{E}[|\pi_n^{\mu}(f) - \pi_n^{\nu}(f)|] = 0$$

In [78], (3.3) is shown to hold uniformly in η for finite or countable state space (for discrete time) under some notion of observability and suitable conditions on transition matrix for X.

In [138], author proved that a general discrete time filter with observation model $Y_n = h(X_n, \xi_n) \in \mathbb{R}^n$ is stable, if h has a uniformly continuous inverse and ξ_n has a density with respect to Lebesgue measure along with a non-vanishing Fourier transform.

In [73], author extended the notion of observability to Polish spaces. In this work, a stronger notion of uniform observability is introduced. In the earlier notion of observability [139], the map from initial condition to the law of Y is only assumed to be one-one whereas in uniform observability, in addition, the map from law of Y to initial condition has to be uniformly continuous. This stronger notion of observability was shown to imply stability, without any rates of convergence.

In [106, 107, 108], authors defined the non-probabilistic version of observability (in contrast to the notion in [139, 73]. Under the assumption of observability and continuity of kernels involved, filter is shown to be stable. Again the rates of convergence were not provided due to the usage of martingale convergence theorems.

In [105], the expectation of total variation distance between incorrectly initialised and optimal filter is upper bound in terms of Dobrushin's ergodic coefficients of signal and observation kernel, under the assumption of non-degeneracy of observation noise.

[2]In general, $\lambda(\cdot,\cdot)$ and R depend on the observation model.

Recall that filter corresponding to linear signal and linear observation model is referred to as linear filter. The stability problem for the model with linear signal and observation model was studied in [85, 33]. The authors of [116] studied the stability in the linear case for non-Gaussian initial conditions using change of probability measures. We refer to [26] for a detailed review of results in the linear case. In [55], small and large observation noise asymptotics are studied (for Kalman-Bucy filter) along with their relation to stability of the filter.

In [118], a robust filter (constructed from the restricting the support of likelihood functions to a compact set) is used to study the stability for discrete time filtering model. Under the assumption of non-degeneracy of kernel of X and surjectivity of h (but can fail injectivity for arguments that are close enough), authors derived the estimates on the distance between these robust filters initialised under different initial conditions and uniform in time estimates on the distance between the robust filter and the corresponding filter. For certain systems, these estimates were used to prove stability of the filter.

As already mentioned in Chapter (1), the filtering model with deterministic signal is very often encountered in geosciences. The stability problem in this regard is relatively new. In the case of deterministic signal, F. Cérou studied the problem of consistency of the filter *i.e.*, error the filter makes in estimating the state variable. In the rest of literature survey, we assume that the signal is deterministic.

In [68, 28], authors studied the non-autonomous linear filter in discrete time (known as Kalman filter). They, in particular, studied the asymptotics of the covariance of the filter under the assumption of uniform observability (a well known notion in the linear context and will be defined in the next chapter for continuous time).

Considering non-autonomous context and assuming uniform observability, long time behavior of the conditional covariance of the Kalman-Bucy filter is studied in [112] and stability of the Kalman-Bucy filter is derived under non-Gaussian initial conditions in [126].

In non-linear context, the stability of filter is derived under suitable observability condition and regularity of the signal in [125].

The results of [125, 126] are part of the doctoral work corresponding to this thesis. We describe these results in detail in the later chapters.

To summarize, we have seen that stability can be achieved either by having strong mixing in the signal or having very rich observations. If we relax one, then we have to strengthen the other to have stability. For more details on the techniques in stability, we refer the reader to [45, 43]. In the above works, Lyapunov techniques in Wonham filter deal with the Kushner-Stratanovich (KS) equation to study stability. The reason is that, in this case, this equation becomes a non-linear stochastic differential equation, rather than a non-linear stochastic partial differential equation. In the case of continuous signal space, KS equation is seldom used to study the stability. Kallianpur-Striebel formula is more suitable in this regard. On the other hand, KS equation is more suitable for computing conditional expectations due to its incremental structure, whereas Kallianpur-Striebel formula requires the entire history of observations to compute the same.

Recalling the reference probability method, at first glance, it may seem that one can study the stability problem under the reference probability measure and transfer it back to original measure after we are done. This approach will not work as the question of stability involves infinite times and the aforementioned measures are not equivalent on the corresponding σ- algebra $\mathcal{F}_\infty$ (see Chapter (2)).

We now give a trivial example of filtering models whose observations are simply not good enough to ensure stability without strong signal mixing. Non-trivial examples can be constructed, but the example illustrates the basic idea of rich observations.

$$\begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix} = \begin{pmatrix} f_1(x_1)dt + \sigma_1(x_1)dB_t^1 \\ f_2(x_2)dt + \sigma_2(x_2)dB_t^2 \end{pmatrix}$$
$$Y_t = \int_0^t x_1(s)ds + W_t.$$

The signal in the above model is badly observed, as the observations do not give any information about the x_2 component. This model is also not observable in the sense of [139, 73]. Indeed, law of observation does not change if we change the initial distribution for the x_2 component. The only way to achieve stability in this extreme case is when the dynamics of x_2 component is asymptotically independent of initial conditions (in other words, ergodic).

Chapter 4

Stability of the Linear Filter

In this chapter, we study the linear filter for stochastic signal and deterministic signal in detail. We derive the Kalman-Bucy equations for Gaussian initial conditions and also derive the evolution equations for non-Gaussian initial conditions. We then state the classical results of stability in the case of stochastic signal and derive in detail, the stability results for the case of deterministic signal. We conclude this chapter by studying the small noise asymptotics for the linear filter in the case of Gaussian initial conditions. Theorems (4.1.13), (4.1.15) and (4.2.3) and Section (5.4) are part of the doctoral work corresponding to this thesis.

Linear filtering model was the first model whose filtering distribution is known to have a simple evolution equation. The reason is as follows: Considering a Gaussian initial condition and linearity in the model (signal + observation), Gaussianity of the filtering distribution is inherited from initial conditions and sustained by the linearity of the model. Therefore, we only need to find the evolution equations for conditional mean and conditional covariance. And the same thing with the stability: we only have to consider the stability of equations for conditional mean and conditional covariance (considering that incorrect initial condition is also Gaussian). Even in the case of non-Gaussian initial conditions, we can get simple evolution equations because one can make appropriate change of measures that turn this problem to a linear filtering problem with Gaussian initial conditions[104]. In the sections to follow, we make precise, the arguments that are given in this paragraph.

Before we begin, let us introduce the filtering model that we work with in this entire chapter. For the sake of completeness, we define the probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ below (we always work with this probability space). For convenience, we define the probability spaces slightly differently for stochastic signal and deterministic signal case. We also note that none of the calculations change due to this difference.

For the case of stochastic signal,

$$\Omega \doteq \mathbb{R}^m \times C([0,\infty), \mathbb{R}^m) \times C([0,\infty), \mathbb{R}^n)$$

$$\mathcal{F} \doteq \mathcal{B}(\Omega)$$
$$\mathbb{P} \doteq \mu \times \mathbb{P}_W^m \times \mathbb{P}_W^n,$$

where, $\mathbb{P}_W^q$ is the Wiener measure on $C([0,\infty),\mathbb{R}^q)$ and μ is a probability measure on $\mathbb{R}^m$. Consider the canonical process on Ω denoted by ω,

$$\omega_{(\cdot)} \doteq (X_0, B_{(\cdot)}, W_{(\cdot)}),$$
$$\mathcal{H}_t \doteq \sigma(\omega_s : \omega \in \Omega, 0 \leq s \leq t)$$
$$\mathcal{F}_t \doteq \mathcal{H}_t \vee \mathcal{N},$$

where, $\mathcal{N}$ is set of all $\mathbb{P}$- null sets. The signal and observation processes are given by

$$dX_t = A_t X_t dt + N_t dB_t,\ \mathcal{L}(X_0) = \mu \tag{4.1}$$
$$dY_t = C_t X_t dt + dW_t,\ Y_0 = 0. \tag{4.2}$$

where, $t \geq 0$, $A_t \in \mathbb{R}^{m\times m}$, $C_t \in \mathbb{R}^{n\times m}$ and $N_t \in \mathbb{R}^{m\times m}$ are uniformly bounded in t. The construction above implies that X_0, B and W are all mutually independent.

For the case of deterministic signal,

$$\Omega \doteq \mathbb{R}^m \times C([0,\infty),\mathbb{R}^n)$$
$$\mathcal{F} \doteq \mathcal{B}(\Omega)$$
$$\mathbb{P} \doteq \mu \times \mathbb{P}_W^n,$$

Consider the canonical process on Ω denoted by ω,

$$\omega_{(\cdot)} \doteq (X_0, W_{(\cdot)}),$$
$$\mathcal{H}_t \doteq \sigma(\omega_s : \omega \in \Omega, 0 \leq s \leq t)$$
$$\mathcal{F}_t \doteq \mathcal{H}_t \vee \mathcal{N},$$

where, $\mathcal{N}$ is set of all $\mathbb{P}$- null sets. The signal and observation processes are given by

$$dX_t = A_t X_t dt,\ \mathcal{L}(X_0) = \mu \tag{4.3}$$
$$dY_t = C_t X_t dt + dW_t,\ Y_0 = 0. \tag{4.4}$$

where, $t \geq 0$, $A_t \in \mathbb{R}^{m\times m}$, $C_t \in \mathbb{R}^{n\times m}$ are uniformly bounded in t. The construction above implies that X_0 and W are mutually independent.

Note that the constructions above differ slightly from the construction given in Chapter (2). In both the cases, $\mathcal{Y}_t$ has the same meaning as it did in earlier chapters. We have now setup the probability space to work with.

4.1 Kalman-Bucy filter

In this section, let us assume that $\mu = \mathcal{N}(M, P)$[1]. The filter with linear filtering model and Gaussian initial conditions is referred to as Kalman-Bucy filter. We are now in a position to derive the properties of π_t^μ. To that end, we have the following result [148, Theorem 9.4].

Theorem 4.1.1. *For any $t \geq 0$, π_t^μ is a Gaussian distribution on $\mathbb{R}^m$.*

Proof. We follow the proof of [148]. For a fixed, consider an increasing sequence of the form $S_N \doteq \{0 = t_1^N < t_2^N < \ldots < t_{a_N}^N = t\}$ such that their union is dense in $[0, t]$. Note that (X, Y) is a Gaussian process. Indeed, consider the augmented process $V \doteq (X^T, Y^T)^T$. Then equation for V is given by

$$dV_t = \begin{pmatrix} A_t & 0 \\ C_t & 0 \end{pmatrix} V_t dt + \begin{pmatrix} N_t & 0 \\ 0 & \mathbb{I}_n \end{pmatrix} d\hat{B}_t,\ V_0 = \begin{pmatrix} X_0 \\ 0 \end{pmatrix},$$

where, $\hat{B}^T \doteq (B^T, W^T)^T$ and $\mathbb{I}_q$ is $q \times q$ identity matrix. To keep the expression short, let us define

$$\hat{A}_t \doteq \begin{pmatrix} A_t & 0 \\ C_t & 0 \end{pmatrix},\ \hat{N}_t \doteq \begin{pmatrix} N_t & 0 \\ 0 & \mathbb{I}_n \end{pmatrix}$$

Applying Ito's formula to $\Gamma_t^{-1} V_t$ with $\dot{\Gamma}_t = \hat{A}_t \Gamma_t,\ \Gamma_0 = \mathbb{I}_n$, we have

$$\begin{aligned}
\Gamma_t^{-1} V_t &= V_0 - \int_0^t \Gamma_s^{-1} \dot{\Gamma}_s \Gamma_s^{-1} V_s ds + \int_0^t \Gamma_s^{-1} A_s V_s ds + \int_0^t \Gamma_s^{-1} \hat{N}_s d\hat{B}_s \\
\Gamma_t^{-1} V_t &= V_0 - \int_0^t \Gamma_s^{-1} A_s V_s ds + \int_0^t \Gamma_s^{-1} A_s V_s ds + \int_0^t \Gamma_s^{-1} \hat{N}_s d\hat{B}_s \\
V_t &= \Gamma_t V_0 + \Gamma_t \int_0^t \Gamma_s^{-1} \hat{N}_s d\hat{B}_s
\end{aligned}$$

Since N_t and Γ_t are non-random, $\Gamma_t \int_0^t \Gamma_s^{-1} \hat{N}_s d\hat{B}_s$ is a Gaussian process. From Gaussianity of V_0, we conclude that V is also a Gaussian process.

Since (X, Y) is a Gaussian process, the conditional distribution $\pi_t^N \doteq \mathbb{P}(X_t \in \cdot | \sigma(Y_{t_i^N} : 1 \leq i \leq a_N))$ is also Gaussian [79, Section 7.3](with say, mean m_t^N and

[1] $\mathcal{N}(x, y)$ denotes m-dimensional normal distribution with mean $x \in \mathbb{R}^m$ and covariance $y \in \mathbb{R}^{m \times m}$.

covariance Q_t^N). Note that $\lim_{N\to\infty}\pi_t^N = \pi_t$, $\mathbb{P}-$ a.s. Now, consider

$$\phi_N(\lambda) \doteq \int_{\mathbb{R}^m} \exp(ix^T\lambda)\pi_t^N(dx) = \mathbb{E}[\exp(i\lambda^T X_t)|\sigma(Y_{t_i^N} : 1 \le i \le a_N)],$$

for every $\lambda \in \mathbb{R}^m$. From the properties of Gaussian, we have

$$\phi_N(\lambda) = \exp(i\lambda^T m_t^N - \frac{1}{2}\lambda^T Q_t^N \lambda)$$

Since $\{S_N\}_{N\ge 1}$ is chosen to be increasing with N, $\{\phi_N(\lambda)\}_{N\ge 1}$ is a uniformly integrable $\mathcal{Y}_N$- martingale. Indeed,

$$\begin{aligned}
&\mathbb{E}[\phi_N(\lambda)|\sigma(Y_{t_i^N} : 1 \le i \le a_{N-1})] \\
&\qquad = \mathbb{E}[\mathbb{E}[\exp(i\lambda^T X_t)|\sigma(Y_{t_i^N} : 1 \le i \le a_N)]|\sigma(Y_{t_i^N} : 1 \le i \le a_{N-1})] \\
&\qquad = \mathbb{E}[\exp(i\lambda^T X_t)|\sigma(Y_{t_i^N} : 1 \le i \le a_{N-1})] = \phi_{N-1}(\lambda),\ \mathbb{P}-\ a.s.
\end{aligned}$$

In the above, we used the properties of conditional expectation. Therefore, from martingale convergence theorem (see Theorem (A.0.12)), $\lim_{N\to\infty}\phi_N(\lambda) = \phi_\infty(\lambda)$, $\mathbb{P}$-a.s. We now show that $\phi_\infty(\lambda)$ is the characteristic function of a Gaussian distribution. To that end, we known that

$$\lim_{N\to\infty} im_t^N - \frac{1}{2}Q_t^N\lambda \text{ exists.}$$

From the arbitrariness of λ, $\lim_{N\to\infty} m_t^N = m_t$ and $\lim_{N\to\infty} Q_t^N = Q_t$. This concludes that π_t is Gaussian distributed [140, Theorem 2.3]. ■

Once we have the above result, we can now only focus on deriving evolution equations of conditional mean and conditional covariance. Define, $\hat{X}_t \doteq \mathbb{E}[X_t|\mathcal{Y}_t]$ and $P_t \doteq \mathbb{E}[(X_t - \hat{X}_t)(X_t - \hat{X}_t)^T|\mathcal{Y}_t]$. We follow the proof of [15, Proposition 6.14] to derive the result below.

Theorem 4.1.2. *If $\mu = \mathcal{N}(M, P)$, then $\hat{X}_t$ and P_t satisfy the following evolution equations:*

$$d\hat{X}_t = A\hat{X}_t dt + P_t C_t^T(dY_t - C_t\hat{X}_t dt),\ \hat{X}_0 = M \tag{4.5}$$

$$\dot{P}_t = A_t P_t + P_t A_t^T - P_t C_t^T C_t P + N_t N_t^T,\ P_0 = P \tag{4.6}$$

Remark 4.1.3. *P_t is a deterministic quantity. And also, the above equations are one way coupled* i.e., *P_t influences $\hat{X}_t$, but not the other way around. This is practically very useful as we can compute the conditional covariance even before any observations are made.*

Proof. To derive the evolution equations for $\hat{X}_t$ and P_t, we just have to write down and simplify the Kushner-Stratanovich (KS) equation (2.1.11) for $g(x) = x^i$ and for

$g(x) = x^i x^j$, where $x = (x^1, x^2, \dots, x^m)$ and $1 \leq i, j \leq m$. But this is not straightforward since x^i and $x^i x^j$ are unbounded and Kushner-Stratanovich equation is derived for $g \in C_b^2(\mathbb{R}^m)$. We therefore, work with the cut-off versions of these functions and then take the limit appropriately in the end. To that end, define the following cut-off function with smooth compactly supported $\Lambda : \mathbb{R}^m \to \mathbb{R}$,

$$\Lambda_k(x) \doteq \Lambda(\frac{x}{k}),\ \forall x \in \mathbb{R}^m$$

with

$$\Lambda(x) \doteq \begin{cases} 1 & \textit{if } \|x\| \leq 1 \\ \exp\left(\frac{\|x\|^2-1}{\|x\|^2-4}\right) & \textit{if } 1 < \|x\| < 2 \\ 0 & \textit{if } \|x\| \geq 2 \end{cases}$$

Before we proceed any further, we note a few important properties of the above defined cut-off function.

1. All the partial derivatives of Λ_k converge to 0 uniformly.
2. For any $g : \mathbb{R}^m \to \mathbb{R}$,

$$\lim_{k\to\infty} g\Lambda_k = g \quad \text{uniformly},\ |g(x)\Lambda_k(x)| \leq |g(x)| \tag{4.7}$$

We now apply KS equation (2.1.11) to $x^i\Lambda_k(x)$, for fixed i (since it is smooth and bounded). In this case, note that

$$\mathcal{A} = \frac{1}{2}\sum_{p,q=1}^{m}(N_t N_t^T)_{pq}\frac{\partial^2}{\partial x^p \partial x^q} + \sum_{p=1}^{m}(A_t x)_p \frac{\partial}{\partial x^p}$$

$$\begin{aligned} \pi_t(x^i\Lambda_k(x)) &= \mu(x^i\Lambda_k(x)) + \int_0^t \pi_s(\mathcal{A}_s(x^i\Lambda_k(x)))ds \\ &+ \int_0^t \left(\pi_s(x^i\Lambda_k(x)(C_s x)^T) - \pi_s(x^i\Lambda_k(x))\pi_s((C_s x)^T)\right) dI_s, \end{aligned} \tag{4.8}$$

where, I_s is the innovations process (see Chapter (2)). We evaluate individual terms separately and take the limit $k \to \infty$. To that end, consider

$$\mathcal{A}_t(x^i\Lambda_k(x)) = \sum_{p,q=1}^{m}(N_t N_t^T)_{pq}\frac{\partial^2}{\partial x^p \partial x^q}x^i\Lambda_k(x) + \sum_{p=1}^{m}(A_t x)_p \frac{\partial}{\partial x^p}x^i\Lambda_k(x)$$

$$= x^i \sum_{p,q=1}^{m} (N_t N_t^T)_{pq} \frac{\partial^2}{\partial x^p \partial x^q} \Lambda_k(x) + \sum_{q=1}^{m} (N_t N_t^T)_{iq} \frac{\partial}{\partial x^q} \Lambda_k(x)$$
$$+ \sum_{p=1}^{m} (N_t N_t^T)_{pi} \frac{\partial}{\partial x^p} \Lambda_k(x) + \sum_{p=1}^{m} x^i (A_t x)_p \frac{\partial}{\partial x^p} \Lambda_k(x)$$
$$+ (A_t x)_i \Lambda_k(x)$$

Taking limit $k \to \infty$ in all the terms other than the stochastic integral term, we have

$$\lim_{k\to\infty} \pi_t(x^i \Lambda_k(x)) = \pi_t(x^i) \tag{4.9}$$
$$\lim_{k\to\infty} \mu(x^i \Lambda_k(x)) = \mu(x^i) = M^i \tag{4.10}$$
$$\lim_{k\to\infty} \int_0^t \pi_s(\mathcal{A}_s(x^i \Lambda_k(x)))ds = \int_0^t \pi_s(\mathcal{A} x^i)ds = \int_0^t \pi_s((A_s x)_i)ds \tag{4.11}$$

where, in (4.9) and (4.10), we used dominated convergence theorem combined with property (4.7). In (4.11), we used the Fubini's theorem and dominated convergence theorem along with uniform (in k) boundedness of partial derivatives of Λ_k. Note that $\pi_t(x^i)$ and $\pi_t(x^i x^j)$ are both finite for any $1 \le i, j \le m$ due to Gaussianity of π_t. This allows us to apply dominated convergence theorem without any trouble.

Finally, consider the stochastic integral term

$$\int_0^t \left(\pi_s \left(x^i \Lambda_k(x) \, (C_s x)^T \right) - \pi_s \left(x^i \Lambda_k(x) \right) \pi_s \left((C_s x)^T \right) \right) dI_s$$

From the uniform convergence of $\Lambda_k \to \Lambda$, we have the following

$$\lim_{k\to\infty} \mathbb{E}[\int_0^t \left(\pi_s \left(x^i \left(\Lambda_k(x) - 1 \right) (C_s x)^T \right) - \pi_s \left(x^i \left(\Lambda_k(x) - 1 \right) \right) \pi_s \left((C_s x)^T \right) \right)^2 ds] = 0$$

This implies that, at least for a subsequence $\{k_n\}_{n\ge1}$, we have

$$\lim_{k\to\infty} \int_0^t \pi_s \left(x^i \Lambda_k(x) \, (C_s x)^T \right) - \pi_s \left(x^i \Lambda_k(x) \right) \pi_s \left((C_s x)^T \right) dI_s$$
$$= \int_0^t \left(\pi_s \left(x^i \, (C_s x)^T \right) - \pi_s \left(x^i \right) \pi_s \left((C_s x)^T \right) \right) dI_s, \ \mathbb{P} - a.s.$$

Now combining the above equations, we have

$$\pi_t(x^i) = \mu(x^i) + \int_0^t \pi_s \left((A_s x)_i \right) ds + \int_0^t \left(\pi_s(x^i (C_s x)^T) - \pi_s(x^i) \pi_s((C_s x)^T) \right) dI_s$$
$$= \mu(x^i) + \int_0^t \pi_s \left((A_s x)_i \right) ds + \sum_{p=1}^{m} \int_0^t \left(\pi_s(x^i (C_s x)^p) - \pi_s(x^i) \pi_s((C_s x)^p) \right) dI_s^p$$

Simplifying the above equation, we have (4.5) in terms of components.

Now to derive the equation for $\pi_t(x^ix^j)$, for a fixed $1 \le i, j \le m$, we follow the same procedure of using a cutoff function and taking the limit. It works again because of properties of Λ_k and the fact that π_t has finite third moment, due to Gaussianity. To avoid the repetition of similar calculations, we directly write down the equation after this limiting procedure.

$$\begin{aligned}
&\pi_t(x^ix^j) \\
&= \mu(x^ix^j) + \int_0^t \pi_s\left(\mathcal{A}(x^ix^j)\right) ds + \int_0^t \left(\pi_s(x^ix^j(C_sx)^T) - \pi_s(x^ix^j)\pi_s((C_sx)^T)\right) dI_s \quad (4.12) \\
&= \mu(x^ix^j) + \int_0^t \left((N_sN_s^T)_{ij} + \sum_{p=1} \pi_s(A_s^{ip}x^px^j + A_s^{jp}x^px^i)\right) ds \\
&+ \int_0^t \sum_{p,q=1}^m \left(\pi_s(x^ix^jx^q) - \pi_s(x^ix^j)\pi_s(x^q)\right) C_s^{pq} dI_s^p
\end{aligned}$$

To simplify any further, we use the following fact related to Gaussianity of π_t:

$$\begin{aligned}
\pi_t(x^ix^jx^p) = &-2\pi_t(x^i)\pi_t(x^j)\pi_t(x^p) + \pi_t(x^i)\pi_t(x^jx^p) + \pi_t(x^j)\pi_t(x^ix^p) \\
&+ \pi_t(x^p)\pi_t(x^jx^i) \qquad (4.13)
\end{aligned}$$

(4.12) becomes

$$\begin{aligned}
&\pi_t(x^ix^j) \\
&= \mu(x^ix^j) + \int_0^t \left((N_sN_s^T)_{ij} + \sum_{p=1} \pi_s\left(A_s^{ip}x^px^j + A_s^{jp}x^px^i\right)\right) ds \\
&+ \int_0^t \sum_{p,q=1}^m \left(-2\pi_s(x^i)\pi_s(x^j)\pi_s(x^p) + \pi_s(x^i)\pi_s(x^jx^p) + \pi_s(x^j)\pi_s(x^ix^p)\right) C_s^{pq} dI_s^p
\end{aligned}$$

$$\begin{aligned}
\pi_t(x^ix^j) = \mu(x^ix^j) &+ \int_0^t \left((N_sN_s^T)_{ij} + \sum_{p=1} \pi_s\left(A_s^{ip}x^px^j + A_s^{jp}x^px^i\right)\right) ds \\
&+ \int_0^t \sum_{p,q=1}^m \left(\pi_s(x^i)P_s^{jp} + \pi_s(x^j)P_s^{ip}\right) C_s^{pq} dI_s^p \qquad (4.14)
\end{aligned}$$

Recall that $dP_t^{ij} = d\pi_t(x^ix^j) - d\left(\pi_t(x^i)\pi_t(x^j)\right)$ and

$$d\left(\pi_t(x^i)\pi_t(x^j)\right) = \pi_t(x^i)d\left(\pi_t(x^j)\right) + d\left(\pi_t(x^i)\right)\pi_t(x^j) + \langle x^i, x^j\rangle_t$$

$$= \sum_{p=1}^{m} \left(\pi_t(x^i) A_t^{jp} \pi_t(x^p) dt + \pi_t(x^i)(P_t C_t^T)^{jp} dI_t^p + \pi_t(x^j) A_t^{ip} \pi_t(x^p) dt \right)$$
$$+ \sum_{p=1}^{m} \left(\pi_t(x^j)(P_t C_t^T)^{ip} dI_t^p \right) + (P_t C_t^T C_t P_t)^{ij} dt \tag{4.15}$$

We then have, from (4.15), (4.14),

$$\begin{aligned}
&dP_t^{ij} \\
&= ((N_t N_t^T)_{ij} + \sum_{p=1} \pi_t (A_t^{ip} x^p x^j + A_t^{jp} x^p x^i)) dt + \sum_{p,q=1}^{m} (\pi_t(x^i) P_t^{jp} + \pi_t(x^j) P_t^{ip}) C_t^{pq} dI_t^p \\
&- \sum_{p=1}^{m} (\pi_t(x^i) A_t^{jp} \pi_t(x^p) dt + \pi_t(x^i)(P_s C_s^T)^{jp} dI_t^p + \pi_t(x^j) A_t^{ip} \pi_t(x^p) dt) \\
&+ \sum_{p=1}^{m} (\pi_t(x^j)(P_t C_t^T)^{ip} dI_t^p) - (P_t C_t^T C_t P_t)^{ij} dt \\
&= ((N_t N_t^T)_{ij} + \sum_{p=1} (A_t^{ip} P_t^{pj} + A_t^{jp} P_t^{pi})) dt + \sum_{p,q=1}^{m} (\pi_t(x^i) P_t^{jp} + \pi_t(x^j) P_t^{ip}) C_t^{pq} dI_t^p \\
&- \sum_{p=1}^{m} (\pi_t(x^i)(P_t C_t^T)^{jp} dI_t^p + \pi_t(x^j)(P_t C_t^T)^{ip} dI_t^p) - (P_t C_t^T C_t P_t)^{ij} dt \\
&= ((N_t N_t^T)_{ij} + \sum_{p=1} (A_t^{ip} P_t^{pj} + A_t^{jp} P_t^{pi})) dt - (P_t C_t^T C_t P_t)^{ij} dt
\end{aligned}$$

This finishes the proof. ■

To summarize, we have shown that if $\mu = \mathcal{N}(M, P)$, then $\pi_t = \mathcal{N}(\hat{X}_t^{M,P}, P_t^P)$, where

$$\begin{aligned}
d\hat{X}_t^{M,P} &= A_t \hat{X}_t^{M,P} + P_t^P C_t^T (dY_s - C_t \hat{X}_t^{M,P} dt), \ \hat{X}_0^{M,P} = M \\
\dot{P}_t^P &= A_t P_t^P + P_t^P A_t^T - P_t^P C_t^T C_t P_t^P + N_t N_t^T, \ P_0^P = P
\end{aligned}$$

Note that we changed the notation to emphasize the dependence on initial condition. We now study the stability of the Kalman-Bucy filter.

4.1.1 Stability of the Kalman-Bucy filter

To study the stability of the Kalman-Bucy filter, we only consider incorrect initial condition ν which is Gaussian.

$$\nu = \mathcal{N}(\bar{M}, \bar{P})$$

We give the stability results for both the cases *viz.*, stochastic signal and deterministic signal. But, we only prove the results for the deterministic case. In the case of filtering model with stochastic signal, asymptotic behavior of (4.5) and (4.6) can be summarized by the following results.

Theorem 4.1.4. *[33, Theorem 4] Under the assumptions of uniform complete observability and uniform complete controllability for the case of stochastic signal, we have*

$$\|P_t^Q - P_t^{\bar{Q}}\| \leq K \exp\left(-\beta t\right),$$

for some $K, \beta > 0$ and $Q, \bar{Q}$ are any symmetric positive semidefinite matrices. K can depend on Q and $\bar{Q}$.

Remark 4.1.5. *In the case of autonomous (time-independent) filtering model, more can be said under weaker versions of observability and controllability* viz., *there exists a symmetric positive definite matrix $\hat{P}$ such that*

$$A\hat{P} + \hat{P}A^T - \hat{P}C^T C\hat{P} + NN^T = 0$$

Note that for autonomous case, $A_t = A,\ C_t = C,\ N_t = N$. From above theorem, consequently, we have $\|P_t^Q - P_t^{\bar{Q}}\| \leq K \exp\left(-\beta t\right)$. See [95, Chapter 4] for more details.

Theorem 4.1.6. *[116, Theorem 2.3] Under the assumptions of uniform complete observability and uniform complete controllability for the case of autonomous filtering model with stochastic signal, there exists $\alpha > 0$ such that*

$$\lim_{t\to\infty} \exp\left(\alpha t\right) \|\hat{X}_t^{M,P} - \hat{X}_t^{\bar{M},\bar{P}}\| = 0,\ \mathbb{P} - \ a.s.$$

Remark 4.1.7. *Even though authors in [116] prove the above result for autonomous case under weaker assumptions, the extension of the same result to non-autonomous (time-dependent) case is straightforward (under the above assumptions).*

We now state and prove the stability of Kalman-Bucy filter for a deterministic signal ($N_t = 0$). In this case, (4.5) and (4.6) become

$$dX_t^{M,P} = A_t X_t^{M,P} dt + P_t^P C_t^T (dY_t - C_t X_t^{M,P} dt), \quad X_0^{M,P} = M, \tag{4.16}$$

$$\dot{P}_t^P = A_t P_t^P + P_t^P A_t^T - P_t^P C_t^T C_t P_t^P, \quad P_0^P = P, \tag{4.17}$$

where $P > 0$[2]. For the rest of this chapter, P_t^P is always considered to be the solution of (4.17) with $P_0^P = P$.

[2] For real symmetric positive semi-definite matrices X and Y of same dimension, we write $X \geq Y$ whenever $\mathbf{x}^T(X - Y)\mathbf{x} \geq 0, \forall\ \mathbf{x} \neq \mathbf{0} \in \mathbb{R}^m$. Notations like $X \leq Y$, $X < Y$ and $X > Y$ are adopted accordingly throughout the chapter.

We shall make the following assumptions for the rest of the chapter.

Assumption 4.1.8. *A_t and C_t are all continuous and uniformly bounded in t and P is invertible.*

Assumption 4.1.9. *The pair $[A_t, C_t]$ is uniformly completely observable (see Definition (A.0.25)).*

Throughout this chapter, we define the norm $\|\cdot\|$ of a $m \times n$ matrix Q as $\|Q\| \doteq \sup_{\|x\|=1} \|Qx\|$, where $\|\cdot\|$ on the right is a norm on the appropriate $\mathbb{R}^k$ for $k = m, n$. To study the behavior of (4.17), we note that, for a symmetric positive semi-definite initial condition $P \geq 0$, solution to (4.17) can be written in an explicit form.

$$P_t^P = \Phi_t \sqrt{P} \left(I + \sqrt{P} \bar{C}_t \sqrt{P} \right)^{-1} \sqrt{P} \Phi_t^T, \tag{4.18}$$

as can be verified explicitly by substituting (4.18) in (4.17), where,

$$\bar{C}_t \doteq \int_0^t \Phi_s^T C_s^T C_s \Phi_s ds.$$

We also note the following result (proved in [112]) that is used later.

Lemma 4.1.10. *Under Assumptions (4.1.8) and (4.1.9), P_t^P is uniformly bounded in t.*

Remark 4.1.11. *Consider the subspace of $\mathbb{R}^m$ defined by $S \doteq \{u : \|\Phi_t^T u\| \to 0 \;\; as \;\; t \to 0\}$. For $v \in S$, it is clear from (4.18) that $v^T P_t^P v \to 0$ as $t \to \infty$ (since $\bar{C}_t$ is positive semi-definite, implying that the uncertainty along S reduces to zero asymptotically in time. This feature is used in data assimilation algorithms in discrete time known as Assimilation in Unstable Subspace (AUS) [38, 119, 137]. This and other properties of the filter covariance (in discrete time) and their relation to Lyapunov vectors and exponents of the dynamics that have been discussed extensively in [68, 28] extend to the filter covariance for the Kalman-Bucy filter (in continuous time).*

To prove the stability of (4.17), we use the well known Lyapunov function approach first used in [33], but also used extensively by later authors. In order to set up our notation, we consider solutions P_t^P and $P_t^{\bar{P}}$ of (4.17) corresponding to two different initial conditions P and $\bar{P}$ ($\bar{P} > 0$), respectively. A straightforward calculation shows that $E_t \doteq P_t^P - P_t^{\bar{P}}$ satisfies

$$\dot{E}_t = B_t^P E_t + E_t \left(B_t^{\bar{P}} \right)^T,$$

where $B_t^Q \doteq \left(A_t - P_t^Q C_t^T C_t \right)$, for any symmetric positive semi-definite matrix Q. Further, it can easily be verified that

$$E_t = \Psi_t^P (P - \bar{P}) (\Psi_t^{\bar{P}})^T, \tag{4.19}$$

with $\dot{\Psi}^P_t = B^P_t \Psi^P_t$, $\Psi^P_0 = \mathbb{I}_m$, $\dot{\Psi}^{\bar{P}}_t = B^{\bar{P}}_t \Psi^{\bar{P}}_t$ and $\Psi^{\bar{P}}_0 = \mathbb{I}_m$. Therefore, stability of the Riccati equation is related to studying the asymptotic properties of Ψ^P_t and $\Psi^{\bar{P}}_t$. Without loss of generality, it is sufficient to study asymptotic properties of Ψ^P_t. To this end, consider a linear system

$$\dot{z}_t = \left(A_t - P^P_t C^T_t C_t\right) z_t \,, \tag{4.20}$$

whose solution is given by $z_t = \Psi^P_t z_0$, where z_0 is the initial condition. The above system (4.20) is said to be asymptotically stable if $\|\Psi^P_t\| \xrightarrow{t\to\infty} 0$ which is equivalent to $\|z_t\| \xrightarrow{t\to\infty} 0$, $\forall\ z_0 \in \mathbb{R}^m$. The following lemma is crucial in proving the asymptotic stability of (4.20).

Lemma 4.1.12. *[130, Lemma 2.5.2] Under Assumptions (4.1.8) and (4.1.9), $[A_t - K_t C_t, C_t]$ is uniformly completely observable, for any K_t that is continuous and bounded in t.*

Proof. Define $v_t,\ w_t,\ x_t,\ y_t$ in the following way:

$$\begin{aligned} \dot{x}_t &= A_t x_t \\ y_t &= C_t x_t \end{aligned}$$

and

$$\begin{aligned} \dot{w}_t &= (A_t - K_t C_t) w_t \\ v_t &= C_t w_t \end{aligned}$$

Assume $x_{t_0} = w_{t_0}$ and consider

$$\begin{aligned} x_t - w_t &= -\int_{t_0}^{t} \Phi(t,s) K_s C_s w_s ds \\ C_t \left(x_t - w_t\right) &= -\int_{t_0}^{t} C_t \Phi(t,s) K_s C_s w_s ds \end{aligned}$$

Define $\hat{K}_t = \frac{K_s C_s w_s}{\|K_s C_s w_s\|}$, if $K_s C_s w_s \neq 0$ and zero otherwise.

$$\begin{aligned} \|C_t \left(x_t - w_t\right)\|^2 &\leq \left(\int_{t_0}^{t} \|C_t \Phi(t,s) \hat{K}_s\| \|K_s C_s w_s\| ds\right)^2 \\ &\leq \left(\int_{t_0}^{t} \|C_t \Phi(t,s) \hat{K}_s\| \|K_s\| \|C_s w_s\| ds\right)^2 \\ &\leq \sup_{s\geq 0} \|K_s\|^2 \left(\int_{t_0}^{t} \|C_t \Phi(t,s) \hat{K}_s\|^2 ds\right) \left(\int_{t_0}^{t} \|C_u w_u\|^2 du\right) \end{aligned}$$

Fix $t = t_0 + \tau$, where corresponds to the pair $[A_t, C_t]$. From triangle inequality, we have

$$\begin{aligned}
&\left(\int_{t_0}^{t_0+\tau} \|C_s w_s\|^2 ds\right)^{\frac{1}{2}} \\
&\geq \left(\int_{t_0}^{t_0+\tau} \|C_s x_s\|^2 ds\right)^{\frac{1}{2}} - \left(\int_{t_0}^{t_0+\tau} \|C_s\left(x_s - w_s\right)\|^2 ds\right)^{\frac{1}{2}} \\
&\geq \sqrt{\rho_1}\|w_{t_0}\| - \left(\int_{t_0}^{t_0+\tau} \|C_s\left(x_s - w_s\right)\|^2 ds\right)^{\frac{1}{2}}, \textit{ From the Assumption } (4.1.9). \\
&\geq \sqrt{\rho_1}\|w_{t_0}\| - \sup_{s\geq 0}\|K_s\| \left(\int_{t_0}^{t_0+\tau} \|C_u w_u\|^2 du\right)^{\frac{1}{2}} \left(\int_{t_0}^{t_0+\tau} \left(\int_{t_0}^{r} \|C_r \Phi(r,s)\hat{K}_s\| ds\right)^2 dr\right)^{\frac{1}{2}} \\
&\geq \sqrt{\rho_1}\|w_{t_0}\| - \sup_{s\geq 0}\|K_s\| \left(\int_{t_0}^{t_0+\tau} \|C_u w_u\|^2 du\right)^{\frac{1}{2}} \sqrt{\tau\rho_2}
\end{aligned}$$

Therefore, we have

$$\int_{t_0}^{t_0+\tau} \|C_s w_s\|^2 ds \geq \frac{\rho_1}{\left(1 + \sup_{s\geq 0}\|K_s\|\sqrt{\tau\rho_2}\right)^2}\|w_{t_0}\|^2 \tag{4.21}$$

To prove the upper bound, we consider

$$\begin{aligned}
\|C_t w_t\|^2 &\leq \|C_t x_t\|^2 + \left(\int_{t_0}^{t} \|C_s w_s\|\|C_t \Phi(t,s)\hat{K}_s\|\|K_s\| ds\right)^2 \\
&\leq \|C_t x_t\|^2 + \int_{t_0}^{t} \|C_s w_s\|^2 ds \int_{t_0}^{t} \|C_r \Phi(t,r)\hat{K}_r\|^2 \|K_r\|^2 dr
\end{aligned}$$

Integrating the above inequality from t_0 to $t_0 + \tau$, we have

$$\begin{aligned}
\int_{t_0}^{t_0+\tau} \|C_t w_t\|^2 dt &\leq \int_{t_0}^{t_0+\tau} \|C_t x_t\|^2 dt + \int_{t_0}^{t_0+\tau} \int_{t_0}^{t} \|C_s w_s\|^2 ds \int_{t_0}^{t} \|C_r \Phi(t,r)\hat{K}_r\|^2 \|K_r\|^2 dr dt \\
&\leq \int_{t_0}^{t_0+\tau} \|C_t x_t\|^2 dt + \sup_{s\geq 0}\|K_s\|^2 \rho_2 \int_{t_0}^{t_0+\tau} \int_{t_0}^{t} \|C_s w_s\|^2 ds dt
\end{aligned}$$

From Gronwall's inequality, we have

$$\int_{t_0}^{t_0+\tau} \|C_t w_t\|^2 dt \leq \rho_2 \|w_{t_0}\|^2 \exp\left(\tau \sup_{s\geq 0}\|K_s\|^2 \rho_2\right) \tag{4.22}$$

The equations (4.21) and (4.22) together imply that the pair $[A - K_t C_t, C_t]$ is uniformly completely observable.

■

Using the above lemma, we can conclude that, from the continuity and boundedness of $P_t^P C_t^T$ in t, $[B_t^P, C_t]$ is uniformly completely observable, i.e., there exist $\tilde{\tau}, \rho_3, \rho_4 > 0$ such that for all $t > 0$ we have,

$$\rho_3 \mathbb{I}_n \leq \int_{t-\tilde{\tau}}^{t} \left(\Psi_t^P\right)^{-T} \left(\Psi_s^P\right)^T C_s^T C_s \Psi_s^P \left(\Psi_t^P\right)^{-1} ds \leq \rho_4 \mathbb{I}_n \,. \tag{4.23}$$

Theorem 4.1.13. *[126, 112] Let P be symmetric positive definite. Under Assumptions (4.1.8) and (4.1.9), (4.20) is asymptotically stable and*

$$\int_0^\infty \left(\Psi_s^P\right)^T \Psi_s^P ds < \frac{\tilde{\tau} P^{-1}}{\rho_3} \tag{4.24}$$

Remark 4.1.14. *Asymptotic stability of (4.20) is proved in [112], whereas the estimate (4.24) is established in [126]. It turns out that (4.24) is sufficient to establish the stability of (4.5) (Asymptotic stability of (4.20) by itself cannot imply stability of (4.5)).*

Proof. We begin with a Lyapunov function

$$V(z_t, t) \doteq z_t^T \left(P_t^P\right)^{-1} z_t. \tag{4.25}$$

Using (4.17) and (4.20), we see that

$$\begin{aligned}
\frac{dV}{dt}(z_t, t) &= -z_t^T (A_t - P^P{}_t C_t^T C_t)^T \left(P_t^P\right)^{-1} z_t \\
&\quad + z_t^T \left(-A_t^T \left(P_t^P\right)^{-1} - \left(P_t^P\right)^{-1} A_t + C_t^T C_t\right) z_t \\
&\quad + z_t^T \left(P_t^P\right)^{-1} (A_t - P^P{}_t C_t^T C_t) z_t \\
&= -z_t^T C_t^T C_t z_t \leq 0\,, \qquad \forall t > 0\,.
\end{aligned} \tag{4.26}$$

Using the relationship $z_s = \Psi_s^P \left(\Psi^P\right)_t^{-1} z_t$, we can write

$$V(z_{t+\tilde{\tau}}, t+\tilde{\tau}) - V(z_t, t) = -z_t^T \int_t^{t+\tilde{\tau}} \left(\Psi_t^P\right)^{-T} \left(\Psi_s^P\right)^T C_s^T C_s \Psi_s^P \left(\Psi_t^P\right)^{-1} ds\, z_t \,.$$

Observe that from (4.23),

$$\rho_3 \|z_t\|^2 \leq V(z_t, t) - V(z_{t+\tilde{\tau}}, t+\tilde{\tau}) \leq \rho_4 \|z_t\|^2, \tag{4.27}$$

which together with the assumption of uniform complete observability of $[A_t, C_t]$ imply that $V(z_t, t) \to 0$, and $\|z_t\| \to 0$, as $t \to \infty$, and that (4.20) is asymptotically stable.

Next, in order to prove (4.24), observe that writing $t = t' + k\tilde{\tau}$, for some $t' \in [0, \tilde{\tau}]$, we have

$$V(z_{t'+(k+1)\tilde{\tau}}, t' + (k+1)\tilde{\tau}) - V(z_{t+k\tilde{\tau}}, t' + k\tilde{\tau}) \leq -\rho_3 \|z_{t'+k\tilde{\tau}}\|^2$$

Adding N such inequalities with $k = 0, 1, 2, ..., N$, we have

$$V(z_{t'+(N+1)\tilde{\tau}}, t' + (N+1)\tilde{\tau}) - V(z_{t'}, t') \leq -\rho_3 \sum_{k=0}^{N} \|z_{t'+k\tilde{\tau}}\|^2$$

Using (4.26), and letting $N \to \infty$,

$$\sum_{k=0}^{\infty} \|z_{t'+k\tilde{\tau}}\|^2 \leq \frac{V(z_{t'}, t')}{\rho_3} \leq \frac{V(z_0, 0)}{\rho_3}. \tag{4.28}$$

Integrating (4.28) with respect to t' in the range $t' \in [0, \tilde{\tau}]$, we have

$$\begin{aligned} \int_0^{\tilde{\tau}} \sum_{k=0}^{\infty} \|z_{t'+k\tilde{\tau}}\|^2 dt' &\leq \int_0^{\tilde{\tau}} \frac{V(z_0, 0)}{\rho_3} dt' \\ \int_0^{\infty} \|z_{t'}\|^2 dt' &\leq \frac{V(z_0, 0)\tilde{\tau}}{\rho_3} \\ z_0^T \Big(\int_0^{\infty} (\Psi_{t'}^P)^T \Psi_{t'}^P dt' \Big) z_0 &\leq \frac{V(z_0, 0)\tilde{\tau}}{\rho_3} = \frac{\tilde{\tau} z_0^T P^{-1} z_0}{\rho_3} \end{aligned} \tag{4.29}$$

Since (4.29) is true for all initial conditions z_0,

$$\int_0^{\infty} (\Psi_{t'}^P)^T \Psi_{t'}^P dt' \leq \frac{\tilde{\tau}}{\rho_3} P^{-1} \tag{4.30}$$

which completes the proof. ■

We now show that (4.5) is stable. To that end, we follow the method of [116].

Theorem 4.1.15. *Under Assumptions* (4.1.8) *and* (4.1.9), $\|\hat{X}_t^{M,P} - \hat{X}_t^{\bar{M},\bar{P}}\| \xrightarrow{t \to \infty} 0$ $\mathbb{P} -$ *a.s*

Proof. Let us begin by noting that innovations process $dI_t = dY_t - C_t \hat{X}_t dt$, is a $\mathcal{Y}_t$-Brownian motion (see Proposition (2.1.15)). Then, using (4.5), we see that the evolution equation for $\hat{X}_t^{M,P} - \hat{X}_t^{\bar{M},\bar{P}}$ is

$$d(\hat{X}_t^{M,P} - \hat{X}_t^{\bar{M},\bar{P}}) = (A_t - P_t^{\bar{P}} C_t^T C_t)(\hat{X}_t^{M,P} - \hat{X}_t^{\bar{M},\bar{P}})dt + (P_t^P - P_t^{\bar{P}})C_t^T (dY_t - C_t \hat{X}_t^{M,P} dt). \tag{4.31}$$

Using a simple application of Ito's formula (see Theorem (A.0.17)), we observe that solution to the above evolution equation is given by

$$\hat{X}_t^{M,P} - \hat{X}_t^{\bar{M},\bar{P}} = \Psi_t^{\bar{P}}(M - \bar{M}) + \int_0^t \Psi_t^{\bar{P}} \left(\Psi_s^{\bar{P}}\right)^{-1} \left(P_s^P - P_s^{\bar{P}}\right) C_s^T d\mathbb{I}_s$$

Next, defining $\hat{V}_t \doteq \int_0^t \left(\Psi_s^{\bar{P}}\right)^{-1} \left(P_s - P_s^{\bar{P}}\right) C_s^T dI_s$, we can express the above solution in a compact form as

$$\hat{X}_t^{M,P} - \hat{X}_t^{\bar{M},\bar{P}} = \Psi_t^{\bar{P}}(M - \bar{M}) + \Psi_t^{\bar{P}} \hat{V}_t. \tag{4.32}$$

Observe now that using (4.19) to write $(P_s^P - P_s^{\bar{P}})$in terms of Ψ_s^P and $\Psi_s^{\bar{P}}$, it is clear that,

$$\mathbb{E}[|\hat{V}_t|^2] = \mathbb{E}\left[\mathrm{tr}\left(\int_0^t (P - \bar{P}) \left(\Psi_s^P\right)^T C_s^T C_s \Psi_s^P (P - \bar{P}) ds\right)\right],$$

where $\mathrm{tr}(A)$ denotes the trace of the square matrix A. Using simple algebra, we can easily conclude that for some $K' > 0$, we have $\|P - \bar{P}\|^2 < K'$. In particular, we could choose K' to be the squared sum of the largest eigenvalues of P and $\bar{P}$. Moreover, we also have $\|C_t^T C_t\| < K$, for some $K > 0$, thus implying

$$\begin{aligned} \mathrm{tr}\left(\int_0^t (P - \bar{P}) \left(\Psi_s^P\right)^T C_s^T C_s \Psi_s^P (P - \bar{P}) ds\right) &\le KK' \mathrm{tr}\left(\int_0^t \left(\Psi_s^P\right)^T \Psi_s^P ds\right) \\ &\le KK' \mathrm{tr}\left(\int_0^\infty \left(\Psi_s^P\right)^T \Psi_s^P ds\right) \\ &< \infty, \end{aligned}$$

where the last inequality follows from Theorem (4.1.13), indicating that $\hat{V}_t$ is a square integrable martingale. Therefore, by martingale convergence theorem (see Theorem (A.0.12)), $\{\hat{V}_t\}_{t \ge 0}$ converges almost surely, as $t \to \infty$, to an integrable random variable, say $\hat{V}_\infty$. Thus, we conclude that $\Psi_t^{\bar{P}} \hat{V}_t \to 0$ $\mathbb{P} - a.s$, since we already know by Theorem (4.1.13) that $\Psi_t^{\bar{P}}$ converges to zero as $t \to \infty$. Similarly, we can deduce that $\Psi_t^{\bar{P}}(M - \bar{M}) \to 0$, as $t \to \infty$, which in view of (4.32) completes the proof. ■

4.2 Linear filter with non-Gaussian initial conditions

In the previous section, the assumption of Gaussian initial condition was crucial to most of the computations. However, this condition can be relaxed. Under different assumptions on the non-Gaussianity of the initial conditions, the linear filtering is shown to exhibit finite dimensionality *i.e.*, only finitely many statistics of π_t are needed to compute the entire π_t [21, 104, 74]. In the following, we describe one such result *viz.*, that of [104], mainly because this approach is used later in filter stability (for non-Gaussian initial conditions). In this section, μ is not necessarily Gaussian.

The main idea behind the approach in [104] is to perform a change of measure under which filtering model takes a simpler form (Kalman-Bucy filter with correlated noise case). Firstly, write (4.1) in the explicit form.

$$X_t = \Phi_t X_0 + \int_0^t \Phi(t,s) N_s dB_s$$

where, Φ_t is such that

$$\dot{\Phi}_t = A_t \Phi_t; \quad \Phi_0 = \mathbb{I}_m$$
$$\Phi(t,s) \doteq \Phi_t \Phi_s^{-1}$$

Now define

$$\bar{X}_t \doteq X_t - \xi_t \; with \; \dot{\xi}_t = A_t \xi_t, \; \xi_0 = X_0$$

From the above definition, it can be seen that

$$\bar{X}_t = \int_0^t \Phi(t,s) N_s dB_s \tag{4.33}$$
$$Y_t = \int_0^t C_s \bar{X}_s ds + \bar{W}_t \tag{4.34}$$

with

$$\bar{W}_t \doteq \int_0^t C_s \xi_s ds + W_t$$

(4.33) and (4.34), look very similar (in form) to (4.1) and (4.2). This is the main observation. Under $\mathbb{P}$, $\bar{W}$ is not a Brownian motion. But, from Girsanov's theorem (see Theorem (A.0.22)), we can make $\bar{W}$ a Brownian motion under a new equivalent probability measure $\bar{\mathbb{P}}$. The probability measure $\bar{\mathbb{P}}$ is defined by

$$\frac{d\bar{\mathbb{P}}}{d\mathbb{P}} \doteq \Delta_{T_0}^{-1},$$

for a fixed $T_0 > 0$. Here, for $t \leq T_0$,

$$\Delta_t \doteq \exp\left(- \int_0^t \xi_s^T C_s^T d\bar{W}_s - \frac{1}{2} \int_0^t \|C_s \xi_s\|^2 ds \right)$$

$\bar{\mathbb{P}}$ is a indeed, a true probability measure in our case(see Lemma (A.0.23)). From Lemma (2.1.4), we write, for $G : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^+ \to \mathbb{R}$ such that $G(X_0, \bar{X}_t, t)$ is integrable,

$$\mathbb{E}[G(X_0, \bar{X}_t, t)|\mathcal{Y}_t] = \frac{\mathbb{E}^{\bar{\mathbb{P}}}[G(X_0, \bar{X}_t, t)\Delta_t|\mathcal{Y}_t]}{\mathbb{E}^{\bar{\mathbb{P}}}[\Delta_t|\mathcal{Y}_t]} \tag{4.35}$$

Notice that the variables X_0, $\bar{W}_t$, $\bar{X}_t$ are all mutually independent under $\bar{\mathbb{P}}$, and that the distribution of X_0, B remains unchanged. Therefore, from (4.33) and (4.34), we can conclude that under $\bar{\mathbb{P}}$, X_0 is independent of $\mathcal{Y}_t$. The other random variables that are remaining in $g(X_0, \bar{X}_t)\Delta_t$ are $\bar{X}_t$ and Δ_t. To get the joint conditional distribution (under $\bar{\mathbb{P}}$) of $\bar{X}_t$ and Δ_t given $\mathcal{Y}_t$, we write Δ_t as

$$\Delta_t = \exp\left(-X_0^T L_t - \frac{1}{2} X_0^T K_t X_0\right),$$

where

$$K_t \doteq \int_0^t \Phi_s C_s^T C_s \Phi_s ds, \; L_t \doteq \int_0^t \Phi_s^T C_s^T d\bar{W}_s.$$

Clearly, instead of joint conditional (under $\bar{\mathbb{P}}$) of $\bar{X}_t$ and Δ_t given $\mathcal{Y}_t$, we can consider the conditional distribution (under $\bar{\mathbb{P}}$) of $\bar{X}_t$ and L_t given $\mathcal{Y}_t$. Define $\eta_t(dx_1, dx_2)$ such that,

$$\bar{\mathbb{P}}[\mathbb{1}_{(\bar{X}_t, L_t)\in B}|\mathcal{Y}_t] = \int_B \eta_t(dx_1, dx_2), \; \forall B \in \mathcal{B}(\mathbb{R}^m \times \mathbb{R}^m)$$

With the above definitions, we have

$$\mathbb{E}^{\bar{\mathbb{P}}}[G(X_0, \bar{X}_t, t)\Delta_t|\mathcal{Y}_t] = \int_{\mathbb{R}^m} \mu(dx) \int_{\mathbb{R}^m \times \mathbb{R}^m} G(x, r_1, t) e^{-\frac{1}{2}x^T K_t x + x^T r_2} \eta_t(dr_1, dr_2), \tag{4.36}$$

The conditional distribution η_t is obtained by studying Kalman-Bucy filter in the framework with correlated observation and system noises for the extended system, $\begin{pmatrix} \bar{X}_t \\ L_t \end{pmatrix}$. It is known that the conditional distribution η_t is again Gaussian [148, Section 9.2], with mean $\begin{pmatrix} \tilde{m}_t \\ \tilde{L}_t \end{pmatrix}$ and covariance $\begin{pmatrix} \tilde{R}_t & S_t \\ S_t^T & R_t \end{pmatrix}$ given by the following set of equations.

$$\tilde{m}_t = X_t^{m', P'} \text{ (solution of (4.5))}, \qquad \tilde{m}_0 = M' = 0, \tag{4.37}$$

$$\begin{aligned}
\tilde{R}_t &= P_t^{P'} \text{ (solution of (4.6))}, & \tilde{R}_0 &= P' = 0\,, \quad (4.38)\\
d\tilde{L}_t &= (\Phi_t + S_t)^T C_t^T (dy_t - C_t \tilde{m}_t dt)\,, & \tilde{L}_0 &= 0\,,\\
\dot{R}_t &= -\Phi_t C_t^T C_t S_t - S_t^T C_t^T C_t \Phi_t - S_t^T C_t^T C_t S_t\,, & R_0 &= 0\,,\\
\dot{S}_t &= A_t S_t - \tilde{R}_t C_t^T C_t S_t - \tilde{R}_t C_t^T C_t \Phi_t\,, & S_0 &= 0\,. \quad (4.39)
\end{aligned}$$

Therefore,

$$\mathbb{E}[G(X_0, \bar{X}_t, t)|\mathcal{Y}_t] = \frac{\int_{\mathbb{R}^m} \mu(dx) \int_{\mathbb{R}^m \times \mathbb{R}^m} G(x, r_1, t) e^{-\frac{1}{2}x^T K_t x + x^T r_2} \eta_t(dr_1, dr_2)}{\int_{\mathbb{R}^m} \mu(dx) \int_{\mathbb{R}^m \times \mathbb{R}^m} e^{-\frac{1}{2}x^T K_t x + x^T r_2} \eta_t(dr_1, dr_2)}$$

For every integrable $g : \mathbb{R}^m \to \mathbb{R}$, choosing G such that $G(X_0, \bar{X}_t, t) = g(\bar{X}_t + \Phi_t X_0) = g(X_t)$ by noting that $X_t = \bar{X}_t + \Phi_t X_0$, we have

$$\pi_t^\mu(g) = \mathbb{E}[g(X_t)|\mathcal{Y}_t] = \frac{\int_{\mathbb{R}^m} \mu(dx) \int_{\mathbb{R}^m \times \mathbb{R}^m} g(r_1 + \Phi_t x) e^{-\frac{1}{2}x^T K_t x + x^T r_2} \eta_t(dr_1, dr_2)}{\int_{\mathbb{R}^m} \mu(dx) \int_{\mathbb{R}^m \times \mathbb{R}^m} e^{-\frac{1}{2}x^T K_t x + x^T r_2} \eta_t(dr_1, dr_2)} \quad (4.40)$$

This completes the computation of π_t^μ for non-Gaussian initial condition. We now give stability results in this case. For the case of stochastic signal, we have the following result due to D. Ocone and E. Pardoux [116].

Theorem 4.2.1. *Under the assumptions of uniform complete observability, uniform complete controllability and* $\mathbb{E}[\|X_0\|^2] < \infty$, *we have*

$$\lim_{t\to\infty} \|\mathbb{E}[X_t|\mathcal{Y}_t] - \hat{X}_t^{\bar{M},\bar{P}}\| = 0 \ \mathbb{P} - \ a.s.$$

and also in L^2 *sense. We also have,*

$$\lim_{t\to\infty} \|\pi_t^\mu(g) - \mathcal{N}(\hat{X}_t^{\bar{M},\bar{P}}, P_t^{\bar{P}})(g)\| = 0, \ \mathbb{P} - \ a.s.,$$

for any bounded uniformly continuous $g : \mathbb{R}^m \to \mathbb{R}$.

Remark 4.2.2. *The above result can be established using the same method that was used by Ocone and Pardoux [116]. In [116], the above result was established under weaker versions of controllability and observability for autonomous case. And also, even though rate of convergence was not studied by the authors in [116], they can be obtained by a careful study of their method (as pointed out by them).*

In the case of deterministic signal, we prove the following result analogous to the above result.

Theorem 4.2.3. *[126, Theorem 4.1] Suppose that Assumptions (4.1.8) and (4.1.9) hold. Let X_0 be square integrable and of the form $X_0 \doteq V_0 + \bar{X}_0$, where $\bar{X}_0$ is a non-degenerate Gaussian random variable independent of V_0. Then for the system given by (4.3) and (4.4),*

$$\lim_{t\to\infty} \|\mathbb{E}[X_t|\mathcal{Y}_t] - \hat{X}_t^{\bar{M},\bar{P}}\| = 0, \;\; \mathbb{P} - \; a.s. \tag{4.41}$$

We also have,

$$\lim_{t\to\infty} \|\pi_t^{\mu}(g) - \mathcal{N}(\hat{X}_t^{\bar{M},\bar{P}}, P_t^{\bar{P}})(g)\| = 0, \;\; \mathbb{P} - \; a.s., \tag{4.42}$$

for any bounded, uniformly continuous $g : \mathbb{R}^m \to \mathbb{R}$.

Remark 4.2.4. *The requirement that the initial condition X_0 be a sum of a Gaussian and a non-Gaussian random variables is not very restrictive. One quite large class of random variables that satisfy this assumption is as follows: for every m-dimensional random variable U with finite second moment and a density f_U, there is a corresponding X_0 satisfying the assumptions of the theorem, where X_0 is defined to be a random variable with density which is a solution of m-dimensional heat equation initialised with f_U.*

Proof. The ideas of our proof are motivated by [104] and by those used in the proof of [116, Theorem 2.6], with certain modifications to accommodate our model with zero noise. $\tilde{m}_t$ in (4.37) is such that $\tilde{m}_0 = \mathbb{E}[\bar{X}_0]$, $\tilde{R}_t$ in (4.38) is such that $\tilde{R}_0 = Cov(\bar{X}_0)$ and $(\tilde{L}_t, S_t, R_t)$ are as defined before. We have seen that non-Gaussian case is given by (4.40) is defined by evolution of finitely many objects *viz.*, K_t, covariance and mean of $\eta(dr_1, dr_2)$ in (4.40). We have

$$\begin{aligned}
\mathbb{E}[g(X_t)|\mathcal{Y}_t] &= \frac{\int_{\mathbb{R}^m} \mu(dx) \int_{\mathbb{R}^m\times\mathbb{R}^m} g(r_1 + \Phi_t x) e^{-\frac{1}{2}x^T K_t x + x^T r_2} \eta_t(dr_1, dr_2)}{\int_{\mathbb{R}^m} \mu(dx) \int_{\mathbb{R}^m\times\mathbb{R}^m} e^{-\frac{1}{2}x^T K_t x + x^T r_2} \eta_t(dr_1, dr_2)} \\
&= \frac{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t} \mu(dx) \int_{\mathbb{R}^m\times\mathbb{R}^m} g(r_1 + \Phi_t x) \tilde{\eta}_t(dr_1, dr_2)}{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t} \mu(dx) \int_{\mathbb{R}^m\times\mathbb{R}^m} \tilde{\eta}_t(dr_1, dr_2)} \\
&= \frac{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t} \mu(dx) \int_{\mathbb{R}^m\times\mathbb{R}^m} g(r_1 + \Phi_t x) \tilde{\eta}_t(dr_1, dr_2)}{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t} \mu(dx)}
\end{aligned} \tag{4.43}$$

where $\tilde{\eta}_t$ is a Gaussian measure with mean $\begin{pmatrix} \tilde{m}_t + S_t x \\ \tilde{L}_t + R_t x \end{pmatrix}$ and covariance $\begin{pmatrix} \tilde{R}_t & S_t \\ S_t^T & R_t \end{pmatrix}$. Setting γ_t to be a Gaussian measure with mean 0 and covariance $\tilde{R}_t$, we have

$$\mathbb{E}[g(X_t)|\mathcal{Y}_t] = \frac{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t} \mu(dx) \int_{\mathbb{R}^m\times\mathbb{R}^m} g(r_1 + \Phi_t x) \tilde{\eta}_t(dr_1, dr_2)}{\int_{\mathbb{R}^m} e^{-\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t} \mu(dx)}$$

$$= \frac{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t-K_t)x+x^T\tilde{L}_t}\mu(dx)\int_{\mathbb{R}^m} g(\Phi_t x+\tilde{m}_t+S_t x+r_3)\gamma_t(dr_3)}{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t-K_t)x+x^T\tilde{L}_t}\mu(dx)}. \tag{4.44}$$

Now setting $g(x) = x$ (this can be done even though g is not bounded because g is integrable with respect to Gaussian measure), we obtain the conditional mean as

$$\mathbb{E}[X_t|\mathcal{Y}_t] = \tilde{m}_t + \mathbb{E}[(\Phi_t + S_t)V_0|\mathcal{Y}_t] = \tilde{m}_t + (\Phi_t + S_t)\mathbb{E}[V_0|\mathcal{Y}_t]$$

Now observe from (4.39) that

$$\frac{d}{dt}(\Phi_t + S_t) = (A_t - \tilde{P}_t C_t^T C_t)(\Phi_t + S_t),$$

which has the same form as (4.20), and thus from Theorem (4.1.13) it follows that $\|\Phi_t + S_t\| \to 0$ as $t \to \infty$.

$$\begin{aligned}\|\mathbb{E}[X_t|\mathcal{Y}_t] - \tilde{m}_t\| &= \|(\Phi_t + S_t)\mathbb{E}[V_0|\mathcal{Y}_t]\| \\ &\leq M_0\|(\Phi_t + S_t)\| \quad \mathbb{P}-a.s. \\ &\xrightarrow{t\to\infty} 0 \quad \mathbb{P}-a.s.,\end{aligned}$$

because $\mathbb{E}[V_0|\mathcal{Y}_t]$ is uniformly integrable (square integrable, in particular). Therefore,

$$\mathbb{E}[X_t|\mathcal{Y}_t] - \tilde{m}_t \to 0 \;\; \mathbb{P}-a.s. \;\; and \;\; in \;\; L^2$$

Now, if we can prove that $(\hat{X}_t^{\bar{M},\bar{P}} - \tilde{m}_t) \to 0,\;\; \mathbb{P}-a.s$ then we shall have shown that

$$\mathbb{E}[X_t|\mathcal{Y}_t] - \hat{X}_t^{\bar{M},\bar{P}} \to 0, \;\; \mathbb{P}-a.s$$

To that end, consider

$$\begin{aligned}d(\tilde{m}_t - \hat{X}_t^{\bar{M},\bar{P}}) = (A_t - P_t^{\bar{P}}C_t^T C_t)(\tilde{m}_t - \hat{X}_t^{\bar{M},\bar{P}})dt + (\tilde{R}_t - P_t^{\bar{P}})C_t^T(dy_t - C_t\mathbb{E}[X_t|\mathcal{Y}_t]) \\ + (\tilde{R}_t - P_t^{\bar{P}})C_t^T C_t(\mathbb{E}[X_t|\mathcal{Y}_t] - \tilde{m}_t)dt\end{aligned}$$

whose solution can be expressed as

$$\begin{aligned}\tilde{m}_t - \hat{X}_t^{\bar{M},\bar{P}} = \Psi_t^{\bar{P}}(\tilde{m}_0 - \bar{M}) + \int_0^t \Psi_t^{\bar{P}}\left(\Psi_s^{\bar{P}}\right)^{-1}(\tilde{R}_s - P_s^{\bar{P}})C_s^T(dY_s - C_s\mathbb{E}[X_s|\mathcal{Y}_s]) \\ + \int_0^t \Psi_t^{\bar{P}}\left(\Psi_s^{\bar{P}}\right)^{-1}(\tilde{R}_s - P_s^{\bar{P}})C_s^T C_s(\mathbb{E}[X_s|\mathcal{Y}_s] - \tilde{m}_s)ds\end{aligned}$$

$$= J_1 + J_2 + J_3,$$

where $J_1 \doteq \Psi_t^{\bar{P}}(\tilde{m}_0 - \bar{M})$, $J_2 \doteq \int_0^t \Psi_t^{\bar{P}} \left(\Psi_s^{\bar{P}}\right)^{-1} (\tilde{R}_s - \bar{P}_s)C_s^T(dY_s - C_s\mathbb{E}[X_s|\mathcal{Y}_s])$, and $J_3 \doteq \int_0^t \Psi_t^{\bar{P}} \left(\Psi_s^{\bar{P}}\right)^{-1} (\tilde{R}_s - \bar{P}_s)C_s^T C_s(\mathbb{E}[X_s|\mathcal{Y}_s] - \tilde{m}_s)ds$. In view of Theorem (4.1.13), it is easy to check that $J_1 \to 0$ and $J_2 \to 0 \;\; \mathbb{P} - a.s.$ Thus, consider

$$\begin{aligned}
J_3 &= \int_0^t \Psi_t^{\bar{P}} \left(\Psi_s^{\bar{P}}\right)^{-1} (\tilde{R}_s - P_s^{\bar{P}})C_s^T C_s(\mathbb{E}[X_s|\mathcal{Y}_s] - \tilde{m}_s)ds \\
&= \Psi_t^{\bar{P}}(P' - \bar{P}) \int_0^t \left(\Psi_s^{P'}\right)^T C_s^T C_s \Psi_s^{P'} \mathbb{E}[V_0|\mathcal{Y}_s]ds \\
&= \Psi_t^{\bar{P}}(P' - \bar{P}) \int_0^t \left(\Psi_s^{P'}\right)^T C_s^T C_s \Psi_s^{P'} (\mathbb{E}[V_0|\mathcal{Y}_s] - \mathbb{E}[V_0|\mathcal{Y}_\infty])ds \\
&+ \Psi_t^{\bar{P}}(P' - \bar{P}) \int_0^t \left(\Psi_s^{P'}\right)^T C_s^T C_s \Psi_s^{P'} ds\, \mathbb{E}[V_0|\mathcal{Y}_\infty] \\
&= L_1 + L_2,
\end{aligned}$$

where,

$$\begin{aligned}
L_1 &\doteq \Psi_t^{\bar{P}}(P' - \bar{P}) \int_0^t \left(\Psi_s^{P'}\right)^T C_s^T C_s \Psi_s^{P'} (\mathbb{E}[V_0|\mathcal{Y}_s] - \mathbb{E}[V_0|\mathcal{Y}_\infty])ds \\
L_2 &\doteq \Psi_t^{\bar{P}}(P' - \bar{P}) \int_0^t \left(\Psi_s^{P'}\right)^T C_s^T C_s \Psi_s^{P'} ds\, \mathbb{E}[V_0|\mathcal{Y}_\infty]
\end{aligned}$$

Again, using the uniform bound on C_s and Theorem (4.1.13), it is clear that $L_2 \to 0 \;\; \mathbb{P} - a.s.$ In order to show that $L_1 \to 0 \;\; \mathbb{P} - a.s$, It suffices to show that

$$\| \int_0^t \left(\Psi_s^{P'}\right)^T C_s^T C_s \Psi_s^{P'} (\mathbb{E}[V_0|\mathcal{Y}_s] - \mathbb{E}[V_0|\mathcal{Y}_\infty])ds \| < \infty$$

To that end, we know that for a given $\epsilon > 0$, there is a $t_\epsilon > 0$ such that for every $t > t_\epsilon$, $\|\mathbb{E}[V_0|\mathcal{Y}_s] - \mathbb{E}[V_0|\mathcal{Y}_\infty]\| < \epsilon$.

$$\begin{aligned}
&\| \int_0^t \left(\Psi_s^{P'}\right)^T C_s^T C_s \Psi_s^{P'} (\mathbb{E}[V_0|\mathcal{Y}_s] - \mathbb{E}[V_0|\mathcal{Y}_\infty])ds \| \\
&\qquad \leq \| \int_0^{t_\epsilon} \left(\Psi_s^{P'}\right)^T C_s^T C_s \Psi_s^{P'} (\mathbb{E}[V_0|\mathcal{Y}_s] - \mathbb{E}[V_0|\mathcal{Y}_\infty])ds \| \\
&\qquad + \| \int_{t_\epsilon}^t \left(\Psi_s^{P'}\right)^T C_s^T C_s \Psi_s^{P'} (\mathbb{E}[V_0|\mathcal{Y}_s] - \mathbb{E}[V_0|\mathcal{Y}_\infty])ds \| \\
&\qquad < \infty \;\; \text{uniformly in } t, \text{ by (4.30)}
\end{aligned}$$

Therefore, the above calculation implies that $(\hat{X}_t^{\bar{M},\bar{P}} - \tilde{m}_t) \to 0, \quad \mathbb{P} - a.s.$ This concludes the proof of (4.41).

We again follow the method of [116] to next prove (4.42). To this end, consider the optimal filtering distribution π_t^μ (recall $\pi_t^\mu(B) = \mathbb{E}[\mathbb{1}_{X_t \in B}|\mathcal{Y}_t]$) and the Gaussian distribution $\tilde{\mathcal{N}}_t \doteq \mathcal{N}(\tilde{m}_t, \tilde{R}_t)$. For a bounded uniformly continuous function $g : \mathbb{R}^n \to \mathbb{R}$, using the expression from (4.44),

$$\begin{aligned} &\int_{\mathbb{R}^n} g(x)\pi_t^\mu(dx) - \int_{\mathbb{R}^n} g(x)\tilde{\mathcal{N}}_t(dx) \\ &= \frac{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t}\mu(dx) \int_{\mathbb{R}^m} g(\Phi_t x + \tilde{m}_t + S_t x + r_3)\gamma_t(dr_3)}{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t}\mu(dx)} - \int_{\mathbb{R}^n} g(x)\tilde{\mathcal{N}}_t(dx) \\ &= \frac{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t}\mu(dx) \int_{\mathbb{R}^m} \left[g(\Phi_t x + \tilde{m}_t + S_t x + r_3) - g(\tilde{m}_t + S_t r_3)\right]\gamma_t(dr_3)}{\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t}\mu(dx)}, \end{aligned} \tag{4.45}$$

where the last line is obtained by using definition of $\gamma_t(dx)$ and by multiplying and the second term in the second line above by $\int_{\mathbb{R}^m} e^{\frac{1}{2}x^T(R_t - K_t)x + x^T\tilde{L}_t}\pi(dx)$. Now, if we partition the $\mu(dx)$ integral into regions $|(\Phi_t + S_t)x| < \delta$ and $|(\Phi_t + S_t)x| \geq \delta$ for a fixed $\delta > 0$, then

$$\begin{aligned} \int_{\mathbb{R}^n} g(x)\pi_t^\mu(dx) - &\int_{\mathbb{R}^n} g(x)\tilde{\mathcal{N}}_t(dx) \\ &\leq \sup_{|z_1 - z_2| < \delta} |g(z_1) - g(z_2)| + 2\sup_z (g(z))\mathbb{E}[\mathbb{1}_{|(\Phi_t + S_t)X_0| > \delta}|\mathcal{Y}_t] \\ &\leq \sup_{|z_1 - z_2| < \delta} |g(z_1) - g(z_2)| + \frac{2\sup_z (g(z))}{\delta^2}\|\Phi_t + S_t\|^2\mathbb{E}[|X_0|^2] \\ &\leq \sup_{|z_1 - z_2| < \delta} |g(z_1) - g(z_2)| \quad \text{as} \quad t \to \infty \end{aligned}$$

where the second inequality follows from Chebyshev's inequality. Observe now that for any $\delta_0 > 0$, we can choose sufficiently small δ, such that

$$\sup_{|z_1 - z_2| < \delta} |g(z_1) - g(z_2)| < \delta_0$$

which implies that $\pi_t^\mu - \tilde{\mathcal{N}}_t \to 0$ $\mathbb{P} - a.s.$ weakly as $t \to \infty$. Using now the fact that $(\hat{X}_t^{\bar{M},\bar{P}} - \tilde{m}_t) \to 0$, $\mathbb{P} - a.s$ and $(P_t^{\bar{P}} - \tilde{R}_t) \to 0$, we conclude that $[\pi_t^\mu(g) - \mathcal{N}(\hat{X}_t^{\bar{M},\bar{P}}, P_t^{\bar{P}})(g)] \xrightarrow{t \to \infty} 0$, $\mathbb{P} - a.s.$ ■

Remark 4.2.5. *In contrast to the results in the case of non-autonomous filtering model with stochastic signal for non-autonomous case, rates of convergence could not be estimated in the case of non-autonomous filtering model with deterministic signal. The main obstruction is in establishing the exponential stability of the* (4.17).

4.3 Small noise analysis

Suppose the underlying signal X^ε is stochastic with $\mathcal{L}(X^\varepsilon_0) = \mu$ and a very small noise (ε denotes the noise strength). Denote its filter by $\pi^\mu_{t,\varepsilon}$, depending on ε. Consider the evolution corresponding to $\pi^\mu_{t,0}$ and let us drive this equation with observations Y^ε that are made on the true system (the one with stochastic signal)[3]. Now $\pi^\mu_{t,0}(g)$ is not optimal estimate of $g(X^\varepsilon_t)$ given Y^ε. The question of interest is: Does this quantity remain close to $\pi^\mu_{t,\varepsilon}(g)$ at all times, for small enough ε. This is desirable because the Kallianpur-Striebel formula in the deterministic signal case involves only finite dimensional integrals (see next chapter). We provide the answer to the question raised above, in the case of Kalman-Bucy filter (linear filter with Gaussian initial condition). To that end, consider the processes given by,

$$X^\varepsilon_t = X^\varepsilon_0 + \int_0^t A_s X^\varepsilon_s ds + \varepsilon \int_0^t F_s dB^\varepsilon_s, \tag{4.46}$$

$$Y^\varepsilon_t = \int_0^t C_s X^\varepsilon_s ds + \int_0^t dW^\varepsilon_s, \tag{4.47}$$

$$\mathcal{L}(X^\varepsilon_0) \doteq \mu = \mathcal{N}(M, P).$$

The probability space and the above model are constructed in a same way as mentioned in the begin of this chapter.As earlier, X^ε_0, V^ε_t and W^ε_t are all mutually independent. Let $\mathcal{Y}^\varepsilon \doteq \sigma\{Y^\varepsilon : 0 \le s \le t\}$ and $\pi^\mu_{t,\varepsilon}$ be the conditional distribution of X^ε_t given $\mathcal{Y}^\varepsilon_t$[4]. Since X^ε_0 is Gaussian, we have $\pi^\mu_{t,\varepsilon} = \mathcal{N}(\hat{X}^{M,P}_{t,\varepsilon}, P^P_{t,\varepsilon})$, where,

$$\begin{aligned} d\hat{X}^{M,P}_{t,\varepsilon} &= A_t \hat{X}^{M,P}_{t,\varepsilon} dt + P^P_{t,\varepsilon} C^T_t (dY^\varepsilon_t - C_t \hat{X}^{M,P}_{t,\varepsilon} dt), \\ \dot{P}^P_{t,\varepsilon} &= A_t P^P_{t,\varepsilon} + P^P_{t,\varepsilon} A^T_t - P^P_{t,\varepsilon} C^T_t C_t P^P_{t,\varepsilon} + \varepsilon^2 F_t F^T_t, \\ \hat{X}^{M,P}_{0,\varepsilon} &= m, \quad P^P_{0,\varepsilon} = P. \end{aligned}$$

We also define the new process $\hat{X}^0_t$ as

$$d\hat{X}^0_t = A_t \hat{X}^0_t dt + P^P_t C^T_t (dY^\varepsilon_t - C_t \hat{X}^0_t dt).$$

Note that above definition involves Y^ε_t, instead of Y^0_t and P^P_t is solution to (4.17). To proceed further, additional assumption is made in this analysis.

[3]Note that evolution equation of $\pi^\mu_{t,0}$ is an algorithm which takes initial condition μ and a driving process as inputs. To get the optimal estimate, the driving process has to be the observation process coming from the corresponding model

[4]These quantities are already defined. But, we now explicitly show the dependence on ε

Assumption 4.3.1. *F_t is uniformly bounded in t and $\dot{z}_t = (A_t - P_t^P C_t^T C_t) z_t$ is exponentially stable.* i.e.,

$$\|\Psi_t^P \left(\Psi_s^P\right)^{-1}\| \leq K_0 \exp(-\alpha(t-s)),$$

for $t \geq s \geq 0$ and for some K_0, $\alpha > 0$. Here P_t^P is solution to (4.17)

Theorem 4.3.2. *[126, Theorem 5.3] Under Assumptions* (4.1.8), (4.1.9) *and* (4.3.1),

$$\mathbb{P}\left(\liminf_{\varepsilon \to 0} \sup_{0 \leq t \leq T_0} ||\hat{X}_{t,\varepsilon}^{M,P} - \hat{X}_t^0|| = 0,\right) = 1, \ \forall T_0 \geq 0$$

Remark 4.3.3. *We note the work of Baras J. S.* et al.*[16] in which the authors studied the limiting finite time behavior of the autonomous Kalman - Bucy filter as the signal and observation noises go to zero. In contrast, we study behavior of non-autonomous Kalman-Bucy filter (corresponding to zero signal noise) over infinite time horizon where the driving observations are from a linear filtering model with vanishing signal noise.*

Proof. Let us begin with observing

$$\begin{aligned}
&\frac{d}{dt}(P_{t,\varepsilon}^P - P_t^P) \\
&= B_t^P(P_{t,\varepsilon}^P - P_t^P) + (P_{t,\varepsilon}^P - P_t^P)(B_t^P)^T - (P_{t,\varepsilon}^P - P_t^P)C_t^T C_t (P_{t,\varepsilon}^P - P_t^P) + \varepsilon^2 F_t F_t^T,
\end{aligned}$$

where, $P_{0,\varepsilon}^P - P_0^P = 0$. Recall that $B_t^P = A_t - P_t^P C_t^T C_t$ and Ψ_t^P is the fundamental matrix solution. If we define $\Delta P_t \doteq P_{t,\varepsilon}^P - P_t^P$, then we have

$$\begin{aligned}
\frac{d}{dt}[(\Psi_t^P)^{-1}\Delta P_t (\Psi_t^P)^{-T}] &= (\Psi_t^P)^{-1}(-\Delta P_t C_t^T C_t \Delta P_t + \varepsilon^2 F_t F_t^T)(\Psi_t^P)^{-T} \\
\Delta P_t &= \Psi_t^P \int_0^t (\Psi_s^P)^{-1}(-\Delta P_s C_s^T C_s \Delta P_s + \varepsilon^2 F_s F_s^T)(\Psi_t^P (\Psi_s^P)^{-1})^T ds \\
&\leq \varepsilon^2 \Psi_t^P \int_0^t (\Psi_s^P)^{-1} F_s F_s^T (\Psi_s^P)^{-T} ds (\Psi_t^P)^T
\end{aligned}$$

From the assumption of exponential stability, we have $\|\Psi_t^P \left(\Psi_s^P\right)^{-1}\| \leq K e^{-\alpha(t-s)}$, for some K, α and for all $t \geq s \geq 0$. Therefore,

$$0 \leq \|\Delta P_t\| \leq \frac{\varepsilon^2 K F}{2\alpha}$$

Now, we consider the evolution equation for $\hat{X}_{t,\varepsilon}^{M,P} - \hat{X}_t^0$

$$d(\hat{X}_{t,\varepsilon}^{M,P} - \hat{X}_t^0) = B_t^P(\hat{X}_{t,\varepsilon}^{M,P} - \hat{X}_t^0)dt + (\Delta P_t)C_t^T(dY_t^\varepsilon - C_t \hat{X}_{t,\varepsilon}^{M,P} dt), \ \ \hat{X}_{0,\varepsilon}^{M,P} - \hat{X}_0^0 = 0,$$

$$\hat{X}^{M,P}_{t,\varepsilon} - \hat{X}^0_t = \int_0^t \Psi^P_t (\Psi^P_s)^{-1} (\Delta P_s) C^T_s (dY^\varepsilon_s - C_s \hat{X}^{M,P}_{s,\varepsilon} ds)$$

Define, $u_t \doteq \int_0^t \left(\Psi^P_s\right)^{-1} (\Delta P_s) C^T_s (dY^\varepsilon_s - C_s \hat{X}^{M,P}_{s,\varepsilon} ds)$ and $\mathcal{I}_t \doteq \sigma\{Y^\varepsilon_r - \int_0^r C_s \hat{X}^{M,P}_{s,\varepsilon} ds : 0 \le r \le t\}$. Clearly, u_t is a $\mathcal{I}_t$-martingale. We know that, from Proposition (A.0.28),

$$\|(\Psi^P_t)^{-1}\| = \|(\Psi^P_t)^{-T}\| \le K' e^{\alpha' t},$$

where, α' can chosen as $\sup_{t \ge 0} \left(\|A_t\| + \|P^P_t C^T_t C_t\|\right)$. Here, we used the fact that

$$\frac{d}{dt}\left(\Psi^P_t\right)^{-T} = -\left(A^T_t - C^T_t C_t P^P_t\right)\left(\Psi^P_t\right)^{-T}$$

Then, for any given $T_0 \ge 0$ and $\lambda > 0$, applying Doob's inequality to submartingale (see Lemma (A.0.13)), $\|u_t\|$, we have

$$\mathbb{P}\left(\sup_{0 \le t \le T_0} \|u_t\| \ge \lambda\right) \le \frac{\mathbb{E}[\|u_{T_0}\|]}{\lambda}$$

$$\mathbb{P}\left(\sup_{0 \le t \le T_0} \| \left(\Psi^P_t\right)^{-1} (\hat{X}^{M,P}_{t,\varepsilon} - \hat{X}^0_t)\| \ge \lambda\right) \le \frac{\mathbb{E}[\| \left(\Psi^P_{T_0}\right)^{-1} (\hat{X}^{M,P}_{T_0,\varepsilon} - \hat{X}^0_{T_0})\|]}{\lambda}$$

$$\mathbb{P}\left(K' e^{\alpha' T_0} \sup_{0 \le t \le T_0} \|(\hat{X}^{M,P}_{t,\varepsilon} - \hat{X}^0_t)\| \ge \lambda\right) \le \| \left(\Psi^P_{T_0}\right)^{-1} \| \frac{\varepsilon\sqrt{KFM}}{2\alpha\lambda}$$

Now, we take $\varepsilon \to 0$ along an arbitrary sequence $\{\varepsilon_n\}_{n \ge 1} \to 0$. Choose a subsequence $\{\varepsilon_{n_k}\}_{k \ge 1}$ such that $\sum_{k \ge 1} \varepsilon_{n_k} < \infty$ and $\lambda = \lambda_0 K' e^{\alpha' T_0}$(arbitrariness of λ is now in λ_0),

$$\mathbb{P}\left(K' e^{\alpha' T_0} \sup_{0 \le t \le T_0} \|(\hat{X}^{M,P}_{t,\varepsilon_{n_k}} - \hat{X}^0_t)\| \ge \lambda_0 K' e^{\alpha' T_0}\right) \le K' e^{\alpha' T_0} \frac{\sqrt{KFM}\varepsilon_{n_k}}{2\alpha\lambda_0 K' e^{\alpha' T_0}}$$

$$\mathbb{P}\left(\sup_{0 \le t \le T_0} \|(\hat{X}^{M,P}_{t,\varepsilon_{n_k}} - \hat{X}^0_t)\| \ge \lambda_0\right) \le \frac{\sqrt{KFM}\varepsilon_{n_k}}{2\alpha\lambda_0}$$

Then, applying Borel-Cantelli lemma (see Lemma (A.0.1)) to $\{S_k\}_{k \ge 1}$, where $S_k \doteq \{\omega \in \Omega : \sup_{0 \le t \le T_0} \|(\hat{X}^{M,P}_{t,\varepsilon_{n_k}} - \hat{X}^0_t)\| \ge \lambda_0\}$, we conclude that

$$\mathbb{P}\left(\lim_{k \to \infty} \sup_{0 \le t \le T_0} \|\hat{X}^{M,P}_{t,\varepsilon_{n_k}} - \hat{X}^0_t\| = 0,\right) = 1, \ \forall T_0 \ge 0.$$

This proves the result. ■

Remark 4.3.4. *The second part of the assumption about exponential stability depends only* *on the matrices A_t, C_t (in particular, on uniform complete observability of the pair) but is not related to the controllability of the pair A_t, F_t.*

The following theorem gives the sufficient condition for the Assumption (4.3.1) to hold.

Theorem 4.3.5. *In the autonomous case ($A_t = A,\ C_t = C$), under the following assumptions,*

1. *A has exponential dichotomy (all eigenvalues have non-zero real part).*
2. *C^TC, when evaluated in the eigenbasis of A (with p unstable and $m-p$ stable eigenvalues), is in the form*

$$C^TC \doteq \begin{pmatrix} K & 0 \\ 0 & \bar{K} \end{pmatrix},$$

where, $K \in \mathbb{R}^{p\times p}$ and invertible. Then

$$\dot{z}_t = (A - P_t^R C^T C) z_t$$

is exponentially stable. Here, $R > 0$ and P_t^R is solution to (4.17).

Proof. Let us work in eigenbasis of A. We first show that we can choose a positive definite matrix P such that there exists a non-zero, non-negative matrix symmetric matrix P^f and $P_t^P - P^f \to 0$ as $t \to \infty$.

Choose an invertible $P = \begin{pmatrix} P^1 & 0 \\ 0 & P^2 \end{pmatrix}$. We guess that P_t^P is of the form $\begin{pmatrix} M_t & 0 \\ 0 & L_t \end{pmatrix}$. Then the equation becomes

$$\begin{pmatrix} \dot{M}_t & 0 \\ 0 & L_t \end{pmatrix} = \begin{pmatrix} A^u & 0 \\ 0 & A^s \end{pmatrix}\begin{pmatrix} M_t & 0 \\ 0 & L_t \end{pmatrix} + \begin{pmatrix} M_t & 0 \\ 0 & L_t \end{pmatrix}\begin{pmatrix} (A^u)^T & 0 \\ 0 & (A_s)^T \end{pmatrix}$$
$$- \begin{pmatrix} M_t & 0 \\ 0 & L_t \end{pmatrix}\begin{pmatrix} K & 0 \\ 0 & \bar{K} \end{pmatrix}\begin{pmatrix} M_t & 0 \\ 0 & L_t \end{pmatrix}.$$

We have, equivalently,

$$\dot{M}_t = A^u M_t + M_t (A^u)^T - M_t K M_t$$
$$\dot{L}_t = A^s L_t + L_t (A^s)^T - L_t K L_t$$

Since we chose the initial condition in the block diagonal form, it will remain in the block diagonal form. We study the above equations separately It is clear that $L_t \to 0$

as $t \to \infty$ (from the corresponding (4.18) and properties of A^s), $\sup_{t\geq 0}|M_t| < \infty$ (from (4.1.10)) and $\inf_{t\geq 0}\sigma_{min}(M_t) > 0$ (from the corresponding (4.18) and properties of A^u), where σ_{min} is the minimum singular value. Therefore, we can study the asymptotic behavior of M_t by studying the asymptotic behavior of M_t^{-1}. One can verify that

$$M_t^{-1} = \phi_{-t}^T M_0^{-1}\phi_{-t} + \int_0^t \phi_{s-t}^T K\phi_{s-t} ds,$$

where, $\dot{\phi}_t = A^u\phi_t$. we know, from the property of A^u, $\|\phi_{-t}\| \to 0$ as $t \to \infty$.

$$\begin{aligned}\lim_{t\to\infty} v_i^T M_t^{-1} v_j &= \lim_{t\to\infty} v_i^T \int_0^t \phi_{s-t}^T K\phi_{s-t} v_j ds \\ &= \int_0^\infty e^{(\lambda_i+\lambda_j)(s-t)} ds v^T K v = \frac{1}{\lambda_i+\lambda_j} v_i^T K v_j.\end{aligned}$$

where, v_i is unstable eigenvector with eigenvalue λ_i (Note that we are working in the eigenbasis and v should be considered as p dimensional vector). Therefore, M_t^{-1} converges to M^{-1} (given by the above equation) and $M_t \to M$ as $t \to \infty$. Also, note that the convergence of M_t^{-1} to M^{-1} is exponential implying exponential convergence of M_t to M. Indeed,

$$\|V_1 - V_2\| = \|V_1(V_1^{-1} - V_2^{-1})V_2\| \leq \|V_1\|\|V_2\|\|V_1^{-1} - V_2^{-1}\|.$$

Therefore, we can conclude that $\lim_{t\to\infty} P_t^P = P^f \doteq \begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix}$. Now, we can show that $A - P^f C^T C$ is asymptotically stable (equivalently, exponentially stable).

$$A - P^f C^T C = \begin{pmatrix} A^u & 0 \\ 0 & A^s \end{pmatrix} - \begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} K & 0 \\ 0 & L \end{pmatrix} = \begin{pmatrix} A^u - MK & 0 \\ 0 & A^s \end{pmatrix}$$

Therefore, if we can show that $A^u - MK$ is asymptotically stable, then we are done. For p-dimensions, from (4.24) and invertibility of K , we have that this indeed is the case.

Now to show that $A - P_t^R C^T C$ is exponentially stable (R is any positive definite matrix), we use the fact that $\|P_t^R - P_t^P\| + \|P_t^P - P^f\| \to 0$ as $t \to \infty$ and we use the following theorem.

Theorem 4.3.6. *[1, Theorem 4.4.6] If $\dot{\psi}_t = \hat{A}_t\psi_t$ with $\psi_0 = \mathbb{I}_p$ is exponentially stable* i.e., $\|\psi_t\psi_s^{-1}\| \leq K_0 \exp(-\alpha(t-s))$, *for $t \geq s \geq 0$ and for some K_0, $\alpha > 0$ and a matrix δ_t is* *such that* $\|\delta_t\| \leq \varepsilon$, *for some* $\varepsilon > 0$, *then*

$$\|\bar{\psi}_t \bar{\psi}_s^{-1}\| \leq K_1 \exp(-\beta(t-s)),\ t \geq s \geq 0,$$

where, $\beta = \alpha - \varepsilon K_0$ *and* $\dot{\bar{\psi}}_t = (A_t + \delta_t)\bar{\psi}_t$ *with* $\bar{\psi}_0 = \mathbb{I}_p$. *In particular, if* ε *is small enough, we have exponential stability for* $\dot{\bar{\psi}}_t = (A_t + \delta_t)\bar{\psi}_t$.

In our case, $\hat{A}_t = A - PC^TC$ and $\delta_t = (P_t^R - P)C^TC$. We know that $\|\delta_t\| \to 0$ as $t \to \infty$. Since, exponential stability is concerned only with the long time behavior, we wait long enough for $\|\delta_t\|$ to become sufficiently small. Therefore, we have the desired exponential stability. ■

4.4 Discussion and Conclusion

The problem that we described is the asymptotic behavior of the distance between an incorrectly initialised linear filter and the correctly initialised linear filter. The classical results on filter stability for the linear case with system noise are established under the assumptions of observability and controllability. These results cannot be trivially extended to the case of linear filter with zero system noise, since such systems are not controllable. It has been shown in this chapter that controllability assumption can be completely discarded. Additionally, we have also seen that system with small noise behaves approximately the same as system with zero noise, *i.e*, small noise limit is non-singular in the case of Gaussian initial conditions.

For general non-autonomous systems, exponential convergence has not been established, even though it is observed numerically in many examples, as noted in [28, 68] and references therein, and as we have already mentioned, the main difficulty is to establish exponential stability of (4.17). The minimal set of assumptions under which exponential convergence of (4.17) is currently under investigation.

Chapter 5

Stability of the Non-linear Filter

In this chapter, we describe the non-linear filtering model in the case of deterministic signal. We shall establish the filter stability in this case, under appropriate conditions on signal and observation model. We consider the filtering problem in this context as a parameter estimation problem (in this case, it is the initial condition of the signal). It is a particular case of smoothing problem. Almost all of the content of this chapter is taken from [125] (which is part of the doctoral work corresponding to this thesis). We describe and prove the results for continuous time case (as we did until now). But, we also state the analogous results in discrete time case.

5.1 Introduction

The techniques in this chapter are inspired from the work of F. Cérou [41]. In [41], F. Cérou studied the problem of accuracy of the filter (which is a measure of deviation of the filter from the signal) for the deterministic dynamics. We use the accuracy of the smoother, *i.e.*, the asymptotic convergence of the conditional distribution of the initial condition given the observations (a particular case of the smoothing problem) to establish stability of the non-linear filter. In contrast to the signal state space that is considered until now (which is $\mathbb{R}^m$), we allow S to be a complete m-dimensional Riemannian manifold with metric d. The probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ is constructed exactly in the same way as it was done in the case of linear filter for the case of deterministic signal (see previous chapter). We give the details of the construction below.

$$
\begin{aligned}
\Omega &\doteq S \times C([0,\infty), \mathbb{R}^n) \\
\mathcal{F} &\doteq \mathcal{B}(\Omega) \\
\mathbb{P} &\doteq \mu \times \mathbb{P}_W^n,
\end{aligned}
$$

Consider the canonical process on Ω denoted by ω,

$$\begin{aligned}\omega_{(\cdot)} &\doteq (X_0, W_{(\cdot)}),\\ \mathcal{H}_t &\doteq \sigma(\omega_s : \omega \in \Omega, 0 \le s \le t)\\ \mathcal{F}_t &\doteq \sigma(\omega_s : \omega \in \Omega, 0 \le s \le t) \vee \mathcal{N},\end{aligned}$$

where, $\mathcal{N}$ is set of all $\mathbb{P}$- null sets. From above, X_0 and W are mutually independent. We consider a continuous time dynamical system $\{\phi_t\}_{t\in\mathbb{R}}$ with initial condition X_0, on S. This dynamical system is observed through the observation process $Y_t \in \mathbb{R}^n$ in the usual way.

$$Y_t = \int_0^t h(s, \phi_s(X_0))ds + W_t\,,$$

where, $h : \mathbb{R}^+ \times S \to \mathbb{R}^n$. As earlier, $\mathcal{Y}_t \doteq \sigma\{Y_s : 0 \le s \le t\}$.

We define the conditional distribution π_t^μ through smoother $\hat{\pi}_t^\mu$, that is the conditional distribution of the initial condition X_0, again conditioned on $\mathcal{Y}_t$. To that end, we define

$$Z\left(t, x, Y_{[0,t]}\right) \doteq \exp\left(\int_0^t h\left(s, \phi_s\left(x\right)\right)^T dY_s - \frac{1}{2}\int_0^t \left\|h\left(s, \phi_s\left(x\right)\right)\right\|^2 ds\right)\,,$$

it follows from Bayes' rule (see Lemma (2.1.4)), that for any bounded continuous function $g : S \to \mathbb{R}$, the smoother is given by

$$\hat{\pi}_t^\mu(g) \doteq \mathbb{E}\left[g\left(X_0\right)|\mathcal{Y}_t\right] = \frac{\int_S g\left(x\right) Z\left(t, x, Y_{[0,t]}\right)\mu\left(dx\right)}{\int_S Z\left(t, x, Y_{[0,t]}\right)\mu\left(dx\right)}\,,$$

Noting that $X_t = \phi_t(X_0)$, the filter is given by

$$\pi_t^\mu(g) \doteq \mathbb{E}\left[g\left(\phi_t(X_0)\right)|\mathcal{Y}_t\right] = \frac{\int_S g\left(\phi_t(x)\right) Z\left(t, x, Y_{[0,t]}\right)\mu(dx)}{\int_S Z\left(t, x, Y_{[0,t]}\right)\mu(dx)}\,.$$

5.2 Stability of the non-linear filter

As earlier, we denote the incorrect initial condition to be ν. Then the corresponding incorrect filter is given by

$$\pi_t^\nu(g) = \frac{\int_S g\left(\phi_t(x)\right) Z\left(t, x, Y_{[0,t]}\right) \nu(dx)}{\int_S Z\left(t, x, Y_{[0,t]}\right) \nu(dx)}.$$

In the following, we prove that π_t^μ and π_t^ν merge weakly (in the sense of [61]) in expectation. This will be established by firstly, proving that smoother is asymptotically accurate, under suitable conditions *i.e.*, $\hat{\pi}_t^\mu$ is supported around the true initial condition X_0 after large times. The asymptotic accuracy of smoother was studied in [41], but under very restrictive conditions. The result we prove will be significant generalisation of the result in [41]. As mentioned in the beginning of this chapter, all results presented in this chapter have discrete time analogs. We only state these results and not prove them (proofs are very similar to those in continuous time case).

5.2.1 Main assumptions

We now state the main assumptions that are made in the analysis. We give the significance of these assumptions later. Consider $\tau > 0$ which will be defined in Assumption (5.2.3).

Assumption 5.2.1. *There exists a bounded open set U with diameter $K < \infty$ such that $\overline{\phi_\tau(U)} \subset U$.*

Assumption 5.2.2. *$\forall x,\ y \in U$, we have $d(\phi_\tau x, \phi_\tau y) \leq C d(x,y)$, for some $C = C(\tau) > 1$.*

Assumption 5.2.3. *There exists $\tau > 0$ such that $\forall t \geq 0$ and $x_1, x_2 \in U$,*

$$\rho_t d(x_1, x_2)^2 \leq \int_t^{t+\tau} \left\| h\left(s, \phi_{s-t}(x_1)\right) - h\left(s, \phi_{s-t}(x_2)\right) \right\|^2 ds \leq R \rho_t d(x_1, x_2)^2, \quad (5.1)$$

where, ρ_t is a positive non-decreasing function such that

$$\lim_{t\to\infty} \rho_t = \infty,\ \lim_{t\to\infty} \frac{\int_0^t \rho_s ds}{\rho_t} = \infty \text{ and } \frac{t\rho_t}{\int_0^t \rho_s ds} \leq C' < \infty,$$

for some $C' > 0$ and $R > 1$.

It follows from this assumption that $\forall x, y \in U$,

$$\sum_{i=0}^{N} \rho_{i\tau} d\left(\phi_{i\tau}(x), \phi_{i\tau}(y)\right)^2 \leq \int_0^t \|h\left(s, \phi_s(x)\right) - h\left(s, \phi_s(y)\right)\|^2 ds$$
$$\leq R \sum_{i=0}^{N+1} \rho_{i\tau} d\left(\phi_{i\tau}(x), \phi_{i\tau}(y)\right)^2 , \tag{5.2}$$

where, $N = \lfloor \frac{t}{\tau} \rfloor$. Define,

$$D_N(x,y) \doteq \left(\sum_{i=0}^{N} \rho_{i\tau} d\left(\phi_{i\tau}(x), \phi_{i\tau}(y)\right)^2 \right)^{\frac{1}{2}} \quad \textit{and} \quad d_N(x,y) \doteq \max_{0 \leq i \leq N-1} d\left(\phi_{i\tau}(x), \phi_{i\tau}(y)\right).$$

It is straightforward to see that $D_N(x,y)$ and $d_N(x,y)$ are metrics on S (for a fixed $N \geq 0$). Moreover, they are such that

$$\rho_0 d_{N+1}(x,y) \leq D_N(x,y) \leq \rho_{N\tau} \sqrt{N+1} d_{N+1}(x,y)$$

It follows from Assumption (5.2.1) that for $x, y \in U$, we have a uniform (in N) bound

$$d_N(x,y) \leq K.$$

Indeed, from the invariance of U, we have $\phi_{i\tau} x, \phi_{i\tau} y \in U$ and hence we get

$$d(\phi_{i\tau} x, \phi_{i\tau} y) \leq K,$$

for all $i \geq 0$.

Assumption 5.2.4. *For $(x,y) \in \mathcal{V} \subset U \times U$, compliment (in $U \times U$) of a zero Lebesgue measure set satisfying $d(x,y) \geq b > 0$, the following holds*

$$D_N^2(x,y) \geq L^2(b) \sum_{i=0}^{N} \rho_{i\tau} ,$$

where, $L(b)$ is a positive constant.

Assumption 5.2.5. $supp(\mu) \subset U$

Before proceeding further, we define the notion the spanning sets [143, Definition 7.8] which plays an important role in the proof of Theorem (5.2.8). It will help us get the estimates of the covering number of a compact set with ϵ-balls (under the metric d_N), for any $\epsilon > 0$.

Definition 5.2.6. *For a given compact set $\mathcal{K}$, $n \geq 0$ and $\epsilon > 0$, the set $F \subset X$ is called (n, ϵ)-spanning set of $\mathcal{K}$ with respect to ϕ_τ if $\forall x \in \mathcal{K}$, $\exists y \in F$ such that*

$$\max_{0 \leq i \leq n-1} d\left(\phi_{i\tau}(x), \phi_{i\tau}(y)\right) \leq \epsilon.$$

Definition 5.2.7. *$r(\mathcal{K}, n, \epsilon, \phi_\tau)$ is defined as the minimum possible cardinality of (n, ϵ)-spanning sets of $\mathcal{K}$.*

Note that for any n, $r(\mathcal{K}, n, \epsilon, \phi_\tau)$ is finite due to compactness of $\mathcal{K}$.

5.2.2 Significance of the assumptions

We now discuss the significance of the above assumptions individually. In the section (5.5), a detailed discussion of some important examples (for which we can explicitly verify or provide strong numerical evidence for these assumptions) is given.

1. Relation to observability: Assumption (5.2.3) resembles closely the well-known observability condition (see Definition (A.0.25)) in the linear case except for the dependence of ρ_t on t satisfying certain conditions. And also, even though the assumed upper bound in Assumption (5.2.3) is satisfied in many of the filtering models, the assumed lower bound is really difficult to come by, in practice. This intuitively implies that observing the signal long enough, we can distinguish between the two initial conditions. As mentioned later, the additional conditions on ρ_t are required precisely in ensuring the positivity of the exponent in (5.17). This can be understood intuitively in the following way. In general, deterministic dynamics has the tendency to lose information which can be attributed to sensitive dependence of the dynamics on initial conditions. Therefore, to establish the accuracy of the smoother, we have to make observations at a rate faster than the rate at which dynamics loses the information in the form of sensitive dependence to initial conditions.

   To express this more precisely, we consider an open ball, denoted by $Q_N(r, x)$, of radius r around $x \in S$ under the metric d_N. It is clear that for $y, y' \in S$, $d_N(y, y') \leq d_{N+1}(y, y')$, $\forall N \geq 0$. Therefore, the volume of $Q_N(r, x)$ is non-increasing in N. Informally, this means that set of all points whose orbits are within a d_N-distance r from the orbit of x can shrink to a zero volume set containing x. If the dynamics is sensitive to initial conditions then the volume of $Q_N(r, x)$ goes to zero as $N \to \infty$. In fact, Assumption (5.2.2) implies that the volume of $Q_N(r, x)$ goes to zero at a rate that is at most exponential, which can be seen in (5.16), leading to the third term in the exponent on the right hand side of (5.17). Considering ρ_t as mentioned in Assumption (5.2.3) leads to the third term going to zero, by ensuring that we are observing at a fast enough rate.

2. Bounded orbits and Attracting set: Assumption (5.2.1) says that if the initial condition lies in the set U, then it signal stays in U for all future times. The assumption also implies the existence of an invariant set (U_I) defined by $U_I = \cap_{t \geq 0} \phi_t(U)$. This is a reasonable assumption since in practice, even though the initial condition may not lie in U, it is plausible to assume that the state being observed (that started inside a bounded set), lies inside U, since a natural system evolving over long enough time would have entered the set U when U is a global attracting set, .

3. Divergence of nearby orbits: Assumption (5.2.4) says that two orbits, started at a given distance away from each other, do not come too close to each other very often. Intuitively, this is reasonable for a system for which the support of the initial condition μ does not contain any stable periodic orbits or fixed points. In Section (5.5), we give examples of general classes of systems that satisfy this assumption.

4. We require the Assumptions (5.2.1) and (5.2.4) to ensure that signal process neither converges to periodic orbits (or fixed points) nor escapes to infinity. This is because we are proving filter stability using the smoother and by observing a signal that either escapes to infinity or converges onto a periodic orbit (or fixed point), we lose all the information of the initial condition (as time goes to infinity). This means that as time progresses, we do not gain any more information about the initial condition than we already have. This can be avoided if we consider a signal that neither converges to a periodic orbit (or fixed point) nor escapes to infinity.

5.2.3 Asymptotic accuracy of the smoother

We now prove that asymptotically, support of the smoother $\hat{\pi}_t^\mu$ is concentrated around the initial condition X_0. And this concentration happens at an exponential rate.

Theorem 5.2.8. *Suppose μ is absolutely continuous with respect to the volume measure σ on S and $\frac{d\mu}{d\sigma}$ is continuous on the support of μ. Under the assumptions (5.2.1), (5.2.2), (5.2.3), (5.2.4) and (5.2.5), for all $a > 0$, there exists $\alpha(a) > 0$ such that the smoother $\hat{\pi}_t^\mu(\cdot) \doteq \mathbb{P}\left[X_0 \in \cdot | \mathcal{Y}_t\right]$ satisfies*

$$e^{\alpha(a)t}\left[1 - \hat{\pi}_t^\mu\left(B_a(X_0)\right)\right] \xrightarrow{t\to\infty} 0 \quad \mathbb{P}\text{-}a.s.,$$

where $B_a(X_0) \doteq \{x \in S : d(x, X_0) \le a\}$ is the ball centred at X_0 and the rate $\alpha(a) > 0$ depends only on the radius a of the ball.

Proof. In order to show that $\hat{\pi}_t^{\mu}(B_a(X_0))$ goes to one, we will show, in Lemma (5.2.11), that $\hat{\pi}_t^{\mu}(B_a(X_0)^c)$ goes to zero at an exponential rate as $t \to \infty$.

Recall that for any measurable set $A \in \mathcal{B}(S)$,

$$\hat{\pi}_t^{\mu}(A) = \frac{\int_A \exp\left(\int_0^t h(s, \phi_s(x))^T dY_s - \frac{1}{2}\int_0^t \|h(s, \phi_s(x))\|^2 ds\right)\mu(dx)}{\int_S \exp\left(\int_0^t h(s, \phi_s(x))^T dY_s - \frac{1}{2}\int_0^t \|h(s, \phi_s(x))\|^2 ds\right)\mu(dx)}$$

We substitute $dY_s = h(s, \phi_s(X_0))\, ds + dW_s$ and multiply the numerator and the denominator by $\exp(\int_0^t h(s, \phi_s(X_0))^T dW_s - \frac{1}{2}\int_0^t \|h(s, \phi_s(X_0))\|^2 ds)$, which is independent of x to get,

$$\hat{\pi}_t^{\mu}(A) = \frac{\int_A \exp\left(\int_0^t A_s(x, X_0)^T dW_s - \frac{1}{2}\int_0^t \|A_s(x, X_0)\|^2 ds\right)\mu(dx)}{\int_S \exp\left(\int_0^t A_s(x, X_0)^T dW_s - \frac{1}{2}\int_0^t \|A_s(x, X_0)\|^2 ds\right)\mu(dx)},$$

where, $A_s(x, X_0) \doteq [h(s, \phi_s(x)) - h(s, \phi_s(X_0))]$. Define the set

$$Q_N(r, x) \doteq \{y \in S : d_N(x, y) < r\},$$

for $r > 0$, and $N \doteq \lfloor \frac{t}{\tau} \rfloor$

We now consider,

$$\begin{aligned}
\hat{\pi}_t^{\mu}(B_a(X_0)^c) &= \frac{\int_{B_a(X_0)^c} \exp\left(\int_0^t A_s(x, X_0)^T dW_s - \frac{1}{2}\int_0^t \|A_s(x, X_0)\|^2 ds\right)\mu(dx)}{\int_S \exp\left(\int_0^t A_s(x, X_0)^T dW_s - \frac{1}{2}\int_0^t \|A_s(x, X_0)\|^2 ds\right)\mu(dx)} \\
&\le \frac{\int_{B_a(X_0)^c} \exp\left(\int_0^t A_s(x, X_0)^T dW_s - \frac{1}{2} D_N(x, X_0)^2\right)\mu(dx)}{\int_{Q_N(r, X_0)} \exp\left(\int_0^t A_s(x, X_0)^T dW_s - \frac{R}{2} D_{N+1}(x, X_0)^2\right)\mu(dx)} \quad \text{using (5.2)} \\
&\le \frac{\int_{B_a(X_0)^c} \exp\left(-D_N(x, X_0)^2\left(-\sup_{x \in B_a(X_0)^c} \frac{\left|\int_0^t A_s(x, X_0)^T dW_s\right|}{D_N(x, X_0)^2} + \frac{1}{2}\right)\right)\mu(dx)}{\int_{Q_N(r, X_0)} \exp\left(\int_0^t A_s(x, X_0)^T dW_s - \frac{R}{2} D_{N+1}(x, X_0)^2\right)\mu(dx)}, \qquad (5.3)
\end{aligned}$$

where we used the fact that

$$\frac{\left|\int_0^t A_s(x, X_0)^T dW_s\right|}{D_N(x, X_0)^2} \le \sup_{x \in B_a(X_0)^c} \frac{\left|\int_0^t A_s(x, X_0)^T dW_s\right|}{D_N(x, X_0)^2}$$

and $D_N(x,X_0)^2 = \sum_{i=0}^N \rho_{i\tau} d\left(\phi_{i\tau}(x), \phi_{i\tau}(X_0)\right)^2$. From (5.3), it is clear that in order to establish our desired result, it is sufficient to find suitable estimates on

$$\sup_{x\in B_a(X_0)^c} \frac{\left|\int_0^t A_s(x,X_0)^T dW_s\right|}{\sum_{i=0}^N \rho_{i\tau} d\left(\phi_{i\tau}(x), \phi_{i\tau}(X_0)\right)^2} \quad \text{and} \quad \sup_{x\in Q_N(r,X_0)} \left|\int_0^t A_s(x,X_0)^T dW_s\right|.$$

These bounds are stated in Lemmas (5.2.9)–(5.2.10).

Lemma 5.2.9. *$\forall a > 0$ and $\forall t \geq \tau$ with $N = \lfloor \frac{t}{\tau} \rfloor$, we have*

$$\mathbb{E}\left[\sup_{B_a(X_0)} \left|\int_0^t A_s(x,X_0)^T dW_s\right|\right] \leq 48K(((N+2)R\rho_{(N+1)\tau})(p(N+2)\log C + \log(qb^p)))^{\frac{1}{2}}$$
$$+ 96\sqrt{Kp}\left((N+2)\rho_{(N+1)\tau}R\right)^{\frac{1}{4}}, \tag{5.4}$$

with b, q being the constants from Lemma (A.0.27), while $C = C(\tau), K, R$ are from assumptions (5.2.3)–(5.2.2).

Proof. Since X_0 and W_t are independent,

$$\mathbb{E}\left[\sup_{B_a(X_0)} \left|\int_0^t A_s(x,X_0)^T dW_s\right|\right] \leq 2\mathbb{E}_{X_0}\mathbb{E}_W\left[\sup_{B_a(X_0)} \int_0^t A_s(x,X_0)^T dW_s\right], \tag{5.5}$$

where, $\mathbb{E}_{X_0}$ and $\mathbb{E}_W$ are, respectively, expectations over distribution of X_0 and Wiener measure corresponding to W_t. Observing that $\int_0^t A_s(x,X_0)^T dW_s$ is a centered Gaussian process, we use the following result (see Theorem (A.0.24))

$$\mathbb{E}_W\left[\sup_{B_a(X_0)} \int_0^t A_s(x,X_0)^T dW_s\right] \leq 24\int_0^\infty \log^{\frac{1}{2}}\left(N\left(B_a(X_0), \bar{d}_t, \epsilon\right)\right) d\epsilon\,, \tag{5.6}$$

where, $N\left(B_a(X_0), \bar{d}_t, \epsilon\right)$ is the minimum number of balls of radius ϵ under the pseudo-metric $\bar{d}_t$ required to cover $B_a(X_0)$ (which is finite for all ϵ due to the compactness of $B_a(X_0)$), where,

$$\bar{d}_t(x,y) \doteq \sqrt{\mathbb{E}_W\left[\left(\int_0^t A_s(x,X_0)^T dW_s - \int_0^t A_s(y,X_0)^T dW_s\right)^2\right]}$$
$$= \sqrt{\int_0^t \|h(s,\phi_s(x)) - h(s,\phi_s(y))\|^2 ds}\,.$$

From (5.2), It is clear that,

$$\bar{d}_t(x,y) \leq \sqrt{R} D_{N+1}(x,y) \leq \sqrt{(N+2)R\rho_{(N+1)\tau}} d_{N+2}(x,y)\,,$$

which implies that

$$N\left(B_a(X_0), \bar{d}_t, \epsilon\right) \leq N(B_a(X_0), \sqrt{R}D_{N+1}, \epsilon) \leq N(B_a(X_0), \sqrt{(N+2)R\rho_{(N+1)\tau}} d_{N+2}, \epsilon)$$

Denoting $\bar{\epsilon}(a,N) \doteq \sqrt{(N+2)R\rho_{(N+1)\tau}} \sup_{x,y \in B_a(X_0)} d_{N+2}(x,y)$, we get the following bound:

$$\begin{aligned}
&\int_0^\infty \log^{\frac{1}{2}}\left(N\left(B_a(X_0), \bar{d}_t, \epsilon\right)\right) d\epsilon \\
&\leq \int_0^{\bar{\epsilon}(a,N)} \log^{\frac{1}{2}}\left(N\left(B_a(X_0), \sqrt{(N+2)R\rho_{(N+1)\tau}} d_{N+2}, \epsilon\right)\right) d\epsilon \\
&= \int_0^{\bar{\epsilon}(a,N)} \log^{\frac{1}{2}}\left(N\left(B_a(X_0), d_{N+2}, \epsilon\left(\sqrt{(N+2)R\rho_{(N+1)\tau}}\right)^{-1}\right)\right) d\epsilon \\
&= \sqrt{(N+2)R\rho_{(N+1)\tau}} \int_0^{\frac{\bar{\epsilon}(a,N)}{\sqrt{(N+2)R\rho_{(N+1)\tau}}}} \log^{\frac{1}{2}}\left(N\left(B_a(X_0), d_{N+2}, \beta\right)\right) d\beta\,.
\end{aligned} \tag{5.7}$$

Note that $N\left(B_a(X_0), d_{N+2}, \beta\right) = r_{N+2}\left(B_a(X_0), \beta, \phi_\tau\right)$, introduced in definition (5.2.7). Hence we will use the bound from lemma (A.0.27) with the choice

$$\mathcal{K} \doteq \overline{\cup_{y \in U}\{x : d(x,y) \leq a\}}\,. \tag{5.8}$$

This is a bound of $r_{N+2}\left(B_a(X_0), \beta, \phi_\tau\right)$ that is valid for all the realisations of X_0. Therefore, all the constants in equations below are also independent of realisations of X_0. Returning to (5.7) and using Lemma (A.0.27) for the chosen $\mathcal{K}$ along with Assumption (5.2.2), we get

$$\begin{aligned}
&\sqrt{(N+2)\rho_{(N+1)\tau}R} \int_0^{\frac{\bar{\epsilon}(a,N)}{\sqrt{(N+2)\rho_{(N+1)\tau}R}}} \log^{\frac{1}{2}}\left(N\left(B_a(X_0), d_{N+2}, \beta\right)\right) d\beta \\
&= \sqrt{(N+2)\rho_{(N+1)\tau}R} \int_0^{\frac{\bar{\epsilon}(a,N)}{\sqrt{(N+2)\rho_{(N+1)\tau}R}}} \log^{\frac{1}{2}}\left(r_{N+2}\left(\beta, B_a(X_0), \phi_\tau\right)\right) d\beta \\
&\leq \sqrt{(N+2)\rho_{(N+1)\tau}R} \int_0^{\frac{\bar{\epsilon}(a,N)}{\sqrt{(N+2)\rho_{(N+1)\tau}R}}} \log^{\frac{1}{2}}\left(q\left(C^{N+2}b\beta^{-1}\right)^p\right) d\beta \\
&\leq \sqrt{(N+2)\rho_{(N+1)\tau}R} \int_0^{\frac{\bar{\epsilon}(a,N)}{\sqrt{(N+2)\rho_{(N+1)\tau}R}}} \log^{\frac{1}{2}}\left(q(C^{N+2}b)^p\left(\beta^{-1}+1\right)^p\right) d\beta \\
&\leq \sqrt{(N+2)\rho_{(N+1)\tau}R} \int_0^{\frac{\bar{\epsilon}(a,N)}{\sqrt{(N+2)\rho_{(N+1)\tau}R}}} \sqrt{\left(\log\left(q(C^{N+2}b)^p\right) + p\log\left(\beta^{-1}+1\right)\right)} d\beta
\end{aligned}$$

$$\leq \sqrt{(N+2)\rho_{(N+1)\tau}R}\int_0^{\frac{\bar{\epsilon}(a,N)}{\sqrt{(N+2)\rho_{(N+1)\tau}R}}}\left(\sqrt{\log\left(q(C^{N+2}b)^p\right)}+\sqrt{p\log\left(\beta^{-1}+1\right)}\right)d\beta$$

$$\leq \bar{\epsilon}(a,N)\sqrt{p\,(N+2)\log\left(C\right)+\log\left(qb^p\right)}+2\sqrt{\bar{\epsilon}(a,N)p}\left(\sqrt{(N+2)\rho_{(N+1)\tau}R}\right)^{\frac{1}{4}}.$$

Here, we used the inequality: $\log(1+\frac{1}{x}) \leq \frac{1}{\sqrt{x}}$ and integrated. From the definition, $\bar{\epsilon}(a,N) \leq \sqrt{(N+2)R\rho_{(N+1)\tau}}K$. Combining the inequalities (5.5), (5.6), and (5.7) with the above inequality gives (5.4), completing the proof of the lemma. ■

As noted earlier, we also need to have estimate on $\sup_{x\in B_a(X_0)^c}\frac{\left|\int_0^t A_s(x,X_0)^T dW_s\right|}{D_N(x,X_0)^2}$ which is given by the lemma below.

Lemma 5.2.10. *$\forall a>0$, $\forall t \geq \tau$ and $N=\lfloor\frac{t}{\tau}\rfloor$, there exists G_a depending only on a such that*

$$\mathbb{E}\left[\sup_{x\in B_a(X_0)^c}\frac{\left|\int_0^t A_s(x,X_0)^T dW_s\right|}{D_N(x,X_0)^2}\right] \leq \frac{S(N)}{\sum_{i=0}^N \rho_{i\tau}}G_a\,,$$

where,

$$S(N)=48K\sqrt{\left((N+2)R\rho_{(N+1)\tau}\right)\left(p\,(N+2)\log\left(C\right)+\log\left(qb^p\right)\right)}$$
$$+96\sqrt{Kp}\left((N+2)\rho_{(N+1)\tau}R\right)^{\frac{1}{4}}.$$

Proof. Consider a sequence $\{a_k\}_{k\in\mathbb{Z}}$ such that $a_k \to 0$ as $k\to -\infty$ and $a_k \to \infty$ as $k\to\infty$. Let k_0 be the largest integer such that $a_{k_0} \leq a$. From the Assumption (5.2.1), there a $k_1 \in \mathbb{Z}$ such that $\forall x, X_0 \in supp(\mu)$, we have

$$a_{k_0} \leq d(x,X_0) \leq a_{k_1},$$

Using this notation, we obtain the required bound as follows:

$$\mathbb{E}\left[\sup_{x\in B_a(X_0)^c}\frac{\left|\int_0^t A_s(x,X_0)^T dW_s\right|}{D_N(x,X_0)^2}\right] \leq \mathbb{E}\left[\sup_{\{x:d(x,X_0)\geq a_{k_0}\}}\frac{\left|\int_0^t A_s(x,X_0)^T dW_s\right|}{\sum_{i=0}^N \rho_{i\tau} d\left(\phi_{i\tau}(x),\phi_{i\tau}(X_0)\right)^2}\right]$$
$$\leq \sum_{k_1\geq k\geq k_0}\mathbb{E}_{X_0}\left[\mathbb{E}_W\left[\sup_{\{x:a_k\leq d(x,X_0)\leq a_{k+1}\}}\frac{\left|\int_0^t A_s(x,X_0)^T dW_s\right|}{\sum_{i=0}^N \rho_{i\tau} d\left(\phi_{i\tau}(x),\phi_{i\tau}(X_0)\right)^2}\right]\right]$$
$$\leq \sum_{k_1\geq k\geq k_0}\mathbb{E}_{X_0}\left[\frac{1}{L^2(a_k)\sum_{i=0}^N \rho_{i\tau}}\mathbb{E}_W\left[\sup_{\{x:a_k\leq d(x,X_0)\leq a_{k+1}\}}\left|\int_0^t A_s(x,X_0)^T dW_s\right|\right]\right]$$

$$
\begin{aligned}
&\le \sum_{k_1 \ge k \ge k_0} \mathbb{E}_{X_0}\left[\frac{1}{L^2(a_k)\sum_{i=0}^N \rho_{i\tau}}\mathbb{E}_W\left[\sup_{\{x:d(x,X_0)\le a_{k+1}\}}\left|\int_0^t A_s(x,X_0)^T dW_s\right|\right]\right] \\
&\le \sum_{k_1 \ge k \ge k_0} \mathbb{E}_{X_0}\left[\frac{1}{L^2(a_k)}\right]\frac{1}{\sum_{i=0}^N \rho_{i\tau}} S(N)K \\
&\le \frac{S(N)}{\sum_{i=0}^N \rho_{i\tau}} \sum_{k_1 \ge k \ge k_0} \frac{1}{L^2(a_k)} K.
\end{aligned} \tag{5.9}
$$

This completes the proof of the lemma. ■

Using the fact that $t \to \infty \Leftrightarrow N \to \infty$ and $\frac{N\rho_{N\tau}}{\sum_{i=0}^N \rho_{i\tau}} \le C'$, it can be seen that

$$
\lim_{t\to\infty} \mathbb{E}\left[\sup_{x\in B_a(X_0)^c} \frac{\left|\int_0^t A_s(x,X_0)^T dW_s\right|}{\sum_{i=0}^N \rho_{i\tau} d\left(\phi_{i\tau}(x),\phi_{i\tau}(X_0)\right)^2}\right] = 0\,. \tag{5.10}
$$

Finally, we need the lemma below to complete the proof of Theorem (5.2.8).

Lemma 5.2.11. *$\forall a > 0, \exists \alpha = \alpha(a) > 0$ such that $\lim_{t\to\infty} e^{\alpha t}\hat{\pi}_t^\mu\left(B_a(X_0)^c\right) = 0,$ a.s.*

Proof. From (5.10), we have

$$
\lim_{t\to\infty} \sup_{x\in B_a(X_0)^c} \frac{\left|\int_0^t A_s(x,X_0)^T dW_s\right|}{\sum_{i=0}^N \rho_{i\tau} d\left(\phi_{i\tau}(x),\phi_{i\tau}(X_0)\right)^2} = 0, \text{ w.p.1}
$$

Recall that $t \to \infty \Leftrightarrow N \to \infty$. In particular, the above equation holds for any subsequence $\{t_j\}$. Therefore, there is sub-subsequence $\{t_{j_q}\}$ such that

$$
\lim_{q\to\infty} \sup_{x\in B_a(X_0)^c} \frac{\left|\int_0^{t_{j_q}} A_s(x,X_0)^T dW_s\right|}{\sum_{i=0}^{N(j,q)} \rho_{i\tau} d\left(\phi_{i\tau}(x),\phi_{i\tau}(X_0)\right)^2} = 0, \text{ a.s,}
$$

where $N(j,q) \doteq \lfloor \frac{t_{j_q}}{\tau} \rfloor$.

From the above, for large enough q, we have

$$
\sup_{x\in B_a(X_0)^c} \frac{\left|\int_0^{t_{j_q}} A_s(x,X_0)^T dW_s\right|}{\sum_{i=0}^{N(j,q)} \rho_{i\tau} d\left(\phi_{i\tau}(x),\phi_{i\tau}(X_0)\right)^2} < \frac{1}{4}
$$

and thereby,

$$
\int_{B_a(X_0)^c} \exp\left(-D_{N(j,q)}(x,X_0)^2\left(-\sup_{x\in B_a(X_0)^c}\frac{\left|\int_0^{t_{j_q}} A_s(x,X_0)^T dW_s\right|}{D_{N(j,q)}(x,X_0)^2} + \frac{1}{2}\right)\right)\mu(dx)
$$

$$\leq \int_{B_a(X_0)^c} \exp\left(-D_{N(j,q)}(x, X_0)^2 \frac{1}{4}\right) \mu(dx)$$
$$\leq \exp\left(-\frac{L^2(a) \sum_{i=0}^{N(j,q)} \rho_{i\tau}}{4}\right). \tag{5.11}$$

Here, we used assumption (5.2.1) and the fact that $\mu\left(B_a(X_0)^c\right) \leq 1$. We now, consider

$$\int_{Q_N(r,X_0)} \exp\left(\int_0^t A_s(x, X_0)^T dW_s - \frac{1}{2}\int_0^t |A_s(x, X_0)|^2 ds\right) \mu(dx)$$
$$\geq \int_{Q_N(r,X_0)} \exp\left(\int_0^t A_s(x, X_0)^T dW_s - \frac{1}{2} R \sum_{i=0}^{N+1} \rho_{i\tau} d\left(\phi_{i\tau}(x), \phi_{i\tau}(X_0)\right)^2\right) \mu(dx)$$
$$\geq \int_{Q_N(r,X_0)} \exp\left(\int_0^t A_s(x, X_0)^T dW_s - \frac{1}{2} R d_{N+1}^2(x, X_0) \sum_{i=0}^{N+1} \rho_{i\tau}\right) \mu(dx)$$
$$\geq \int_{Q_N(r,X_0)} \exp\left(-\sum_{i=0}^{N+1} \rho_{i\tau}\left(-\frac{\int_0^t A_s(x, X_0)^T dW_s}{\sum_{i=0}^{N+1} \rho_{i\tau}} + \frac{R}{2} r\right)\right) \mu(dx), \tag{5.12}$$

In the last inequality, we used the definition of $Q_N(r, X_0)$. And also, from the definition of $Q_N(r, x)$, it is clear that $Q_N(r, x) \subset B_r(x) \subset B_a(X_0)$. Therefore,

$$\mathbb{E}\left[\sup_{Q_N(r,X_0)} \left|\int_0^t A_s(x, X_0)^T dW_s\right|\right] \leq \mathbb{E}\left[\sup_{B_r(X_0)} \left|\int_0^t A_s(x, X_0)^T dW_s\right|\right]$$
$$\leq \mathbb{E}\left[\sup_{B_a(X_0)} \left|\int_0^t A_s(x, X_0)^T dW_s\right|\right]$$

From Lemma (5.2.9), it follows that

$$\frac{1}{\sum_{i=0}^{N+1} \rho_{i\tau}} \mathbb{E}\left[\sup_{Q_N(r,X_0)} \left|\int_0^t A_s(x, X_0)^T dW_s\right|\right] \leq \frac{S(N) G_a}{\sum_{i=0}^{N+1} \rho_{i\tau}}$$

Again, since $\frac{N \rho_{N\tau}}{\sum_{i=0}^{N} \rho_{i\tau}} \leq C'$, it converges to zero as $t \to \infty$ which again implies that

$$\lim_{t\to\infty} \frac{\sup_{B_r(X_0)} \left|\int_0^t A_s(x, X_0)^T dW_s\right|}{\sum_{i=0}^{N+1} \rho_{i\tau}} = 0, \text{ w.p.1.}$$

In particular, it converges to zero in probability on subsequence t_j. Therefore, we can choose a sub-subsequence, $\{t_{j_q}\}$ (that works for the previous scenario) such that

$$\lim_{q\to\infty} \frac{\sup_{Q_{N(j,q)}(r,X_0)} \left|\int_0^{t_{j_q}} A_s(x,X_0)^T dW_s\right|}{\sum_{i=0}^{N(j,q)+1} \rho_{i\tau}} = 0, \quad \text{a.s.}$$

For large enough q,

$$\frac{\sup_{Q_{N(j,q)}(r,X_0)} |\int_0^{t_{j_q}} A_s(x,X_0)^T dW_s|}{\sum_{i=0}^{N(j,q)+1} \rho_{i\tau}} < \frac{Rr}{2}$$

Therefore, (5.12) becomes

$$\begin{aligned}
\int_{Q_{N(j,q)}(r,X_0)} & \exp\left(-\sum_{i=0}^{N(j,q)+1} \rho_{i\tau}\left(-\frac{\int_0^t A_s(x,X_0)^T dW_s}{\sum_{i=0}^{N(j,q)+1} \rho_{i\tau}} + \frac{Rr}{2}\right)\right)\mu(dx) \\
&\geq \int_{Q_{N(j,q)}(r,X_0)} \exp\left(-\sum_{i=0}^{N(j,q)+1} \rho_{i\tau} Rr\right)\mu(dx) \\
&\geq \exp\left(-\sum_{i=0}^{N(j,q)+1} \rho_{i\tau} Rr\right)\mu(Q_{N(j,q)}(r,X_0)) \qquad (5.13)
\end{aligned}$$

Combining inequalities (5.13) and (5.11), we have

$$\hat{\pi}^{\mu}_{t_{j_q}}(B_a(X_0)^c) \leq \frac{\exp\left(-\sum_{i=0}^{N(j,q)} \rho_{i\tau}\left(\frac{L^2(a)}{4} - Rr\right) + \rho_{\tau(N(j,q)+1)} Rr\right)}{\mu\left(Q_{N(j,q)}(r,X_0)\right)} \qquad (5.14)$$

As mentioned in Section (5.2.2), in general, the set $Q_n(r,X_0)$ will shrink to a set containing X_0 (which is not open) as $n \to \infty$. This is because of sensitive dependence on the initial conditions. We will see that $\mu\left(Q_{N_{j_q}}(r,X_0)\right)$ goes to zero at most at an exponential rate.

From the assumption of absolute continuity of μ with respect to σ , we have $\frac{d\mu}{d\sigma}(X_0) > 0$ $\mu-$a.s. From the continuity of $\frac{d\mu}{d\sigma}$, there exist $r_1 > 0$ and $C_1 > 0$ such that $\frac{d\mu}{d\sigma}(x) > C_1$, for any $x \in B_{r_1}(X_0)$. Therefore, with the help of Radon-Nikodym Theorem and choosing $r < r_1$, we have

$$\mu\left(Q_{N(j,q)}(r,X_0)\right) > C_1\sigma\left(Q_{N(j,q)}(r,X_0)\right). \qquad (5.15)$$

From the Assumption (5.2.2), we have the following:

$$\begin{aligned} d_N(x,y) &\leq C^N d(x,y) \\ \implies B_{\frac{r}{C^N}}(X_0) &\subset Q_N(r, X_0) \end{aligned}$$

(5.15) becomes

$$\begin{aligned} \mu\left(Q_{N(j,q)}(r, X_0)\right) > C_1\sigma\left(Q_{N(j,q)}(r, X_0)\right) > C_1\sigma\left(B_{\frac{r}{C^{N(j,q)}}}(X_0)\right) \\ > C_1C_2\left(\frac{r}{C^{N(j,q)}}\right)^p, \end{aligned} \tag{5.16}$$

for some $C_2 = C_2(p, \mathcal{K})$ (with $\mathcal{K}$ defined in (5.8)) and (5.14) becomes

$$\begin{aligned} &\hat{\pi}^{\mu}_{t_{j_q}}(B_a(X_0)^c) \\ &\leq \frac{1}{C_1C_2r^p}\exp\left(-\sum_{i=0}^{N(j,q)}\rho_{i\tau}\left(\left(\frac{L^2(a)}{4} - Rr\right) - \frac{\rho_{\tau(N(j,q)+1)}Rr}{\sum_{i=0}^{N(j,q)}\rho_{i\tau}} - \frac{N(j,q)\log_e C}{\sum_{i=0}^{N(j,q)}\rho_{i\tau}}\right)\right) \end{aligned} \tag{5.17}$$

Choosing r small enough such that $\frac{L^2(a)}{4} - Rr > 0$ and from Assumption (5.2.3) ($\lim_{t\to\infty}\rho_t = \infty$ and $\lim_{t\to\infty}\frac{\int_0^t \rho_s ds}{\rho_t} = \infty$), for large enough q, the exponent can be made positive which results in $\hat{\pi}^{\mu}_{t_{j_q}}(B_a(X_0)^c)$ converging exponentially to zero almost surely as $q \to \infty$. Since, the subsequence t_j is arbitrary, it implies that $\hat{\pi}^{\mu}_t(B_a(X_0)^c)$ converges exponentially to zero almost surely as $t \to \infty$. ■

From Lemma (5.2.11), it is clear that the assertion of the Theorem (5.2.8) follows. ■

Remark 5.2.12. *Note that until (5.17), all the computations go through if we assumed ρ_t is constant. We used the assumed behavior of ρ_t only in (5.17). Let us assume, for the sake of this remark, that $\rho_t = \rho$. Then for the exponent in (5.17) to be negative, we require that $L(a)$ is sufficiently large compared to the other terms viz., Rr and $\log_e C$. Even though Rr can be made small (by choosing r small, as r can be chosen freely) and $\log_e C$ is known beforehand, the estimates on $L(a)$ are difficult to come by for a dynamical system. In fact, even showing existence of $L(a)$ is seen to be difficult. This is the reason behind the assumption on the behavior of ρ_t.*

In the previous theorem, we established that conditional distribution of X_0 given observations is asymptotically supported only on closed balls around X_0 of arbitrary radius. In the following, we extend the previous statement to any measurable set, $A \in \mathcal{B}(S)$.

Proposition 5.2.13. *Under the hypothesis of Theorem* (5.2.8), $\lim_{t\to\infty} \hat{\pi}_t^{\mu}(A) = 0, \quad \forall A \in \mathcal{B}(S), \quad X_0 \notin A$

Proof. It can be seen easily that the conclusion of the theorem holds even if $d(x, X_0) \leq a$ is replaced with $d(x, X_0) < a$. Indeed, as $B_{a-\rho}(X_0) \subset \{x \in S : d(x, X_0) < a\} \subset B_a(X_0)$ holds for $\rho < a$ and $\forall \gamma > 0, \quad \forall t > 0$, we have

$$\begin{aligned} e^{\gamma t}\left(\hat{\pi}_t^{\mu}\left(B_{a-\rho}(X_0)\right) - 1\right) &\leq e^{\gamma t}\left(\hat{\pi}_t^{\mu}\left(\{x \in S : d(x, X_0) < a\}\right) - 1\right) \\ &\leq e^{\gamma t}\left(\hat{\pi}_t^{\mu}\left(B_a(X_0)\right) - 1\right), \end{aligned}$$

We can clearly see that ρ can be chosen small enough such that there exists $\gamma > 0$ such that

$$\lim_{t\to\infty} e^{\gamma t}(\hat{\pi}_t^{\mu}(B_{a-\rho}(X_0)) - 1) = 0, \quad \textit{and} \quad \lim_{t\to\infty} e^{\gamma t}(\hat{\pi}_t^{\mu}(B_a(X_0)) - 1) = 0$$

Indeed, the desired value of γ is minimum of the $\alpha(a)$ and $\alpha(a-\rho)$. Therefore,

$$\lim_{t\to\infty} e^{\gamma t}\left(\hat{\pi}_t^{\mu}\left(\{x \in S : d(x, X_0) < a\}\right) - 1\right) = 0 \ \text{ a.s}, \quad \forall a > 0,$$

Writing the above in a concrete way, we have $\forall b > 0, \ z \in S$,

$$\begin{aligned} \lim_{t\to\infty} \hat{\pi}_t^{\mu}\left(\{x \in S : d(z, x) < b\}\right) &= 1 \text{ a.s}, \ : \ d(X_0, z) \leq b \\ &= 0 \text{ a.s}, \ : \ d(X_0, z) > b \end{aligned}$$

Extending this to all open sets, we have for any open U

$$\begin{aligned} \lim_{t\to\infty} \hat{\pi}_t^{\mu}(U) &= 1 \text{ a.s}, \ : \ X_0 \in U \\ &= 0 \text{ a.s}, \ : \ X_0 \notin U \end{aligned}$$

This can be done since open balls form a base of the usual topology of S. And also, for any closed set C

$$\begin{aligned} \lim_{t\to\infty} \hat{\pi}_t^{\mu}(C) &= 1 \text{ a.s}, \ : \ X_0 \in C \\ &= 0 \text{ a.s}, \ : \ X_0 \notin C \end{aligned}$$

Finally, to extend it to all measurable sets, we use the property of regular probability measure with Borel σ-algebra of a metric space [25][Theorem 1.1].

By [25, Theorem 1.1], for every measurable set $A \in \mathcal{B}(S)$, there exist closed set C_0, open set U_0 such that $C_0 \subset A \subset U_0$ and $\hat{\pi}_t^{\mu}(U_0/C_0) < \frac{1}{2}$.

Let A be such that $X_0 \in A$ which implies that $X_0 \in U_0$. Choose $0 < \eta < \frac{1}{4}$ and t large enough such that $\hat{\pi}_t^{\mu}(U_0) > 1 - \eta$. Considering C_0, if $X_0 \notin C_0$ then again by choosing t large enough, we have $\hat{\pi}_t^{\mu}(C_0) < \eta$. But this is a contradiction. Indeed, as $\hat{\pi}_t^{\mu}(U_0) = \hat{\pi}_t^{\mu}(C_0) + \hat{\pi}_t^{\mu}(U_0/C_0)$ and $\hat{\pi}_t^{\mu}(U_0) < \eta + \frac{1}{2} < 1 - \eta$. Therefore, $X_0 \in C_0$.

This implies that $\lim_{t\to\infty} \hat{\pi}_t^{\mu}(A) = 0$ ■

5.2.4 Stability of the filter

We now state and prove the filter stability.

Theorem 5.2.14. *Under the hypothesis of Theorem (5.2.8), If $\mu \sim \nu$ then for any bounded continuous $g : S \to \mathbb{R}$,*

$$\lim_{t\to\infty} \mathbb{E}\left[\left|\pi_t^{\mu}(g) - \pi_t^{\nu}(g)\right|\right] = 0$$

.

Proof. From the Proposition (5.2.13), for any measurable $A \in \mathcal{B}(S)$

$$\begin{aligned} \hat{\pi}_{\infty}^{\mu}(A) \doteq \lim_{t\to\infty} \hat{\pi}_t^{\mu}(A) &= 1 \text{ a.s,} \ : \ X_0 \in A \\ &= 0 \text{ a.s,} \ : \ X_0 \notin A \end{aligned}$$

This is by definition the Dirac measure at X_0. Therefore, for any integrable function $f : S \to \mathbb{R}$, $\mathbb{E}\left[f(X_0)|\mathcal{Y}_{\infty}\right] = f(X_0)$.

Suppose $J \doteq \frac{d\nu}{d\mu}$ and $\sup\limits_{x\in S} |g(x)| < M$. From the assumption on μ and ν, we can express $\pi_t^{\nu}(g)$ as

$$\begin{aligned} \pi_t^{\nu}(g) &= \frac{\int_S g\left(\phi_t(x)\right) Z\left(t, x, Y_{[0,t]}\right) \nu(dx)}{\int_S Z\left(t, x, Y_{[0,t]}\right) \nu(dx)} \\ &= \frac{\int_S g\left(\phi_t(x)\right) Z\left(t, x, Y_{[0,t]}\right) J(x)\mu(dx)}{\int_S Z\left(t, x, Y_{[0,t]}\right) J(x)\mu(dx)} \frac{\int_S Z\left(t, x, Y_{[0,t]}\right) \mu(dx)}{\int_S Z\left(t, x, Y_{[0,t]}\right) \mu(dx)} \\ &= \frac{\mathbb{E}\left[g(\phi_t(X_0)J(X_0)|\mathcal{Y}_t\right]}{\mathbb{E}\left[J(X_0)|\mathcal{Y}_t\right]} \end{aligned}$$

$$\mathbb{E}\left[\left|\pi_t^{\mu}(g) - \pi_t^{\nu}(g)\right|\right] = \mathbb{E}\left[\frac{\left|\mathbb{E}\left[g\left(\phi_t(X_0)\right)\left(\mathbb{E}\left[J(X_0)|\mathcal{Y}_t\right] - J(X_0)\right)|\mathcal{Y}_t\right]\right|}{\mathbb{E}\left[J(X_0)|\mathcal{Y}_t\right]}\right]$$

$$\leq \mathbb{E}\left[\frac{\mathbb{E}\left[|g\left(\phi_t(X_0)\right)\left(\mathbb{E}\left[J(X_0)|\mathcal{Y}_t\right]-J(X_0)\right)|\,|\mathcal{Y}_t\right]}{\mathbb{E}\left[J(X_0)||\mathcal{Y}_t\right]}\right]$$

$$\begin{aligned}
&\leq M\mathbb{E}\left[\frac{\mathbb{E}\left[|\mathbb{E}\left[J(X_0)|\mathcal{Y}_t\right]-J(X_0)|\,|\mathcal{Y}_t\right]}{\mathbb{E}\left[J(X_0)|\mathcal{Y}_t\right]}\right] \\
&\leq M\mathbb{E}\left[\frac{|\mathbb{E}\left[J(X_0)|\mathcal{Y}_t\right]-J(X_0)|}{\mathbb{E}\left[J(X_0)|\mathcal{Y}_t\right]}\right]
\end{aligned} \tag{5.18}$$

Due to integrability of J, martingale convergence theorem implies

$$\lim_{t\to\infty}\mathbb{E}\left[J(X_0)|\mathcal{Y}_t\right]=\mathbb{E}\left[J(X_0)|\mathcal{Y}_\infty\right]=J(X_0)\ \text{ a.s}$$

Choose a subsequence $t_n \uparrow \infty$. Apply the Lemma (A.0.6) for $f_n \doteq \frac{J(X_0)}{\mathbb{E}[J(X_0)|\mathcal{Y}_{t_n}]}$ (Note that $J(X_0) > 0$ $\mathbb{P}$- a.s) and $f \doteq 1$, to get the desired result. ■

Remark 5.2.15. *We show below that Theorem (5.2.8), Proposition (5.2.13) and Theorem (5.2.14) together imply that Assumptions (5.2.3), (5.2.1) and (5.2.2) together form a sufficient condition for the notion of observability defined in [139, Definition 2]. Since* $\mathbb{E}\left[f(X_0)|\mathcal{Y}_\infty\right] = f(X_0)$ *for any integrable function* $f : S \to \mathbb{R}$, X_0 *is measurable with respect to* $\mathcal{Y}_\infty$. *It implies that there exists a function* F, *that is measurable with respect to* $\mathcal{Y}_\infty$ *such that* $F : C([0,\infty), \mathbb{R}^n) \to X$ *and* $X_0 = F\left(Y_{[0,\infty)}\right)$. *Therefore, we arrive at the conclusion that law of observation process determines the law of* X_0 *uniquely which is exactly the definition of observability in [139].*

5.3 Discrete time nonlinear filter

In this section, we study the stability of the nonlinear filter in discrete time setting. We will setup the discrete time filter in the form where the filter at any time instant depends on the entire observation sequence up to that instant, which is unlike the recursive form of the filter that is useful in applications.

5.3.1 Setup

Again let the state space, S be p-dimensional complete Riemannian manifold with metric d. On S, we have a homeomorphism $T : S \to S$ along with initial condition X_0, whose distribution is μ. We denote discrete time with k. These dynamics are observed partially in the following way.

$$Y_k = \sum_{i=1}^{k} h\left(i, T^i(X_0)\right) + W_k,$$

where, $h : \mathbb{Z}^+ \times S \to \mathbb{R}^n$ and $Y_k \in \mathbb{R}^n$ is the observation process and $W_k \in \mathbb{R}^n$ is the position of an i.i.d random walk with standard Gaussian increment after k steps, starting at origin. Moreover, X_0 and $W_{k+1} - W_k$ are assumed to be independent for any $k \geq 1$. Therefore,

$$\left\{ S \times (\mathbb{R}^n)^{\mathbb{Z}^+}, \mathcal{B}\left(S \times \mathbb{R}^n)^{\mathbb{Z}^+} \right), \mathbb{P} = \mu \times \mathbb{P}_W \right\}$$

is considered to be our probability space. Here, $\mathbb{P}_W$ is the probability measure of W. Let $\mathcal{Y}_k = \sigma \{Y_i : 0 \leq s \leq k,\ i \in \mathbb{Z}^+\}$, the observation process filtration. We shall see that the results of stability for the case of continuous time extend to the discrete time case with very minor changes. Noting this, we denote all the quantities that appear in both continuous and discrete time cases by same symbols.

$\hat{\pi}_k^\mu$, π_k^μ, ν and π_k^ν have similar meanings to what they mean in continuous time case. Define,

$$Z(k, x, Y_{0:k}) \doteq \exp \left(\sum_{i=1}^{k} h\left(i, T^i(x)\right)^T (Y_i - Y_{i-1}) - \frac{1}{2} \sum_{i=1}^{k} \left\| h\left(i, T^i(x)\right) \right\|^2 \right),$$

with the convention that $\sum_1^0 \doteq 0$. From Bayes' rule, for any bounded continuous function g,

$$\hat{\pi}_k^\mu(g) = \mathbb{E}\left[g(X_0)|\mathcal{Y}_k\right] = \frac{\int_S g(x) Z(k, x, Y_{0:k}) \mu(dx)}{\int_S Z(k, x, Y_{0:k}) \mu(dx)} \tag{5.19}$$

For a fixed k, the filter is given by

$$\pi_k^\mu(g) = \mathbb{E}\left[g(T^k(X_0))|\mathcal{Y}_k\right] = \frac{\int_S g(T^k(x)) Z(k, x, Y_{0:k}) \mu(dx)}{\int_S Z(k, x, Y_{0:k}) \mu(dx)}$$

The expression for the corresponding incorrect filter is given by

$$\pi_k^\mu(g) = \frac{\int_S g\left(T^k(x)\right) Z(k, x, Y_{0:k}) \nu(dx)}{\int_S Z(k, x, Y_{0:k}) \nu(dx)} \tag{5.20}$$

5.3.2 Stability of the filter

We show that the above mentioned filter is stable in the following sense:

$$\lim_{k \to \infty} \mathbb{E}\left[\left|\pi_k^\mu(g) - \pi_k^\nu(g)\right|\right] = 0,$$

for a bounded continuous $g : S \to \mathbb{R}$.

To establish the above, we need a discrete analog of Theorem (5.2.8). This can be done under the following discrete analogs of Assumptions (5.2.3), (5.2.1), (5.2.2). Again note that we use same symbols for the quantities that appear in both the cases.

Assumption 5.3.1. *There exists a bounded open set U such that $\overline{TU} \subset U$.*

Assumption 5.3.2. *$\forall x,\ y \in U$, we have $d(Tx, Ty) \leq Cd(x, y)$, for some $C > 1$.*

Assumption 5.3.3. *There exists $\rho_k,\ R,\ k_0 > 0$ such that*

$$\forall k \geq 0,\ \rho_k d(x_1, x_2)^2 \leq \sum_{i=k}^{k+k_0} \left\| h\left(i, T^{i-k}(x_1)\right) - h\left(i, T^{i-k}(x_2)\right) \right\|^2 \leq R\rho_k d(x_1, x_2)^2, \tag{5.21}$$

where, ρ_k is a positive non-decreasing function such that $\forall x_1, x_2 \in U$ $\lim_{k\to\infty} \frac{\sum_{i=0}^{k} \rho_i}{\rho_k} = \infty$, $\frac{N\rho_N}{\sum_{i=0}^{k} \rho_i} \leq C'$ (for some $C' > 0$) and $R > 1$.

Assumption 5.3.4. *For $(x, y) \in \mathcal{V} \subset U \times U$, compliment (in $U \times U$) of a zero Lebesgue measure set satisfying $d(x, y) \geq b > 0$, the following holds*

$$D_N^2(x, y) \geq L^2(b) \sum_{i=0}^{N} \rho_{i\tau},$$

where, $L(b)$ is a positive constant.

Assumption 5.3.5. *$supp(\mu) \subset U$*

Remark 5.3.6. *From the Assumption (5.3.1), it follows that for $x \in supp(\mu)$ and $y \in supp(\mu)$, $d_N(x, y) \leq K$.*

It follows from the Assumption (5.3.3) that

$$\begin{aligned}\sum_{i=1}^{N} \rho_{ik_0} d\left(T^i(x), T^i(y)\right)^2 &\leq \sum_{i=0}^{k} \left\| h\left(T^i(x)\right) - h\left(T^i(y)\right) \right\|^2 \\ &\leq R \sum_{i=1}^{N+1} \rho_{ik_0} d\left(T^i(x), T^i(y)\right)^2, \quad \forall x, y \in S, \end{aligned} \tag{5.22}$$

where, $N = \lfloor \frac{k}{k_0} \rfloor$.

Remark 5.3.7. *The significance of the above assumptions is exactly the same as that of the assumptions in Section (5.2).*

Now we state the discrete analogs of Theorem (5.2.8), Proposition (5.2.13) and Theorem (5.2.14).

Theorem 5.3.8. *Suppose μ is absolutely continuous with respect to volume, σ of S and $\frac{d\mu}{d\sigma}$ is continuous on the support of μ. Under the assumptions (5.3.1), (5.3.2), (5.3.3), (5.3.4) and (5.3.5),*

$$\lim_{k\to\infty} e^{\alpha k}\left(\hat{\pi}_k^{\mu}\left(\{x \in S : d(x, X_0) \leq a\}\right) - 1\right) = 0 \;\; a.s, \;\; \forall a > 0,$$

and for some $\alpha \doteq \alpha(a) > 0$ which depends only on a.

Proof. The proof of this theorem follows exactly in the same lines as that of Theorem (5.2.8). So the proof is omitted. ■

Proposition 5.3.9. *Under the hypothesis of Theorem (5.3.8),*

$$\lim_{k\to\infty} \hat{\pi}_k^{\mu}(A) = 0, \;\; \forall A \in \mathcal{B}(S), \;\; X_0 \notin A$$

Proof. We observe that the proof of Proposition (5.2.13) remains unchanged if the continuous time is replaced with discrete time. ■

Theorem 5.3.10. *Under the hypothesis of Theorem (5.2.8), If $\mu \sim \nu$ then for any bounded continuous $g : S \to \mathbb{R}$,*

$$\lim_{k\to\infty} \mathbb{E}\left[\left|\pi_k^{\mu}(g) - \pi_k^{\nu}(g)\right|\right] = 0$$

.

Proof. Proof is again omitted as it is exactly in the same lines as that of Theorem (5.2.14). ■

Remark 5.3.11. *Remarks analogous to (5.2.15) and the rest of the remarks of the previous section follow in the case of discrete time.*

5.4 Structure of the conditional distribution

In this section, we will see that the conditional distribution after large times is supported nearly on the topological attractor. Recall that topological attractor is defined as $\Lambda \doteq \cap_{n\geq 0} T^n U$, where U is an open set such that $\overline{TU} \subset U$ [87, Pg. 128].

Assumption 5.4.1. *Assume that there is an open set U such that $\overline{TU} \subset U$ and $\forall x \in S$, there exists $n(x) \geq 0$ such that $T^{n(x)}x \in U$.*

We restrict ourselves to the case of discrete time filter and adopt the notation of Section (5.3) in this entire section. Let μ be equivalent to volume.

From (5.19), for any $A \in \mathcal{B}(S)$, we have

$$\begin{aligned}\hat{\pi}_k^\mu(A) &= \mathbb{E}[\mathbb{1}_{\{X_0 \in A\}}|\mathcal{Y}_k] \\ &= \frac{\int_A Z(k,x,Y_{0:k})\mu(dx)}{\int_S Z(k,x,Y_{0:k})\mu(dx)} \\ &= \frac{\int_A \exp\left(\sum_{i=1}^k \Delta h(i,x,X_0)^T (W_i - W_{i-1}) - \frac{1}{2}\sum_{i=1}^k \|\Delta h(i,x,X_0)\|^2\right)\mu(dx)}{\int_S \exp\left(\sum_{i=1}^k \Delta h(i,x,X_0)^T (W_i - W_{i-1}) - \frac{1}{2}\sum_{i=1}^k \|\Delta h(i,x,X_0)\|^2\right)\mu(dx)},\end{aligned}$$

where, $\Delta h(i,x,X_0) \doteq h(i,T^i(x)) - h(i,T^i(X_0))$From (5.20), for any $A \in \mathcal{B}(S)$, we have

$$\begin{aligned}\pi_k^\mu(A) = \mathbb{E}\left[\mathbb{1}_{\{T^k(X_0)\in A\}}|\mathcal{Y}_k\right] &= \frac{\int_{\{T^k(x)\in A\}} Z(k,x,Y_{0:k})\mu(dx)}{\int_S Z(k,x,Y_{0:k})\mu(dx)} \\ &= \frac{\int_A Z(k,T^{-k}y,Y_{0:k})\mu\circ T^{-k}(dy)}{\int_S Z(k,T^{-k}y,Y_{0:k})\mu\circ T^{-k}(dy)}\end{aligned} \tag{5.23}$$

Therefore, support of π_k^μ is always contained in the support of $\mu \circ T^{-k}$. So, it is sufficient to show that the asymptotically the support of $\mu \circ T^{-k}$ is near the topological attractor to conclude that after large times, π_k^μ puts negligible mass far away from the topological attractor.

To that end, we define the following disjoint family of sets, $\{U_m\}_{\mathbb{Z}^+}$:

$$U_l^m \doteq \left\{x \in S : \inf\left\{k \in \mathbb{Z}^+ : T^k \in \Lambda_m\right\} = l\right\},$$

where, $\Lambda_m \doteq \cap_{n=0}^m T^n U$. From the assumption (5.4.1), for any given $m \geq 0$, it follows that

$$S = \cup_{l\geq 0} U_l^m$$

Now, for a given $m \geq 0$ and $k \geq m$, consider

$$\begin{aligned}\mu\circ T^{-k}(\Lambda_m) &= \mu\left(\left\{x \in S : T^k x \in \Lambda_m\right\}\right) \\ &= \mu\left(\left\{x \in S : \inf\left\{n \in \mathbb{Z}^+ : T^n x \in \Lambda_m\right\} \leq k\right\}\right) \\ &= \mu\left(\cup_{n=0}^k U_l^m\right)\end{aligned}$$

From above, we have $\lim_{k\to\infty} \mu \circ T^{-k}(\Lambda_m) = 1, \ \forall m \geq 0$. Note that this is not a uniform limit in $m \geq 0$. This concludes that asymptotically π_k^μ is supported on Λ_m for every $m \geq 0$. Informally, it means that dynamical system started with initial condition far away from the attractor will lie in some arbitrary small neighbourhood of attractor after sufficiently long time.

As $\mu \circ T^{-k}$ is also asymptotically supported on Λ_m for every $m \geq 0$, it is reasonable to assume that initial condition of the system is supported on Λ.

5.5 Examples and Discussions

In the following, we describe the filtering models which satisfy the assumptions in the Sections (5.2) and (5.3).

5.5.1 Examples with compact state space

We consider (S, d) to be compact and $h(.,.) : \mathbb{R}^+ \times S \to \mathbb{R}^p$ is such that $h(t,.)$ is bi-lipshitz for every $t \geq 0$ that satisfies the following:

$$K(t)d(x,y) \leq \|h(t,x) - h(t,y)\| \leq RK(t)d(x,y),$$

for some $\alpha > 0, R > 1$, $K(t)$ such that $K(t) = O(t^\alpha)$ and is increasing in t. Since any dynamical system $\{\phi_t\}_{t\in\mathbb{R}}$ with ϕ_t being a $C^{1+\alpha}$ diffeomorphism on S (with $\alpha > 0$, for every $t \in \mathbb{R}$) is such that ϕ_t is bi-lipshitz, we have

$$\frac{1}{MC^t} d(x,y) \leq d(\phi_t x, \phi_t y) \leq MC^t d(x,y),$$

$\forall t \in \mathbb{R}$ and for some $C, M > 1$. Now consider the following expression:

$$\int_t^{t+\tau} |h(s, \phi_{s-t}(x_1)) - h(s, \phi_{s-t}(x_2))|^2 \, ds$$

From the above, we have

$$\begin{aligned}\int_t^{t+\tau} |h(s, \phi_{s-t}(x_1)) - h(s, \phi_{s-t}(x_2))|^2 \, ds &\leq \int_t^{t+\tau} R^2 K^2(s) d(\phi_{s-t}(x_1), \phi_{s-t}(x_2))^2 \, ds \\ &\leq M^2 R^2 d(x,y)^2 \int_t^{t+\tau} K^2(s) C^{2(s-t)} ds\end{aligned}$$

Similarly we can obtain the following lower bound:

$$\int_t^{t+\tau} |h(s, \phi_{s-t}(x_1)) - h(s, \phi_{s-t}(x_2))|^2 \, ds \geq \frac{1}{M^2} d(x,y)^2 \int_t^{t+\tau} K^2(s) C^{-2(s-t)} ds$$

We consider $K(t)$ to be of the form $= Bt^q$, for some $q > 0$. Define

$$\rho_t^1 \doteq B^2 \int_t^{t+\tau} t^{2q} C^{-2(s-t)} ds \text{ and } \rho_t^2 \doteq B^2 \int_t^{t+\tau} t^{2q} C^{2(s-t)} ds.$$

It can be seen from computing the integrals that

$$1 \leq \frac{\rho_t^2}{\rho_t^1} \leq \bar{M},$$

for some $\bar{M} > 1$ independent of $t \geq 0$. It can be seen that $\rho_t^1 \sim O(t^{2q})$. Therefore, by defining ρ_t in Assumption (5.2.3) as $\rho_t \doteq \frac{1}{M}\rho_t^1$, we can conclude that the above model satisfies both Assumptions (5.2.3) and (5.2.2). Since S is compact, Assumptions (5.2.1) hold trivially by choosing U in Assumption (5.2.1) as S. In the above, we presented only continuous time models. Models in discrete time can be constructed similarly.

In the following, we give sufficient conditions for Assumption (5.2.4) to hold. Recall that Assumption (5.2.4) says that there is a set $\mathcal{V} \subset S \times S$ that is of full measure under $\sigma \otimes \sigma$ such that for $x, y \in \mathcal{V}$ satisfying $d(x, y) \geq b > 0$, the following holds

$$D_N^2(x, y) \geq L^2(b) \sum_{i=0}^{N} \rho_{i\tau}, \tag{5.24}$$

where, $L(b)$ is a positive constant. In the following, we show that (5.24) holds for a particular type of dynamical systems *viz.,* uniformly hyperbolic systems [132, Definition 4.1]. The arguments made are independent of whether time is discrete or continuous. So without loss in generality, let us suppose that the time is discrete with T being the homeomorphism. Suppose T is a $C^{1+\alpha}$ uniformly hyperbolic diffeomorphism with $\alpha > 0$. From [132, Proposition 7.4], T is expansive, *i.e.,* there exists $\epsilon > 0$ such that for every $x, y \in S$ with $x \neq y$, there exists $n \in \mathbb{Z}$ such that $d(T^n x, T^n y) > 2\epsilon$. From the continuity of T and compactness of S, we have the following:

Lemma 5.5.1. [1] *For any $\delta > 0$ and for some $\epsilon > 0$, if $x, y \in S$ such that $d(x, y) \geq \delta$ then there exists $J \in \mathbb{N}$ such that for some $n \in \mathbb{Z}$ with $|n| \leq J$, we have*

$$d(T^n x, T^n y) > \epsilon$$

[1] https://planetmath.org/UniformExpansivity

Proof. Consider the compact set, $K \doteq \{z = (x, y) \in S \times S : d(x, y) \geq \delta\}$. Choose $x, y \in S$ such that $d(x, y) \geq \delta$. From expansivity, there exists $n(x, y) \in \mathbb{Z}$ such that $d(T^{n(x,y)}x, T^{n(x,y)}y) > \epsilon$.

Define, $G(.,.) : S \times S \to S \times S$ by

$$G(u, v) \doteq (T^{n(x,y)}u, T^{n(x,y)}v).$$

It is clear that G is continuous on $S \times S$ and from the continuity of G, there is a neighbourhood $U(\bar{z})$ around $\bar{z} = (x, y)$ such that $d(T^{n(x,y)}u, T^{n(x,y)}v) > \epsilon$, $\forall (u, v) \in U(\bar{z})$. Since $\bar{z} = (x, y)$ is an arbitrary point in K, we can cover K by a family of open sets given by $\{U(z)\}_{z \in K}$. From compactness of K, there is a finite set $\{z_i\}_{i=1}^{k_0} \subset K$ such that $K \subset \cup_{i=1}^{k_0} U(z_i)$. Now, defining

$$J \doteq \max_{i=1,..,k_0} \left(|n(x_i, y_i)| : z_i = (x_i, y_i)\right),$$

we have the result. ∎

In particular, if we choose $\delta < \epsilon$, $d(T^n x, T^n y) > \epsilon$ for infinitely many $n \in \mathbb{Z}$. Suppose, x is in the global unstable manifold of y such that $d(x, y) > \epsilon$, *i.e.*,

$$d(T^n x, T^n y) \leq B\lambda^n d(x, y),$$

where, $n \leq 0$, $B > 0$ and $\lambda > 1$ (independent of x and y). It is clear that there exists $\bar{N}$ such that $d(T^n x, T^n y) < \epsilon$, $\forall n \leq -\bar{N}$. Therefore, from the above lemma, it is clear that if $|n| > \bar{N}$ and $d(T^n x, T^n y) \geq \epsilon$ then $n > 0$. Let $\{n_k(x, y)\}_{k \in \mathbb{N}}$ be a subsequence such that $d(T^{n_k(x,y)}x, T^{n_k(x,y)}y) \geq \epsilon$. From the above discussion, it is clear that $\{n_k(x, y)\}_{k \in \mathbb{N}}$ is an infinite set and in particular, $n_k(x, y) > \bar{N}$ infinitely many times. Therefore, without loss in generality, let us restrict the attention to $\{n_k(x, y)\}_{k \in \mathbb{N}}$ such that $n_k(x, y) \geq \bar{N}$, $\forall k \in \mathbb{N}$. From Lemma (5.5.1) and above discussion, we have the following:

$$n_{k+1}(x, y) - n_k(x, y) \leq J.$$

Note that J is independent of x and y as long as $d(x, y) \geq \epsilon$. Therefore, the cardinality of the set $\{n_k(x, y)\}_{k \in \mathbb{N}} \cap [\bar{N} + 1, 2, 3, ..., \bar{N} + \hat{N}]$ is at least $\lfloor \frac{\hat{N}}{J} \rfloor$, for any $\hat{N} \in \mathbb{N}$. As a result, we have the following for $N > \bar{N}$:

$$D_N^2(x, y) \geq \epsilon \sum_{\substack{k \in \mathbb{N}, \\ \bar{N} < n_k(x,y) \leq N}} \rho_{n_k(x,y)\tau} + \sum_{i=0}^{\bar{N}} d(T^i x, T^i y)\rho_{i\tau} \geq \epsilon \sum_{i=\bar{N}+1}^{\lfloor \frac{N-\bar{N}}{J} \rfloor + \bar{N} + 1} \rho_{i\tau}$$

$$
\begin{aligned}
&+\sum_{i=0}^{\bar{N}} d(T^i x, T^i y)\rho_{i\tau} \\
&\geq \min\left(\epsilon, \inf_{\substack{x,y\in S,\\ d(x,y)>\epsilon}} \left(\min_{i\leq \bar{N}}\left(d(T^i x, T^i y)\right)\right)\right) \sum_{i=0}^{\lfloor \frac{N-\bar{N}}{J}\rfloor+\bar{N}+1} \rho_{i\tau} \\
&\geq \min\left(\epsilon, \inf_{\substack{x,y\in S,\\ d(x,y)>\epsilon}} \left(\min_{i\leq \bar{N}}\left(d(T^i x, T^i y)\right)\right)\right) \sum_{i=0}^{\lfloor \frac{N}{J}\rfloor} \rho_{i\tau} \qquad (5.25) \\
&\geq G(J)\min\left(\epsilon, \inf_{\substack{x,y\in S,\\ d(x,y)>\epsilon}} \left(\min_{i\leq \bar{N}}\left(d(T^i x, T^i y)\right)\right)\right) \sum_{i=0}^{N} \rho_{i\tau}, \qquad (5.26)
\end{aligned}
$$

where, $G(J) > 0$ depends only on J. Inequality (5.25) follows from non-decreasing property of ρ_t, applying the lowest bound to any sum up to first $\lfloor \frac{N}{J} \rfloor$ terms of an subsequence of a non-decreasing sequence and inequality (5.26) follows from the form of ρ_t. And also, from uniform hyperbolicity, bi-lipshitz property of T and $d(x,y) > \epsilon$, for $n \leq \bar{N}$, we have

$$
\begin{aligned}
d(T^n x, T^n y) &\geq \frac{1}{C^n} d(x,y) \\
&\geq \frac{1}{C^{\bar{N}}} d(x,y) \\
&> \frac{1}{C^{\bar{N}}} \epsilon,
\end{aligned}
$$

for some $C > 1$. Therefore, we have

$$
\inf_{\substack{x,y\in S,\\ d(x,y)>\epsilon}} \left(\min_{i\leq \bar{N}}\left(d(T^i x, T^i y)\right)\right) > \frac{1}{C^{\bar{N}}}\epsilon
$$

and we have shown that if x lies in the unstable manifold of y and $d(x,y) > \epsilon$, we have

$$
D_N^2(x,y) \geq \min\left(G(\epsilon, J), \frac{1}{C^{\bar{N}}}\epsilon\right) \sum_{i=0}^{N} \rho_{i\tau}
$$

Now, we extend the above inequality, to x and y when x does not lie in either global stable or unstable manifolds of y. To that end, from [17], it is known that global stable manifolds form a foliation of S and global unstable manifold through a given point in S is their transversal. Therefore, for a given x and y such that $d(x,y) > \epsilon$ and x that does not lie in the stable manifold of y, there is a point $z \in X$ contained in the global unstable manifold of y such that x is the global stable manifold of z and we have

$$d(T^n z, T^n y) \leq d(T^n z, T^n x) + d(T^n x, T^n y)$$

From the property of global stable manifold and Lemma (5.5.1), there exists J_1 such that $d(T^n z, T^n x) \leq \frac{\epsilon}{2}$, $\forall n \geq J_1$. If $J_1 > J$, we replace J by J_1. Choosing $n = n_k(x,y)$, we get

$$\epsilon < d(T^n z, T^n y) \leq \frac{\epsilon}{2} + d(T^n x, T^n y)$$
$$\frac{\epsilon}{2} < d(T^n x, T^n y).$$

Therefore, we have

$$D_N^2(x,y) \geq \min\left(G\left(\frac{\epsilon}{2}, J\right), \frac{1}{C^{\bar{N}}}\epsilon \right) \sum_{i=0}^{N} \rho_{i\tau}.$$

Since the global stable manifold is strictly a lower dimensional manifold due to uniform hyperbolicity, we proved that (5.24) holds on a full measure set under measure $\sigma \times \sigma$ (σ is the Riemannian volume), which is sufficient for Theorem (5.2.8) to hold.

5.5.2 Examples with non-compact state space

We now consider $S = \mathbb{R}^p$ (which is non-compact) and continuous time models only. Choose $h(t,x) \doteq K(t)\bar{h}(x) : \mathbb{R}^+ \times S \to \mathbb{R}^p$ with any bi-lipshitz $\bar{h} : S \to \mathbb{R}^p$ and $K(t) = O(t^q)$. In the following, we show that the class of dynamical systems given by (5.27) along with the chosen observation model satisfy Assumptions (5.2.3), (5.2.1) and (5.2.2). To that end, let ϕ_t be the solution of the ordinary differential equation given below

$$\frac{d}{dt}\phi_t + A\phi_t + B\left(\phi_t, \phi_t\right) = f, \tag{5.27}$$

where, $B(.,.) : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^p$ is symmetric bi-linear operator such that $u^T B(u,u) = 0$, $\forall u \in \mathbb{R}^p$ and A is $p \times p$ matrix such that $u^T A u > \lambda |u|^2$, $\forall u \neq 0$. Observe that we have $|u^T B(v,w)| \leq H|u||v||w|$, for some H. From [88, Remark 2.4], we have the existence of bounded open set U such that $\overline{\phi_t U} \subset U$. And also, from [88, Lemma 2.6], we have the following:

$$|\phi_\tau(u) - \phi_\tau(v)| \leq e^{\gamma\tau}|u - v|, \tag{5.28}$$

$\forall u \in U, \forall v \in \mathbb{R}^p$ and for some $\gamma > 0$. Defining, $e_t \doteq \phi_t(u) - \phi_t(v)$, we have

$$\begin{aligned}
\frac{d}{dt}e_t + Ae_t + B(\phi_t(u), \phi_t(u)) - B(\phi_t(v), \phi_t(v)) &= 0 \\
e_t^T \frac{d}{dt}e_t + e_t^T Ae_t + e_t^T \left(B\left(\phi_t(u), \phi_t(u)\right) - B\left(\phi_t(v), \phi_t(u)\right)\right) &= 0 \\
\frac{1}{2}\frac{d}{dt}|e_t|^2 + e_t^T Ae_t + 2e_t^T \left(B\left(\phi_t(u), e_t\right) - B\left(e_t, e_t\right)\right) &= 0 \\
\frac{1}{2}\frac{d}{dt}|e_t|^2 + |A||e_t|^2 - 2H|e_t|^2|\phi_t(u)| &\geq 0 \\
\frac{d}{dt}|e_t|^2 + \left(2|A| + 4HR_U\right)|e_t|^2 &\geq 0,
\end{aligned}$$

where, $R_U \doteq \sup_{u \in U} \sup_{t \geq 0} |\phi_t(u)|$ and we used the properties of A and $B(.,.)$. We integrate the above equation to get,

$$|e_t|^2 \geq |e_0| - (2|A| + 4HR_U) \int_0^t |e_s|^2 ds$$

Applying the inequality from [67, Lemma 2], we have

$$|\phi_\tau(u) - \phi_\tau(v)| \geq \exp\left(-\left(|A| + 2HR_U\right)\tau\right)|u - v| \tag{5.29}$$

From the above, it is clear that Assumptions (5.2.1) and (5.2.2) hold. From the calculations similar to those in Section (5.5.1), we can conclude that Assumptions (5.2.3) holds when either of x_1 or x_2 in Assumption (5.2.3) lie in U. Note that this is always true in the setup of the Theorem (5.2.8).

In the following, we provide two well-known models, *viz.*, Lorenz 96 model and Lorenz 63 model and give numerical evidence that these models satisfy Assumption (5.2.4). To that end, we will show from numerical computations that $D_N(x, y) \doteq \sum_{i=0}^N \rho_{i\tau} d(T^i x, T^i y) \geq H \sum_{i=0}^N \rho_{i\tau}$, for some $H > 0$ and x, y are such that $d(x, y) > b$, for some b.

Lorenz 63 model:[101] In this case, $S = \mathbb{R}^3$, $\phi_t(u) = [x_t(u), y_t(u), z_t(u)]^T$ with

$$\begin{aligned}
\frac{d}{dt}x_t &= a(y_t - x_t) \\
\frac{d}{dt}y_t &= x_t(b - z_t) - y_t \\
\frac{d}{dt}z_t &= x_t y_t - cz_t,
\end{aligned}$$

where, we dropped the dependence of u. For $a = 10$, $b = 28$ and $c = \frac{8}{3}$, it is known that the above model exhibits a chaotic behavior. In the Figure (5.1), we can see that

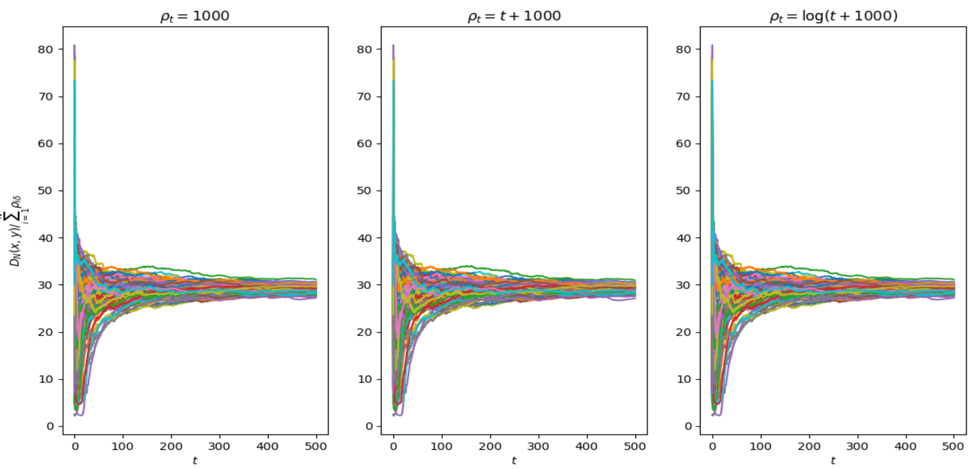


FIGURE 5.1: (Lorenz 63 model) Dependence of $\frac{D_N(x,y)}{\sum_{i=0}^{N} \rho_{i\delta}}$ *vs* $t = N\delta$ with $\delta = 0.01$ for 100 samples. We have plotted for $\rho_t = 1000,\ t + 1000,\ \log(t + 1000)$. The initial conditions for the samples are randomly chosen from uniform distribution on $[-10, 10]^3$.

$E_N(x,y) \doteq \sum_{i=0}^{N} d(T^i x, T^i y) \geq HN$, for some $H > 0$.

Lorenz 96 model:[102] For this model, $S = \mathbb{R}^p$, $\phi_t(u) = [x_t^1(u), x_t^2(u), ..., x_t^p(u)]^T$ with

$$\frac{d}{dt} x_t^i = (x_t^{i+1} - x_t^{i-2}) x_t^{i-1} - x_t^i + F$$

where, it is assumed that $x_t^{-1} = x_t^{p-1}, x_t^0 = x_t^p, x_t^1 = x_t^{p+1}$ and we again dropped the dependence of u. For $F = 8$, this model is known to exhibit chaotic behavior. In the Figure (5.2), we can see that $E_N(x,y) \doteq \sum_{i=0}^{N} d(T^i x, T^i y) \geq HN$, for some $H > 0$. Note that, in both the Figures (5.1) and (5.2), the three plots with different choices of ρ_t look very similar. The differences in the plots can only be seen at a much finer scale (not shown). Since their main significance is only to show that all the curves (in these plots) are bounded below, the finer structure of these plots is omitted.

5.5.3 Qualitative understanding of Assumptions (5.2.4) and (5.3.4)

In the remainder of the section, we try to explain the validity of the Assumptions (5.2.4) and (5.3.4) in a qualitative way. We restrict ourselves to the discrete time setup and to that end, we consider a bi-lipshitz homeomorphism, $T : S \to S$. We will see that the sensitive dependence and positiveness of Lyapunov exponent is used in order to argue the validity of these assumptions.

To that end, we assume that $T : S \to S$ satisfies the following properties:

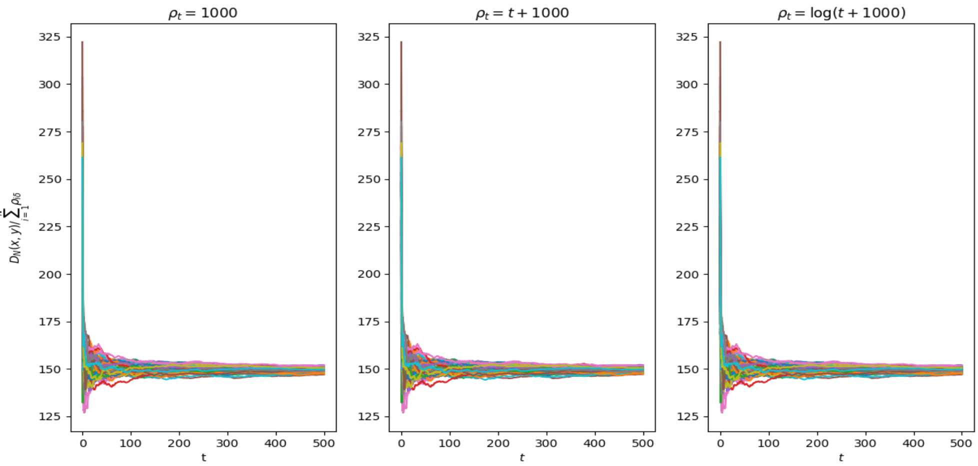


FIGURE 5.2: (Lorenz 96 model with $N = 36$) Dependence of $\frac{D_N(x,y)}{\sum_{i=0}^{N} \rho_{i\delta}}$ *vs* $t = N\delta$ with $\delta = 0.01$ for 100 samples. We have plotted for $\rho_t = 1000,\ t + 1000,\ \log(t + 1000)$. The initial conditions for the samples are randomly chosen from uniform distribution on $[-10, 10]^{36}$.

1. Sensitivity to initial conditions: There exists $\delta > 0$ such that for $x \in S$, $\forall \epsilon > 0$, there exists a σ-null (zero volume) set $\mathcal{V}(x)$ such that for all $y \in B_\epsilon(x) \backslash \mathcal{V}$, there is $n(x, y) \in \mathbb{N}$ such that $d(T^{n(x,y)}x, T^{n(x,y)}y) > \delta$. And for $y \in \mathcal{V}(x)$, $d(T^n x, T^n y) \to 0$ as $n \to \infty$. (Note that this is a stronger notion than the one given in [66]).

2. Positive Lyapunov exponent: If $y \in B_r^c(x) \backslash \mathcal{V}$ then $d(T^i x, T^i y) > \delta$ for $i \sim \frac{1}{\lambda} \log \frac{\delta}{r}$, where $\lambda > 0$ plays the role of Lyapunov exponent (Note that this property is qualitative in nature).

We give an informal argument using these properties to show that (5.3.4) holds. Choose $r > 0$ and fix x and y such that $d(x, y) > r$. And also, define $a_n = d(T^n x, T^n y)$. We assume that $\inf_n(a_n) = 0$, otherwise (5.24) trivially holds for a given x, y and of T . And also, we assume that $\limsup_{n\to\infty} a_n > 0$.

Let $\mathcal{C}$ be the set of all subsequences $\{n_k\}_{k\in\mathbb{N}}$ of $\mathbb{N}$ such that $a_{n_k} \to 0$ as $k \to \infty$. Defining $\mathcal{D} \doteq \mathcal{C}^c$, we see that for any $\{n_k\}_{k\in\mathbb{N}} \in \mathcal{D}$, $\inf_k (a_{n_k}) > 0$. Choose $\{n_k\}_{k\in\mathbb{N}} \in \mathcal{D}$ such that for any $\{p_k\}_{k\in\mathbb{N}}$ (such that $\{n_k\}_{k\in\mathbb{N}} \cap \{p_k\}_{k\in\mathbb{N}}$ is an infinite set), there exists a sub-subsequence $\{q_k\}_{k\in\mathbb{N}}$ of $\{n_k\}_{k\in\mathbb{N}} \cup \{p_k\}_{k\in\mathbb{N}}$ with the property $a_{q_k} \to 0$ as $k \to \infty$ ($\{n_k\}_{k\in\mathbb{N}}$ can be seen to exist). Suppose that $n_{k+1} - n_k \to \infty$ as $k \to \infty$. From the definition of $\mathcal{C}$ and $\mathcal{D}$, we can see that there exists an element, $\{m_k\}_{k\in\mathbb{N}} \in \mathcal{C}$ given by $\{m_k\}_{k\in\mathbb{N}} = \mathbb{N} \backslash \{n_k\}_{k\in\mathbb{N}}$ (From the assumption that $\inf_n(a_n) = 0$, it is an infinite set). From the assumption on $\{n_k\}_{k\in\mathbb{N}}$, it is clear that by choosing k becomes large enough, cardinality of the set $[n_k, n_{k+1}] \cap \{m_k\}_{k\in\mathbb{N}}$ can be made as larger than any desired integer.

In other words, for every $\rho > 0$, $M \in \mathbb{N}$, there exists k_0 such that for all $k \geq k_0$, we have

$$n_{k+1} - n_k > M \text{ and } a_m < \rho, \ \forall n_k < m < n_{k+1}.$$

Choosing $\bar{x} \doteq T^{n_k+1}x$ and $\bar{y} \doteq T^{n_k+1}y$, we see that this violates the assumptions on the dynamical system. Indeed, for $i \sim \frac{1}{\lambda}\log\frac{\delta}{r}$, we have $d(T^i\bar{x}, T^i\bar{y}) > \delta$ which contradicts the statement that $a_m = d(T^{m-n_k+1}\bar{x}, T^{m-n_k+1}\bar{y}) < \rho$, $\forall n_k < m < n_{k+1}$. Therefore, the supposition that $n_{k+1} - n_k \to \infty$ as $k \to \infty$ is false and there exists a positive constant, J such that $n_{k+1} - n_k \leq J$ for any k. This implies that cardinality of the set $\{n_k\}_{k\in\mathbb{N}} \cap [1, 2, 3, ..., N]$ is at least $\lfloor \frac{N}{J} \rfloor$. As a result, we have the following

$$D_N^2(x,y) \geq \delta \sum_{\substack{k\in\mathbb{N}, \\ n_k<N}} \rho_{n_k\tau} \geq \delta \sum_{i=0}^{\lfloor \frac{N}{J} \rfloor} \rho_{i\tau} \geq \delta G(\alpha, J) \sum_{i=0}^{N} \rho_{i\tau},$$

where $G(\alpha, J) > 0$ depends only on α and J. The above inequalities follow from non-decreasing property of ρ_t, applying the lowest bound to any sum up to first $\lfloor \frac{N}{J} \rfloor$ terms of an subsequence of a non-decreasing sequence and the form of ρ_t.

To summarize, in the current section, we studied various filtering models that satisfy the assumptions of Sections (5.2) and (5.3).

5.6 Conclusion

The problem that we studied is the stability of the nonlinear filter with deterministic dynamics. In order to establish stability, we first proved, in Theorem (5.2.8), an accuracy or consistency result for the smoother, *i.e.*, the convergence of the conditional distribution of the initial condition given observations. We used this result to prove the stability of the filter in Theorem (5.2.14). Using essentially identical methods, we also established the accuracy of the smoother (Theorem (5.3.8)) and the stability of the filter (Theorem (5.3.10)) in the case of discrete time.

The main assumptions used in proving these results are quite natural as discussed in Section (5.2.2), and are indeed satisfied by two classes of dynamical systems, as discussed in Section (5.5). In particular, these assumptions are valid for a class of diffeomorphisms of compact manifolds with appropriate enough observation function, as well as a class of nonlinear differential equations that includes models such as the Lorenz models (using numerical evidence for Assumption (5.2.4)). It will be of interest to find examples of non-trivial observational model (that is $h(\cdot,\cdot)$ is not invertible) that satisfy the assumptions.

There are various possible directions for further studies. Theorem (5.2.14) and Theorem (5.3.10) do not give any rate of convergence, because of the use of Martingale convergence theorem, and it would be interesting to find finer methods that may give the rate of convergence, such as those [28, Section 4.3] available for the convergence of covariance of the filter for linear models. Further, partly because of the use of convergence of the smoother to prove filter stability, our results do not give much information about the structure of the asymptotic filtering distribution, such as that which is available [28, Sections 4.3, 5], [126, Remark 3.2] for the linear filter.

Appendix A

Elements of Probability, Filtering Theory and Dynamical Systems

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete filtered probability space such that $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$ and $\mathcal{F}_0$ contains all $\mathbb{P}$- null sets. Until mentioned otherwise, on any complete metric space, we always work with the Borel σ- algebra as the underlying σ- algebra.

Lemma A.0.1 (Borel-Cantelli lemma). *Let $\{A_n\}_{n\geq 1} \subset \mathcal{F}$ such that*

$$\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$$

Then

$$\mathbb{P}(\{\omega \in \Omega : \omega \in A_n \textit{ for infinitely many } n\}) = 1$$

Theorem A.0.2. *[117, Lemma 2.1.2] Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(Z, \mathcal{Z})$ be a measurable space. Let $X,\ Y : \Omega \to Z$ be two random variables and $\mathcal{G}_Y$ be the $\sigma-$ algebra generated by Y. If X is measurable with respect to $\mathcal{G}_Y$, then there exists a $\mathcal{Z}$-measurable function $g : Z \to Z$ such that*

$$X = g(Y)$$

Theorem A.0.3 (Radon-Nikodym Theorem). *[29, Theorem A.30] Let m_1 and m_2 be two probability measures on $(\Omega, \mathcal{F})$ such that $m_1 \ll m_2$* i.e., *$\forall A \in \mathcal{F}$, $m_2(A) = 0 \implies m_1(A) = 0$. Then there exists a unique (upto a m_2- null set) non-negative measurable function f such that*

$$m_1(A) = \int_A f dm_2.$$

f is denoted by $\frac{dm_1}{dm_2}$ and called density of m_1 with respect to m_2.

Definition A.0.4 (Conditional expectation)**.** *Let* $X : \omega \to \mathbb{R}$ *be an integrable random variable* i.e., $\mathbb{E}[|X|] < \infty$ *and suppose* $\mathcal{G}$ *is sub* σ*- algebra of* $\mathcal{F}$*. Then a random variable* $Y : X \to \mathbb{R}$ *is defined as the conditional expection of* X *with respect to* $\mathcal{G}$ *if*

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}, \; \forall A \in \mathcal{G}$$

and Y *is* $\mathcal{G}$*- measurable.*

Remark A.0.5. *Using Radon-Nikodym theorem, it can be shown that such a* Y *(denoted by* $\mathbb{E}[X|\mathcal{G}]$*) exists uniquely almost surely.*

Here are few important properties of conditional expectation. Let $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ be sub σ- algebras such that $G \subset H$ and X, Y be two integrable random variables

1. For constants α and β, $\mathbb{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha\mathbb{E}[X|\mathcal{G}] + \beta\mathbb{E}[Y|\mathcal{G}], \; \mathbb{P}$- a.s.

2. $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$

3. If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X], \; \mathbb{P}$- a.s.

4. If X is $\mathcal{G}$- measurable, then $\mathbb{E}[X|\mathcal{G}] = X, \; \mathbb{P}$- a.s.

5. $\mathbb{E}[\mathbb{E}[X|H]|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}], \; \mathbb{P}$- a.s.

6. Analogous dominated convergence theorem, montone convergence theorem, Fatou's lemma, Jensen's Inequaity hold for conditional expectations.

7. If X is $\mathcal{G}$- measurable, then

$$\mathbb{E}[XY|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}], \; \mathbb{P}\text{- } a.s.$$

Lemma A.0.6 (Scheffe's Lemma)**.** *[146, Pg. 55] Suppose* f_n *and* f *are non-negative integrable functions in* $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ *and* $f_n \xrightarrow{n\to\infty} f$ *a.s. And also, suppose that* $\mathbb{E}[f_n] \xrightarrow{n\to\infty} \mathbb{E}[f]$*. Then* $\mathbb{E}\left[|f_n - f|\right] \xrightarrow{n\to\infty} 0$

Definition A.0.7 (Regular conditional distribution)**.** *A function* $\tilde{P} : \Omega \times \mathcal{F} \to [0, 1]$ *is called regular conditional distribution of random variable* X *with respect to sub* σ*- algebra* $\mathcal{G}$ *if the following is satisfied:*

1. *For any* $B \in \mathcal{F}$, $\tilde{P}(. , B)$ *is* $\mathcal{G}$*- measurable.*

2. *For any* $\omega \in \mathcal{F}$, $\tilde{P}(\omega, \, .)$ *is a probability measure.*

3. *For any* $B \in \mathcal{F}$, $\tilde{P}(. , B)$ *is a version of* $\mathbb{E}[\mathbb{1}_{X\in A}|\mathcal{G})(.)$

Definition A.0.8 (Stopping time). *A random variable $T : (\omega, \mathcal{F}) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ is called $\mathcal{F}_t$- stopping time if $\{T \leq t\} \in \mathcal{F}_t, \forall t \geq 0$.*

Definition A.0.9 (Progressive measurability). *A process $X_t : (\Omega, \mathcal{F}) \to (\Gamma, \mathcal{H})$, $t \in \mathbb{R}^+$ is progressively measurable with respect to a filtration $\mathcal{G}_t$ if the mapping $(t, \omega) \longmapsto X_t(\omega) : ([0, T] \times \Omega, \mathcal{B}([0, T]) \otimes \mathcal{G}_t) \to (\Gamma, \mathcal{H})$ is measurable, for all $T \geq 0$.*

In the following, all the processes involved are $\mathbb{R}$- valued until mentioned otherwise. Appropriate multi-dimensional versions of the given results hold true.

Definition A.0.10 (Gaussian process). *A process $\{G_t\}_{t \geq 0}$ is Gaussian if for any finite collection $\{t_i\}_{i=1}^n$, the law of the random variable $\{G_{t_1}, G_{t_2}, \ldots, G_{t_n}\}$ is Gaussian.*

Definition A.0.11 (Martingale). *Let X_t be an $\mathcal{F}_t$- adapted process. Then X_t is called $\mathcal{F}_t$- martingale if it satisfies the following conditions:*

1. *X_t is integrable for all $t \geq 0$*
2. *$\forall 0 \leq s \leq t$, we have $\mathbb{E}[X_t | \mathcal{F}_s] = X_s$, $\mathbb{P}-$ a.s.*

If we replace '$=$' in the above condition with '$\geq$' ('$\leq$'), then it is called $\mathcal{F}_t$- submartingale ($\mathcal{F}_t$- supermartingale).

Theorem A.0.12 (Submartingale Convergence theorem). *[86, Theorem 1.3.15]Let X_t be a $\mathcal{F}_t$- right continuous submartingale and let $\sup_{t \geq 0} \mathbb{E}[max\{X_t, 0\}] < \infty$. Then $X_\infty \doteq \lim_{t \to \infty} X_t$ exists $\mathbb{P}$- a.s. and $\mathbb{E}[X_\infty] < \infty$. Moreover, if $\{X_t\}_{t \geq 0}$ is a family uniformly integrable random variables, then X_t converges to X_∞ in L^1.*

Lemma A.0.13 (Doob's submartingale inequality). *[86, Theorem 1.3.8(i)] Let X_t be a $\mathcal{F}_t$-submartingale. For $T > 0$, we have*

$$\mathbb{P}(\sup_{0 \leq t \leq T} X_t \geq \lambda) \leq \frac{\mathbb{E}[\max\{X_T, 0\}]}{\lambda}$$

Definition A.0.14 (Local Martingale). *A $\mathcal{F}_t$- adapted process X_t is called $\mathcal{F}_t$- local martingale if there exists an increasing sequence of $\mathcal{F}_t$- stopping times $\{T_n\}_{n \geq 1} \uparrow \infty$ such that $X_{t \wedge T_n}$ is $\mathcal{F}_t$- martingale for every $n \geq 1$ and $X_0 = 0$, $\mathbb{P}-$ a.s. $x \wedge y$ stands for minimum of x and y.*

Note that a martingale is trivially a local martingale.

Definition A.0.15 (Quadratic variation process). *Let X_t be $\mathcal{F}_t$- square integrable continuous local martingale. Then a process denoted by $\langle X \rangle_t$ is called quadratic variation process if $X_t^2 - \langle X \rangle_t$ is a $\mathcal{F}_t$- martingale. Existence and uniqueness of such a process is a consequence of Doob-Meyer decomposition.*

Let us recall few properties of Ito's integral of f_t with respect to a square integrable continuous martingale X_t and f_t is such that it is progressively measurable with respect to $\mathcal{F}_t$ and

$$\mathbb{E}[\int_0^t f_s^2 d\langle X\rangle_s] < \infty$$

Define, $I_t^X(f) = \int_0^t f_s dX_s$

1. (Ito's isometry) $\mathbb{E}[\left(\int_0^t f_s dX_s\right)^2] = \mathbb{E}[\int_0^t f_s^2 d\langle X\rangle_s]$.
2. (Linearity) $I_t^X(\alpha f + \beta g) = \alpha I_t^X + \beta I_t^X$, for $\alpha,\ \beta \in \mathbb{R}$.
3. (Martingale property) $\mathbb{E}[I_t^X(f)|\mathcal{F}_s] = I_s^X(f)$, $\mathbb{P}$- a.s.

Definition A.0.16 (Semimartingale). *A process X_t is a continuous $\mathcal{F}_t$- semimartingale if it has the following decomposition,*

$$X_t = X_0 + B_t^v + M_t,\ \mathbb{P} - a.s.$$

where, M_t is continuous $\mathcal{F}_t$- local martingale and B_t^v is a continuous bounded variation process.

Theorem A.0.17 (Ito's formula). *Let $f : \mathbb{R} \to \mathbb{R}$ be twice differentiable and X_t be a $\mathcal{F}_t$-semimartingale with decomposition as above. Then $f(X_t)$ is also a $\mathcal{F}_t$- semimartingale with the decomposition,*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dM_s + \int_0^t f'(X_s)dB_s^v + \frac{1}{2}\int_0^t f''(X_s)d\langle M\rangle_s,$$

where, $\langle M\rangle_t$ is a process such that $M_t^2 - \langle M\rangle_t$ is a continuous local martingale.

In particular, a smooth function of a semimartingale is also a semimartingale. We now consider the following process that is of interest in the context of filtering theory:

$$\mathcal{E}_t \doteq \exp\left(X_t - \frac{1}{2}\langle X\rangle_t\right),$$

where, X_t is a continuous $\mathcal{F}_t$- local martingale.

Proposition A.0.18. *Process $\mathcal{E}_t$ is a $\mathcal{F}_t$- supermartingale.*

Proof. Applying Ito's formula to $\mathcal{E}_t$, we get

$$\mathcal{E}_t = \mathcal{E}_0 + \int_0^t \mathcal{E}_s dX_s + \frac{1}{2}\int_0^t \mathcal{E}_s d\langle X\rangle_s - \frac{1}{2}\int_0^t \mathcal{E}_s d\langle X\rangle_s$$

$$= 1 + \int_0^t \mathcal{E}_s dX_s$$

This implies that $\mathcal{E}_t$ is continuous local martingale. Then there exists a sequence of stopping times $T_n \uparrow \infty$ such $\mathcal{E}_{t\wedge T_n}$ is a martingale, for $n \geq 1$. Since $\mathcal{E}_t \geq 0$, from martingale property and Fatou's lemma, we have

$$\lim_{n\to\infty} \mathbb{E}[\mathcal{E}_{t\wedge T_n}|\mathcal{F}_s] = \lim_{n\to\infty} \mathcal{E}_{s\wedge T_n}$$
$$\mathbb{E}[\lim_{n\to\infty} \mathcal{E}_{t\wedge T_n}|\mathcal{F}_s] \leq \mathcal{E}_s$$
$$\mathbb{E}[\mathcal{E}_t|\mathcal{F}_s] \leq \mathcal{E}_s$$

Therefore, $\mathcal{E}_t$ is a $\mathcal{F}_t$- supermartingale. ■

Definition A.0.19 (Optional process)**.** *The optional σ- algebra $\mathcal{O}$ is defined as the σ-algebra on $[0,\infty) \times \Omega$ generated by $\mathcal{F}_t$- adapted right continuous (with left limit) processes. A process is called optional if it is $\mathcal{O}$- measurable.*

Theorem A.0.20 (Optional projection)**.** *[54, Theorem 43] Let Z_t be a bounded measurable process (not necessarily adapted to $\mathcal{F}_t$). There exists an optional process denoted by ${}^\circ Z_t$ such that for every stopping time T, we have*

$$\mathbb{E}[Z_T \mathbb{1}_{T<\infty}|\mathcal{F}_T] =^\circ Z_T, \ \mathbb{P}- \ a.s.$$

Any other process ${}^\circ Z_t'$ satisfying above condition will be such that

$$\mathbb{P}\left(\{{}^\circ Z_t =^\circ Z_t', \forall 0 \leq t < \infty\}\right) = 1.$$

Here, $\mathcal{F}_T$ is collection of all sets $A \in \mathcal{F}$ such that $A \cap T \leq t \in \mathcal{F}_t$, for each $t \geq 0$. Above statement hold also for unbounded non-negative measurable process.

Theorem A.0.21 (Levy's characterization theorem)**.** *Let X_t be a continuous square integrable $\mathcal{F}_t$- local martingale. If the quadratic process of X_t is*

$$\langle X\rangle_t = t, \ \forall t \geq 0,$$

then X_t is $\mathcal{F}_t$- Brownian motion.

Theorem A.0.22 (Girsanov's Theorem)**.** *[86, Theorem 3.5.1] Let $V \doteq \{V_t\}_{t\geq 0}$ be a $\mathcal{F}_t$-Brownian motion (of dimension n) and $H \doteq \{H_t\}_{t\geq 0}$ be a n-dimensional $\mathcal{F}_t$- progressively measurable process such that*

$$\mathbb{P}[\int_0^t \|H_s\|^2 ds < \infty] = 1, \ \forall t \geq 0.$$

And also, assume $Z_t \doteq \exp\left(\int_0^t H_s^T dV_s - \frac{1}{2}\int_0^t \|H_s\|^2 ds\right)$ *is a* $\mathcal{F}_t$*- martingale. Then, for a fixed* $T > 0$*,*

$$\tilde{V}_t \doteq V_t - \int_0^t H_s ds$$

is a $\mathcal{F}_t$*- Brownian motion for* $0 \leq t \leq T$ *on* $(\Omega, \mathcal{F}_T, \mathbb{Q})$*, where*

$$\mathbb{Q}(A) \doteq \mathbb{E}[\mathbb{1}_A Z_T]$$

We now give a sufficient condition for Z_t defined above to be a martingale. We already know from Proposition (A.0.18) that Z_t is a $\mathcal{F}_t$- supermartingale.

Lemma A.0.23 (Beneš condition)**.** *[86, Pg. 200] Assume same conditions and definitions as in Theorem (A.0.22). Let process H be of the form*

$$H_t = \hat{H}(t, V.),$$

where, $\hat{H}(t, \cdot)$ *is a progressively measurable functional on* $C([0, \infty), \mathbb{R}^n)$*, for each* $t \geq 0$*. Suppose that for each* $T \geq 0$*, there exist some* $K_T > 0$ *such that*

$$\|\hat{H}(t, x)\| \leq K_T \left(1 + \sup_{0 \leq s \leq t} \|x_s\|\right), \forall 0 \leq t \leq T \text{ and } \forall x \in C([0, \infty), \mathbb{R}^n).$$

Then Z_t *is a* $\mathcal{F}_t$*- martingale.*

Theorem A.0.24. *[98, Theorem 6.1] Let U be the topological space and* $K \subset U$ *be a compact set in U. Suppose* $\{G_x\}_{x \in K}$ *is a Gaussian random field on* $(\Omega, \mathcal{F}, \mathbb{P})$ i.e.*, for any finite collection* $\{x_1, x_2, x_3, \ldots, x_n\}$*, the law of random variable* $\{G_{x_1}, G_{x_2}, G_{x_3}, \ldots, G_{x_n}\}$ *is centered Gaussian on* $\mathbb{R}^n$*. Define the following quantity:*

$$F(K) = \mathbb{E}[\sup_{x \in K} G_x]$$

Then we have

$$F(K) \leq 24 \int_0^\infty \log^{\frac{1}{2}}\left(N(K, d_G, \varepsilon)\right) d\varepsilon,$$

where, $d_G(x, y) \doteq \mathbb{E}[|G_x - G_y|^2]^{\frac{1}{2}}$ *and* $N(K, d_G, \varepsilon)$ *is the minimum number of* ε*- open balls (under metric* d_G*) needed to cover K.*

Definition A.0.25 (Uniform complete observability)**.** *Let* $A_t : \mathbb{R}^m \to \mathbb{R}^m$ *and* $C_t : \mathbb{R}^m \to \mathbb{R}^n$ *be two matrices continuous in t. Then the pair* $[A_t, C_t]$ *is said to uniformly* *complete observable if there exist* $\rho_1, \rho_2, \tau > 0$ *such that* $\forall t \geq 0$, *we have*

$$\rho_1 \mathbb{I}_m \leq \int_{t-\tau}^{t} (\Phi_s)^{-T} (\Phi_t)^T C_s^T C_s \Phi_t (\Phi_s)^{-1} ds \leq \rho_2 \mathbb{I}_m,$$

where, $\dot{\Phi}_t = A_t \Phi_t$ *with* $\Phi_0 = \mathbb{I}_m$.

Definition A.0.26 (Uniform complete controllability). *Let* A_t *and* Φ_t *be as defined above. Suppose that* $C_t : \mathbb{R}^n \to \mathbb{R}^m$ *be a matrix continuous in* t. *Then the pair* $[A_t, C_t]$ *is said to be uniformly complete controllable if there exist* $\rho_1, \rho_2, \tau > 0$ *such that* $\forall t \geq 0$, *we have*

$$\rho_1 \mathbb{I}_m \leq \int_{t-\tau}^{t} (\Phi_t) (\Phi_s)^{-1} C_s C_s^T (\Phi_s)^{-T} (\Phi_t)^{-T} ds \leq \rho_2 \mathbb{I}_m,$$

Lemma A.0.27. *[143, Pg.181] For a given compact set* $\mathcal{K}$, *there exist* $q = q_{\mathcal{K}}$ *and* $b = b_{\mathcal{K}}$ *such that the following holds for all* $n \geq 0$.

$$r_n(\epsilon, \mathcal{K}, \phi_\tau) \leq q(C^n b \epsilon^{-1})^p,$$

where, ϕ_τ *is as defined in Chapter* (5), $r_n(\cdot,\cdot,\cdot)$ *is as defined in Definition* (5.2.7) *and* C *is as defined in Assumption* (5.2.2).

Proposition A.0.28. *[1, Theorem 2.3.1] Let* $\mathbb{X}_t$ *be a* $m \times m$ *matrix such that*

$$\dot{\mathbb{X}}_t = \tilde{A}_t \mathbb{X}_t, \ \mathbb{X}_0 = \mathbb{I}_m,$$

where, $t \geq 0$, $\tilde{A}_t \in \mathbb{R}^{m \times m}$ *such that* $\sup_{t \geq 0} \|\tilde{A}_t\| < \infty$. *Then*

$$\|\mathbb{X}_t\| \leq K \exp(\alpha t),$$

for some $K, \alpha > 0$. α *can be taken to be* $\sup_{t \geq 0} \|\tilde{A}_t\|$.

Proof of Theorem (2.1.2). We follow the proof of [15, Theorem 2.35]. We already know that $\mathcal{Y}_t \subset \mathcal{Y}_{t+}$. To prove the other way, fix $t \geq 0$ and fix $A \in \mathcal{Y}_{t+}$ to contruct the process M_s defined as

$$M_s \doteq \begin{cases} \mathbb{1}_A - \mathbb{E}\left[\mathbb{1}_A | \mathcal{Y}_t\right] & \text{if } s \geq t \\ 0 & \text{if } s < t \end{cases}$$

From the definition of M_s, M_s is right continuous. And also, M_s is a bounded $\mathcal{Y}_{s+}$-martingale, from definition. From Theorem (A.0.12), There exists M_∞ such that

$$\lim_{s \to \infty} M_s = M_\infty, \ \mathbb{P} - a.s.$$

and it can be seen that $M_s = \mathbb{E}[M_\infty|\mathcal{Y}_{s+}]$. Now M_∞ is $\mathcal{Y}_\infty$- measurable. From [15, Proposition 2.31], M_∞ has the following representation.

$$M_\infty = \mathbb{E}[M_\infty] + \int_0^\infty \eta_u^T dI_u,$$

where, η_u is progressively measurable with respect to $\mathcal{Y}_s$ and $\mathbb{E}[\int_0^\infty \|\eta_u\|^2 du] < \infty$. Conditioning the above equation with respect to $\mathcal{Y}_{s+}$, we have

$$\begin{aligned}\mathbb{E}[M_\infty|\mathcal{Y}_{s+}] &= \mathbb{E}[\mathbb{E}[M_\infty]|\mathcal{Y}_{s+}] + \mathbb{E}[\int_0^\infty \eta_u^T dI_u|\mathcal{Y}_{s+}] \\ M_s &= \mathbb{E}[M_0] + \int_0^t \eta_u dI_u\end{aligned}$$

This concludes that M_s is continous and $\mathcal{Y}_s$- adapted (from Theorem (2.1.13)). In particular, $A \in \mathcal{Y}_t$. From arbitrariness of A, we have $\mathcal{Y}_{t+} \subset \mathcal{Y}_t$ ∎

Bibliography

[1] L Ya Adrianova. *Introduction to linear systems of differential equations*. American Mathematical Soc., 1995.

[2] Deborah F Allinger and Sanjoy K Mitter. "New results on the innovations problem for non-linear filtering". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 4.4 (1981), pp. 339–348.

[3] Brian DO Anderson. "Stability properties of Kalman-Bucy filters". In: *Journal of the Franklin Institute* 291.2 (1971), pp. 137–144.

[4] Jeffrey L. Anderson and Stephen L. Anderson. "A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts". In: *Monthly Weather Review* 127.12 (1999), pp. 2741–2758.

[5] Amit Apte et al. "Sampling the posterior: An approach to non-Gaussian data assimilation". In: *Physica D: Nonlinear Phenomena* 230 (June 2007), pp. 50–64. DOI: 10.1016/j.physd.2006.06.009.

[6] Mark Asch, Marc Bocquet, and Maëlle Nodet. *Data Assimilation: Methods, Algorithms, and Applications*. SIAM, 2016.

[7] Rami Atar. "Exponential decay rate of the filter's dependence on the initial distribution". In: *The Oxford Handbook of Nonlinear Filtering* (2011), pp. 299–318.

[8] Rami Atar. "Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain". In: *The Annals of Probability* 26.4 (1998), pp. 1552–1574.

[9] Rami Atar and Ofer Zeitouni. "Exponential stability for nonlinear filtering". In: *Annales de l'Institut Henri Poincare (B) Probability and Statistics* 33.6 (1997), pp. 697–725.

[10] Rami Atar and Ofer Zeitouni. "Lyapunov exponents for finite state nonlinear filtering". In: *SIAM Journal on Control and Optimization* 35.1 (1997), pp. 36–55.

[11] Nattapol Aunsri. "A particle filtering approach for noisy seismic events tracking". In: *2016 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*. IEEE. 2016, pp. 1–1.

[12] B. Azimi-Sadjadi and P. S. Krishnaprasad. "Change detection for nonlinear systems; a particle filtering approach". In: *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*. Vol. 5. 2002, 4074–4079 vol.5.

[13] Tim Bailey and Hugh Durrant-Whyte. "Simultaneous localization and mapping (SLAM): Part II". In: *IEEE robotics & automation magazine* 13.3 (2006), pp. 108–117.

[14] J. Baillieul and J. C. Willems, eds. *Mathematical Control Theory*. Berlin, Heidelberg: Springer-Verlag, 1998.

[15] Alan Bain and Dan Crisan. *Fundamentals of stochastic filtering*. Vol. 60. Springer Science & Business Media, 2008.

[16] JS Baras, A Bensoussan, and MR James. "Dynamic observers as asymptotic limits of recursive filters: Special cases". In: *SIAM Journal on Applied Mathematics* 48.5 (1988), pp. 1147–1158.

[17] Luis Barreira and Ya B Pesin. *Introduction to smooth ergodic theory*. Vol. 148. American Mathematical Soc., 2013.

[18] Peter Baxendale, Pavel Chigansky, and Robert Liptser. "Asymptotic stability of the Wonham filter: ergodic and nonergodic signals". In: *SIAM journal on control and optimization* 43.2 (2004), pp. 643–669.

[19] VE Beneš. "Exact finite-dimensional filters for certain diffusions with nonlinear drift". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 5.1-2 (1981), pp. 65–92.

[20] VE Beneš. "Nonexistence of strong nonanticipating solutions to stochastic DEs: implications for functional DEs, filtering, and control". In: *Stochastic Processes and their applications* 5.3 (1977), pp. 243–263.

[21] Václav E. Beneš and Ioannis Karatzas. "Estimation and control for linear, partially observable systems with non-gaussian initial distribution". In: *Stochastic Processes and their Applications* 14.3 (1983), pp. 233 –248.

[22] Alain Bensoussan. *Stochastic control of partially observable systems*. Cambridge University Press, 1992.

[23] H Benzerrouk. "Modern approaches in nonlinear filtering theory applied to original problems of aerospace integrated navigation systems with non-Gaussian noises". In: *Saint Petersburg State University: Sankt-Peterburg, Russia* (2014).

[24] Abhay G Bhatt, Amarjit Budhiraja, and Rajeeva L Karandikar. "Markov property and ergodicity of the nonlinear filter". In: *SIAM Journal on Control and Optimization* 39.3 (2000), pp. 928–949.

[25] Patrick Billingsley. *Convergence of probability measures*. Second. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., 1999.

[26] Adrian N Bishop and Pierre Del Moral. "On the Stability of Kalman–Bucy Diffusion Processes". In: *SIAM Journal on Control and Optimization* 55.6 (2017), pp. 4015–4047.

[27] T. Blackburn and D. Vaughan. "Application of linear optimal control and filtering theory to the Saturn V launch vehicle". In: *IEEE Transactions on Automatic Control* 16.6 (1971), pp. 799–806.

[28] Marc Bocquet et al. "Degenerate Kalman filter error covariances and their convergence onto the unstable subspace". In: *SIAM/ASA Journal on Uncertainty Quantification* 5.1 (2017), pp. 304–333.

[29] L Breiman. "Probability". In: *Classics in Applied Mathematics. SIAM* 7 (1992).

[30] Jochen Bröcker and Gianluigi Del Magno. "Asymptotic stability of the optimal filter for random chaotic maps". In: *Nonlinearity* 30.5 (2017), p. 1809.

[31] R. W. Brockett. "Nonlinear Systems and Nonlinear Estimation Theory". In: *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*. Ed. by Michiel Hazewinkel and Jan C. Willems. Dordrecht: Springer Netherlands, 1981, pp. 441–477.

[32] R. W. Brockett and Clark J.M.C. "The geometry of the conditional density functions". In: Academic Press, New York, 1980, pp. 299–309.

[33] Richard S Bucy. "Global theory of the Riccati equation". In: *Journal of computer and system sciences* 1.4 (1967), pp. 349–361.

[34] RS Bucy and PD Joseph. "Filtering for Stochastic Processes with Applications to Guidance. 1968". In: *Interscience-Wiley, New York* ().

[35] Amarjit Budhiraja. "Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter". In: *Annales de l'IHP Probabilités et statistiques* 39.6 (2003), pp. 919–941.

[36] Amarjit Budhiraja and Daniel Ocone. "Exponential stability in discrete-time filtering for non-ergodic signals". In: *Stochastic processes and their applications* 82.2 (1999), pp. 245–257.

[37] Amarjit Budhiraja and Daniel Ocone. "Exponential stability of discrete-time filters for bounded observation noise". In: *Systems & Control Letters* 30.4 (1997), pp. 185–193.

[38] Alberto Carrassi et al. "Controlling instabilities along a 3DVar analysis cycle by assimilating in the unstable subspace: a comparison with the EnKF". In: *arXiv:0804.2136* (2008).

[39] Alberto Carrassi et al. "Data assimilation in the geosciences: An overview of methods, issues, and perspectives". In: *WIREs Clim Change* 9.5 (2018), e535.

[40] H Carvalho et al. "Optimal nonlinear filtering in GPS/INS integration". In: *IEEE Transactions on Aerospace and Electronic Systems* 33.3 (1997), pp. 835–850.

[41] Frédéric Cérou. "Long time behavior for some dynamical noise free nonlinear filtering problems". In: *SIAM Journal on Control and Optimization* 38.4 (2000), pp. 1086–1101.

[42] Yiping Chen and Nozer D. Singpurwalla. "A NON-GAUSSIAN KALMAN FILTER MODEL FOR TRACKING SOFTWARE RELIABILITY". In: *Statistica Sinica* 4.2 (1994), pp. 535–548.

[43] P Chigansky, R Liptser, and R Van Handel. "Intrinsic methods in filter stability". In: *Handbook of Nonlinear Filtering* (2009).

[44] Pavel Chigansky. "An ergodic theorem for filtering with applications to stability". In: *Systems & control letters* 55.11 (2006), pp. 908–917.

[45] Pavel Chigansky. *Stability of nonlinear filters: A survey, 2006. Mini-course lecture notes, Petropolis, Brazil*.

[46] Pavel Chigansky. "Stability of the nonlinear filter for slowly switching Markov chains". In: *Stochastic processes and their applications* 116.8 (2006), pp. 1185–1194.

[47] Pavel Chigansky, Robert Liptser, et al. "On a role of predictor in the filtering stability". In: *Electronic Communications in Probability* 11 (2006), pp. 129–140.

[48] Pavel Chigansky, Robert Liptser, et al. "Stability of nonlinear filters in nonmixing case". In: *The Annals of Applied Probability* 14.4 (2004), pp. 2038–2056.

[49] JMC Clark. "The design of robust approximations to the stochastic differential equations of nonlinear filtering". In: *Communication systems and random process theory* 25 (1978), pp. 721–734.

[50] JMC Clark, Daniel Ocone, and C Coumarbatch. "Relative entropy and error bounds for filtering of Markov processes". In: *Mathematics of Control, Signals and Systems* 12.4 (1999), pp. 346–360.

[51] Dan Crisan. "The stochastic filtering problem: a brief historical account". In: *Journal of Applied Probability* 51.A (2014), pp. 13–22.

[52] F. Daum. "Exact finite-dimensional nonlinear filters". In: *IEEE Transactions on Automatic Control* 31.7 (1986), pp. 616–622.

[53] Mark HA Davis. "A pathwise solution of the equations of nonlinear filtering". In: *Theory of Probability & Its Applications* 27.1 (1982), pp. 167–175.

[54] Claude Dellacherie and Paul-André Probabilities Meyer. "potential. B. Theory of martingales. Translated from the French by JP Wilson". In: *North-Holland Mathematics Studies* 72 ().

[55] B Delyon and O Zeitouni. *Lyapunov exponents for filtering problems, Applied Stochastic Analysis, edited by Davis, MHA and Elliot, RJ*. 1991.

[56] R. Desai, T. Lele, and F. Viens. "A Monte-Carlo method for portfolio optimization under partially observed stochastic volatility". In: *2003 IEEE International Conference on Computational Intelligence for Financial Engineering, 2003. Proceedings.* 2003, pp. 257–263.

[57] Randal Douc, Eric Moulines, Yaacov Ritov, et al. "Forgetting of the initial condition for the filter in general state-space hidden Markov chain: a coupling approach". In: *Electronic Journal of Probability* 14 (2009), pp. 27–49.

[58] Randal Douc et al. "Forgetting of the initial distribution for nonergodic hidden Markov chains". In: *The Annals of Applied Probability* 20.5 (2010), pp. 1638–1662.

[59] Tyrone E Duncan. "Evaluation of likelihood functions". In: (1968).

[60] Hugh Durrant-Whyte and Tim Bailey. "Simultaneous localization and mapping: part I". In: *IEEE robotics & automation magazine* 13.2 (2006), pp. 99–110.

[61] Anthony D'Aristotile, Persi Diaconis, and David Freedman. "On merging of probabilities". In: *Sankhyā Ser. A* 50.3 (1988), pp. 363–380.

[62] Steven Fletcher. *Data Assimilation for the Geosciences: From Theory to Application*. Elsevier, 2017.

[63] W. Fong et al. "Monte Carlo smoothing with application to audio signal enhancement". In: *IEEE Transactions on Signal Processing* 50.2 (2002), pp. 438–449.

[64] P Frost and Thomas Kailath. "An innovations approach to least-squares estimation–Part III: Nonlinear estimation in white Gaussian noise". In: *IEEE Transactions on Automatic Control* 16.3 (1971), pp. 217–226.

[65] Masakiyo Fujimoto and Satoshi Nakamura. "Particle filter based non-stationary noise tracking for robust speech recognition". In: *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.* Vol. 1. IEEE. 2005, pp. I–257.

[66] Eli Glasner and Benjamin Weiss. "Sensitive dependence on initial conditions". In: *Nonlinearity* 6.6 (1993), pp. 1067–1075.

[67] H. E. Gollwitzer. "A Note on a Functional Inequality". In: *Proceedings of the American Mathematical Society* 23.3 (1969), pp. 642–647.

[68] Karthik S Gurumoorthy et al. "Rank deficiency of Kalman error covariance matrices in linear time-varying system with deterministic evolution". In: *SIAM Journal on Control and Optimization* 55.2 (2017), pp. 741–759.

[69] Fredrik Gustafsson. "Particle filter theory and practice with positioning applications". In: *IEEE Aerospace and Electronic Systems Magazine* 25.7 (2010), pp. 53–82.

[70] Fredrik Gustafsson et al. "Particle filters for positioning, navigation, and tracking". In: *IEEE Transactions on signal processing* 50.2 (2002), pp. 425–437.

[71] Ramon van Handel. "On the exchange of intersection and supremum of σ-fields in filtering theory". In: *Israel Journal of Mathematics* 192.2 (2012), pp. 763–784.

[72] Ramon van Handel. "The stability of conditional Markov processes and Markov chains in random environments". In: *Ann. Probab.* 37 (2009), pp. 1876–1925.

[73] Ramon van Handel. "Uniform observability of hidden Markov models and filter stability for unstable signals". In: *Ann. Appl. Probab.* 19 (2009), pp. 1172–1199.

[74] U. G. Haussmann and E. Pardoux. "A conditionally almost linear filtering problem with non-gaussian initial condition". In: *Stochastics* 23.2 (1988), pp. 241–275.

[75] M. Hazewinkel, S. I. Marcus, and H. J. Sussmann. "Nonexistence of finite dimensional filters for conditional statistics of the cubic sensor problem". In: *Filtering and Control of Random Processes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1984, pp. 76–103.

[76] M. Hazewinkel, S.I. Marcus, and H.J. Sussmann. "Nonexistence of finite-dimensional filters for conditional statistics of the cubic sensor problem". In: *Systems and Control Letters* 3.6 (1983), pp. 331 –340.

[77] J. L. Hibey and C. D. Charalambous. "Conditional densities for continuous-time nonlinear hybrid systems with applications to fault detection". In: *IEEE Transactions on Automatic Control* 44.11 (1999), pp. 2164–2169.

[78] Ze-Chun Hu and Wei Sun. "A note on exponential stability of the nonlinear filter for denumerable Markov chains". In: *Systems & control letters* 55.11 (2006), pp. 955–960.

[79] Andrew H Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007.

[80] G Kallianpur, Masatoshi Fujisaki, and Hiroshi Kunita. "Stochastic differential equations for the non linear filtering problem". In: *Osaka Journal of Mathematics* 9.1 (1972), pp. 19–40.

[81] G Kallianpur and Charlotte Striebel. "Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors". In: *The Annals of Mathematical Statistics* 39.3 (1968), pp. 785–801.

[82] Gopinath Kallianpur. *Stochastic filtering theory*. Vol. 13. Springer Science & Business Media, 1980.

[83] Gopinath Kallianpur and Charlotte Striebel. "Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes". In: *Theory of Probability & Its Applications* 14.4 (1969), pp. 567–594.

[84] R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: *Journal of Basic Engineering* 82.1 (Mar. 1960), pp. 35–45.

[85] Rudolph E Kalman and Richard S Bucy. "New results in linear filtering and prediction theory". In: *Journal of basic engineering* 83.1 (1961), pp. 95–108.

[86] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*. Vol. 113. Springer Science & Business Media, 2012.

[87] Anatole Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*. Vol. 54. Cambridge university press, 1996.

[88] David TB Kelly, KJH Law, and Andrew M Stuart. "Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time". In: *Nonlinearity* 27.10 (2014), p. 2579.

[89] M. Kleptsyna and A. Veretennikov. "On discrete time ergodic filters with wrong initial data". In: *Probability Theory and Related Fields* 141 (July 2008), pp. 411–444.

[90] Andrei Nikolaevitch Kolmogorov. "Stationary sequences in Hilbert space". In: *Bull. Math. Univ. Moscow* 2.6 (1941), pp. 1–40.

[91] Zvonko Kostanjcar, Branko Jeren, and Jurica Cerovec. "Particle Filters in Decision Making Problems under Uncertainty". In: *Automatika* 50 (Apr. 2009), pp. 245–251.

[92] Hiroshi Kunita. "Asymptotic behavior of the nonlinear filtering errors of Markov processes". In: *Journal of Multivariate Analysis* 1.4 (1971), pp. 365–393.

[93] Hiroshi Kunita. "Ergodic properties of nonlinear filtering processes". In: *Spatial stochastic processes*. Springer, 1991, pp. 233–256.

[94] Harold J Kushner. "Dynamical equations for optimal nonlinear filtering". In: *Journal of Differential Equations* 3.2 (1967), pp. 179–190.

[95] H Kwakernak and R Sivan. *Linear Optimal Control Systems. Willey & Sons*. 1972.

[96] Kody Law, Andrew Stuart, and Kostas Zygalakis. *Data Assimilation*. Springer, 2015.

[97] François Le Gland, Nadia Oudjane, et al. "Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters". In: *The Annals of Applied Probability* 14.1 (2004), pp. 144–187.

[98] Michel Ledoux. "Isoperimetry and Gaussian analysis". In: *Lectures on probability theory and statistics*. Springer, 1996, pp. 165–294.

[99] Robert S Liptser and Albert N Shiryaev. *Statistics of random processes: I. General theory*. Springer Science & Business Media, 2001.

[100] Edward N Lorenz. "Deterministic and stochastic aspects of atmospheric dynamics". In: *Irreversible phenomena and dynamical systems analysis in geosciences*. Springer, 1987, pp. 159–179.

[101] Edward N Lorenz. "Deterministic nonperiodic flow". In: *Journal of the atmospheric sciences* 20.2 (1963), pp. 130–141.

[102] Edward N Lorenz. "Predictability: A problem partly solved". In: *Proc. Seminar on predictability*. Vol. 1. 1996.

[103] Wei-Lwun Lu, Kenji Okuma, and J.J. Little. "Tracking and recognizing actions of multiple hockey players using the boosted particle filter". In: *Image and Vision Computing* 27 (Jan. 2009), pp. 189–205. DOI: 10.1016/j.imavis.2008.02.008.

[104] Armand M Makowski. "Filtering formulae for partially observed linear systems with non-Gaussian initial conditions". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 16.1-2 (1986), pp. 1–24.

[105] Curtis McDonald and Serdar Yüksel. "Exponential Filter Stability via Dobrushin's Coefficien". In: *arXiv preprint arXiv:1910.08463* (2019).

[106] Curtis McDonald and Serdar Yüksel. "Observability and Filter Stability for Partially Observed Markov Processes". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE. 2019, pp. 1623–1628.

[107] Curtis McDonald and Serdar Yuksel. "Stability of Non-Linear Filters and Observability of Stochastic Dynamical Systems". In: *arXiv preprint arXiv:1812.01772* (2018).

[108] Curtis McDonald and Serdar Yüksel. "Stability of Non-Linear Filters, Observability and Relative Entropy". In: *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2018, pp. 110–114.

[109] S.K. Mitter. "On the analogy between mathematical problems of non-linear filtering and quantum physics". In: 10. 1979, pp. 163–216.

[110] Michael Montemerlo et al. "FastSLAM: A factored solution to the simultaneous localization and mapping problem". In: *Aaai/iaai* 593598 (2002).

[111] Aleem Mushtaq, Yu Tsao, and Chin Hui-Lee. "A particle filter feature compensation approach to robust speech recognition". In: *Eleventh Annual Conference of the International Speech Communication Association*. 2010.

[112] Boyi Ni and Qinghua Zhang. "Stability of the Kalman filter for continuous time output error systems". In: *Systems & Control Letters* 94 (2016), pp. 172–180.

[113] Daniel Ocone. "Asymptotic stability of Beneš filters". In: *Stochastic analysis and applications* 17.6 (1999), pp. 1053–1074.

[114] Daniel Ocone. "Entropy Inequalities and Entropy Dynamics in Nonlinear Filtering of Diffusion Processes". In: *Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W.H. Fleming*. Boston, MA: Birkhäuser Boston, 1999, pp. 477–496.

[115] Daniel Ocone and Etienne Pardoux. "A Lie algebraic criterion for non-existence of finite dimensionally computable filters". In: 1989, pp. 197–204.

[116] Daniel Ocone and Etienne Pardoux. "Asymptotic stability of the optimal filter with respect to its initial condition". In: *SIAM Journal on Control and Optimization* 34.1 (1996), pp. 226–243.

[117] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2003.

[118] Nadia Oudjane and Sylvain Rubenthaler. "Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals". In: *Stochastic Analysis and applications* 23.3 (2005), pp. 421–448.

[119] Luigi Palatella, Alberto Carrassi, and Anna Trevisan. "Lyapunov vectors and assimilation in the unstable subspace: theory and applications". In: *Journal of Physics A: Mathematical and Theoretical* 46.25 (2013), p. 254020.

[120] TN Palmer. "Stochastic weather and climate models". In: *Nature Reviews Physics* 1.7 (2019), pp. 463–471.

[121] Anastasia Papavasiliou. "Parameter estimation and asymptotic stability in stochastic filtering". In: *Stochastic processes and their applications* 116.7 (2006), pp. 1048–1065.

[122] L. R. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition". In: *Proceedings of the IEEE* 77.2 (1989), pp. 257–286.

[123] Yogesh Rathi, Namrata Vaswani, and Allen Tannenbaum. "A generic framework for tracking using particle filter with dynamic shape prior". In: *IEEE Transactions on Image Processing* 16.5 (2007), pp. 1370–1382.

[124] Yogesh Rathi et al. "Tracking deforming objects using particle filtering for geometric active contours". In: *IEEE transactions on pattern analysis and machine intelligence* 29.8 (2007), pp. 1470–1475.

[125] Anugu Sumith Reddy and Amit Apte. "Stability of Non-linear Filter for Deterministic Dynamics". In: *arXiv preprint arXiv:1910.14348* (2019).

[126] Anugu Sumith Reddy, Amit Apte, and Sreekar Vadlamani. "Asymptotic properties of linear filter for deterministic processes". In: *Systems & Control Letters* 139 (2020), p. 104676.

[127] Sebastian Reich and Colin Cotter. *Probabilistic forecasting and Bayesian data assimilation*. Cambridge University Press, 2015.

[128] Wolfgang J Runggaldier. "Estimation via stochastic filtering in financial market models". In: *Contemporary mathematics* 351 (2004), pp. 309–318.

[129] Simo Särkkä. *Bayesian filtering and smoothing*. Vol. 3. Cambridge University Press, 2013.

[130] Shankar Sastry and Marc Bodson. *Adaptive control: stability, convergence and robustness*. Courier Corporation, 2011.

[131] Subhamoy Sen et al. "Seismic-induced damage detection through parallel force and parameter estimation using an improved interacting Particle-Kalman filter". In: *Mechanical Systems and Signal Processing* 110 (2018), pp. 231–247.

[132] Michael Shub. *Global stability of dynamical systems*. Springer Science & Business Media, 1986.

[133] Louis Shue, Brian DO Anderson, and Subhrakanti Dey. "Exponential stability of filters and smoothers for hidden Markov models". In: *IEEE Transactions on Signal Processing* 46.8 (1998), pp. 2180–2194.

[134] Wilhelm Stannat. "Stability of the Filter Equation for a Time-Dependent Signal on $\mathbb{R}^d$". In: *Applied Mathematics and Optimization* 52.1 (2005), pp. 39–71.

[135] Wilhelm Stannat. "Stability of the Optimal Filter via Pointwise Gradient Estimates". In: *Stochastic Partial Differential Equations and Applications-VII* (2005), pp. 281–293.

[136] Łukasz Stettner. "On invariant measures of filtering processes". In: *Stochastic Differential Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989, pp. 279–292.

[137] A Trevisan and L Palatella. "On the Kalman filter error covariance collapse into the unstable subspace". In: *Nonlinear Processes in Geophysics* 18.2 (2011), pp. 243–250.

[138] Ramon Van Handel. "Discrete time nonlinear filters with informative observations are stable". In: *Electronic Communications in Probability* 13 (2008), pp. 562–575.

[139] Ramon Van Handel. "Observability and nonlinear filtering". In: *Probability theory and related fields* 145.1-2 (2009), pp. 35–74.

[140] S.R.S. Varadhan. *Probability Theory*. Vol. 7. American Mathematical Soc., 2001.

[141] M. Vellekoop and J. Clark. "A Nonlinear Filtering Approach to Changepoint Detection Problems: Direct and Differential-Geometric Methods". In: *SIAM J. Control and Optimization* 42 (Jan. 2003), pp. 469–494. DOI: 10.1137/050647438.

[142] A. Yu. Veretennikov and M. L. Kleptsyna. "On Continuous Time Ergodic Filters with Wrong Initial Data". In: *Theory of Probability and Its Applications* 53.2 (2009), pp. 269–300.

[143] Peter Walters. *An introduction to ergodic theory*. Vol. 79. Springer Science & Business Media, 1982.

[144] Norbert Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964. ISBN: 0262730057.

[145] Darren J Wilkinson. *Stochastic modelling for systems biology*. CRC press, 2018.

[146] David Williams. *Probability with Martingales.* Cambridge mathematical textbooks. Cambridge University Press, 1991.

[147] W Murray Wonham. "Some applications of stochastic differential equations to optimal nonlinear filtering". In: *Journal of the Society for Industrial and Applied Mathematics, Series A: Control* 2.3 (1964), pp. 347–369.

[148] Jie Xiong. *An introduction to stochastic filtering theory*. Vol. 18. Oxford University Press on Demand, 2008.

[149] Moshe Zakai. "On the optimal filtering of diffusion processes". In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 11.3 (1969), pp. 230–243.

[150] Ofer Zeitouni. "Approximate and limit results for nonlinear filters with small observation noise: the linear sensor and constant diffusion coefficient case". In: *IEEE transactions on automatic control* 33.6 (1988), pp. 595–599.