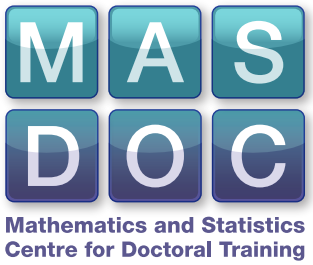


Assimilating Data into Mathematical Models

by

Daniel Sanz-Alonso

Thesis

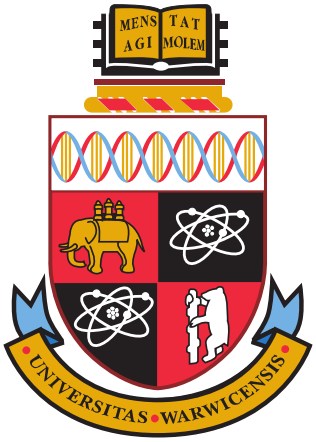
Submitted for the degree

Doctor of Philosophy

Mathematics Institute

The University of Warwick

June 2016



Contents

Acknowledgments

First of all I would like to thank Andrew Stuart; I could not have asked for a better supervisor either at a mathematical or a personal level. Thanks also to my second supervisor Gareth Roberts from whom I have learned so much, and to my examiners Adam Johansen and Arnaud Doucet, who took the interest and time to read this thesis from cover to cover. I am particularly thankful to Sergios Agapiou and Omiros Papaspiliopoulos for their continuous support. The memory of our time together at Warwick, Barcelona, and Savannah will stay with me forever, as will my friendship with you. I am also grateful to my collaborators Kody Law and Abhishek Shukla, to my first year mentor David Kelly, and to Rodrigo Targino and Gareth Peters for their warm hospitality in my visits to London.

I thank everyone who has made so enjoyable my years at Warwick. Thanks to my tennis friends and rivals (Andy, Atanas, James, Lukasz, Lynn, Tony, and Yuan) and to Faz, Felipe, Gianmarco and Rachele. My biggest thanks to Cyril, Kyung, and Neil.

Finalmente, gracias a mi familia y muy especialmente a mis padres Jesús y Mercedes, hermano Carlos y abuelas Amelia y Felicia. Estuvisteis conmigo en los momentos malos. Es hora de compartir los buenos.

Declarations

This thesis has four chapters. Chapter 1 is a brief overview of the Bayesian approach to blending mathematical models with data. For this introductory chapter, I do not claim any originality in the material itself, but only in the presentation, and in the choice of contents. Chapters 2, 3 and 4 are transcripts of published and submitted papers, with minimal cosmetic modifications. I now detail my contributions to each of these papers.

Chapter 2 is a transcript of the published paper "Long-time Asymptotics of the Filtering Distribution for Partially Observed Chaotic Dynamical Systems" [Sanz-Alonso and Stuart, 2015] written in collaboration with Andrew Stuart. The idea of building a unified framework for studying filtering of chaotic dissipative dynamical systems is from Andrew. My ideas include the truncation of the 3DVAR algorithm that allows for unbounded observation noise, using the squeezing property as the unifying arch across all models, and most of the links with control theory. I stated and proved all the results of the paper. I also wrote the first version of the paper, which was subsequently much improved with Andrew's input.

Chapter 3 is a transcript of the published paper "Filter Accuracy for the Lorenz 96 Model: Fixed Versus Adaptive Observation Operators" [Law et al., 2016], written in collaboration with Kody Law, Abhishek Shukla, and Andrew Stuart. My contribution to this paper was in proving most of the theoretical results. I did not contribute to the numerical experiments. The idea of using adaptive observation operators is from Abhishek.

Chapter 4 is a transcript of the submitted paper "Importance Sampling: Computational Complexity and Intrinsic Dimension" [Agapiou et al., 2015], writ-

ten in collaboration with Sergios Agapiou, Omiros Papaspiliopoulos, and Andrew Stuart. The idea of relating the two notions of intrinsic dimension described in the paper is from Omiros. Sergios stated and proved Theorem 4.2.3. Andrew's input was fundamental in making the paper well structured, and in the overall writing style. The paper was written very collaboratively among the four of us, and some of the results were the fruit of many discussions involving different subsets of authors. Some of my inputs include: the idea of using metrics between probability measures to study the performance of importance sampling, establishing connections to tempering, the analysis of singular limits both for inverse problems and filtering, most of the filtering section and in particular the use of the theory of inverse problems to analyze different proposals in the filtering set-up, the proof of Theorem 4.2.1, and substantial input in the proof of all the results of the paper not mentioned before. This paper aims to bring cohesion and new insights into a topic with a vast literature, and I helped towards this goal by doing most of the literature review involved.

Chapter 1

Introduction

This thesis touches on a number of subjects, including inverse problems, nonlinear filtering, data assimilation, uncertainty quantification, and computational statistics. The recurrent theme is that of combining mathematical models (often differential equations) with observed data. I have aimed to improve the current understanding of the theoretical and computational challenges that arise when doing so. Identifying these challenges precisely is fundamental in developing new algorithms that tackle them.

1.1 A Bird's-Eye View

The question underlying most of this thesis is how to improve on the incomplete understanding that mathematical models provide about physical systems by using noisy, indirect observations of the system. This is a problem with many faces. In this thesis I use a wide set of mathematical tools that includes the theory of (chaotic) ordinary and partial (stochastic) differential equations, control theory, variational data assimilation algorithms, Kalman filtering and generalizations, importance sampling and particle filters, probability theory in infinite dimensional Banach spaces, and the Bayesian formulation of inverse problems in function space.

I focus on two classes of problems: inverse problems and filtering. In the former the data is assimilated all at once, while in the latter the data is assimilated sequentially in time, as it becomes available. The *sequential* nature of filtering poses distinct computational challenges, especially if online algorithms are required.

I have mostly worked within the Bayesian paradigm, where uncertain quantities are treated as random variables. Under the Bayesian formulation the aim is to update (sequentially in the case of filtering) the distribution suggested by expert

knowledge and the mathematical model (the *prior* distribution), into the conditional distribution given the observed data (the *posterior* or *filtering* distribution). At the heart of this update is Bayes' formula.

In almost all cases of applied interest some algorithm is needed in order to approximate the updated measures. This is because Bayes' formula often involves analytically intractable integrals. Crucially, there are two settings where Bayes' formula can be readily computed: finite-state problems, and problems with a linear and Gaussian structure. Both are of central applied importance, but their key role goes beyond that. Indeed, most algorithms used outside the finite-state or linear-Gaussian set-up proceed by *imposing* some particle or linear-Gaussian approximation, and then computing Bayes' formula exactly with an approximate prior and observation model. We remark that other closed conjugate analyses outside the finite-state or linear Gaussian set-ups are possible but of less applied relevance, see [Bernardo and Smith, 1994] and [Vidoni, 1999].

In this thesis I address a number of theoretical and computational questions arising from the Bayesian formulation, and from the algorithmic approximations thereof. From a theoretical viewpoint, it is of the essence to have theoretical guarantees that as more data (generated from some "truth") are incorporated into the prior, the updated measures concentrate around the underlying truth. I have studied this question in the mathematically rich, and practically important framework of filtering deterministic chaotic dynamical systems. From a computational viewpoint, I have investigated three main sources of challenges: the high dimensional nature of the systems arising in applications, small noise regimes, and highly nonlinear problems. I have aimed to give a precise meaning to these challenges. This is important in order to build new algorithms that tackle them, and to compare the performance of different algorithms on a sound basis. Moreover, I have contributed towards a better understanding of how these precisely defined challenges affect the performance of algorithms based on importance sampling.

This thesis has four chapters. Chapter 2 is a transcript of the published paper "Long-time Asymptotics of the Filtering Distribution for Partially Observed Chaotic Dynamical Systems" [Sanz-Alonso and Stuart, 2015]. Chapter 3 is a transcript of the published paper "Filter Accuracy for the Lorenz 96 Model: Fixed Versus Adaptive Observation Operators" [Law et al., 2016]. Chapter 4 is a transcript of the submitted paper "Importance Sampling: Computational Complexity and Intrinsic Dimension" [Agapiou et al., 2015]. The remainder of this introductory Chapter 1 is organized as follows. Section 1.2 introduces inverse problems and filtering by means of simple but rich and paradigmatic examples. Section 1.3 describes two classes of algorithms

widely used in the Bayesian assimilation of data: Gaussian methods (which will be largely studied in Chapters 2 and 3), and particle methods (which will be studied in Chapter 4). Section 1.4 loosely presents three situations where the algorithmic approximation of measures is particularly challenging: large dimensional systems, small noise regimes, and highly nonlinear problems. I show how these challenging settings impact in different ways the performance of Gaussian and particle methods. I also give a brief account of the relative advantages of these algorithms. Section 1.5 summarizes the main contributions of this thesis, and Section 1.6 closes with some future research directions.

1.2 Inverse Problems and Filtering: Guiding Examples

We illustrate the Bayesian approach by means of examples. Subsection 1.2.1 introduces Bayesian inverse problems, and Subsection 1.2.2 Bayesian filtering. In each subsection a motivating example is followed by a discussion and a literature review. The examples show the advantages and limitations of the Bayesian approach in concrete settings. The discussion and literature review aim to provide a wider picture, and to highlight the paradigmatic features of the examples.

1.2.1 Inverse Problems: an Example

1.2.1.1 Mathematical Model, Prior, Data and Posterior

Consider the mathematical model of an elliptic partial differential equation, defined in a bounded domain $D \subset \mathbb{R}^d$ with Lipschitz boundary ∂D,

$$\begin{aligned} -\nabla \cdot (\kappa \nabla p) &= f, \quad x \in D, \\ p &= 0, \quad x \in \partial D, \end{aligned} \tag{1.2.1}$$

where f is given and assumed to belong to the dual of $V := H_0^1(D)$. It is well known that, for given $\kappa \in L^\infty(D)$ with $\operatorname{ess\,inf}_{x \in D} \kappa(x) = \kappa_{\min} > 0$, problem (1.2.1) has a unique weak solution $p \in V$. The partial differential equation in (1.2.1) is an example of a balance law and models many different phenomena. To be concrete, let us say that p represents the pressure of a fluid in a porous medium at steady state. Then (1.2.1) combines the constitutive equation known as Darcy's law (which, informally, asserts that fluids tend to flow from high to low pressure regions) with conservation of mass. The function κ represents the permeability of the medium, and quantifies the intensity with which the fluid flows from higher to lower pressure regions. There is central applied interest, notably in the oil industry, in understanding subsurface

flows. In such scenarios it is difficult to determine by measurements the permeability function $\kappa : D \to (\kappa_{\min}, \infty)$ that should be plugged into (1.2.1). Thus, even if the toy model (1.2.1) were a perfect model for the pressure p of interest, there is uncertainty in the solution coming from the uncertainty on the input κ.

The first step of the Bayesian approach is to acknowledge the uncertainty in κ by viewing it as a random variable whose distribution is called the *prior*. In practice the prior should encode all the available knowledge on κ. For instance, it may be known that the function κ has certain smoothness. Alternatively, it may be known that the medium is composed of two different rocks, each having known constant permeability. In this later case, the uncertainty reduces to determining the surface that separates the rocks. Both examples show a common trait: although the aim is to recover the infinite dimensional object κ, and thereby the pressure p, the problem is heavily constrained by the prior.

The second step of the Bayesian approach is to use indirect data in order to reduce the uncertainty in κ. Let us assume that we have access to J noisy measurements of the pressure p

$$y_j = l_j(p) + \eta_j, \quad 1 \leq j \leq J,$$

where the l_j are linear functionals on V, and η_j is some random noise. Noting that p is a nonlinear function of κ, this can be rewritten as

$$y = \mathcal{G}(\kappa) + \eta,$$

where $y := [y_1, \ldots, y_J]^T \in \mathbb{R}^J$, $\eta := [\eta_1, \ldots, \eta_J]^T \in \mathbb{R}^J$, and $\mathcal{G}_j(\kappa) = l_j(p)$. The key object of interest in the Bayesian formulation is the *posterior* distribution, which is the conditional distribution of κ given y. The posterior can be pushed forward through the mathematical model (1.2.1) to produce a distribution on p. This distribution may be used to compute the 'most likely' value of p, known as the MAP (maximum a posteriori) estimator. Importantly, it also contains quantitative and qualitative information on the uncertainty remaining in the quantity of interest p.

1.2.1.2 Discussion and Theoretical Challenges

As described in the example, Bayesian inversion proceeds by (i) specifying a prior on the unknown input (permeability) (ii) updating it based on data, and (iii) propagating it through the mathematical model to obtain a distribution on the unknown of interest (pressure). This should be contrasted to plain vanilla uncertainty quantification, that consists of steps (i) and (iii), and omits step (ii). The objective is to

reduce the uncertainty in the values of the pressure by the use of data. Practical implementation of step (iii) may be problematic in complex models arising in applications. This issue will not be considered in this thesis, as we focus on steps (i) and (ii).

In our example the uncertainty stems from the input parameter κ; the data consists of observations of the true pressure. Further examples include (i) the inverse problem of determining the solution of an evolution equation at time $T > 0$ with uncertain initial condition, by use of partial and noisy observations of the solution at time $T^* > 0$. This problem will be discussed in the next section in a filtering set-up; and (ii) Electrical Impedance Tomography (EIT), where the conductivity of a body is inferred from electrode measurements on its surface.

All these inverse problems are clearly *underdetermined*: they aim to recover an infinite dimensional object (the function κ, and hence p in our example) from a finite number of noisy observations. The Bayesian formulation acts as a natural *regularization* —with a clear probabilistic interpretation— of the inverse problem through the introduction of a prior distribution on the unknown. The prior is fundamental in two ways. First, it reduces dramatically the number of parameters that are effectively estimated by imposing structure and correlations on the unknown. We will give an insightful interpretation of this in Chapter 4, where we relate the dimension of the Bayesian inverse problem to the notion of effective number of parameters from statistics and machine learning. Second, introducing probabilistic information on the unknown results naturally —through the Bayesian machinery— in a posterior distribution on the unknown as the solution to the inverse problem. The Bayesian solution to the inverse problem is a probability distribution that accounts for the uncertainty in the reconstruction due to the underdetermined nature of the problem. This is in contrast with traditional approaches where the solution is typically limited to a point estimate. The choice of the prior is, however, critical for the success and validity of the Bayesian approach. Some interesting theoretical research questions are: How should the prior be chosen so that the posterior contracts at optimal rate around the unknown as more data are collected? How to quantify the errors introduced by discretization of function space inverse problems, accounting for both discretization of the stochastic input parameter space, and temporal and spatial discretizations of the underlying mathematical model?

1.2.1.3 Literature Review

Methods used for vanilla uncertainty quantification include stochastic finite element methods and stochastic collocation methods [Ghanem and Spanos, 2003]. A good

introductory text with a focus on methods based on generalized polynomial chaos expansions is [Xiu, 2010].

A review of the classical approach to inverse problems is given in [Engl et al., 1996], [Mueller and Siltanen, 2012]. Excellent surveys of the Bayesian approach are [Kaipio and Somersalo, 2005], and [Stuart, 2010]. The latter deals with the formulation of the problems in function space.

The classical elliptic inverse problem is studied in [Banks and Kunisch, 2012] and [Richter, 1981], and the Bayesian formulation in [Dashti et al., 2012] and [Dashti and Stuart, 2011]. Posterior consistency results were established in [Vollmer, 2013].

The regularizing effect of the prior in the Bayesian approach can be interpreted as a Tikhonov-Phillips regularization in the classical formulation [Kaipio and Somersalo, 2005]. The connection between MAP estimators and the classical regularized solution of the inverse problem is also discussed in [Kaipio and Somersalo, 2005]. [Dashti et al., 2013] extended the results to Banach separable spaces.

The inverse EIT problem was formulated and studied in [Somersalo et al., 1992], and in function space setting in [Dunlop and Stuart, 2015]. See [Iglesias et al., 2015] for a Bayesian level set formulation of geometric inverse problems, where the goal is to recover the surfaces separating different media.

1.2.2 Filtering: an Example

1.2.2.1 Mathematical Model, Data and Filtering Distribution

Consider as mathematical model the following differential equation, defined in a Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle, |\cdot|)$,

$$\frac{dv}{dt} + Av + B(v,v) = f. \tag{1.2.2}$$

We will be interested in filtering dissipative systems of the form (1.2.2) with energy conserving nonlinearity. Thus we will impose that, for some $\lambda > 0$, and for all $v \in \mathcal{H}$, $\langle Av, v\rangle \geq \lambda|v|^2$ and $\langle B(v,v), v\rangle = 0$. Dissipative models of the form (1.2.2) include the Lorenz '63 model, the Lorenz '96 model, and the Navier Stokes equation on a two-dimensional torus. The underlying spaces are $\mathcal{H} = \mathbb{R}^3$ for the Lorenz '63 model, $\mathcal{H} = \mathbb{R}^d$, $d \geq 3$ for the Lorenz '96 model, and certain infinite dimensional space of functions for the Navier Stokes equation (details and motivation for studying these models are given in Chapter 2). Under mild assumptions, satisfied by all the aforementioned models, equation (1.2.2) supplemented with an initial condition $v(0) = v_0 \in \mathcal{H}$, has a unique solution in any finite time interval. We will be interested in the scenario where there is uncertainty in the initial condition. The aim will be

to estimate the solution $v(t)$ at time $t > 0$ based on the data available at that given time.

The first step of the Bayesian approach to filtering is, in analogy with Section 1.2.1, to assume that the initial condition is only known statistically, $v_0 \sim \mu_0$. Thus μ_0 can be thought of as a prior measure on the initial condition of the system. The prior μ_0, together with the model (1.2.2), suggest a prior measure ν_t on any $v(t)$ by pushing-forward μ_0 by the solution semigroup Ψ_t associated with (1.2.2). That is, $\nu_t(\cdot) = \mu_0\big(\Psi_t^{-1}(\cdot)\big)$.

The second step is to use indirect data in order to reduce the uncertainty in $v(t)$. We consider the most realistic scenario where the data arrives discretely in time, every $h > 0$ units of time. Denote $v(jh) = v_j$, $j \geq 1$, and let

$$y_j := h(v_j) + \eta_j, \tag{1.2.3}$$

where h is some observation function, and η_j is some random noise. The aim of filtering is to compute sequentially in time the *filtering distributions*,

$$\mu_j(\cdot) := \mathbb{P}(v_j \in \cdot | y_1, \ldots, y_j), \quad j \geq 1, \tag{1.2.4}$$

as new observations become available.

For later reference we now introduce some terminology and a concrete mathematical setting for filtering. The process of interest is called the *signal process*, and is denoted $\{v_j\}_{j\geq 0}$. The *observation process* is denoted $\{y_j\}_{j\geq 0}$, where $y_0 = 0$ so that there is effectively no observation at discrete time $j = 0$. We denote $Y_j := \{y_i\}_{i=1}^j$. The signal and observation processes satisfy

$$\begin{aligned} v_{j+1} &= \Psi(v_j) + \xi_j, \quad j \geq 0, \\ y_{j+1} &= Hv_{j+1} + \eta_{j+1}, \quad j \geq 0, \end{aligned} \tag{1.2.5}$$

where $\Psi : \mathbb{R}^{d_v} \to \mathbb{R}^{d_v}$ and $H \in \mathbb{R}^{d_y \times d_v}$. We make the following statistical assumptions:

(i) The initial condition v_0 is only known statistically, $v_0 \sim N(m_0, C_0)$.

(ii) The ξ_j's form an i.i.d. sequence, independent of v_0, with $\xi_0 \sim N(0, Q)$.

(iii) The η_j's form an i.i.d. sequence, independent of $\big(v_0, \{\xi_j\}_{j\geq 0}\big)$, with $\eta_1 \sim N(0, R)$.

The assumption that the signal is finite dimensional, $v_j \in \mathbb{R}^{d_v}$, is made for ease of exposition. It will be dropped in Chapters 2, 3 and 4. The assumption that the *observation operator* H is linear is often restrictive. More general observations

of the form $y_{j+1} = h(v_{j+1}) + \eta_j$ are of applied importance. When the signal arises from a differential equation (such as (1.2.2)), the map Ψ is the Δt time solution semigroup, with $\Delta t > 0$ the time between observations. Note that deterministic dynamics corresponds to the degenerate case $Q = 0$.

1.2.2.2 Discussion and Theoretical Challenges

The filtering setting (1.2.5) can be cast into the more general language of hidden Markov models. Our interest in models of the specific form (1.2.5) comes from geophysical applications, in particular weather forecasting. Distinct features of these filtering problems are (i) the huge dimension of the unknown and the data (currently of the order of 10^9 and 10^6, respectively, in operating weather forecasting models in the United Kingdom), (ii) the turbulent character of the dynamics, and (iii) that observations are often sparse in time and space, and subject to small noise. The term *data assimilation* is often used in the geosciences to refer to filtering problems with these features. The mathematical model (1.2.2) is flexible and allows for models containing these three features in different degrees. For this reason it has been often used to test filtering algorithms. Filtering has many applications outside the geosciences, some of which will be briefly reviewed in the next subsection.

The filtering distributions (1.2.4) depend on the choice of the initial prior distribution μ_0. It is of crucial theoretical and applied importance to determine whether, for sufficiently large discrete time j, the distribution μ_j essentially 'forgets' the prior μ_0. This question has given rise to an extensive literature on *nonlinear filtering stability*. On the theoretical side, filtering stability provides a justification for the Bayesian approach: the choice of the prior becomes, for sufficiently large j, irrelevant for the Bayesian solution μ_j. On the applied side, filtering stability is essential for algorithms to successfully approximate the filtering distributions: it guarantees that the approximation error at a given time step is not amplified at later times. A further key question is whether in the long-time asymptotic the filtering distributions concentrate around the realization of the signal that underlies the data. This is called *filter accuracy*. In this thesis we encounter scenarios where the signal dynamics are unpredictable and only sparse observations of them are available, but still the filtering distributions provide useful information on the state of the signal. In the terminology of the previous section, we will study partially and noisily observed chaotic dynamics where the ν_t's contain scarce information on the state of the signal, while the μ_t are peaked around the true value of the signal that underlies the observations.

1.2.2.3 Literature Review

An excellent survey of nonlinear filtering, that makes apparent the depth and breadth of the subject, is [Crisan and Rozovskii, 2011]. Recent books on Bayesian filtering and data assimilation are [Law et al., 2015], [Majda and Harlim, 2012], [Särkkä, 2013], [Reich and Cotter, 2015], and [Cappé et al., 2009]. Other fundamental books include [Jazwinski, 2007] and [Kalnay, 2003]. As mentioned before, the filtering problems studied in this thesis are motivated mainly by geophysical applications. The list of applications of nonlinear filtering is endless, with an enormous body of literature associated with each of them. To give the reader a flavour of the breadth of applications we replicate here the (incomplete) list in [Van Handel, 2006]: navigation and target tracking, changepoint detection, stochastic control, finance, audio and image enhancement, biology, quantum optics, speech recognition, and communication theory. We refer to [Van Handel, 2006] for pointers to the literature on each of those applications.

The starting point of the study of stability of nonlinear filters is the seminal paper [Kunita, 1971], which addressed the related question of filter ergodicity. The development and popularization of particle filters in the 1990s gave rise to a renewed interest in the subject, which resulted in numerous papers built on [Kunita, 1971]. Unfortunately, Kunita's paper had a mistake in one of the proofs, where an unjustified exchange of supremum and intersection of σ-fields was performed. This mistake was inherited by many papers in the 1990s. A key contribution that helped to clarify these issues is [Budhiraja, 2003], where several desirable properties of filters —that include filter stability, filter ergodicity and permissibility of the interchange of supremum and intersection of certain σ-fields— are proved to be equivalent. Recent state-of-the-art studies on filter stability include [Kleptsyna and Veretennikov, 2008], [Douc et al., 2009], and [Tong and Van Handel, 2014]. See also [Crisan and Rozovskii, 2011] for more references.

A pioneering work on filter accuracy is [Cérou, 2000]. Some of the questions investigated in the second and third chapters of this thesis are motivated by this paper, and more directly by the subsequent body of work that studied the accuracy of suboptimal filters that approximate the filtering distributions: [Brett et al., 2013], [Law et al., 2014], [Kelly et al., 2014].

1.3 Algorithms

We now introduce two classes of algorithms that will be extensively studied in this thesis: Gaussian and particle approximation methods. We will describe the methods

in the filtering setting (1.2.5). Our presentation is biased toward algorithms that are *sequential*, meaning that the cost of approximating the filtering distribution at time $j+1$ given an approximation at time j does not depend on j. Some of the algorithms described here have also been recently used to approximate posterior distributions arising from non-sequential Bayesian inverse problems. Gaussian approximation methods will be studied in much more detail in Chapters 2 and 3, and particle approximation methods in Chapter 4. Both classes bypass the intractability of Bayes' formula by invoking different approximations. For convenience we recall Bayes' formula, which can be non-rigorously written as

$$\underbrace{\mathbb{P}(\theta|y)}_{\text{posterior}} \propto \underbrace{\mathbb{P}(\theta)}_{\text{prior}} \underbrace{\mathbb{P}(y|\theta)}_{\text{likelihood}} . \tag{1.3.1}$$

Here θ should be interpreted as the quantity of interest, and y as data. $\mathbb{P}(\theta)$ is the distribution of θ before the data y is assimilated. $\mathbb{P}(y|\theta)$ represents the likelihood of the data y given a value θ of the unknown. $\mathbb{P}(\theta|y)$ is the distribution of interest, i.e. the distribution of the quantity of interest θ given data y. The normalizing constant that makes the right-hand side of (1.3.1) a probability distribution is typically hard to compute.

Gaussian and particle methods rely, respectively, on the following key facts:

- If $\mathbb{P}(\theta)$ is Gaussian, and $\mathbb{P}(y|\theta)$ is Gaussian, then $\mathbb{P}(\theta|y)$ is Gaussian.
- If $\mathbb{P}(\theta)$ is a particle measure with equal weights, informally written as $\mathbb{P}(d\theta) = \frac{1}{N}\sum_{=1}^{N} \delta_{\theta^n}(d\theta)$, then $\mathbb{P}(\theta|y)$ is a weighted particle measure. Moreover,

$$\mathbb{P}(d\theta|y) \propto \sum_{n=1}^{N} \mathbb{P}(y|\theta^n)\delta_{\theta^n}(d\theta). \tag{1.3.2}$$

A third class of algorithms, which we do not study in this thesis, is that of Markov Chain Monte Carlo (MCMC) methods. They rely on the following key fact:

- It is often possible to construct a Markov Chain with unique invariant distribution $\mathbb{P}(\theta|y)$, and so this distribution can be approximated by the occupation measure of the Markov Chain over some time interval.

Gaussian and MCMC algorithms can approximate the posterior without computing the normalizing constant in (1.3.1), while particle methods often produce an approximation to this constant. MCMC methods are arguably the gold standard for computationally challenging Bayesian inverse problems. However, only recent

methodological developments have suggested the potential of these methods in sequential filtering problems. Both Gaussian and particle methods have also their own merits and caveats, which we will explore in the next section. It is noteworthy that much of the methodological progress in the last ten years has revolved around two ideas. The first one is to combine algorithms from the three classes above that retain their respective advantages while mitigating their disadvantages. The second related idea is to develop sequential algorithms for static inverse problems, and modifying algorithms well suited for static problems for their use in sequential ones.

The remainder of this section is organized as follows. Subsection 1.3.1 describes Gaussian approximation methods, and Subsection 1.3.2 describes particle approximation methods. The section closes with a literature review in Subsection 1.3.3.

We work in the framework of (1.2.5). The aim will be to approximate the filtering distributions (recall (1.2.4))

$$\mu_j(\cdot) := \mathbb{P}(v_j \in \cdot | Y_j), \quad j \geq 1.$$

Some of the algorithms approximate, as an intermediate step, the *predictive* distributions

$$\widehat{\mu}_{j+1}(\cdot) := \mathbb{P}(v_{j+1} \in \cdot | Y_j), \quad j \geq 0. \tag{1.3.3}$$

1.3.1 Gaussian Approximation Algorithms

These algorithms rely on the key fact that Bayes' formula can be easily computed when the prior is Gaussian and the observations are linear and Gaussian. More precisely, *suppose* that the predictive distribution was Gaussian, say

$$\widehat{\mu}_{j+1} = N(\widehat{m}_{j+1}, \widehat{C}_{j+1}).$$

Then, using the linearity and Gaussianity of the observation model (1.2.5), is easy to check that the filtering distribution is again Gaussian, say $\mu_{j+1} = N(m_{j+1}, C_{j+1})$. Moreover, its mean m_{j+1} and covariance C_{j+1} are given by the Kalman formulae

$$\begin{aligned} C_{j+1} &= (I - K_{j+1}H)\widehat{C}_{j+1}, \\ m_{j+1} &= \widehat{m}_{j+1} + K_{j+1}(y_{j+1} - H\widehat{m}_{j+1}), \end{aligned} \tag{1.3.4}$$

where K_{j+1} is the so-called Kalman gain

$$K_{j+1} := \widehat{C}_{j+1}H^T(H\widehat{C}_{j+1}H^T + \Gamma)^{-1}. \tag{1.3.5}$$

The predictive distributions $\{\widehat{\mu}_j\}$ are *not* Gaussian except in the case of linear dynamics $\Psi(v) = Mv$. Gaussian approximation algorithms proceed by *approximating* the predictive distribution by a Gaussian,

$$\widehat{\mu}_{j+1} \approx N(\widehat{m}_{j+1}, \widehat{C}_{j+1}). \tag{1.3.6}$$

and then applying Bayes' rule exactly with this (possibly uncontrolled) approximation of the prior. As implied by the previous discussion, this procedure results in a Gaussian approximation of the filtering distribution, with mean and covariance given by (1.3.4)

Different algorithms in the family differ in the way the Gaussian approximation to the predictive distributions (1.3.6) is performed. A very basic description of some of these algorithms follows. The starting point for all of them is a Gaussian approximation of the filtering distribution at time j, $\mu_j \approx N(m_j, C_j)$. The output is a Gaussian approximation, $\mu_{j+1} \approx N(m_{j+1}, C_{j+1})$, of the filtering distribution at time $j+1$.

- Three dimensional variational (3DVAR) method:
  1. Set $\widehat{m}_{j+1} := \Psi(m_j)$, $\widehat{C}_{j+1} := C$, for some C independent of time j.
  2. Compute m_{j+1} and C_{j+1} using (1.3.4) and (1.3.5).
- Extended Kalman filter (ExKF):
  1. Set $\widehat{m}_{j+1} := \Psi(m_j)$, $\widehat{C}_{j+1} := D\Psi(m_j)C_j D\Psi(m_j)^T + Q$.
  2. Compute m_{j+1} and C_{j+1} using (1.3.4) and (1.3.5).
- Ensemble Kalman filter (EnKF):
  1. Generate N samples, $\{v_j^n\}_{n=1}^N$, from $N(m_j, C_j)$.
  2. Propagate them: $\widehat{v}_{j+1}^n := \Psi(v_j^n) + \xi_j^n$, $1 \le n \le N$, $\xi_j^n \sim N(0, Q)$.
  3. Set $\widehat{m}_{j+1} := \frac{1}{N}\sum_{n=1}^N \widehat{v}_{j+1}^n$, $\widehat{C}_{j+1} := \frac{1}{N}\sum_{n=1}^N (\widehat{v}_{j+1}^n - \widehat{m}_{j+1})(\widehat{v}_{j+1}^n - \widehat{m}_{j+1})^T$.
  4. Compute m_{j+1} and C_{j+1} using (1.3.4) and (1.3.5).

Note that the Kalman gain $K = K_{j+1}$ is independent of time j. More elaborate forms of these algorithms are used in practice. In particular, the above description of the EnKF is far from operational implementations. We have two motivations for presenting the algorithms in the basic form above. First, their common underlying structure is highlighted. Second, it will make it easier to explain some central ideas.

The mere formulation of the algorithms above suggests that the main dilemma facing Gaussian approximation algorithms is how to propagate the covariance matrix through the nonlinear dynamics. The 3DVAR method provides the simplest solution: it essentially ignores the problem. The ExKF propagates the covariance through the linearized dynamics. The EnKF uses an empirical approximation.

We now close this brief description of Gaussian filters. We refer to Subsection 1.3.3 for literature review, to Section 1.4 for a discussion on their strengths and weaknesses, and to Chapters 2 and 3 for new theoretical and numerical results on these algorithms.

1.3.2 Particle Approximation Algorithms

These algorithms rely on the key fact that Bayes' formula can be easily computed when the prior is a particle measure and the likelihood can be evaluated. Again we aim for a simple but clear presentation. The input will be a particle approximation of the filtering distribution at time j, $\mu_j \approx \frac{1}{N}\sum_{n=1}^{N}\delta_{v_j^n}$. The output is a particle approximation of the filtering distribution at time $j+1$, $\mu_{j+1} \approx \frac{1}{N}\sum_{n=1}^{N}\delta_{v_j^n}$.

The most simple algorithm is as follows:

1. Generate a particle approximation of the predictive distribution $\widehat{\mu}_{j+1}$:

$$\widehat{v}_{j+1}^n \sim N\big(\Psi(v_j^n), Q\big), \quad 1 \le n \le N, \quad \widehat{\mu}_{j+1} \approx \frac{1}{N}\sum_{n=1}^{N}\delta_{\widehat{v}_{j+1}^n}.$$

2. Use (1.3.2) with prior $\frac{1}{N}\sum_{n=1}^{N}\delta_{\widehat{v}_{j+1}^n}$ and likelihood $N(y_{j+1}; Hv_{j+1}, R)$ to produce a weighted particle approximation of the filtering distribution μ_{j+1}.

3. Sample N times from the distribution obtained in 2 to produce a particle approximation with equal weights.

This algorithm is called the *bootstrap filter*, and is a simple form of Sequential Importance Resampling (SIR) method. The bootstrap filter, as the Gaussian methods of the previous subsection, split the filtering step into two steps:

$$\underbrace{\mu_j}_{:=\mathbb{P}(v_j|Y_j)} \xrightarrow{prediction} \underbrace{\widehat{\mu}_{j+1}}_{:=\mathbb{P}(v_{j+1}|Y_j)} \xrightarrow{analysis} \underbrace{\mu_{j+1}}_{:=\mathbb{P}(v_{j+1}|Y_{j+1})}.$$

First the predictive distribution is approximated, and then a closed form of Bayes' formula is used to assimilate the new observation y_{j+1}. These two steps are called prediction and analysis, respectively. They are naturally built into the Gaussian

approach, since for these methods Bayes' formula needs to be employed with Gaussian prior and linear-Gaussian observations. In contrast, Bayes' formula for particle measures can be used as long as the corresponding likelihood can be evaluated. This opens the possibility of studying particle algorithms associated with other decompositions of the filtering step, such as

$$\underbrace{\mu_j}_{:=\mathbb{P}(v_j|Y_j)} \;\rightarrow\; \mathbb{P}(v_j|Y_{j+1}) \;\rightarrow\; \underbrace{\mu_{j+1}}_{:=\mathbb{P}(v_{j+1}|Y_{j+1})} \quad .$$

The well-known resulting algorithm is often called SIR with optimal proposal. It can be implemented in simple settings such as (1.2.5). However, in more general problems it is usually not implementable, since it requires evaluation of $\mathbb{P}(y_{j+1}|v_j)$ and propagation of conditioned dynamics. We will clarify this point in Chapter 4, where we will also provide new mathematical understanding as to why the optimal proposal algorithm is advantageous when implementable.

1.3.3 Literature Review

Our description of the algorithms is inspired by [Law et al., 2015]. Some novel features of our presentation are: (i) we emphasize how Gaussian and particle algorithms are built on the observation that Bayes' formula is tractable for linear Gaussian models and particle measures, and (ii) our unified description of Gaussian methods, which goes one step further than [Law et al., 2015] by resorting to the basic formulation of the EnKF in [Vanden-Eijnden and Weare, 2012].

Despite its simplicity, the 3DVAR algorithm was successfully used in operational weather forecasting in the 90s and early 00s. Now it has been replaced by a combination of the 4DVAR method [Dimet and Talagrand, 1986] –which includes a temporal dimension as well as the three spatial dimenions in its variational formulation– with the EnKF. 3DVAR was first described in the meteorology and data assimilation literature in [Lorenc, 1986], [Parrish and Derber, 1992]. In control theory terminology 3DVAR can be interpreted as a *nonlinear observer*. The analysis of nonlinear observers in [Thau, 1973], [Tarn and Rasis, 1976] is central for the study of 3DVAR, as described in Chapters 2 and 3 . Excellent references for the ExKF and the EnKF are [Jazwinski, 2007] and [Evensen, 2003], respectively. A popular subclass of Gaussian methods are based on the Unscented transformation [Julier and Uhlmann, 1997]. These methods allow to compute the first moments of the predictive distribution with third order accuracy, by means of a deterministic transformation of *sigma-points.* Their use in large dimensional geophysical applica-

tions has been limited due to the costly growth of the required sigma-points with the dimension of the problem.

Accuracy of the 3DVAR algorithm in tracking the signal has been studied in [Law et al., 2014], [Brett et al., 2013], [Bloemker et al., 2014]. Chapters 2 and 3 will contribute to this topic. The analysis of the EnKF is more subtle: the low-rank empirical approximation of the prediction covariance may cause catastrophic filter divergence, whereby ensemble-state estimates explode to machine infinity. This can occur even for dissipative systems satisfying an absorbing ball property [Harlim and Majda, 2010], [Gottwald and Majda, 2013]. A better understanding of the mechanism that leads to such a catastrophic behaviour of ensemble methods is given in [Kelly et al., 2015], and [Tong et al., 2015]. Catastrophic divergence arises through amplification of energy in the analysis step in situations where the observations lie in the complement of the subspace generated by the ensemble. It can be avoided by using some inflation of the covariance which makes the filter stable [Tong et al., 2015], see also [Kelly et al., 2014].

The particle algorithms in Subsection 1.3.2 are the basis of particle filters, also known as sequential Monte Carlo (SMC) methods. The popularity of these methods has only grown since the introduction of the bootstrap filter in [Gordon et al., 1993]. We refer to [Doucet and Johansen, 2009] for an excellent survey, and to [Doucet et al., 2000]. SMC algorithms have given rise to an extensive body of deep mathematics [Del Moral, 2004]. They have been often claimed to perform poorly in high dimensions [Bengtsson et al., 2008], [Bickel et al., 2008], but are extremely successful in small or moderate dimensional highly nonlinear problems arising in engineering, tracking, finance, etc. In Chapter 4 we will explore their potential use in geophysical applications. Moreover, we will give new mathematical understanding on how the dimensionality of the problem affects the performance of these methods.

Markov Chain Monte Carlo algorithms [Metropolis et al., 1953], [Hastings, 1970], [Liu, 2008] are not treated in this thesis. Contrary to the Gaussian and particle methods described above, MCMC are typically *not* sequential, meaning that the cost of updating the distribution at time j depends on j. A key recent development of the MCMC machinery in this respect is the particle MCMC method proposed in [Andrieu et al., 2010], which combines in nontrivial fashion MCMC and SMC methods. A further development in this line is the SMC^2 method [Chopin et al., 2013]. Note also that many SMC methods make use of MCMC as a building block, see for instance [Kantas et al., 2014].

Particle MCMC methods and SMC^2 are part of the recent trend of combining algorithms. Other examples include: EnKF and particle filters [Frei and Künsch,

2013], [Chustagulprom et al., 2015], [Stordal et al., 2011], 3DVAR and EnKF [Hamill and Snyder, 2000], 4DVAR and EnKF [Zhang et al., 2009], etc. Another trend is to modify sequential methods for their use in static problems: [Chopin, 2002], [Schillings and Stuart, 2016], [Beskos et al., 2015], and [Iglesias et al., 2013].

1.4 Identifying Challenges and Choosing Algorithms

In this section we identify three sources of *computational* challenges in the Bayesian approach: large dimensional systems, small noise regimes, and highly nonlinear problems. This thesis aims to provide a better understanding of, and give a precise meaning to, all these challenges, and to determine conditions under which certain algorithms can overcome them. Computational challenges should be distinguished from theoretical ones. For instance, it is intuitively clear that small observational noise is in theory desirable, since it helps the concentration of the posterior around the unknown of interest. However, small observational noise causes the posterior to be far from the prior, which poses a challenge for many algorithms. Subsection 1.4.1 introduces and briefly describes these challenges, and Subsection 1.4.2 discusses how they affect Gaussian and particle methods. More broadly, Subsection 1.4.2 provides a comparison between these two classes of algorithms, and their relative merits.

1.4.1 Large Dimension, Small Noise, Nonlinearities

Arguably the main computational challenge facing the Bayesian approach today is how to deal with the increasingly large dimension of unknowns of interest, and data sets available. The sheer dimension makes it costly to perform basic operations (e.g.. evaluating functions, storing matrices), and computing gradients is often unfeasible. When implementing discretized versions of problems defined in function spaces, it is important to understand the changes in computational cost and errors of the algorithms as the level of the discretizations are refined. It is also desirable to combine this analysis with a theoretical understanding of the errors introduced by the discretization itself.

A second way in which the dimensionality affects in particular the performance of particle methods is by typically causing the updated measures to be far from the prior measures. This is part of a deep story, and will be the subject of Chapter 4. There we show how a suitable notion of distance between the prior and the posterior affects the performance of importance sampling, which is at the heart of particle methods. Large nominal dimension (defined as the minimum of the dimension of the unknown, and the dimension of the data), small observational noise,

and high regularity of the prior all contribute to moving the posterior away from the prior. In Chapter 4 we define a notion of intrinsic dimension for Bayesian inverse problems that combines all of these ingredients. We will study how the intrinsic dimension, and each of its ingredients, impacts the distance between posterior and prior, and hence the performance of algorithms based on importance sampling.

As well as contributing to increased intrinsic dimension, small noise regimes pose a different algorithmic challenge in filtering settings. Degenerate observation noise, and small or degenerate noise in the dynamics, can compromise the ergodicity, controllability, and observability of the system. As already noted in Subsection 1.2.2.2, ergodicity acts as a dissipation mechanism for algorithmic errors.

Lastly, nonlinearities pose a problem to the propagation of covariances in Gaussian methods. The current theoretical understanding of the errors introduced by these Gaussian approximations in nonlinear settings is far from satisfactory. Moreover, nonlinear systems typically result in multimodal updated measures, which are often harder to approximate.

1.4.2 Brief Comparison of Algorithms

The Gaussian ansatz underlying Gaussian methods is not justified in nonlinear settings. In such scenarios, these methods cannot hope to recover the distribution of interest, but only the first two moments. In Chapters 2 and 3 we will show, however, that even though statistical information is lost when using these methods in highly nonlinear settings, they can reliably *track* the signal, i.e. find the mean of the filtering distribution. The accuracy of particle methods in approximating the whole filtering distribution is –contrary to that of Gaussian methods– not dependent on linear or Gaussian assumptions on the filtering set-up. Indeed, under mild assumptions, it can be shown that these algorithms provide a consistent approximation to the updated measures: any desired accuracy can be in principle achieved by using a large enough number of particles in the approximation. Particle methods are however not well suited for problems were the updated measures are far from the prior. This often happens in large dimensional problems or small noise regimes, as will be made precise in Chapter 4 through the notion of intrinsic dimension.

Among Gaussian methods, the EnKF has two main advantages. First, the nonlinear character of the dynamics is not ignored while propagating the covariance. Second, the empirical approach provides a low-rank approximation of the covariance. This is particularly well suited for high dimensional problems, where the number of samples is usually much lower than the dimension d_v of the signal. For instance, EnKF has proved to be successful in weather forecasting applications, where $\widehat{C}_{j+1}$ is

a matrix of size $d_v \times d_v \approx 10^9 \times 10^9$ and the number N of samples is of the order $N = 10^2$. Thus, $\widehat{C}_{j+1}$ is approximated by a matrix of rank $N \leq 10^2$. When the algorithm is carefully implemented it only needs to store and propagate matrices of size $N \times d_v$ rather than the full covariance of size $10^{d_v \times d_v}$. In contrast, the ExKF requires to store and propagate the full covariance, and to compute a gradient which is often not feasible in high dimensional problems. Our results In Chapters 2 and 3 show that for accurate tracking of the signal it is often enough to assimilate data containing information on the unstable directions of the dynamics. In particular, Chapter 3 shows that tracking can be successful even with extremely small dimensional observation space, provided that at each time step the observations are allowed to align with the unstable parts of the dynamics. In the language of Chapter 4, this implies a dramatic reduction of the intrinsic dimension of the problem.

1.5 Main Contributions

In the following two subsections I summarize some of the contributions of this thesis.

1.5.1 Filtering Chaotic Dynamical Systems

Trajectories of chaotic systems tend to diverge exponentially fast. Any uncertainty on the initial conditions is thus rapidly amplified by the dynamics. However, when observations of the system are available, they may be used to ameliorate this growth in uncertainty and potentially lead to accurate estimates of the state of the system. I have developed [Sanz-Alonso and Stuart, 2015] a unified theory that gives sufficient conditions on the observations of a wide class of dissipative chaotic differential equations that guarantee long-time accuracy of the estimated state variables. Examples include the Lorenz 63 and 96 models, as well as the Navier Stokes equations on a two-dimensional torus. The importance of these model problems within geophysical applications is highlighted in [Majda and Wang, 2006], and their use for testing the efficacy of filtering algorithms is exemplified in [Law and Stuart, 2012], [Majda and Harlim, 2012]. A key ingredient in proving the results was to introduce a new modification of the 3DVAR algorithm from meteorology [Lorenc, 1986], specially tailored to dissipative systems. One of the main aims of that paper was to help to bridge the gap between the control theory and data assimilation communities. I have also contributed to an in-depth numerical and theoretical study of the Lorenz '96 model [Law et al., 2016], and towards exploring the advantages of using adaptive observation operators, that align with the unstable directions of the system. The theory includes discrete time data assimilation and continuous limits. The advan-

tages of using adaptive observations within the 3DVAR algorithm and the ExKF was numerically investigated.

1.5.2 Importance Sampling: Computational Complexity and Intrinsic Dimension

Importance sampling is a simple building block of many state-of-the art sampling algorithms, which has often been claimed to deteriorate in high dimensions and in small noise regimes [Bengtsson et al., 2008], [Ades and Van Leeuwen, 2013]. It is however unclear what should be understood by 'dimension' in the Bayesian framework. The prior infuses information and correlations on the components of the unknown, reducing the number of parameters that are estimated [Agapiou et al., 2015]. I have brought ideas from machine learning and statistics [Zhang, 2002], [Lu and Mathé, 2014] in order to define an *intrinsic* dimension of inverse problems, which can be interpreted as the number of components in which the data substantially changes the prior. It is the intrinsic dimension that affects importance sampling. Crucially, I show that Bayesian inverse problems defined in infinite dimensional Hilbert spaces can potentially have small intrinsic dimension, and importance sampling can be successfully used, as long as the posterior is absolutely continuous with respect to the prior. Further, I established a direct link between absolute continuity and intrinsic dimension, and I found precise rates of degeneracy of importance sampling in terms of intrinsic dimension under various parameter regimes relevant to practitioners. I have explored the implications that these insights have in the context of particle filters, and I have linked our intrinsic dimension to other notions of dimension, such as those in [Chorin and Morzfeld, 2013], [Bickel et al., 2008].

1.6 Ongoing and Future Research

Here is a sketch of some ideas I am currently exploring, and some possible future lines:

- *Gaussian ansatzs.* Laying a mathematical foundation and justification for the Gaussian ansatzs employed by many data assimilation algorithms. In this direction I have started to investigate Gaussian approximations to nonlinear stochastic differential equations. A thorough investigation of this subject will provide a much needed understanding of the validity of these algorithms, possibly in terms of the frequency of the observations, the nonlinearities of the system, and the size of the observation noise.

- *Large dimension, small noise, nonlinearities.* It would be interesting to generalize our notion of intrinsic dimension, which currently assumes linearity, to nonlinear systems. A related idea is the design of algorithms that automatically detect the coordinates of the system in which the data is informative. This is of key importance in the function space setting, where truncating the measures is needed in numerical computations, and there is a risk that by truncating in the wrong place much information is lost.

- *Continuous time limits.* Investigating the long-time behaviour of the filtering distributions for partially observed chaotic dynamical systems under continuous-time observation limits.

Chapter 2

Filter Accuracy for Chaotic Dynamical Systems: a General Framework

2.1 Introduction

The evolution of many physical systems can be successfully modelled by a deterministic dynamical system for which the initial conditions may not be known exactly. In the presence of chaos, uncertainty in the initial conditions will be dramatically amplified even in short time-intervals. However, when observations of the system are available, they may be used to ameliorate this growth in uncertainty and potentially lead to accurate estimates of the state of the system. In this work we provide sufficient conditions on the observations of a wide class of dissipative chaotic differential equations that guarantee long-time accuracy of the estimated state variables. The equations covered by our theory include the Lorenz '63 and '96 models as well as the Navier Stokes equation on a two-dimensional torus. The importance of these model problems within geophysical applications is highlighted in [Majda and Wang, 2006], and their use for testing the efficacy of filtering algorithms is exemplified in [Majda and Harlim, 2012; Law and Stuart, 2012].

It is often natural to acknowledge the uncertainty on the initial condition by viewing it as a probability distribution which is propagated by the dynamics. Whenever a new observation of the state variables becomes available, this distribution is updated to incorporate it, reducing uncertainty. This process is performed sequentially in what is known as filtering [Crisan and Rozovskii, 2011]. Unfortunately, in almost all situations of applied relevance —with the exception of finite state

signals and the linear Gaussian case— the analytical expression for these *filtering distributions* involves integrals that cannot be computed in closed form. It is thus necessary to employ a numerical algorithm to sequentially approximate the filtering distributions. In order to develop good algorithms a thorough understanding of the properties of these distributions is desirable. The interplay between properties of the filtering distributions and those of their numerical approximations is perhaps best exemplified by the case of filter stability and particle filtering: the long-time behaviour of particle filtering algorithms depends crucially on the sensitivity of filtering distributions to their initial condition [Del Moral, 2004], [Crisan and Heine, 2008], [Lei and Bickel, 2013]. The main result of this paper shows long-time concentration of the filtering distributions towards the true underlying signal for partially observed chaotic dynamics. The proofs combine the asymptotic boundedness of a new suboptimal filter with the mean-square optimality of the mean of the filtering distribution as an estimator of the signal [Williams, 1991]. All our examples rely on synchronization properties of dynamical systems. This tool underlies the study of noise-free data assimilation initiated in [Hayden et al., 2011] for the Lorenz '63 and the Navier Stokes equation. The paper [Hayden et al., 2011] motivated studies of the 3DVAR filter (three-dimensional variational method) for a variety of dissipative chaotic dynamical systems, conditioned on noisy observations, in [Brett et al., 2013] (Navier Stokes), [Law et al., 2014] (Lorenz '63) and [Law et al., 2016] (Lorenz '96). The 3DVAR filter from meteorology [Lorenc, 1986; Parrish and Derber, 1992] is a method which, iteratively in time, solves a quadratic minimization problem representing a compromise between matching the model and the data. Here we study the filtering distribution itself, using modifications of the 3DVAR filter which exploit dissipativity to obtain upper bounds on the error made by the optimal filter. We also provide a unified methodology for the analysis. Furthermore, whereas previous work in [Brett et al., 2013], [Law et al., 2016] required the observation noise to have bounded support, here only finite variance is assumed.

The suboptimal modified 3DVAR filter that we use in our analysis can also be interpreted using ideas from nonlinear observer theory [Thau, 1973], [Tarn and Rasis, 1976]. Its asymptotic boundedness is proved by a Lyapunov-type argument. Although more sophisticated suboptimal filters could be used to gain insight on the filtering distributions, our choice of modified nonlinear observers is particularly well-suited to deal with high (possibly infinite) dimensional signals, as indicated by the fact that the theory includes the Navier-Stokes equation. Filtering in high dimensions is not, in general, well-understood. For example, the question of whether some form of particle filtering could be robust with respect to dimension has received

much recent attention [Snyder et al., 2008], [Rebeschini and van Handel, 2015], [Beskos et al., 2014a]. By understanding properties of the filtering distribution in high and infinite dimensions we provide insight that may inform future development of particle filters.

The chapter is organized as follows. In Section 2.2 we set up the notation and formulate the questions we address in the rest of the chapter. Section 2.3 reviews the 3DVAR algorithm from data assimilation and its relation to more general nonlinear observers from the control theory literature. A new truncated nonlinear observer is also introduced. In Section 2.4 we prove long-time asymptotic results for these suboptimal filters, and thereby deduce long-time accuracy of the filtering distributions. Section 2.5 contains some applications to relevant models and we close in Section 2.6.

2.2 Set-up

Filtering problems are naturally formulated within the framework of Hidden Markov Models. The general setting that we consider is that of a Markov chain $\{v_j, y_j\}_{j\geq 0}$, where $\{v_j\}_{j\geq 0}$ is the *signal* process, and $\{y_j\}_{j\geq 0}$ is the *observation* process. We assume throughout $y_0 = 0$ so that y_0 gives no information on the initial value of the signal and that, for each $j \geq 1$, y_j is a noisy observation of v_j. We are interested in the value of the signal, but have access only to outcomes of the observation process. We suppose that both take values in a separable Hilbert space $\mathcal{H} = (\mathcal{H}, \langle\cdot,\cdot\rangle, |\cdot|)$ and that the signal is randomly initialized with distribution μ_0, $v_0 \sim \mu_0$. We assume further that there is a *deterministic* map Ψ such that

$$v_{j+1} = \Psi(v_j), \quad \text{for } j \geq 0, \tag{2.2.1}$$

and therefore all the randomness in the signal comes from its initialization.

The observation process is given by

$$y_j = Pv_j + \epsilon w_j, \quad \text{for } j \geq 1, \tag{2.2.2}$$

where P denotes some linear operator that projects the signal onto a proper subspace of $\mathcal{H}$, $\{w_j\}_{j\geq 1}$ is an i.i.d. noise sequence —independent of v_0— and $\epsilon > 0$ quantifies the strength of the noise. We define $Q = I - P$. For mathematical convenience, and contrary to usual convention, we see both observations and noise as taking values in the same space $\mathcal{H}$ as the signal, with the standing assumptions $Qy_j = 0$, $Qw_1 = 0$

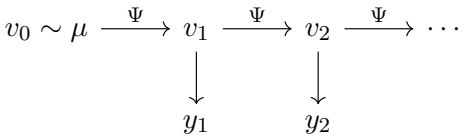


Figure 2.1: *Graphic representation of the dependence structure assumed throughout this chapter. Conditional on $v_0, \ldots, v_j$, the distribution of v_{j+1} is completely determined by v_j via a deterministic map Ψ; therefore the signal process forms a Markov chain. Similarly, conditional on $\{v_j\}_{j\geq 0}$, $\{y_j\}_{j\geq 1}$ is a sequence of independent random variables such that the conditional distribution of y_j depends only on v_j.*

and $Pw_1 = w_1$ a.s.[1] Thus Q is a projection operator onto the unobserved part of the system. For $j \geq 0$, we let $Y_j := \sigma(y_i,\, i \leq j)$ be the σ–algebra generated by the observations up to the discrete time j.

Note that the law of $\{v_j, y_j\}_{j\geq 0}$ is completely determined by four elements: the law of v_0, the map Ψ, the law of w_1 and the observation operator P. We will denote by $\mathbb{P}$ the law of $\{v_j, y_j\}_{j\geq 0}$ and by $\mathbb{E}$ the corresponding expectation. It will be assumed throughout that $\mathbb{E}|v_0|^2 < \infty$ and that the observation noise satisfies $\mathbb{E}w_1 = 0$ and $\mathbb{E}|w_1|^2 < \infty$. For convenience and without loss of generality we normalize the latter so that $\mathbb{E}|w_1|^2 = 1$.

The main object of interest in filtering theory are the conditional distributions of the signal at discrete time $j \geq 1$ given all observations up to time j. These are known as *filtering distributions* and will be denoted by

$$\mu_j(\cdot) := \mathbb{P}\big[v_j \in \cdot | Y_j\big].$$

The mean $\widehat{v}_j$ of the filtering distribution μ_j is known as the *optimal filter*

$$\widehat{v}_j := \mathbb{E}[v_j|Y_j] = \int_{\mathcal{H}} v\, \mu_j(dv).$$

By the mean-square minimization property of the conditional expectation [Williams, 1991], this filter is optimal in the sense that, among all Y_j-measurable random variables, it is the only one —up to equivalence— that minimizes the L^2 distance

[1] More generally, when an operator $T : \mathcal{H} \to \mathcal{H}$ acts on the observations it should be implicitly understood that $T : \mathcal{H} \to \mathcal{H}$ satisfies $T = PT$. Moreover it will often be assumed that $T\big|_P : P\mathcal{H} \to P\mathcal{H}$ is positive definite and then the operator $T^{-1} : \mathcal{H} \to \mathcal{H}$ should be interpreted as satisfying $PT^{-1} = T^{-1}\big|_P$, $QT^{-1} \equiv 0$.

to the signal v_j:

$$\mathbb{E}|v_j - \widehat{v}_j|^2 \leq \mathbb{E}|v_j - z_j|^2, \quad \text{for all } Y_j\text{–measurable } z_j. \tag{2.2.3}$$

In words, $\widehat{v}_j$ is the best possible estimator (in the mean-square sense) of the state of the signal at time j given information up to time j. The optimal filter is usually, like the filtering distributions, not analytically available. However, by studying suitable suboptimal filters $\{z_j\}_{j\geq 0}$ and using (2.2.3) we can find sufficient conditions under which the optimal filter is close to the signal in the long-time horizon. We thus provide sufficient conditions under which the observations counteract the potentially chaotic behaviour of the dynamical system, and allow predictability on infinite time-horizons.

The main objective of this chapter is to investigate the long-time asymptotic behaviour of the filtering distribution for discrete-time chaotic signals, arising from the solution to a dissipative quadratic system with energy-conserving nonlinearity

$$\frac{dv}{dt} + Av + B(v, v) = f, \tag{2.2.4}$$

which is observed at discrete times $t_j = jh$, $j \geq 1$, $h > 0$. The bilinear form $B(\cdot, \cdot)$ will be assumed throughout to be symmetric. We denote by Ψ_t the one-parameter solution semigroup associated with (2.2.4), i.e. for $v_0 \in \mathcal{H}$, $\Psi_t(v_0)$ is the value at time t of the solution to (2.2.4) with initial condition v_0. Furthermore we introduce the abbreviation $\Psi = \Psi_h$.

Our theory —developed in Section 2.4— relies on two assumptions that we now state and explain.

Assumptions 2.2.1.

*1. (***Absorbing ball property***.) There are constants $r_0, r_1 > 0$ such that*

$$|\Psi_t(v_0)|^2 \leq \exp(-r_1 t)|v_0|^2 + r_0\big(1 - \exp(-r_1 t)\big), \qquad t \geq 0. \tag{2.2.5}$$

Therefore, setting $r = \sqrt{2r_0}$, the ball $\mathcal{B} := \{u \in \mathcal{H} : |u| \leq r\}$ is absorbing and forward invariant for the dynamical system (2.2.1).

*2. (***Squeezing property***.) There is a function $V : \mathcal{H} \to [0, \infty)$ such that $V(\cdot)^{1/2}$ is a Hilbert norm equivalent to $|\cdot|$, a bounded operator D, an absorbing set $\mathcal{B}_V = \{u \in \mathcal{H} : V(u)^{1/2} \leq R\} \supset \mathcal{B}$, and a constant $\alpha \in (0, 1)$ such that, for all $u \in \mathcal{B}$, $v \in \mathcal{B}_V$,*

$$V\Big((I - DP)\big(\Psi(v) - \Psi(u)\big)\Big) \leq \alpha V(v - u).$$

The absorbing ball property concerns only the signal dynamics. It is satisfied by many dissipative models of the form (2.2.4) —see Section 2.5. The squeezing property involves both the signal dynamics and the observation operator P. It is satisfied by several problems of interest provided that the assimilation time h is sufficiently small and that the 'right' parts of the system are observed; see again Section 2.5 for examples. We remark that several forms of the squeezing property can be found in the dissipative dynamical systems literature. They all refer to the existence of a contracting part of the dynamics. Their importance for filtering has been explored in [Hayden et al., 2011], [Brett et al., 2013] and [Chueshov, 2014]. It also underlies the analysis in [Kelly et al., 2014] and [Law et al., 2016], as we make apparent here. We have formulated the squeezing property to suit our analyses and with the intention of highlighting the similar role that it plays to detectability for linear problems, as explained in Subsection 2.4.2. The function V will represent a Lyapunov type function in Section 2.4. For all the chaotic examples in Section 2.5 the operator D will be chosen as the identity, but other choices are possible. As we shall see, the absorbing ball property is not required when a global form of the squeezing property, as may arise for linear problems, is satisfied.

We will construct suboptimal filters $\{m_j\}_{j\geq 0}$ that are forced to lie in $\mathcal{B}_V$. By the absorbing ball property the signal v_j is contained, for large j and with high probability, in the forward-invariant ball $\mathcal{B}$. Therefore, intuitively, the squeezing property can be applied, for large j, to $m_j \in \mathcal{B}_V$, $v_j \in \mathcal{B}$.

The main result of the chapter, Theorem 2.4.8, shows that, when Assumption 2.2.1 holds, the optimal filter accurately tracks the signal. Specifically we show that there is a constant $c > 0$, independent of the noise strength ϵ, such that

$$\limsup_{j\to\infty} \mathbb{E}|v_j - \widehat{v}_j|^2 \leq c\epsilon^2. \tag{2.2.6}$$

Note that (2.2.6) not only guarantees that in the low noise regime the optimal filter (i.e. the mean of the filtering distribution) is —on average— close to the signal, but also that the variance of the filtering distribution is —on average— small. Indeed, since

$$\mathrm{var}[v_j\big|Y_j] = \mathbb{E}\left[(v_j - \widehat{v}_j)\otimes(v_j - \widehat{v}_j)\middle|Y_j\right]$$

it follows by taking expectations and using linearity of the trace operator that

$$\mathrm{Trace}\,\mathbb{E}\,\mathrm{var}[v_j\big|Y_j] = \mathbb{E}|v_j - \widehat{v}_j|^2,$$

and therefore (2.2.6) implies

$$\limsup_{j\to\infty} \text{Trace}\, \mathbb{E}\, \text{var}[v_j|Y_j] \leq c\epsilon^2.$$

We hence see that (2.2.6) guarantees that the variance of the filtering distributions scales as the size of the observation noise, like $\mathcal{O}(\epsilon^2)$. Thus the initial uncertainty in the initial condition which is $\mathcal{O}(1)$ is reduced, in the large-time asymptotic, to uncertainty of $\mathcal{O}(\epsilon)$: the observations have overcome the effect of chaos. Small variance of the long-time filtering distribution had been previously proposed as a condition for successful data assimilation [Chorin and Morzfeld, 2013].

2.3 Suboptimal Filters

The aim of this section is to introduce a suboptimal filter, designed to track dynamics satisfying Assumption 2.2.1. This filter is based on the 3DVAR algorithm from data assimilation, and nonlinear observers from control applications. We give the necessary background on these in Subsection 2.3.1 before introducing the new filter in Subsection 2.3.2.

2.3.1 3DVAR Filter

The 3DVAR filter approximates the filtering distribution μ_{j+1} by a Gaussian $N(z_{j+1}, C)$ whose mean can be found recursively starting from a deterministic point $z_0 \in \mathcal{H}$ by solving the variational problem

$$z_{j+1} := \text{argmin}_z \left\{ \frac{1}{2} \left| C_\sharp^{-1/2}\big(z - \Psi(z_j)\big) \right|^2 + \frac{1}{2\epsilon^2} \left| \Gamma^{-1/2}(y_{j+1} - Pz) \right|^2 \right\}, \quad (2.3.1)$$

where $C_\sharp$ is a fixed model covariance that represents the lack of confidence in the model Ψ, and Γ is the covariance operator of the observation noise w_1.

The covariance C of the 3DVAR filter is determined by the Kalman update formula

$$C^{-1} = C_\sharp^{-1} + P^T \Gamma^{-1} P.$$

It is immediate from (2.3.1) that z_j is Y_j-measurable for all $j \geq 0$, and it can be shown [Law et al., 2015] that the solution z_{j+1} to this variational problem satisfies

$$z_{j+1} = (I - KP)\Psi(z_j) + Ky_{j+1}, \quad (2.3.2)$$

where K is the Kalman gain

$$K = C_\sharp P^T (P C_\sharp P^T + \epsilon^2 \Gamma)^{-1}.$$

The 3DVAR filter was introduced, and has been widely applied, in the meteorological sciences [Parrish and Derber, 1992; Lorenc, 1986]. Long-time asymptotic stability and accuracy properties —that guarantee that the means z_j become close to the signal v_j— have recently been studied for the Lorenz '63 model [Law et al., 2014] subject to additive Gaussian noise, and the Lorenz '96 and Navier-Stokes equation observed subject to bounded noise [Law et al., 2016], [Brett et al., 2013].

It will be convenient to allow for other choices of operator K in the above definition, and consider the more general recursion

$$z_{j+1} = (I - DP)\Psi(z_j) + D y_{j+1}, \tag{2.3.3}$$

where D is some linear operator that we are free to choose as desired. Filters of the form (2.3.3) are known as nonlinear observers [Thau, 1973], [Tarn and Rasis, 1976]. The 3DVAR filter can be seen as an instance of these where the operator D is determined by model and noise covariances, and by the observation operator. We now derive a recursive formula for the error made by nonlinear observers when approximating the signal. To that end note, firstly, that the signal $\{v_j\}_{j\geq 0}$ satisfies

$$v_{j+1} = (I - DP)\Psi(v_j) + DP\Psi(v_j).$$

Secondly, using (2.2.2) at time $j+1$, combined with the assumption that $Pw_{j+1} = w_{j+1}$,

$$z_{j+1} = (I - DP)\Psi(z_j) + DP\Psi(v_j) + \epsilon DP w_{j+1}$$

Therefore, substracting the previous two equations, we obtain that the error $\delta_j := v_j - z_j$ satisfies

$$\delta_{j+1} = (I - DP)\big(\Psi(v_j) - \Psi(z_j)\big) - \epsilon DP w_{j+1}. \tag{2.3.4}$$

Despite their simplicity nonlinear observers are known to accurately track the signal under suitable conditions [Tarn and Rasis, 1976; Thau, 1973]. Equation (2.3.4) plays a central role in such analysis, and will underlie our analysis too. It demonstrates the importance of the operator $(I - DP)\Psi$ in the propagation of error; this operator combines the properties of the dynamical system, encoded in Ψ, with the properties of the observation operator P.

2.3.2 Nonlinear Observers and Truncated Nonlinear Observers

In the remainder of this section we introduce a truncated nonlinear observer that is especially tailored to exploit the absorbing ball property of the underlying dynamics.

Given a non-empty closed convex subset $\mathcal{C} \subset \mathcal{H}$, take $m_0 \in \mathcal{C}$ and, for $j \geq 0$, define the *$\mathcal{C}$-truncated nonlinear observer* m_{j+1} by

$$m_{j+1} := P_{\mathcal{C}}\Big((I - DP)\Psi(m_j) + Dy_{j+1}\Big), \tag{2.3.5}$$

where $P_{\mathcal{C}}$ is the orthogonal (with respect to a suitable inner product) projection operator onto the set $\mathcal{C}$; this is well-defined for any non-empty closed convex set [Rudin, 1987]. In the next section we will analyze the long-time behaviour of this filter when $\mathcal{C}$ is chosen as $\mathcal{B}_V$ and the inner product is the one induced by $V^{1/2}$ (see Assumption 2.2.1). The main advantage of this truncated filter is that $m_j \in \mathcal{B}_V$ for all $j \geq 0$, and large uninformative observations y_j corresponding to large realizations of the observation noise w_j will not hinder the performance of the filter. Examples of other truncated stochastic algorithms can be found in [Kushner and Yin, 2003].

2.4 Stochastic Stability of Suboptimal Filters and Filter Accuracy

In this section we prove long-time accuracy of certain suboptimal filters under different assumptions on the underlying dynamics and observation model. These results are used to establish long-time concentration of the filtering distributions. We start in Subsection 2.4.1 by recalling the Lyapunov method for proving asymptotic boundedness of stochastic algorithms. In Subsection 2.4.2 we employ this method to show asymptotic accuracy of nonlinear observers when a global form of the squeezing property is satisfied, as happens for certain linear problems. Finally, in Subsection 2.4.3 we use truncated nonlinear observers to deal with chaotic models where only the weaker Assumption 2.2.1 holds.

2.4.1 The Lyapunov Method for Stability of Stochastic Filters

Consider a Markov chain $\{\delta_j\}_{j\geq 0}$ and think of it as the random sequence of errors made by some filtering procedure. The next result, from [Tarn and Rasis, 1976], underlies much of the analysis in the following subsections.

Lemma 2.4.1. *Let δ_j^1 and δ_j^2 be two realizations of the $\mathcal{H}$-valued random variable δ_j and set $\Delta_j = \delta_j^1 - \delta_j^2$. Suppose that there is a function $V : \mathcal{H} \to [0, \infty)$ such that*

1. $V(0) = 0, \qquad V(x) \geq \theta|x|^2$ *for all* $x \in \mathcal{H}$ *and some* $\theta > 0$.

2. *There are real numbers* $K > 0$ *and* $\alpha \in (0,1)$ *such that, for all* $\Delta_j \in \mathcal{H}$,

$$\mathbb{E}[V(\Delta_{j+1})|\Delta_j] \leq K + \alpha V(\Delta_j).$$

Then, for any $a \in \mathcal{H}$,

$$\theta\mathbb{E}[|\Delta_j|^2|\Delta_0 = a] \leq \alpha^j V(a) + K\sum_{i=0}^{j-1}\alpha^i.$$

Therefore, regardless of the initial state Δ_0,

$$\limsup_{j\to\infty}\mathbb{E}|\Delta_j|^2 \leq \frac{K}{\theta(1-\alpha)}.$$

2.4.2 Filter Accuracy with Global Squeezing Property

The following results show that if, for some suitable operator D, the map $(I-DP)\Psi$ satisfies a global Lipschitz condition, then it is possible to use nonlinear observers to deduce long-time accuracy of the filtering distributions. Although such a global condition does not typically hold for dissipative chaotic dynamical systems arising in applications, the following discussion serves as a motivation for the more general theory in Subsection 2.4.3. Moreover, the results in this subsection are of interest in their own right. In particular they are enough to deal with the important case of linear signal dynamics.

Theorem 2.4.2. *Assume that there is a Hilbert norm* $V(\cdot)^{1/2}$ *in* $\mathcal{H}$, *equivalent to* $|\cdot|$, *and a bounded operator* D *and constant* $\alpha \in (0,1)$ *such that*

$$V\Big((I-DP)\big(\Psi(v)-\Psi(u)\big)\Big) \leq \alpha V(v-u) \quad \forall u, v \in \mathcal{H}.$$

Define $\{z_j\}_{j\geq 0}$ *by (2.3.3). Then there is a constant* $c > 0$, *independent of the noise strength* ϵ, *such that*

$$\limsup_{j\to\infty}\mathbb{E}|v_j - z_j|^2 \leq c\epsilon^2.$$

Proof. By assumption V satisfies the first condition in Lemma 4.1. Set $\delta_j = v_j - z_j$.

Then, using equation (2.3.4) and the independence structure,

$$\begin{aligned}\mathbb{E}[V(\delta_{j+1})|\delta_j] &= \mathbb{E}\Big[V\Big((I-DP)\big(\Psi(v_j)-\Psi(z_j)\big)-\epsilon Dw_{j+1}\Big)\Big|\delta_j\Big] \\ &= \mathbb{E}\Big[V\Big((I-DP)\big(\Psi(v_j)-\Psi(z_j)\big)\Big)\Big|\delta_j\Big] + \epsilon^2\mathbb{E}V(Dw_{j+1}) \\ &\leq \mathbb{E}\Big[V\Big((I-DP)\big(\Psi(v_j)-\Psi(z_j)\big)\Big)\Big|\delta_j\Big] + C\epsilon^2 \\ &\leq \alpha V(\delta_j) + C\epsilon^2,\end{aligned}$$

where $C > 0$ is independent of ϵ and to obtain the first inequality we used equivalence of norms and the fact that D is bounded. Thus the second condition in Lemma 4.1 holds and the proof is complete. □

The following corollary is an immediate consequence of the L^2 optimality property of the optimal filter (2.2.3).

Corollary 2.4.3. *Under the hypothesis of the previous theorem*

$$\limsup_{j\to\infty}\mathbb{E}|v_j-\widehat{v}_j|^2 \leq c\epsilon^2, \qquad \limsup_{j\to\infty}\mathrm{Trace}\,\mathbb{E}\,\mathrm{var}[v_j\big|Y_j] \leq c\epsilon^2.$$

In the remainder of this subsection we apply, for the sake of motivation, the previous theorem to the case of linear finite dimensional dynamics. We take $\mathcal{H} = \mathbb{R}^d$ and let the signal be given by

$$v_{j+1} = Lv_j, \quad j \geq 1, \qquad v_0 \sim \mu_0. \tag{2.4.1}$$

This framework has been widely studied within the control theory community, mostly —but not exclusively— in the case where both the initial distribution of the signal and the observation noise are Gaussian. Other than its modelling appeal, this *linear Gaussian* setting has the exceptional feature that the filtering distributions are themselves again Gaussian. Moreover, their means and covariances can be iteratively computed using the Kalman filter [Kalman, 1960]. Since the optimal filter is the mean of the filtering distribution, the explicit characterization of the Kalman filter yields an explicit characterization of the optimal filter. Suppose that, for some given $\widehat{v}_0 \in \mathbb{R}^d$ and $C_0 \in \mathbb{R}^{d\times d}$, $\mu_0 = N(\hat{v}_0, C_0)$ and suppose further that $w_1 \sim N(0,\Gamma)$. Then the filtering distributions are Gaussian, $\mu_j = N(\widehat{v}_j, C_j)$, $j \geq 1$, and the means and covariances satisfy the recursion (see [Law et al., 2015])

$$\begin{aligned}\widehat{v}_{j+1} &= (I-K_{j+1}P)L\widehat{v}_j + K_{j+1}y_{j+1}, \\ C_{j+1}^{-1} &= C_{j+1|j}^{-1} + \epsilon^{-2}P^T\Gamma^{-1}P,\end{aligned} \tag{2.4.2}$$

where the *predictive Kalman covariance* $C_{j+1|j}$ and Kalman gain K_{j+1} are given by

$$\begin{aligned} C_{j+1|j} &= LC_jL^T, \\ K_{j+1} &= C_{j+1|j}P^T(PC_{j+1|j}P^T + \epsilon^2\Gamma)^{-1}. \end{aligned}$$

Similar formulae are available when the covariance operator Γ is not invertible in the observation space [Law et al., 2015].

Remark 2.4.4. *It is clear from (2.4.2) that the Kalman filter covariance C_j, which is the covariance of the filtering distribution μ_j, is deterministic and in particular does not make use of the observations. It follows from the discussion in Section 2.2 that in the linear Gaussian setting*

$$\limsup_{j\to\infty} \mathbb{E}|v_j - \widehat{v}_j|^2 \leq c\epsilon^2$$

implies

$$\limsup_{j\to\infty} \text{Trace}\, C_j \leq c\epsilon^2.$$

In the linear setting the global squeezing property in Theorem 2.4.2 reduces to the control theory notion of detectability, that we now recall.

Definition 2.4.5. *The pair (L, P) is called* detectable *if there exists a matrix D such that $\rho(L - DP) < 1$, where $\rho(\cdot)$ denotes spectral radius.*

We remark that the condition $\rho(L - DP) < 1$ guarantees the existence of a Hilbert norm in $\mathbb{R}^d$ in which the linear map defined by the matrix $L - DP$ is contractive. It therefore yields a global form of the squeezing property. Note that detectability may hold for unstable dynamics with $\rho(L) > 1$. However the observations need to contain information on the unstable directions. It is not necessary that these are directly observed, but only that we can retrieve information from them by exploiting any rotations present in the dynamics. This is the interpretation of the matrix D in the definition. The next result states the abstract global theorem of the previous section in the setting of linear dynamics. Our aim in including it here is to make apparent the connection between classical control theory [Lancaster and Rodman, 1995], ideas from data assimilation concerning the 3DVAR filter [Brett et al., 2013], [Kelly et al., 2014], [Law et al., 2016], and the new results for chaotic systems observed with unbounded noise in Section 2.4.3.

Theorem 2.4.6. *Assume that $\mathcal{H} = \mathbb{R}^d$ and $\Psi(v) = Lv$ with $L \in \mathbb{R}^{d\times d}$. Then if the pair (L, P) is detectable there is a constant $c > 0$ independent of the noise strength*

ϵ, *such that*

$$\limsup_{j\to\infty} \mathbb{E}|v_j - \widehat{v}_j|^2 \le c\epsilon^2,$$

and consequently in the linear Gaussian setting

$$\limsup_{j\to\infty} \operatorname{Trace} C_j \le c\epsilon^2.$$

Proof. By the Hautus lemma [Sontag, 1998] the pair (L, P) is detectable if and only if

$$\operatorname{Rank} \begin{pmatrix} \lambda I - L \\ P \end{pmatrix} = d$$

for all λ with $|\lambda| \ge 1$ or, in other words, if $\operatorname{Ker}(\lambda I - L) \cap \operatorname{Ker}(P) = \{0\}$ for all λ with $|\lambda| \ge 1$. Using this characterization of detectability it is immediate from the identity

$$\operatorname{Ker}(\lambda I - L) \cap \operatorname{Ker}(PL) = \operatorname{Ker}(\lambda I - L) \cap \operatorname{Ker}(P), \quad \lambda \neq 0,$$

that (L, P) is detectable iff (L, PL) is detectable. Now by hypothesis (L, P) is detectable and so there exists a matrix D such that $\rho\big((I - DP)L\big) < 1$. Hence the linear map defined in $\mathbb{R}^d$ by the matrix $(I - DP)L$ is globally contractive in some Hilbert norm. The result follows from Theorem 2.4.2 and Corollary 2.4.3. □

2.4.3 Filter Accuracy for Chaotic Deterministic Dynamics

In this section we study filter accuracy for signals satisfying Assumption 2.2.1. Our analysis now makes use of *truncated* nonlinear observers (2.3.5), which are forced to lie in the absorbing ball $\mathcal{B}_V$. The idea is that, once the signal gets into the absorbing ball, projecting the filter into $\mathcal{B}_V$ reduces the distance from the signal, as measured by the Lyapunov function V. This is the content of the following lemma. $P_{\mathcal{B}_V} x$ denotes the closest point (in the $V^{1/2}$ norm) to $x \in \mathcal{H}$ in the set $\mathcal{B}_V$. Therefore, $P_{\mathcal{B}_V} x = R^{1/2} \frac{x}{V^{1/2}(x)}$ for $x \notin \mathcal{B}_V$.

Lemma 2.4.7. *Let $V^{1/2}(\cdot)$ be a Hilbert norm and let $R > 0$ be arbitrary. Set $\mathcal{B}_V := \{b \in \mathcal{H} : V(b) \le R\}$ similarly as in Assumption 2.2.1. Then,*

$$V(P_{\mathcal{B}_V} x - b) \le V(x - b), \qquad x \in \mathcal{H}, b \in \mathcal{B}_V. \tag{2.4.3}$$

Proof. The case $x \in \mathcal{B}_V$ is obvious so we assume $V(x) > R$. Let $\langle \cdot, \cdot \rangle_V$ denote the inner product associated with the norm $V^{1/2}$. We claim that

$$\langle P_{\mathcal{B}_V} x - b,\, x - P_{\mathcal{B}_V} x \rangle_V \ge 0. \tag{2.4.4}$$

Indeed we have

$$\begin{aligned}\langle P_{\mathcal{B}_V}x-b,\,x-P_{\mathcal{B}_V}x\rangle_V &= \left\langle R^{1/2}\frac{x}{V^{1/2}(x)}-b,\,x-R^{1/2}\frac{x}{V^{1/2}(x)}\right\rangle_V\\ &= \left(1-\frac{R^{1/2}}{V^{1/2}(x)}\right)\left\langle R^{1/2}\frac{x}{V^{1/2}(x)}-b,x\right\rangle_V\\ &= \left(1-\frac{R^{1/2}}{V^{1/2}(x)}\right)\left(R^{1/2}V^{1/2}(x)-\langle b,x\rangle_V\right)\\ &\geq \left(1-\frac{R^{1/2}}{V^{1/2}(x)}\right)\left(R^{1/2}V^{1/2}(x)-V^{1/2}(b)V^{1/2}(x)\right).\end{aligned}$$

Now, $R \geq V(b)$ because $b \in \mathcal{B}_V$ and the claim is proved.

Finally note that (2.4.4) implies $V(P_{\mathcal{B}_V}x-b) \leq V(x-b)$. To see this recall the elementary fact that for arbitrary $x_1, x_2 \in \mathcal{H}$ we have that $\langle x_1, x_2\rangle_V \geq 0$ implies $V(x_1) \leq V(x_1+x_2)$ and choose $x_1 := P_{\mathcal{B}_V}x-b$ and $x_2 := x-P_{\mathcal{B}_V}x$. □

Using the fact established in Lemma 2.4.7 we are now in a position to prove positive results about the truncated nonlinear observer, and hence the optimal filter, in the long-time asymptotic regime.

Theorem 2.4.8. *Suppose that Assumption 2.2.1 holds. Let $\{m_j\}_{j\geq 0}$ be the sequence of $\mathcal{B}_V$-truncated nonlinear observers given by (2.3.5). Then there is a constant $c>0$, independent of the noise strength ϵ, such that*

$$\limsup_{j\to\infty}\mathbb{E}|v_j-m_j|^2 \leq c\epsilon^2.$$

Consequently,

$$\limsup_{j\to\infty}\mathbb{E}|v_j-\widehat{v}_j|^2 \leq c\epsilon^2, \qquad \limsup_{j\to\infty}\mathrm{Trace}\,\mathbb{E}\,\mathrm{var}[v_j\big|Y_j] \leq c\epsilon^2.$$

Proof. By Lemma 2.4.9 below, for arbitrary $\delta>0$ there is $J>0$ such that, for every $j \geq J$,

$$\int_{\{v_J\notin\mathcal{B}\}} V(v_j-m_j)d\mathbb{P} < \delta. \tag{2.4.5}$$

Now, for $j \geq J$ we have by the absorbing ball property that $v_J \in \mathcal{B}$ implies that

$v_{j+1} \in \mathcal{B}$, and hence by Lemma 2.4.7

$$\begin{aligned}
&\int_{\{v_J \in \mathcal{B}\}} V(v_{j+1} - m_{j+1}) d\mathbb{P} \\
&\qquad \leq \int_{\{v_J \in \mathcal{B}\}} V\Big((I - DP)\big(\Psi(v_j) - \Psi(m_j)\big) - \epsilon D w_{j+1}\Big) d\mathbb{P} \\
&\qquad = \int_{\{v_J \in \mathcal{B}\}} V(\epsilon D w_{j+1}) d\mathbb{P} + \int_{\{v_J \in \mathcal{B}\}} V\Big((I - DP)\big(\Psi(v_j) - \Psi(m_j)\big)\Big) d\mathbb{P} \\
&\qquad\qquad - 2 \int_{\{v_J \in \mathcal{B}\}} \Big\langle \epsilon w_{j+1}, (I - DP)\big(\Psi(v_j) - \Psi(m_j)\big) \Big\rangle_V d\mathbb{P}.
\end{aligned}$$

Using the independence structure the last term vanishes, and for the second term we can employ the squeezing property with $v_j \in \mathcal{B}$, $m_j \in \mathcal{B}_V$ to deduce

$$\int_{\{v_J \in \mathcal{B}\}} V(v_{j+1} - m_{j+1}) d\mathbb{P} \leq c\epsilon^2 + \alpha \int_{\{v_J \in \mathcal{B}\}} V(v_j - m_j) d\mathbb{P}.$$

Since $\alpha \in (0, 1)$, Gronwall's lemma starting from J gives (for a different constant $c > 0$)

$$\limsup_{j \to \infty} \int_{\{v_J \in \mathcal{B}\}} V(v_{j+1} - m_{j+1}) d\mathbb{P} \leq c\epsilon^2. \tag{2.4.6}$$

Finally, combining (2.4.5) and (2.4.6) yields

$$\limsup_{j \to \infty} \mathbb{E} V(v_j - m_j) \leq c\epsilon^2 + \delta,$$

and since $\delta > 0$ was arbitrary and the norms $V(\cdot)^{1/2}$ and $|\cdot|$ are assumed equivalent the proof is complete. □

The following lemma is used in the preceding proof.

Lemma 2.4.9. *Let $\delta > 0$. Then, with the notation and assumptions of the previous theorem, there is $J = J(\delta)$ such that, for every $j \geq J$,*

$$\int_{\{v_J \notin \mathcal{B}\}} V(v_j - m_j) d\mathbb{P} < \delta.$$

Proof. Firstly, by the assumed equivalence of norms there is $\theta > 0$ such that $V(\cdot)^{1/2} \leq \theta |\cdot|$. Secondly, using the absorbing ball property it is easy to check that $\mathbb{P}[v_J \notin \mathcal{B}]$ can be made arbitrarily small by choosing J large enough. Therefore, since we work with the standing assumption that $\mathbb{E}|v_0|^2 < \infty$, it is possible to

choose J large enough so that

$$\int_{\{v_J \notin \mathcal{B}\}} \theta^2 |v_0|^2 + R^2 + 2R\theta |v_0| d\mathbb{P} \leq \delta.$$

Then, for $j > J$,

$$\begin{aligned}
\int_{\{v_J \notin \mathcal{B}\}} V(v_j - m_j) d\mathbb{P} &\leq \int_{\{v_J \notin \mathcal{B}\}} V(v_j) + V(m_j) + 2V(v_j)^{1/2} V(m_j)^{1/2} d\mathbb{P} \\
&\leq \int_{\{v_J \notin \mathcal{B}\}} V(v_j) + R^2 + 2RV(v_j)^{1/2} d\mathbb{P} \\
&\leq \int_{\{v_J \notin \mathcal{B}\}} \theta^2 |v_j|^2 + R^2 + 2R\theta |v_j| d\mathbb{P} \\
&\leq \int_{\{v_J \notin \mathcal{B}\}} \theta^2 |v_0|^2 + R^2 + 2R\theta |v_0| d\mathbb{P} \leq \delta,
\end{aligned}$$

where we used that, for $j > J$ and $v_J \notin \mathcal{B}$, $|v_j| \leq |v_0|$ by (2.2.5). □

2.5 Application to Relevant Models

2.5.1 Finite Dimensions (Lorenz '63 and '96 Models)

We study first the finite dimensional case $\mathcal{H} = \mathbb{R}^d$. Our aim is to introduce a general setting for which Assumption 2.2.1 holds, and thus the theory of the previous section can be applied. In order to do so we need to introduce suitable norms, and some conditions on the general nonlinear dissipative equation (2.2.4). We start by setting $|\cdot|$ to be the Euclidean norm, and $V(\cdot) = |P\cdot|^2 + |\cdot|^2$.

Next we introduce a set of hypotheses on the general system (2.2.4), and the observation matrix P.

Assumptions 2.5.1.

1. $\langle Au, u\rangle \geq |u|^2, \quad \forall u \in \mathcal{H}$.
2. $\langle B(u,u), u\rangle = 0, \quad \forall u \in \mathcal{H}$. *(Energy conserving nonlinearity.)*
3. *There is* $c_1 > 0$ *such that* $2|\langle B(u,\tilde{u}), \tilde{u}\rangle| \leq c_1 |P\tilde{u}||u||\tilde{u}|, \quad \forall u, \tilde{u} \in \mathcal{H}$.
4. *There is* $c_2 > 0$ *such that* $|B(u,\tilde{u})| \leq c_2 |u||\tilde{u}|, \quad \forall u, \tilde{u} \in \mathcal{H}$.
5. *There are* $c_3 > 0$ *and* $c_4 \geq 0$ *such that* $\langle Au, Pu\rangle \geq c_3 |Pu|^2 - c_4 |u|^2$.

Assumptions 2.5.1.1, 2.5.1.2 and 2.5.1.4 are satisfied by various important dissipative equations, including the Lorenz '63 [Hayden et al., 2011] (and used in [Law et al., 2014]), and Lorenz '96 models [Law et al., 2016]. Assumptions 2.5.1.3 and 2.5.1.5 are fulfilled when the 'right' parts of the system are observed. Examples

of observation matrices P that fit into our theory are given —both for the Lorenz '63 and '96 models— in Subsections 2.5.1.1 and 2.5.1.2.

The first two items of Assumption 2.5.1 are enough to ensure the absorbing ball property Assumption 2.2.1. Indeed, if these conditions hold then taking the inner product of (2.2.4) with v gives

$$\frac{1}{2}\frac{d}{dt}|v|^2 + \langle Av, v\rangle + \langle B(v,v), v\rangle = \langle f, v\rangle,$$

or

$$\frac{d}{dt}|v|^2 + |v|^2 \leq |f|^2.$$

Finally, Gronwall's lemma yields Assumption 2.2.1.1 with $r_0 = |f|^2$ and $r_1 = 1$, and the absorbing ball

$$\mathcal{B} := \{u \in \mathcal{H} : |u| \leq r := \sqrt{2}|f|\}. \tag{2.5.1}$$

We now show that the squeezing property is also satisfied provided that the time h between observations is sufficiently small. The proof is based on the analysis of the Lorenz '63 model in [Hayden et al., 2011]. Recall that $Q = I - P$ is the operator that projects onto the unobserved part of the system.

Lemma 2.5.2. *Suppose that Assumption 2.5.1 holds and let $r' > 0$. Then there is $h^\star > 0$ with the property that for all $h < h^\star$, $v \in \mathcal{B}$, and $u \in \mathcal{H}$ with $|u - v| \leq r'$, there exists $\alpha = \alpha(r') \in (0,1)$ such that*

$$V\Big(Q\big(\Psi(v) - \Psi(u)\big)\Big) \leq \alpha V(v - u).$$

Proof. Denote $\delta_0 = u - v$ and $\delta(t) = \Psi_t(u) - \Psi_t(v)$. Lemma 2.5.3 below shows that

$$|\delta(t)|^2 \leq b_1(t)|\delta_0|^2 + b_2(t)|P\delta_0|^2,$$

where $b_1(t)$ and $b_2(t)$ are also defined in Lemma 2.5.3. Therefore, noting that $V\big(Q\delta(t)\big) = |Q\delta(t)|^2 \leq |\delta(t)|^2$,

$$V\big(Q\delta(t)\big) \leq \max\{b_1(t), b_2(t)\}V(\delta_0).$$

Since $b_1(0) = 1$, $b_2(0) = 0$, and $b_1'(0) = -1 < 0$ it follows that, for all sufficiently small t, $\max\{b_1(t), b_2(t)\} \in (0,1)$ and the lemma is proved. □

The following result has been used in the proof.

Lemma 2.5.3. *Suppose that the notation and assumptions of the previous lemma are in force, and that $|\delta_0| \leq r'$. Then, for $t \in [0, h)$,*

$$|P\delta(t)|^2 \leq |P\delta_0|^2 + \left(k_4(e^{kt} - 1) + k_5(e^{2kt} - 1)\right)|\delta_0|^2,$$

and

$$\begin{aligned}|\delta(t)|^2 &\leq k_1(1 - e^{-t})|P\delta_0|^2 \\ &+ \left(e^{-t} + k_2\left[\frac{e^{kt} - e^{-t}}{k+1} - (1 - e^{-t})\right] + k_3\left[\frac{e^{2kt} - e^{-t}}{2k+1} - (1 - e^{-t})\right]\right)|\delta_0|^2,\end{aligned}$$

where k and k_i, $1 \leq i \leq 5$, are constants defined in the proof, and k_3 and k_5 depend on r'. Therefore,

$$|P\delta(t)|^2 \leq a_1(t)|\delta_0|^2 + |P\delta_0|^2 \tag{2.5.2}$$

and

$$|\delta(t)|^2 \leq b_1(t)|\delta_0|^2 + b_2(t)|P\delta_0|^2, \tag{2.5.3}$$

where the functions a_1, b_1 and b_2 are defined in the obvious way from the expressions above.

Proof. Firstly, it is not difficult to check (see for example [Kelly et al., 2014]) that Assumptions 2.5.1.1, 2.5.1.2 and 2.5.1.3 imply that there exists a constant $k > 0$ such that, for $u \in \mathcal{H}$, $v \in \mathcal{B}$ and $t > 0$,

$$|\delta(t)|^2 \leq e^{kt}|\delta_0|^2.$$

Next, using the definition of δ and the symmetry of $B(\cdot, \cdot)$ it is possible to derive [Law et al., 2014] the error equation

$$\frac{d\delta}{dt} + A\delta + 2B(v, \delta) + B(\delta, \delta) = 0. \tag{2.5.4}$$

Taking the inner product with δ we obtain

$$\frac{1}{2}\frac{d}{dt}|\delta|^2 + \langle A\delta, \delta\rangle + 2\langle B(v, \delta), \delta\rangle = 0,$$

and therefore

$$\frac{1}{2}\frac{d}{dt}|\delta|^2 + |\delta|^2 \leq c_1 r|\delta||P\delta| \leq \frac{1}{2}|\delta|^2 + \frac{1}{2}c_1^2 r^2|P\delta|^2,$$

i.e.

$$\frac{d|\delta|^2}{dt} + |\delta|^2 \leq c_1^2 r^2 |P\delta|^2. \tag{2.5.5}$$

We now bound $|P\delta|^2$. Taking the inner product of (2.5.4) with $P\delta$

$$\frac{1}{2}\frac{d}{dt}|P\delta|^2 + \langle A\delta, P\delta\rangle + 2\langle B(v,\delta), P\delta\rangle + \langle B(\delta,\delta), P\delta\rangle = 0.$$

Hence,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}|P\delta|^2 + \langle A\delta, P\delta\rangle &\leq 2|\langle B(v,\delta), P\delta\rangle| + |\langle B(\delta,\delta), P\delta\rangle| \\ &\leq 2c_2 r|\delta||P\delta| + c_2|\delta|^2|P\delta|\end{aligned}$$

and

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}|P\delta|^2 + c_3|P\delta|^2 &\leq c_4|\delta|^2 + 2c_2 r|\delta||P\delta| + c_2|\delta|^2|P\delta| \\ &\leq c_4|\delta|^2 + 2c_2 r|\delta||P\delta| + c_2|\delta|e^{kt/2}r'|P\delta| \\ &\leq c_4|\delta|^2 + \frac{2}{c_3}c_2^2 r^2|\delta|^2 + \frac{c_3}{2}|P\delta|^2 + \frac{1}{2c_3}c_2^2 e^{kt}r'^2|\delta|^2 + \frac{c_3}{2}|P\delta|^2\end{aligned}$$

i.e.

$$\frac{d}{dt}|P\delta|^2 \leq \left(2c_4 + \frac{4}{c_3}c_2^2 r^2 + \frac{1}{c_3}c_2^2 e^{kt}r'^2\right)|\delta|^2.$$

On integrating from 0 to t and using that $|\delta(t)|^2 \leq |\delta_0|^2 e^{kt}$:

$$\begin{aligned}|P\delta(t)|^2 &\leq |P\delta_0|^2 + \left(\frac{2c_4 + \frac{4}{c_3}c_2^2 r^2}{k}(e^{kt}-1) + \frac{c_2^2 r'^2}{2kc_3}(e^{2kt}-1)\right)|\delta_0|^2 \\ &= |P\delta_0|^2 + \Big(k_4(e^{kt}-1) + k_5(e^{2kt}-1)\Big)|\delta_0|^2,\end{aligned}$$

where the last equality defines k_4 and k_5. This proves (2.5.2). Then, going back to (2.5.5),

$$\frac{d}{dt}|\delta|^2 + |\delta|^2 \leq c_1^2 r^2 \left\{|P\delta_0|^2 + \Big(k_4(e^{kt}-1) + k_5(e^{2kt}-1)\Big)|\delta_0|^2\right\}.$$

After denoting $k_1 = c_1^2 r^2$, $k_2 = k_1 k_4$ and $k_3 = k_1 k_5$ the inequality above becomes

$$\frac{d}{dt}|\delta|^2 + |\delta|^2 \leq k_1|P\delta_0|^2 + \Big(k_2(e^{kt}-1) + k_3(e^{2kt}-1)\Big)|\delta_0|^2.$$

Finally, Gronwall's lemma gives (2.5.3). □

The previous lemmas show that Assumption 2.5.1 implies the squeezing property Assumption 2.2.1.2 provided that the assimilation time h is sufficiently small. Indeed taking

$$\mathcal{B}_V := \{u \in \mathcal{H} : V(u)^{1/2} \leq \sqrt{2}r\} \tag{2.5.6}$$

with r as in (2.5.1) we have that $|u-v| \leq (1+\sqrt{2})r$ for $u \in \mathcal{B}$, $v \in \mathcal{B}_V$, and we are in the setting of Lemma 2.5.2 with $r' = (1+\sqrt{2})r$. Moreover, the requirement $\mathcal{B} \subset \mathcal{B}_V$ in 2.2.1.2 is also fulfilled. Therefore the following result is a direct application of Theorem 2.4.8.

Theorem 2.5.4. *Assume that the signal dynamics are defined via a general dissipative differential equation on $\mathbb{R}^d$ with quadratic energy-conserving nonlinearity of the form (2.2.4), and that Assumption 2.5.1 is satisfied. Then there is $h^\star > 0$ such that Assumption 2.2.1 is also satisfied for all $h < h^\star$. Therefore, if $\{m_j\}_{j\geq 0}$ denotes the sequence of $\mathcal{B}_V$-truncated nonlinear observers given by (2.3.5) and* (2.5.6)*, then there is a constant $c > 0$, independent of the noise strength ϵ, such that, for all discrete assimilation time $h < h^\star$,*

$$\limsup_{j\to\infty} \mathbb{E}|v_j - m_j|^2 \leq c\epsilon^2.$$

Consequently

$$\limsup_{j\to\infty} \mathbb{E}|v_j - \widehat{v}_j|^2 \leq c\epsilon^2, \qquad \limsup_{j\to\infty} \mathrm{Trace}\, \mathbb{E}\, \mathrm{var}[v_j|Y_j] \leq c\epsilon^2.$$

2.5.1.1 Lorenz '63 Model

A first example of a system of the form (2.2.4) is the Lorenz '63 model, which corresponds to a three dimensional problem defined by (2.2.4) with

$$A = \begin{bmatrix} a & -a & 0 \\ a & 1 & 0 \\ 0 & 0 & b \end{bmatrix},$$

$$B(u,\tilde{u}) = \begin{bmatrix} 0 \\ (u_1\tilde{u}_3 + u_3\tilde{u}_1)/2 \\ -(u_1\tilde{u}_2 + u_2\tilde{u}_1)/2 \end{bmatrix}, \qquad f = \begin{bmatrix} 0 \\ 0 \\ -b(r+a) \end{bmatrix}.$$

ϵ	MSE
1	1.59
0.1	1.3×10^{-2}
0.01	4.93×10^{-4}

Table 2.1: *Bounds in the MSE given by the truncated nonlinear observer for the Lorenz '63 model. Only the first coordinate is observed. The assimilation time step is* $h = 0.01$ *and the signal was filtered up to time* $T = 5$. *The MSE was computed using* 20 *initializations of* $v_0 \sim N(0, I)$; *for each of these initializations five observation sequences were generated using Gaussian noise. The MSE shown is the Monte Carlo average of the filter error at time* T *over all these simulations.*

The standard parameter values are $(a, b, r) = (10, 8/3, 28)$. Define the projection matrix

$$P := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

It is then immediate from the definitions that the first, second and fourth items of Assumption 2.5.1 are satisfied [Hayden et al., 2011]. A verification of the third and fifth items can be found in the proof of Theorem 2.5 of [Hayden et al., 2011].

To provide insight, in Table 1 we show a Monte Carlo estimate of the mean square error (MSE) $\mathbb{E}|m_j - v_j|^2$ made by a truncated nonlinear observer with different values of the observation noise strength ϵ. The results suggest that the MSE of this suboptimal filter decreases as $\mathcal{O}(\epsilon^2)$, in agreement with our theoretical analyses. This provides an upper bound for the error made by the optimal filter. It is worth mentioning that the values of h for which we observe accurate signal reconstruction are often far larger than the upper limits required by our theory.

Remark 2.5.5. *An accuracy result for the Lorenz '63 model, similar to Theorem 2.5.4 above, was established in [Law et al., 2014] using the 3DVAR algorithm. Indeed truncation is not needed for this model since a global form of the squeezing property Assumption 2.2.1.2 holds (with* $v \in \mathcal{B}$, $u \in \mathcal{H}$*).*

2.5.1.2 Lorenz '96 Model

Another system that satisfies the assumptions introduced in this section is the Lorenz '96 model, which is of the form (2.2.4) with the choices $A = I_{d\times d}$, where we

ϵ	MSE
1	1.11
0.1	1.08×10^{-2}
0.01	3.36×10^{-4}

Table 2.2: *Same experiment as in Table* 1, *now for the Lorenz '96 model with the observation operator* (2.5.7).

assume $d \in 3\mathbb{N}$, forcing term

$$f = \begin{bmatrix} 8 \\ \vdots \\ 8 \end{bmatrix},$$

and bilinear form

$$B(u, \tilde{u}) = -\frac{1}{2} \begin{bmatrix} \tilde{u}_2 u_d + u_2 \tilde{u}_d - \tilde{u}_d u_{d-1} - u_d \tilde{u}_{d-1} \\ \vdots \\ \tilde{u}_{i-1} u_{i+1} + u_{i-1} \tilde{u}_{i+1} - \tilde{u}_{i-2} u_{i-1} - u_{i-2} \tilde{u}_{i-1} \\ \vdots \\ \tilde{u}_{d-1} u_1 + u_{d-1} \tilde{u}_1 - \tilde{u}_{d-2} u_{d-1} - u_{d-2} \tilde{u}_{d-1} \end{bmatrix}.$$

Define the projection matrix P by replacing every third column of the identity matrix $I_{d \times d}$ by the zero column vector

$$P = \begin{bmatrix} e_1, & e_2, & 0, & e_4, & e_5, & 0, & \cdots \end{bmatrix}. \tag{2.5.7}$$

For a proof that the first, second and fourth items of Assumption 2.5.1 are satisfied see Property 2.1.1 in [Law et al., 2016]. The third item results from combining Property 2.1.1 and Property 2.2.2 in [Law et al., 2016]. Finally, since $A = I$ the fifth item holds with $c_3 = 1, c_4 = 0$.

As for the Lorenz '63 model, we show a Monte Carlo estimate of the error made by a truncated nonlinear observer in Table 2. Again the error decreases as ϵ^2.

2.5.2 Infinite Dimensions (Navier-Stokes Equation)

It is well known [Bloemker et al., 2013] that the incompressible Navier-Stokes equation on the torus $\mathbb{T}^2 = [0, l] \times [0, l]$ can be written in the form (2.2.4) as we now recall.

Let $\hat{H}$ be the space of zero-mean, divergence-free, vector-valued polynomials

u from $\mathbb{T}^2$ to $\mathbb{R}^2$. Let H be the closure of $\hat{H}$ with respect to the L^2 norm. Finally, let $P_H : (L^2(\mathbb{T}^2))^2 \to H$ be the Leray-Helmholtz orthogonal projector [Temam, 1995]. Then, the operator A and the symmetric bilinear form B in (2.2.4) are given by

$$Au = -\nu P_H \Delta, \qquad B(u,v) = \frac{1}{2} P_H[u \cdot \nabla v] + \frac{1}{2} P_H[v \cdot \nabla u],$$

were ν is the viscosity.

We assume that $f \in H$ so that $P_H f = f$. In the periodic case considered here $A = -\nu\Delta$ with domain $\mathcal{D}(A) = H^2(\mathbb{T}^2) \cap H$. Moreover, the solution to the Navier-Stokes equation (see below for the precise definition) can be written as a Fourier series

$$v = \sum_{k \in \mathcal{K}} v_k e^{ikx}, \quad \mathcal{K} = \left\{ \frac{2\pi}{L}(n_1, n_2) : n_i \in \mathbb{Z}, (n_1, n_2) \neq (0,0) \right\}.$$

The Fourier coefficients encode the divergence-free property and hence may be written as $v_k = v'_k k^\perp / |k|$ for scalar coefficients v'_k, where $|\cdot|$ is the Euclidean norm, and for $k = (k_1, k_2)$ $k^\perp = (k_2, -k_1)$. We now define the observation operator $P = P_\lambda$ in the general observation model (2.2.2) as

$$P_\lambda u = \sum_{|k|^2 \leq \lambda} u_k e^{ikx},$$

and set $Q_\lambda = I - P_\lambda$. Several choices of noise fit into our theory, and a natural one is given by

$$w_1 = \sum_{|k|^2 \leq \lambda} \xi_k e^{ikx}, \tag{2.5.8}$$

where $\xi_k \sim N\left(0, \left(k^2 n(\lambda)\right)^{-1}\right)$ and $n(\lambda) := \#\{k : k^2 \leq \lambda\}$.

We have already defined L^2 divergence-free functions as an appropriate closure of $\hat{H}$, and denoted this space by H; we now define H^1 divergence-free functions as the closure of $\hat{H}$ with respect to the H^1 norm, and we denote this space $\mathcal{H}$. It is in $\mathcal{H}$ that we will apply our general theory. We define a norm in $\mathcal{H}$

$$\|u\|^2_{H^1} := \sum_{k \in \mathcal{K}} k^2 |u_k|^2,$$

which is equivalent to the H^1 norm. Note that with this definition $\mathbb{E}\|w_1\|^2_{H^1} = 1$.

The following theorem —see [Temam, 1995], [Constantin and Foias, 1988] or [Robinson, 2001]— guarantees the existence and uniqueness of strong solutions to this problem with initial conditions in $\mathcal{H}$.

Proposition 2.5.6. *Let $u_0 \in \mathcal{H}$ and $f \in H$. Then (2.2.4) has a unique strong solution*

$$u \in L^\infty\big((0,T);\mathcal{H}\big) \cap L^2\big((0,T);\mathcal{D}(A)\big) \quad \text{and} \quad \frac{du}{dt} \in L^2\big((0,T),H\big)$$

for any $T > 0$. Furthermore, this solution is in $C([0,T];\mathcal{H})$ and depends continuously on the initial data u_0 in the $\mathcal{H}$ norm.

Take $|\cdot| = V(\cdot)^{1/2} = \|\cdot\|_{H^1}$. It is not difficult to prove the absorbing ball property for the Navier-Stokes equation with initial conditions in $\mathcal{H}$ [Robinson, 2001]. Indeed there is $\theta = \theta(\nu) > 0$ such that, for every $u \in \mathcal{H}$, $|Au|^2 \geq \theta|u|^2$. Then Assumption 2.2.1.1 is satisfied with $r_0 = |f|^2\theta^2$ and $r_1 = \theta$. We hence set

$$\mathcal{B} = \{u \in \mathcal{H} : |u| \leq r := \sqrt{2}\frac{|f|}{\theta}\}. \tag{2.5.9}$$

The following squeezing property is taken from [Brett et al., 2013], which uses the analysis in [Hayden et al., 2011].

Lemma 2.5.7. *For every $r' > 0$ there are constants $\alpha = \alpha(r') \in (0,1)$ and $\lambda_\star = \lambda_\star(r') > 0$ with the property that, for $\lambda > \lambda_\star$, there exists $h^\star = h^\star(r', \lambda)$ such that, for all $u, v \in B(r') := \{x \in \mathcal{H} : V(x)^{1/2} \leq r'\}$, and assimilation time $h < h^\star$,*

$$V\Big(Q_\lambda\big(\Psi(v) - \Psi(u)\big)\Big) \leq \alpha V(v-u).$$

The previous lemma yields Assumption 2.2.1.2. for sufficiently small assimilation time h by choosing $\mathcal{B}_V = \mathcal{B}$ and $r' = 2r$. The next result is then a straightforward application of Theorem 2.4.8.

Theorem 2.5.8. *Take $|\cdot|$ and V as above, and let $\{m_j\}_{j\geq 0}$ be the sequence of $\mathcal{B}_V$-truncated nonlinear observers with $\mathcal{B}_V = \mathcal{B}$ given by (2.5.9). Then there are $h^\star$, $\lambda_\star > 0$, such that for all $h < h^\star$ and $\lambda > \lambda_\star$ Assumption 2.2.1 is satisfied and therefore there exists a constant $c > 0$, independent of the noise strength ϵ, such that*

$$\limsup_{j\to\infty} \mathbb{E}|v_j - m_j|^2 \leq c\epsilon^2,$$

Consequently,

$$\limsup_{j\to\infty} \mathbb{E}|v_j - \widehat{v}_j|^2 \leq c\epsilon^2, \qquad \limsup_{j\to\infty} \operatorname{Trace} \mathbb{E}\operatorname{var}[v_j|Y_j] \leq c\epsilon^2.$$

2.6 Conclusions

We conclude by summarizing our work and highlighting future directions.

- Noisy observations can be used to compensate, in the long-time asymptotic regime, for uncertainty in the initial conditions of unstable or chaotic dynamical systems.
- It would be interesting to study similar questions in continuous time, and to investigate the impact of other sources of uncertainty, such as those arising from incomplete knowledge of the parameters in the model.
- We have determined conditions on the dynamics and observations under which the optimal filter accurately tracks the signal (and the variance of the filtering distributions becomes small) in the long-time asymptotic.
- These properties of the true filtering distribution are potentially useful for the design of improved algorithmic approximations of the filtering distributions.
- We have introduced a modification of the 3DVAR filter as a tool to prove our results. This new filter is potentially of interest in its own right as a practical algorithm.

Chapter 3

Filter accuracy for the Lorenz '96 Model

3.1 Introduction

Data assimilation is concerned with the blending of data and dynamical mathematical models, often in an online fashion where it is known as filtering; motivation comes from applications in the geophysical sciences such as weather forecasting [Kalnay, 2003], oceanography [Bennett, 2003] and oil reservoir simulation [Oliver et al., 2008]. Over the last decade there has been a growing body of theoretical understanding which enables use of the theory of synchronization in dynamical systems to establish desirable properties of these filters. This idea is highlighted in the recent book [Abarbanel, 2013] from a physics perspective and, on the rigorous mathematical side, has been developed from a pair of papers by Olson, Titi and co-workers [Olson and Titi, 2003; Hayden et al., 2011], in the context of the Navier-Stokes equation in which a finite number of Fourier modes are observed. This mathematical work of Olson and Titi concerns perfect (noise-free) observations, but the ideas have been extended to the incorporation of noisy data for the Navier-Stokes equation in the papers [Bloemker et al., 2014; Brett et al., 2013]. Furthermore the techniques used are quite robust to different dissipative dynamical systems, and have been demonstrated to apply in the Lorenz '63 model [Hayden et al., 2011; Law et al., 2014], and also to point-wise in space and continuous time observations [Azouani et al., 2014] by use of a control theory perspective similar to that which arises from the derivation of continuous time limits of discrete time filters [Bloemker et al., 2014]. A key question in the field is to determine relationships between the underlying dynamical system and the observation operator which are sufficient to ensure that the signal

can be accurately recovered from a chaotic dynamical system, whose initialization is not known precisely, by the use of observed data. Our purpose is to investigate this question theoretically and computationally. We work in the context of the Lorenz '96 model, widely adopted as a useful test model in the atmospheric sciences data assimilation community [Majda and Harlim, 2012; Ott et al., 2004].

The primary contributions of this chapter are: (i) to theoretically demonstrate the robustness of the methodology proposed by Olson and Titi, by extending it to the Lorenz '96 model; (ii) to highlight the gap between such theories and what can be achieved in practice, by performing careful numerical experiments; and (iii) to illustrate the power of allowing the observation operator to adapt to the dynamics as this leads to accurate reconstruction of the signal based on very sparse observations. Indeed our approach in (iii) suggests highly efficient new algorithms where the observation operator is allowed to adapt to the current state of the dynamical system. The question of how to optimize the observation operator to maximize information was first addressed in the context of atmospheric science applications in [Lorenz and Emanuel, 1998]. The adaptive observation operators that we propose are not currently practical for operational atmospheric data assimilation, but they suggest a key principle which should underlie the construction of adaptive observation operators: to learn as much as possible about modes of instability in the dynamics at minimal cost.

The outline of the chapter is as follows. In Section 3.2 we introduce the model set up and a family of Kalman-based filtering schemes which include as particular cases the Three-dimensional Variational method (3DVAR) and the Extended Kalman Filter (ExKF) used in this chapter. All of these methods may be derived from sequential application of a minimization principle which encodes the trade-off between matching the model and matching the data. In Section 3.3 we describe the Lorenz '96 model and discuss its properties that are relevant to this work. In Section 3.4 we introduce a fixed observation operator which corresponds to observing two thirds of the signal and study theoretical properties of the 3DVAR filter, in both a continuous and a discrete time setting. In Section 3.5 we introduce an adaptive observation operator which employs knowledge of the linearized dynamics over the assimilation window to ensure that the unstable directions of the dynamics are observed. We then numerically study the performance of a range of filters using the adaptive observations. In Subsection 3.5.1 we consider the 3DVAR method, whilst Subsection 3.5.2 focuses on the Extended Kalman Filter (ExKF). In Subsection 3.5.2 we also compare the adaptive observation implementation of the ExKF with the AUS (Assimilation in Unstable Space) scheme [Trevisan and Uboldi, 2004]

which motivates our work. The AUS scheme projects the model covariances into the subspaces governed by the unstable dynamics, whereas we use this idea on the observation operators themselves, rather than on the covariances. In Section 3.6 we summarize the work and draw some brief conclusions. In order to maintain a readable flow of ideas, the proofs of all properties, propositions and theorems stated in the main body of the text are collected in an appendix.

Throughout the chapter we denote by $\langle \cdot, \cdot \rangle$ and $|\cdot|$ the standard Euclidean inner-product and norm. For positive-definite matrix C we define $|\cdot|_C := |C^{-\frac{1}{2}} \cdot|$.

3.2 Set Up

We consider the ordinary differential equation (ODE)

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \mathcal{F}(v), \quad v(0) = v_0, \tag{3.2.1}$$

where the solution to (3.2.1) is referred to as the *signal*. We denote by $\Psi : \mathbb{R}^J \times \mathbb{R}^+ \to \mathbb{R}^J$ the solution operator for the equation (3.2.1), so that $v(t) = \Psi(v_0; t)$. In our discrete time filtering developments we assume that, for some fixed $h > 0$, the signal is subject to observations at times $t_k := kh$, $k \geq 1$. We then write $\Psi(\cdot) := \Psi(\cdot; h)$ and $v_k := v(kh)$, with slight abuse of notation to simplify the presentation. Our main interest is in using partial observations of the discrete time dynamical system

$$v_{k+1} = \Psi(v_k), \quad k \geq 0, \tag{3.2.2}$$

to make estimates of the state of the system. To this end we introduce the family of linear observation operators $\{H_k\}_{k\geq 1}$, where $H_k : \mathbb{R}^J \to \mathbb{R}^J$ is assumed to have rank (which may change with k) less than or equal to $M \leq J$. We then consider data $\{y_k\}_{k\geq 1}$ given by

$$y_k = H_k v_k + \nu_k, \quad k \geq 1, \tag{3.2.3}$$

where we assume that the random and/or systematic error ν_k (and hence also y_k) is contained in $H_k\mathbb{R}^J$. If $Y_k = \{y_\ell\}_{\ell=1}^k$ then the objective of filtering is to estimate v_k from Y_k given incomplete knowledge of v_0; furthermore this is to be done in a sequential fashion, using the estimate of v_k from Y_k to determine the estimate of v_{k+1} from Y_{k+1}. We are most interested in the case where $M < J$, so that the observations are partial, and $H_k\mathbb{R}^J$ is a strict subset of $\mathbb{R}^J$; in particular we address the question of how small M can be chosen whilst still allowing accurate recovery of the signal over long time-intervals.

Let m_k denote our estimate of v_k given Y_k. The discrete time filters used in this chapter have the form

$$m_{k+1} = \text{argmin}_m \left\{ \frac{1}{2}\big|m - \Psi(m_k)\big|^2_{\widehat{C}_{k+1}} + \frac{1}{2}\big|y_{k+1} - H_{k+1}m\big|^2_{\Gamma} \right\}. \tag{3.2.4}$$

The norm in the second term is only applied within the M-dimensional image space of H_{k+1}, where y_{k+1} lies; then Γ is realized as a positive-definite $M \times M$ matrix in this image space, and $\widehat{C}_{k+1}$ is a positive-definite $J \times J$ matrix. The minimization represents a compromise between respecting the model and respecting the data, with the covariance weights $\widehat{C}_{k+1}$ and Γ determining the relative size of the two contributions; see [Law et al., 2015] for more details. Different choices of $\widehat{C}_{k+1}$ give different filtering methods. For instance, the choice $\widehat{C}_{k+1} = C_0$ (constant in k) corresponds to the 3DVAR method. More sophisticated algorithms, such as the ExKF, allow $\widehat{C}_{k+1}$ to depend on m_k.

All the discrete time algorithms we consider proceed iteratively in the sense that the estimate m_{k+1} is determined by the previous one, m_k, and the observed data y_{k+1}; we are given an initial condition m_0 which is an imperfect estimate of v_0. It is convenient to see the update $m_k \mapsto m_{k+1}$ as a two-step process. In the first one, known as the *forecast step*, the estimate m_k is evolved with the dynamics of the underlying model yielding a prediction $\Psi(m_k)$ for the current state of the system. In the second step, known as the *analysis step*, the forecast is used in conjunction with the observed data y_{k+1} to produce the estimate m_{k+1} of the true state of the underlying system v_{k+1}, using the minimization principle (3.2.4).

In Section 3.4 we study the continuous time filtering problem for fixed observation operator, where the goal is to estimate the value of a continuous time signal

$$v(t) = \Psi(v_0, t), \quad t \geq 0,$$

at time $T > 0$. As in the discrete case, it is assumed that only incomplete knowledge of v_0 is available. In order to estimate $v(T)$ we assume that we have access, at each time $0 < t \leq T$, to a (perhaps noisily perturbed) projection of the signal given by a fixed, constant in time, observation matrix H. The continuous time limit of 3DVAR with constant observation operator H, is obtained by setting $\Gamma = h^{-1}\Gamma_0$ and $\widehat{C}_{k+1} = C$ and letting $h \to 0$. The resulting filter, derived in [Bloemker et al., 2014], is given by

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \mathcal{F}(m) + CH^*\Gamma_0^{-1}\Big(\frac{\mathrm{d}z}{\mathrm{d}t} - Hm\Big), \tag{3.2.5}$$

where the observed data is now z – formally the time-integral of the natural con-

tinuous time limit of y – which satisfies the stochastic differential equation (SDE)

$$\frac{\mathrm{d}z}{\mathrm{d}t} = Hv + H\Gamma_0^{\frac{1}{2}}\frac{\mathrm{d}w}{\mathrm{d}t}, \tag{3.2.6}$$

for w a unit Wiener process. This filter has the effect of nudging the solution towards the observed data in the H-projected direction. A similar idea is used in [Azouani et al., 2014] to assimilate pointwise observations of the Navier-Stokes equation.

For the discrete and continuous time filtering schemes as described we address the following questions:

- how does the filter error $|m_k - v_k|$ behave as $k \to \infty$ (discrete setting)?
- how does the filter error $|m(t) - v(t)|$ behave as $t \to \infty$ (continuous setting)?

We answer these questions in Section 3.4 in the context of the Lorenz '96 model: for a carefully chosen fixed observation operator we determine conditions under which the large time filter error is small – this is filter accuracy. We then turn to the adaptive observation operator and focus on the following lines of enquiry:

- how much do we need to observe to obtain filter accuracy? (in other words what is the minimum rank of the observation operator required?)
- how does adapting the observation operator affect the answer to this question?

We study both these questions numerically in Section 3.5, again focussing on the Lorenz '96 model to illustrate ideas.

3.3 Lorenz '96 Model

The Lorenz '96 model is a lattice-periodic system of coupled nonlinear ODE whose solution $u = (u^{(1)}, \ldots, u^{(J)})^T \in \mathbb{R}^J$ satisfies

$$\frac{\mathrm{d}u^{(j)}}{\mathrm{d}t} = u^{(j-1)}(u^{(j+1)} - u^{(j-2)}) - u^{(j)} + F \quad \text{for } j = 1, 2, \cdots, J, \tag{3.3.1}$$

subject to the periodic boundary conditions

$$u^{(0)} = u^{(J)}, \qquad u^{(J+1)} = u^{(1)}, \qquad u^{(-1)} = u^{(J-1)}. \tag{3.3.2}$$

Here F is a forcing parameter, constant in time. For our numerical experiments we will choose F so that the dynamical system exhibits sensitive dependence on initial conditions and positive Lyapunov exponents. For example, for $F = 8$ and $J = 60$

the system is chaotic. Our theoretical results apply to any choice of the parameter F and to arbitrarily large system dimension J.

It is helpful to write the model in the following form, widely adopted in the analysis of geophysical models as dissipative dynamical systems [Temam, 1997]:

$$\frac{\mathrm{d}u}{\mathrm{d}t} + Au + B(u,u) = f, \quad u(0) = u_0 \tag{3.3.3}$$

where

$$A = I_{J\times J}, \quad f = \begin{pmatrix} F \\ \vdots \\ F \end{pmatrix}_{J\times 1}$$

and for $u, \tilde{u} \in \mathbb{R}^J$

$$B(u,\tilde{u}) = -\frac{1}{2}\begin{pmatrix} \tilde{u}^{(2)}u^{(J)} + u^{(2)}\tilde{u}^{(J)} - \tilde{u}^{(J)}u^{(J-1)} - u^{(J)}\tilde{u}^{(J-1)} \\ \vdots \\ \tilde{u}^{(j-1)}u^{(j+1)} + u^{(j-1)}\tilde{u}^{(j+1)} - \tilde{u}^{(j-2)}u^{(j-1)} - u^{(j-2)}\tilde{u}^{(j-1)} \\ \vdots \\ \tilde{u}^{(J-1)}u^{(1)} + u^{(J-1)}\tilde{u}^{(1)} - \tilde{u}^{(J-2)}u^{(J-1)} - u^{(J-2)}\tilde{u}^{(J-1)} \end{pmatrix}_{J\times 1}.$$

We will use the following properties of A and B, proved in the appendix:

Properties 3.3.1. *For $u, \tilde{u} \in \mathbb{R}^J$*

1. $\langle Au, u\rangle = |u|^2$.
2. $\langle B(u,u), u\rangle = 0$.
3. $B(u,\tilde{u}) = B(\tilde{u},u)$.
4. $|B(u,\tilde{u})| \leq 2|u||\tilde{u}|$.
5. $2\langle B(u,\tilde{u}), u\rangle = -\langle B(u,u), \tilde{u}\rangle$.

Property (1) shows that the linear term induces dissipation in the model, whilst property (2) shows that the nonlinear term is energy-conserving. Balancing these two properties against the injection of energy through f gives the existence of an absorbing, forward-invariant ball for equation (3.3.3), as stated in the following proposition, proved in the appendix.

Proposition 3.3.2. *Let $K = 2JF^2$ and define $\mathcal{B} := \{u \in \mathbb{R}^J : |u|^2 \leq K\}$. Then $\mathcal{B}$ is an absorbing, forward-invariant ball for equation (3.3.3): for any $u_0 \in \mathbb{R}^J$ there is time $T = T(|u_0|) \geq 0$ such that $u(t) \in \mathcal{B}$ for all $t \geq T$.*

3.4 Fixed Observation Operator

In this section we consider filtering the Lorenz '96 model with a specific choice of fixed observation matrix P (thus $H_k = H = P$) that we now introduce. First, we let $\{e_j\}_{j=1}^J$ be the standard basis for the Euclidean space $\mathbb{R}^J$ and assume that $J = 3J'$ for some integer $J' \geq 1$. Then the projection matrix P is defined by replacing every third column of the identity matrix $I_{J\times J}$ by the zero vector:

$$P = \Big(\; e_1,\, e_2,\, 0,\, e_4,\, e_5,\, 0, \ldots \; \Big)_{J\times J}. \tag{3.4.1}$$

Thus P has rank $M = 2J'$. We also define its complement Q as

$$Q = I_{J\times J} - P.$$

Remark 3.4.1. *Note that in the definition of the projection matrix P we could have chosen either the first or the second column to be set to zero periodically, instead of choosing every third column this way; the theoretical results in the remainder of this section would be unaltered by doing this.*

The matrix P provides sufficiently rich observations to allow the accurate recovery of the signal in the long-time asymptotic regime, both in continuous and discrete time settings. The following property of P, proved in the appendix, plays a key role in the analysis:

Properties 3.4.2. *The bilinear form $B(\cdot,\cdot)$ as defined after* (3.3.3) *satisfies $B(Qu, Qu) = 0$ and, furthermore, there is a constant $c > 0$ such that*

$$|\langle B(u,u), \tilde{u}\rangle| \leq c|u||\tilde{u}||Pu|.$$

All proofs in the following subsections are given in the appendix.

3.4.1 Continuous Assimilation

In this subsection we assume that the data arrives continuously in time. Subsection 3.4.1.1 deals with noiseless data, and the more realistic noisy scenario is studied in Subsection 3.4.1.2. We aim to show that, in the large time asymptotic, the filter is

close to the truth. In the absence of noise our results are analogous to those for the partially observed Lorenz '63 and Navier-Stokes models in [Olson and Titi, 2003]; in the presence of noise the results are similar to those proved in [Bloemker et al., 2014] for the Navier-Stokes equation and in [Law et al., 2014] for the Lorenz '63 model, and generalize the work in [Tarn and Rasis, 1976] to non-globally Lipschitz vector fields.

3.4.1.1 Noiseless Observations

The true solution v satisfies the following equation

$$\frac{\mathrm{d}v}{\mathrm{d}t} + v + B(v, v) = f, \quad v(0) = v_0. \tag{3.4.2}$$

Suppose that the projection Pv of the true solution is perfectly observed and continuously assimilated into the approximate solution m. The *synchronization filter* m has the following form:

$$m = Pv + q, \tag{3.4.3}$$

where v is the true solution given by (3.4.2) and q satisfies the equation (3.3.3) projected by Q to obtain

$$\frac{\mathrm{d}q}{\mathrm{d}t} + q + QB(Pv + q, Pv + q) = Qf, \quad q(0) = q_0. \tag{3.4.4}$$

Equations (3.4.3) and (3.4.4) form the continuous time synchronization filter. The following theorem shows that the approximate solution converges to the true solution asymptotically as $t \to \infty$.

Theorem 3.4.3. *Let m be given by the equations* (3.4.3), (3.4.4) *and let v be the solution of the equation* (3.4.2) *with initial data $v_0 \in \mathcal{B}$, the absorbing ball in Proposition 3.3.2, so that* $\sup_{t\geq 0}|v(t)|^2 \leq K$. *Then*

$$\lim_{t\to\infty} |m(t) - v(t)|^2 = 0.$$

The result establishes that in the case of high frequency in time observations the approximate solution converges to the true solution even though the signal is observed partially at frequency 2/3 in space. We now extend this result by allowing for noisy observations.

3.4.1.2 Noisy Observations: Continuous 3DVAR

Recall that the continuous time limit of 3DVAR is given by (3.2.5) where the observed data z, the integral of y, satisfies the SDE (3.2.6). We study this filter in the case where $H = P$ and under small observation noise $\Gamma_0 = \epsilon^2 I$. The 3DVAR model covariance is then taken to be of the size of the observation noise. We choose $C = \sigma^2 I$, where $\sigma^2 = \sigma^2(\epsilon) = \eta^{-1}\epsilon^2$, for some $\eta > 0$. Then equations (3.2.5) and (3.2.6) can be rewritten as

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \mathcal{F}(m) + \frac{1}{\eta}\Big(\frac{\mathrm{d}z}{\mathrm{d}t} - Pm\Big) \tag{3.4.5}$$

where

$$\frac{\mathrm{d}z}{\mathrm{d}t} = Pv + \epsilon P\frac{\mathrm{d}w}{\mathrm{d}t}, \tag{3.4.6}$$

and w is a unit Wiener process. Note that the parameter ϵ represents both the size of the 3DVAR observation covariance and the size of the noise in the observations.

The reader will notice that the continuous time synchronization filter is obtained from this continuous time 3DVAR filter if ϵ is set to zero and if the (singular) limit $\eta \to 0$ is taken. The next theorem shows that the approximate solution m converges to a neighbourhood of the true solution v where the size of the neighbourhood depends upon ϵ. Similarly as in [Law et al., 2014] and [Bloemker et al., 2014] it is required that η, the ratio between the size of observation and model covariances, is sufficiently small. The next theorem is thus a natural generalization of Theorem 3.4.3 to incorporate noisy data.

Theorem 3.4.4. *Let (m, z) solve the equations (3.4.5), (3.4.6) and let v solve the equation (3.4.2) with the initial data $v(0) \in \mathcal{B}$, the absorbing ball of Proposition 3.3.2, so that $\sup_{t\geq 0}|v(t)|^2 \leq K$. Then for the constant c as given in the Property 3.4.2, given $\eta < \frac{4}{c^2 K}$ we obtain*

$$\mathbb{E}|m(t) - v(t)|^2 \leq e^{-\lambda t}|m(0) - v(0)|^2 + \frac{2J\epsilon^2}{3\lambda\eta^2}(1 - e^{-\lambda t}), \tag{3.4.7}$$

where λ is defined by

$$\lambda = 2\left(1 - \frac{c^2\eta K}{4}\right). \tag{3.4.8}$$

Thus

$$\mathrm{limsup}_{t\to\infty}\mathbb{E}|m(t) - v(t)|^2 \leq a\epsilon^2,$$

where $a = \frac{2J}{3\lambda\eta^2}$ does not depend on the strength of the observation noise, ϵ.

3.4.2 Discrete Assimilation

We now turn to discrete data assimilation. Recall that filters in discrete time can be split into two steps: forecast and analysis. In this section we establish conditions under which the corrections made at the analysis steps overcome the divergence inherent due to nonlinear instabilities of the model in the forecast stage. As in the previous section we study first the case of noiseless data, generalizing the work of [Hayden et al., 2011] from the Navier-Stokes and Lorenz '63 models to include the Lorenz '96 model, and then study the case of 3DVAR, generalizing the work in [Brett et al., 2013; Law et al., 2014], which concerns the Navier-Stokes and Lorenz '63 models respectively, to the Lorenz '96 model.

3.4.2.1 Noiseless Observations

Let $h > 0$, and set $t_k := kh$, $k \geq 0$. For any function $g : \mathbb{R}^+ \to \mathbb{R}^J$, continuous in $[t_{k-1}, t_k)$, we denote $g(t_k^-) := \lim_{t \uparrow t_k} g(t)$. Let v be a solution of equation (3.4.2) with $v(0)$ in the absorbing forward-invariant ball $\mathcal{B}$. The discrete time synchronization filter m of [Hayden et al., 2011] may be expressed as follows:

$$\frac{\mathrm{d}m}{\mathrm{d}t} + m + B(m, m) = f, \quad t \in (t_k, t_{k+1}), \tag{3.4.9a}$$

$$m(t_k) = Pv(t_k) + Qm(t_k^-). \tag{3.4.9b}$$

Thus the filter consists of solving the underlying dynamical model, by resetting the filter to take the value $Pv(t)$ in the subspace $P\mathbb{R}^J$ at every time $t = t_k$. The following theorem shows that the filter m converges to the true signal v.

Theorem 3.4.5. *Let v be a solution of the equation (3.4.2) with $v(0) \in \mathcal{B}$. Then there exists $h^* > 0$ such that for any $h \in (0, h^*]$ the approximating solution m given by (3.4.9) converges to v as $t \to \infty$.*

3.4.2.2 Noisy Observations: Discrete 3DVAR

Now we consider the situation where the data is noisy and $H_k = P$. We employ the 3DVAR filter which results from the minimization principle (3.2.4) in the case where $\widehat{C}_{k+1} = \sigma^2 I$ and $\Gamma = \epsilon^2 I$. Recall the true signal is determined by the equation (3.2.2) and the observed data by the equation (3.2.3), now written in terms of the

true signal $v_k = v(t_k)$ solving the equation (3.3.3) with $v_0 \in \mathcal{B}$. Thus

$$\begin{aligned} v_{k+1} &= \Psi(v_k), \quad v_0 \in \mathcal{B}, \\ y_{k+1} &= Pv_{k+1} + \nu_{k+1}. \end{aligned}$$

If we define $\eta := \frac{\epsilon^2}{\sigma^2}$ then the 3DVAR filter can be written as

$$m_{k+1} = \Big(\frac{\eta}{1+\eta}P + Q\Big)\Psi(m_k) + \frac{1}{1+\eta}y_{k+1},$$

after noting that $Py_{k+1} = y_{k+1}$ because P is a projection and ν_{k+1} is assumed to lie in the image of P. In fact the data has the following form:

$$\begin{aligned} y_{k+1} &= Pv_{k+1} + P\nu_{k+1} \\ &= P\Psi(v_k) + \nu_{k+1}. \end{aligned}$$

Combining the two equations gives

$$m_{k+1} = \Big(\frac{\eta}{1+\eta}P + Q\Big)\Psi(m_k) + \frac{1}{1+\eta}\Big(P\Psi(v_k) + \nu_{k+1}\Big). \tag{3.4.10}$$

We can write the equation for the true solution v_k, given by (3.2.2), in the following form:

$$v_{k+1} = \Big(\frac{\eta}{1+\eta}P + Q\Big)\Psi(v_k) + \frac{1}{1+\eta}P\Psi(v_k). \tag{3.4.11}$$

Note that $v_k = v(t_k)$ where $v(\cdot)$ solves (3.4.2). We are interested in comparing the output of the filter, m_k, with the true signal v_k. Notice that if the noise ν_k is set to zero and if the limit $\eta \to 0$ is taken then the filter becomes

$$m_{k+1} = P\Psi(v_k) + Q\Psi(m_k)$$

which is precisely the discrete time synchronization filter. Theorem 3.4.6 below will reflect this observation, constituting a noisy variation on Theorem 3.4.5.

We will assume that the ν_k are independent random variables that satisfy the bound $|\nu_k| \leq \epsilon$, thereby linking the scale of the covariance Γ employed in 3DVAR to the size of the noise. We let $\|\cdot\|$ be the norm defined by $\|z\| := |z| + |Pz|$, $z \in \mathbb{R}^J$.

Theorem 3.4.6. *Let v be the solution of the equation (3.4.2) with $v(0) \in \mathcal{B}$. Assume that $\{\nu_k\}_{k\geq 1}$ is a sequence of independent bounded random variables such that, for every k, $|\nu_k| \leq \epsilon$. Then there are choices (detailed in the proof in the appendix) of assimilation step $h > 0$ and parameter $\eta > 0$ sufficiently small such that, for some*

$\alpha \in (0,1)$ and provided that the noise $\epsilon > 0$ is small enough, the error satisfies

$$\|m_{k+1} - v_{k+1}\| \leq \alpha \|m_k - v_k\| + 2\epsilon. \tag{3.4.12}$$

Thus, there is $a > 0$ such that

$$\limsup_{k\to\infty} \|m_k - v_k\| \leq a\epsilon.$$

3.5 Adaptive Observation Operator

The theory in the previous section demonstrates that accurate filtering of chaotic models is driven by observing enough of the dynamics to control the exponential separation of trajectories in the dynamics. However the fixed observation operator P that we analyze requires observation of $2/3$ of the system state vector. Even if the observation operator is fixed our numerical results will show that observation of this proportion of the state is not necessary to obtain accurate filtering. Furthermore, by adapting the observations to the dynamics, we will be able to obtain the same quality of reconstruction with even fewer observations. In this section we will demonstrate these ideas in the context of noisy discrete time filtering, and with reference to the Lorenz '96 model.

The variational equation for the dynamical system (3.2.1) is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} D\Psi(u,t) = D\mathcal{F}\big(\Psi(u,t)\big) \cdot D\Psi(u,t); \qquad D\Psi(u,0) = I_{J\times J}, \tag{3.5.1}$$

using the chain rule. The solution of the variational equation gives the derivative matrix of the solution operator Ψ, which in turn characterizes the behaviour of Ψ with respect to small variations in the initial value u. Let $L_{k+1} := L(t_{k+1})$ be the solution of the variational equation (3.5.1) over the assimilation window (t_k, t_{k+1}), initialized at $I_{J\times J}$, given as

$$\frac{\mathrm{d}L}{\mathrm{d}t} = D\mathcal{F}\big(\Psi(m_k, t-t_k)\big)L, \quad t \in (t_k, t_{k+1}); \qquad L(t_k) = I_{J\times J}. \tag{3.5.2}$$

Let $\{\lambda_k^j, \psi_k^j\}_{j=1}^J$ denote eigenvalue/eigenvector pairs of the matrix $L_{k+1}^T L_{k+1}$, where the eigenvalues (which are, of course, real) are ordered to be non-decreasing, and the eigenvectors are orthonormalized with respect to the Euclidean inner-product $\langle\cdot,\cdot\rangle$. We define the adaptive observation operator H_k to be

$$H_k := H_0(\psi_k^1, \cdots, \psi_k^J)^T \tag{3.5.3}$$

where

$$H_0 = \begin{pmatrix} 0 & 0 \\ 0 & I_{M\times M} \end{pmatrix}. \tag{3.5.4}$$

Thus H_0 and H_k both have rank M. Defined in this way we see that for any given $v \in \mathbb{R}^J$ the projection $H_k v$ is given by the vector

$$\left(0, \cdots, 0, \langle \psi_k^{J-M+1}, v\rangle, \cdots, \langle \psi_k^{J}, v\rangle\right)^T,$$

that is the projection of v onto the M eigenvectors of $L_{k+1}^T L_{k+1}$ with largest modulus.

Remark 3.5.1. *In the following work we consider the leading eigenvalues and corresponding eigenvectors of the matrix $L_k^T L_k$ to track the unstable (positive Lyapunov growth) directions. To leading order in h it is equivalent to consider the matrix $L_k L_k^T$ in the case of frequent observations (small h) as can be seen by the following expressions*

$$\begin{aligned} L_k^T L_k &= (I + hD\mathcal{F}_k)^T (I + hD\mathcal{F}_k) + \mathcal{O}(h^2) \\ &= I + h(D\mathcal{F}_k^T + D\mathcal{F}_k) + \mathcal{O}(h^2) \end{aligned}$$

and

$$\begin{aligned} L_k L_k^T &= (I + hD\mathcal{F}_k)(I + hD\mathcal{F}_k)^T + \mathcal{O}(h^2) \\ &= I + h(D\mathcal{F}_k + D\mathcal{F}_k^T) + \mathcal{O}(h^2), \end{aligned}$$

where $D\mathcal{F}_k = D\mathcal{F}(m_k)$.

Of course for large intervals h, the above does not hold, and the difference between $L_k^T L_k$ and $L_k L_k^T$ may be substantial. It is however clear that these operators have the same eigenvalues, with the eigenvectors of $L_k L_k^T$ corresponding to λ_k^j given by $L_k \psi_k^j$ for the corresponding eigenvector ψ_k^j of $L_k^T L_k$. That is to say, for the linearized deformation map L_k, the direction ψ_k^j is the pre-deformation principle direction corresponding to the principle strain λ_k^j induced by the deformation. The direction $L_k \psi_k^j$ is the post-deformation principle direction corresponding to the principle strain λ_k^j. The dominant directions chosen in Eq. (3.5.3) are those directions corresponding to the greatest growth over the interval (t_k, t_{k+1}) of infinitesimal perturbations to the predicting trajectory, $\Psi(m_{k-1}, h)$ at time t_k. This is only one sensible option. One could alternatively consider the directions corresponding to the

greatest growth over the interval (t_{k-1}, t_k), or over the whole interval (t_{k-1}, t_{k+1}). Investigation of these alternatives is beyond the scope of this work and is therefore deferred to later investigation.

We make a small shift of notation and now consider the observation operator H_k as a linear mapping from $\mathbb{R}^J$ into $\mathbb{R}^M$, rather than as a linear operator from $\mathbb{R}^J$ into itself, with rank M; the latter perspective was advantageous for the presentation of the analysis, but differs from the former which is sometimes computationally advantageous and more widely used for the description of algorithms. Recall the minimization principle (3.2.4), noting that now the first norm is in $\mathbb{R}^J$ and the second in $\mathbb{R}^M$.

3.5.1 3DVAR

Here we consider the minimization principle (3.2.4) with the choice $\widehat{C}_{k+1} = C_0 \in \mathbb{R}^{J\times J}$, a strictly positive-definite matrix, for all k. Assuming that $\Gamma \in \mathbb{R}^{M\times M}$ is also strictly positive-definite, the filter may be written as

$$m_{k+1} = \Psi(m_k) + G_{k+1}\Big(y_{k+1} - H_{k+1}\Psi(m_k)\Big) \tag{3.5.5a}$$

$$G_{k+1} = C_0 H_{k+1}^T (H_{k+1} C_0 H_{k+1}^T + \Gamma)^{-1}. \tag{3.5.5b}$$

As well as using the choice of H_k defined in (3.5.3), we also employ the fixed observation operator where $H_k = H$, including the choice $H = P$ given by (3.4.1). In the last case $J = 3J'$, $M = 2J'$ and P is realized as a$2J' \times 3J'$ matrix.

We make the choices $C_0 = \sigma^2 I_{J\times J}$, $\Gamma = \epsilon^2 I_{M\times M}$ and define $\eta = \epsilon^2/\sigma^2$. Throughout our experiments we take $h = 0.1$, $\epsilon^2 = 0.01$ and fix the parameter $\eta = 0.01$ (i.e. $\sigma = 1$). We use the Lorenz '96 model (3.3.1) to define Ψ, with the parameter choices $F = 8$ and $J = 60$. The system then has 19 positive Lyapunov exponents which we calculate by the methods described in [Benettin et al., 1976]. The observational noise is i.i.d Gaussian with respect to time index k, with distribution $\nu_1 \sim N(0, \epsilon^2)$.

Throughout the following we show (approximation) to the expected value, with respect to noise realizations around a single fixed true signal solving (3.4.2), of the error between the filter and the signal underlying the data, in the Euclidean norm, as a function of time. We also quote numbers which are found by time-averaging this quantity. The expectation is approximated by a Monte Carlo method in which I realizations of the noise in the data are created, leading to filters $m_k^{(i)}$,

with k denoting time and i denoting realization. Thus we have, for $t_k = kh$,

$$\text{RMSE}(t_k) = \frac{1}{I}\sum_{i=1}^{I}\sqrt{\frac{\|m_k^{(i)} - v_k\|^2}{J}}.$$

This quantity is graphed, as a function of k, in what follows. Notice that similar results are obtained if only one realization is used ($I = 1$) but they are more noisy and hence the trends underlying them are not so clear. We take $I = 10^4$ throughout the reported numerical results. When we state a number for the RMSE this will be found by time-averaging after ignoring the initial transients ($t_k < 40$):

$$\text{RMSE} = \underset{t_k>40}{\text{mean}}\{\text{RMSE}(t_k)\}.$$

In what follows we will simply refer to RMSE ; from the context it will be clear whether we are talking about the function of time, RMSE(t_k), or the time-averaged number RMSE.

Figures 3.1, 3.2 and 3.3 exhibit, for fixed observation 3DVAR and adaptive observation 3DVAR, the RMSE as a function of time. The Figure 3.1 shows the RMSE for fixed observation operator where the observed space is of dimension 60 (complete observations), 40 (observation operator defined as in the equation (3.4.1)), 36 and 24 respectively. For values $M = 60$, 40 and 36 the error decreases rapidly and the approximate solution converges to a neighbourhood of the true solution where the size of the neighbourhood depends upon the variance of the observational noise. For the cases $M = 60$ and $M = 40$ we use the identity operator $I_{J\times J}$ and the projection operator P as defined in the equation (3.4.1) as the observation operators respectively. The observation operator for the case $M = 36$ can be given as

$$P_{36} = \Big(\ e_1,\, e_2,\, 0,\, e_4,\, 0,\, e_6, e_7,\, 0,\, e_9,\, 0,\, e_{11},\, e_{12},\, 0,\, e_{14}, \dots \ \Big)_{J\times J} \tag{3.5.6}$$

where we observe 3 out of 5 directions periodically. The RMSE , averaged over the trajectory, after ignoring the initial transients, is 1.30×10^{-2} when $M = 60$, 1.14×10^{-2} when $M = 40$ and 1.90×10^{-2} when $M = 36$; note that this is on the scale of the observational noise. The rate of convergence of the approximate solution to the true solution in the case of partial observations is lower than the rate of convergence when full observations are used however the RMSE is lower in the case when $M = 40$ due to fewer noisy inputs in stable directions in comparison to the case when all directions are observed. The convergence of the approximate solution to the true solution for the case when $M = 36$ shows that the value $M = 40$,

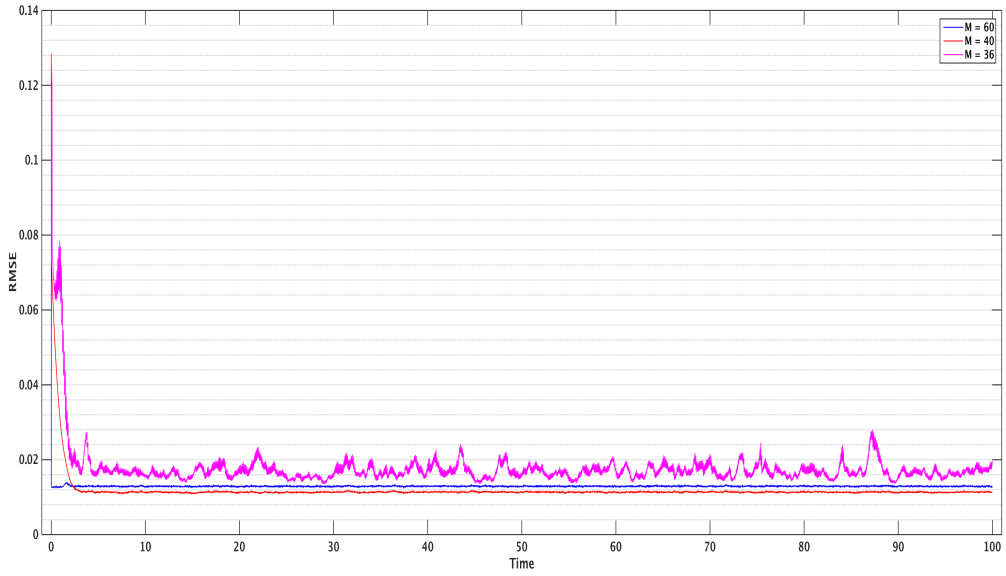


Figure 3.1: Fixed Observation Operator 3DVAR. RMSE values averaged over the trajectory for $M = 60$, 40 and 36 are 1.30×10^{-2}, 1.14×10^{-2} and 1.90×10^{-2} respectively.

for which theoretical results have been presented in Section 3.4, is not required for small error ($\mathcal{O}(\epsilon)$) consistently over the trajectory. We also consider the case when $24 = 40\%$ of the modes are observed using the following observation operator:

$$P_{24} = \Big(\; e_1,\, 0,\, 0,\, e_4,\, 0,\, 0, e_7,\, 0,\, 0,\, e_{10},\, e_{11},\, 0,\, 0,\, e_{14}, \ldots \; \Big)_{J \times J}. \tag{3.5.7}$$

Thus we observe 4 out of 10 directions periodically; this structure is motivated by the work reported in [Abarbanel, 2013; Kostuk, 2012] where it was demonstrated that observing 40% of the modes, with the observation directions chosen carefully and with observations sufficiently frequent in time, is sufficient for the approximate solution to converge to the true underlying solution. The Figure 3.2 shows that, in our observational set-up, observing 24 of the modes only allows marginally successful reconstruction of the signal, asymptotically in time; the RMSE makes regular large excursions and the time-averaged RMSE over the trajectory is (5.73×10^{-2}), which is an order of magnitude larger than for 36, 40 or 60 observations.

Figure 3.3 shows the RMSE for adaptive observation 3DVAR. In this case we notice that the error is consistently small, uniformly in time, with just 9 or more modes observed. When $M = 9$ (15% observed modes) the RMSE averaged

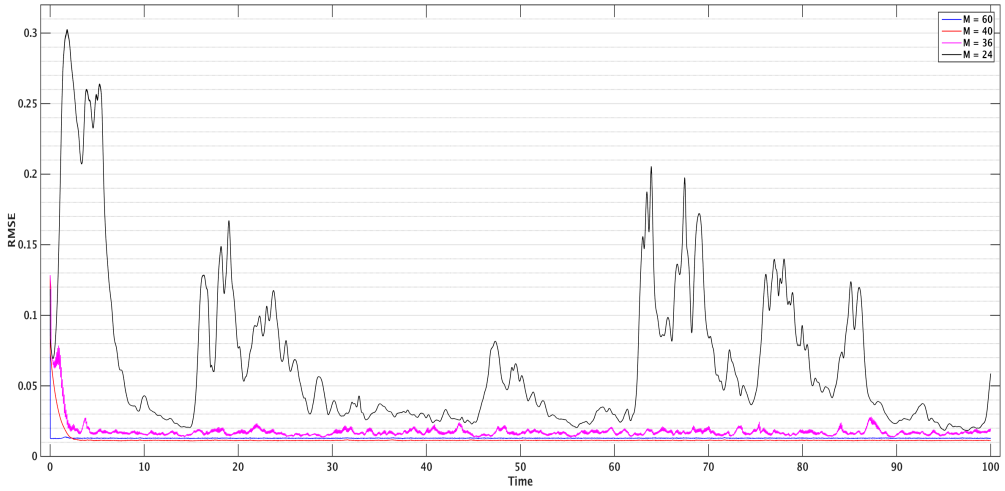


Figure 3.2: Fixed Observation Operator 3DVAR. Comparison with the case when $M = 24$. RMSE value averaged over the trajectory for $M = 24$ is 5.73×10^{-2}.

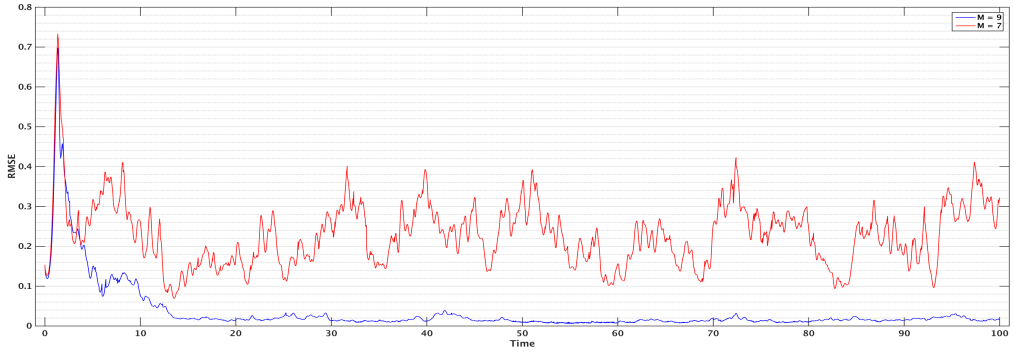
over the trajectory is 1.35×10^{-2} which again is of the order of the observational noise variance. For $M \geq 9$ the error is similar – see Figure 3.3b. On the other hand, for smaller values of M the error is not controlled as shown in Figure 3.3a where the RMSE for $M = 7$ is compared with that for $M = 9$; for $M = 7$ it is an order of magnitude larger than for $M = 9$. It is noteworthy that the number of observations necessary and sufficient for accurate reconstruction is approximately half the number of positive Lyapunov exponents.

3.5.2 Extended Kalman Filter

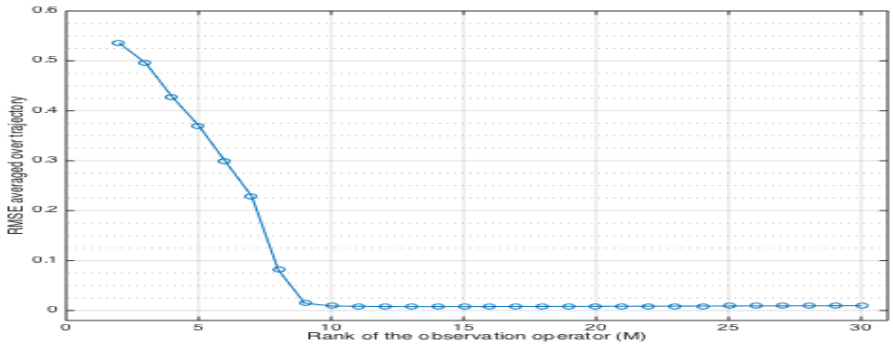
In the Extended Kalman Filter (ExKF) the approximate solution evolves according to the minimization principle (3.2.4) with C_k chosen as a covariance matrix evolving in the forecast step according to the linearized dynamics, and in the assimilation stage updated according to Bayes' rule based on a Gaussian observational error covariance. This gives the method

$$\begin{aligned} m_{k+1} &= \Psi(m_k) + G_{k+1}\Big(y_{k+1} - H_{k+1}\Psi(m_k)\Big), \\ \widehat{C}_{k+1} &= D\Psi(m_k) C_k D\Psi(m_k)^T, \\ C_{k+1} &= (I_{J\times J} - G_{k+1}H_{k+1})\widehat{C}_{k+1}, \\ G_{k+1} &= \widehat{C}_{k+1}H_{k+1}^T(H_{k+1}\widehat{C}_{k+1}H_{k+1}^T + \Gamma)^{-1}. \end{aligned}$$

We first consider the ExKF scheme with a fixed observation operator $H_k =$



(a) Comparison of RMSE between M = 7 and M = 9. RMSE values averaged over trajectory are 2.25×10^{-1}, 1.35×10^{-2} respectively.

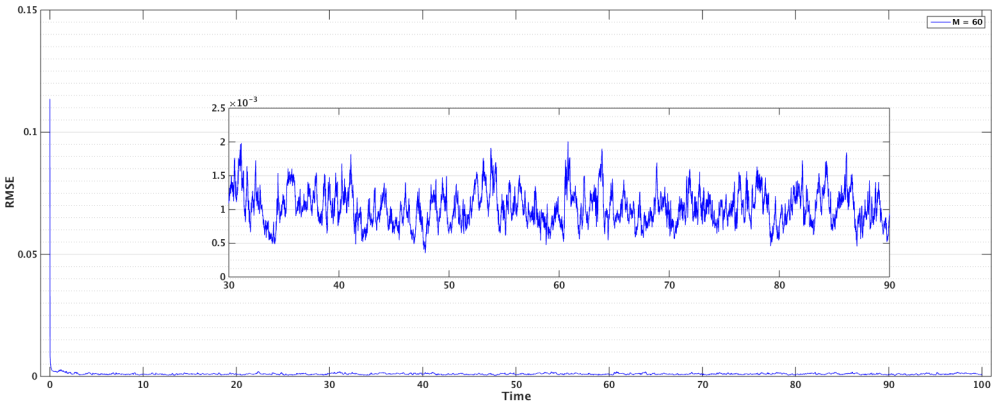


(b) Averaged RMSE for different choices of M

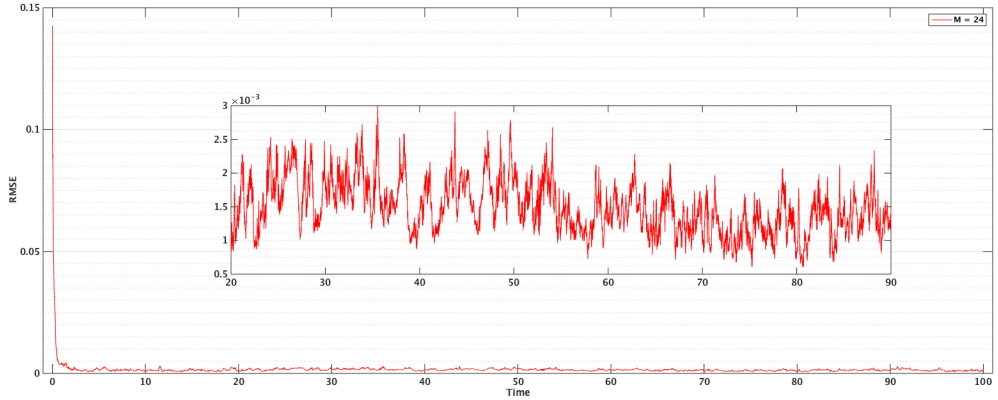
Figure 3.3: Adaptive Observation 3DVAR

H. We make two choices for H: the full rank identity operator and a partial observation operator given by (3.5.7) so that 40% of the modes are observed. For the first case the filtering scheme is the standard ExKF with all the modes being observed. The approximate solution converges to the true solution and the error decreases rapidly as can be seen in the Figure 3.4a. The RMSE is 9.49×10^{-4} which is an order of magnitude smaller than the analogous error for the 3DVAR algorithm when fully observed which is, recall, 1.30×10^{-2}. For the partial observations case with $M = 24$ we see that again the approximate solution converges to the true underlying solution as shown in the Figure 3.4b. Furthermore the solution given by the ExKF with $M = 24$ is far more robust than for 3DVAR with this number of observations. The RMSE is also lower for ExKF (2.68×10^{-3}) when compared with the 3DVAR scheme (5.73×10^{-2}).

We now turn to adaptive observation within the context of the ExKF. The



(a) Percentage of components observed = 100%. RMSE value averaged over trajectory 9.49×10^{-4}.



(b) Percentage of components observed = 40%. RMSE value averaged over trajectory 1.39×10^{-3}.

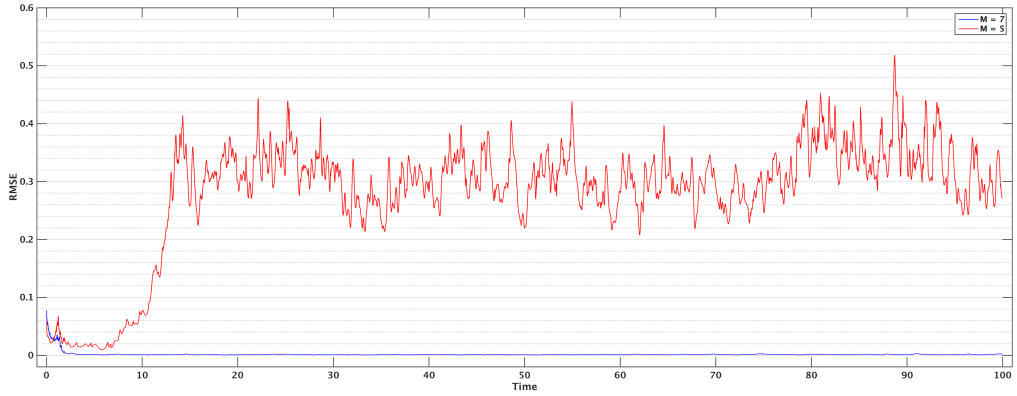
Figure 3.4: Fixed Observation ExKF. The zoomed in figures shows the variability in RMSE between time $t = 20$ and $t = 90$.

Figure 3.5 shows that it is possible to obtain an RMSE which is of the order of the observational error, and is robust over long time intervals, using only a 7 dimensional observation space, improving marginally on the 3DVAR situation where 9 dimensions were required to attain a similar level of accuracy.

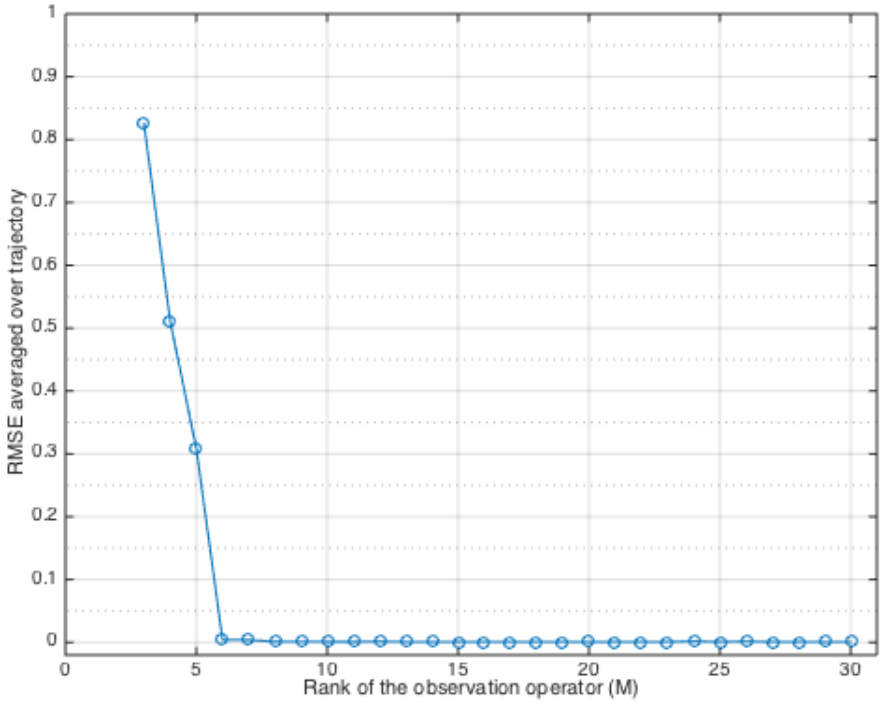
The AUS scheme, as proposed by Trevisan and co-workers [Trevisan and Uboldi, 2004], is an ExKF method which operates by confining the analysis update to the subspace spanned by a finite number of directions, ideally designed to capture the instabilities in the dynamics. This is typically achieved by choosing to work in the subspace of the linear dynamics spanned by the M largest growth directions; furthermore M is fixed as the number (precomputed) of non-negative Lyapunov exponents. Asymptotically this method with $H = I_{J\times J}$ behaves similarly to the adaptive ExKF with observation operator of rank M. To understand the intuition behind the AUS method we plot in Figure 3.6a the rank (computed by truncation to zero of eigenvalues below a threshold) of the covariance matrix C_k from standard ExKF based on observing 60 and 24 modes. Notice that in both cases the rank approaches a value of 19 or 20 and that 19 is the number of non-negative Lyapunov exponents. This means that the covariance is effectively zero in 40 of the observed dimensions and that, as a consequence of the minimization principle (3.2.4), data will be ignored in the 40 dimensions where the covariance is negligible. It is hence natural to simply confine the update step to the subspace of dimension 19 given by the number of positive Lyapunov exponents, right from the outset. This is exatcly what AUS does by reducing the rank of the error covariance matrix C_k. Numerical results are given in Figure 3.6b which shows the RMSE over the trajectory for the ExKF-AUS assimilation scheme with time. After initial transients the error is mostly of the numerical order of the observational noise. Occasional jumps outside this error bound are observed but the approximate solution converges to the true solution each time. The RMSE for ExKF-AUS is 1.49×10^{-2}. However, if the rank of the error covariance matrix C_0 in AUS is chosen to be less than the number of unstable modes for the underlying system, then the approximate solution does not converge to the true solution.

3.6 Conclusions

In this chapter we have studied the long-time behaviour of filters for partially observed dissipative dynamical systems, using the Lorenz '96 model as a canonical example. We have highlighted the connection to synchronization in dynamical systems, and shown that this synchronization theory, which applies to noise-free data,

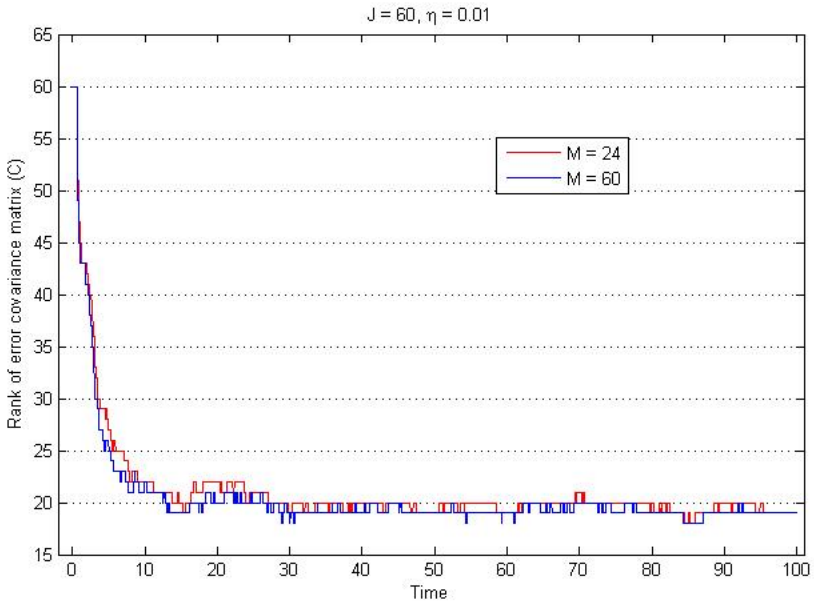


(a) Comparison of RMSE between M = 5 and M = 7. RMSE values averaged over trajectory are 2.84×10^{-1}, 1.31×10^{-3} respectively.

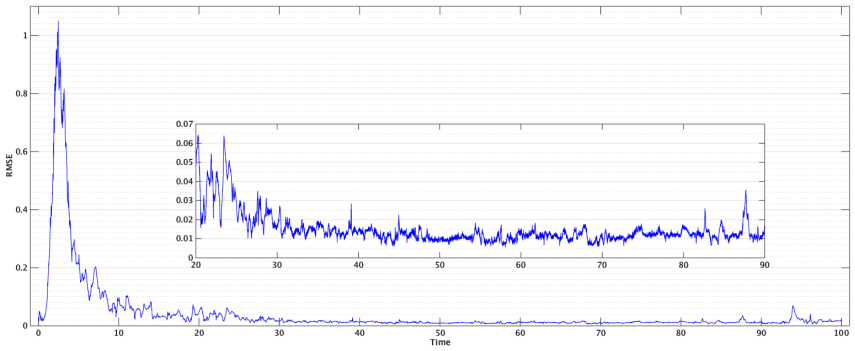


(b) Averaged RMSE for different choices of M.

Figure 3.5: Adaptive Observation ExKF



(a) Standard ExKF with 60 and 24 observed modes. The rank of the error covariance matrix C_k decays to (approximately) the number of unstable Lyapunov modes in the underlying system, namely 19.



(b) RMSE value averaged over trajectory: 1.49×10^{-2}. The zoomed in figures shows the variability in RMSE between time $t = 20$ and $t = 90$.

Figure 3.6: Rank of error covariance and ExKF-Assimilation in Unstable Space

is robust to the addition of noise, in both the continuous and discrete time settings. In so doing we are studying the 3DVAR algorithm. In the context of the Lorenz '96 model we have identified a fixed observation operator, based on observing 2/3 of the components of the signal's vector, which is sufficient to ensure desirable long-time properties of the filter. However it is to be expected that, within the context of fixed observation operators, considerably fewer observations may be needed to ensure such desirable properties. Ideas from nonlinear control theory will be relevant in addressing this issue. We also studied adaptive observation operators, targeted to observe the directions of maximal growth within the local linearized dynamics. We demonstrated that with these adaptive observers, considerably fewer observations are required. We also made a connection between these adaptive observation operators, and the AUS methodology which is also based on the local linearized dynamics, but works by projecting within the model covariance operators of ExKF, whilst the observation operators themselves are fixed; thus the model covariances are adapted. Both adaptive observation operators and the AUS methodology show the potential for considerable computational savings in filtering, without loss of accuracy.

In conclusion our work highlights the role of ideas from dynamical systems in the rigorous analysis of filtering schemes and, through computational studies, shows the gap between theory and practice, demonstrating the need for further theoretical developments. We emphasize that the adaptive observation operator methods may not be implementable in practice on the high dimensional systems arising in, for example, meteorological applications. However, they provide conceptual insights into the development of improved algorithms and it is hence important to understand their properties.

Appendix: Proofs

Proof of Properties 3.3.1. Properties 1, 2 and 3 are straightforward and we omit the proofs. We start showing 4. For any $u \in \mathbb{R}^J$ set

$$\|u\|_\infty = \max_{1\le j\le J} |u^{(j)}|$$

and recall that $|u|^2 \ge \|u\|_\infty^2$. Then, for $u, \tilde{u} \in \mathbb{R}^J$, and for $1 \le j \le J$, we have that

$$2|B(u,\tilde{u})^{(j)}| \;\le\; \|u\|_\infty(|\tilde{u}^{(j+1)}| + |\tilde{u}^{(j-2)}|) + \|\tilde{u}\|_\infty(|u^{(j+1)}| + |u^{(j-2)}|),$$

and so

$$\begin{aligned} 4|B(u,\tilde{u})|^2 &\le 2\|u\|_\infty^2 \sum_{j=1}^{J}(|\tilde{u}^{(j+1)}| + |\tilde{u}^{(j-2)}|)^2 + 2\|\tilde{u}\|_\infty^2 \sum_{j=1}^{J}(|u^{(j+1)}| + |u^{(j-2)}|)^2 \\ &\le 8\|u\|_\infty^2|\tilde{u}|^2 + 8\|\tilde{u}\|_\infty^2|u|^2 \\ &\le 16|u|^2|\tilde{u}|^2. \end{aligned}$$

Hence

$$|B(u,\tilde{u})| \le 2|u||\tilde{u}|.$$

For 5 we use rearrangement and periodicity of indices under summation as follows:

$$\begin{aligned} 2\langle B(u,\tilde{u}),u\rangle &= \sum_{j=1}^{J}\Big(u^{(j)}(u^{(j-1)}\tilde{u}^{(j+1)} + \tilde{u}^{(j-1)}u^{(j+1)} - \tilde{u}^{(j-1)}u^{(j-2)} - u^{(j-1)}\tilde{u}^{(j-2)})\Big) \\ &= \sum_{j=1}^{J}(u^{(j)}u^{(j-1)}\tilde{u}^{(j+1)} - u^{(j)}\tilde{u}^{(j-1)}u^{(j-2)}) \\ &= \sum_{j=1}^{J}(u^{(j-1)}u^{(j-2)}\tilde{u}^{(j)} - u^{(j+1)}\tilde{u}^{(j)}u^{(j-1)}) \\ &= \sum_{j=1}^{J}\Big(\tilde{u}^{(j)}(u^{(j-1)}u^{(j-2)} - u^{(j+1)}u^{(j-1)})\Big) \\ &= -\langle B(u,u),\tilde{u}\rangle. \end{aligned}$$

□

Proof of Proposition 3.3.2. Taking the Euclidean inner product of $u(t)$ with equa-

tion (3.3.3) and using properties 1 and 2 we get

$$\frac{1}{2}\frac{\mathrm{d}|u|^2}{\mathrm{d}t} = -|u|^2 + \langle f, u\rangle.$$

Using Cauchy-Schwartz and Young's inequalities for the last term gives

$$\frac{\mathrm{d}|u|^2}{\mathrm{d}t} + |u|^2 \leq JF^2.$$

Therefore, using Gronwall's lemma,

$$|u(t)|^2 \leq |u_0|^2 e^{-t} + JF^2(1 - e^{-t}),$$

and the result follows. □

Proof of Property 3.4.2. The first part is automatic since, if $q := Qu$, then for all j either $q^{(j-1)} = 0$ or $q^{(j-2)} = q^{(j+1)} = 0$. Since $B(Qu, Qu) = 0$ and $B(\cdot,\cdot)$ is a bilinear operator we can write

$$\begin{aligned} B(u,u) &= B(Pu + Qu, Pu + Qu) \\ &= B(Pu, Pu) + 2B(Pu, Qu). \end{aligned}$$

Now using property 4, and the fact that there is $c > 0$ such that $|Pu| + 2|Qu| \leq \frac{c}{2}|u|$,

$$\begin{aligned} |\langle B(u,u), \tilde{u}\rangle| &\leq |B(u,u)||\tilde{u}| \\ &\leq |B(Pu,Pu) + 2B(Pu,Qu)||\tilde{u}| \\ &\leq 2|Pu||\tilde{u}|(|Pu| + 2|Qu|) \\ &\leq c|Pu||\tilde{u}||u|. \end{aligned}$$

□

Proof of Theorem 3.4.3. Define the error in the approximate solution as $\delta = m - v = q - Qv$. Note that $Q\delta = \delta$. The error satisfies the following equation

$$Q\frac{\mathrm{d}\delta}{\mathrm{d}t} + Q\delta + Q\big(B(Pv + q, Pv + q) - B(v,v)\big) = 0.$$

Splitting $v = Pv + Qv$ and noting, from Properties 3.4.2, that $B(Qv, Qv) = 0$ and $B(q,q) = 0$, yields

$$\frac{\mathrm{d}Q\delta}{\mathrm{d}t} + Q\delta + 2QB(Pv, Q\delta) = 0.$$

Taking the inner product with $Q\delta$ gives

$$\frac{1}{2}\frac{\mathrm{d}|Q\delta|^2}{\mathrm{d}t} + |Q\delta|^2 + 2\langle B(Pv, Q\delta), Q\delta\rangle = 0.$$

Note that from the Properties 3.3.1, 3 and 5, and Property 3.4.2, we have

$$\begin{aligned} 2\langle B(u, Q\delta), Q\delta\rangle &= -\langle B(Q\delta, Q\delta), u\rangle \\ &= 0. \end{aligned}$$

Thus since $Q\delta = \delta$ we have

$$\frac{\mathrm{d}|\delta|^2}{\mathrm{d}t} + 2|\delta|^2 = 0,$$

and so

$$|\delta(t)|^2 = |\delta(0)|^2 e^{-2t}.$$

As $t \to \infty$ the error $\delta(t) \to 0$. □

Proof of Theorem 3.4.4. From (3.4.5) and (3.4.6)

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \mathcal{F}(m) + \frac{1}{\eta}\Big(Pv + \epsilon P\frac{\mathrm{d}w}{\mathrm{d}t} - Pm\Big).$$

Thus

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -m - B(m, m) + f + \frac{1}{\eta}P(v - m) + \frac{\epsilon}{\eta}P\frac{\mathrm{d}w}{\mathrm{d}t}.$$

The signal is given by

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -v - B(v, v) + f,$$

and so the error $\delta = m - v$ satisfies

$$\frac{\mathrm{d}\delta}{\mathrm{d}t} = -\delta - 2B(v, \delta) - B(\delta, \delta) - \frac{1}{\eta}P\delta + \frac{\epsilon}{\eta}P\frac{\mathrm{d}w}{\mathrm{d}t}.$$

Lemma 3.6.2 below, Properties 3.3.1 and Itô's formula give

$$\frac{1}{2}\mathrm{d}|\delta|^2 + \Big(1 - \frac{c^2 K\eta}{4}\Big)|\delta|^2\mathrm{d}t \leq \frac{\epsilon}{\eta}\langle P\mathrm{d}w, \delta\rangle + \frac{J}{3}\frac{\epsilon^2}{\eta^2}\mathrm{d}t.$$

Integrating and taking expectations

$$\frac{\mathrm{d}\mathbb{E}|\delta|^2}{\mathrm{d}t} \leq -\lambda\mathbb{E}|\delta|^2 + \frac{2J\epsilon^2}{3\eta^2}.$$

Use of the Gronwall inequality gives the desired result. □

We now turn to discrete-time data assimilation, where the following lemma plays an important role:

Lemma 3.6.1. *Consider the Lorenz '96 model* (3.3.3) *with* $F > 0$ *and* $J \geq 3$. *Let* v *and* u *be two solutions in* $[t_k, t_{k+1})$, *with* $v(t_k) \in \mathcal{B}$. *Then there exists a* $\beta \in \mathbb{R}$ *such that*

$$|u(t) - v(t)|^2 \leq |u(t_k) - v(t_k)|^2 e^{\beta(t-t_k)} \quad t \in [t_k, t_{k+1}).$$

Proof. Let $\delta = m - v$. Then δ satisfies

$$\frac{1}{2}\frac{\mathrm{d}|\delta|^2}{\mathrm{d}t} + |\delta|^2 + 2\langle B(v,\delta),\delta\rangle + \langle B(\delta,\delta),\delta\rangle = 0 \tag{3.6.1}$$

so that, by Property 3.3.1, item 2,

$$\frac{1}{2}\frac{\mathrm{d}|\delta|^2}{\mathrm{d}t} + |\delta|^2 - 2|\langle B(v,\delta),\delta\rangle| \leq 0.$$

Using Properties 3.3.1 items 4 and 5 gives $|\langle B(v,\delta),\delta\rangle| \leq K^{\frac{1}{2}}|\delta|^2$, where K is defined in Proposition 3.3.2, so that

$$\frac{1}{2}\frac{\mathrm{d}|\delta|^2}{\mathrm{d}t} \leq (2K^{\frac{1}{2}} - 1)|\delta|^2.$$

Integrating the differential inequality gives

$$|\delta(t)|^2 \leq |\delta(t_k)|^2 e^{\beta(t-t_k)}. \tag{3.6.2}$$

□

Note if $F < \frac{1}{2\sqrt{2}J}$ then $\beta = 2(2K^{\frac{1}{2}} - 1) < 0$ and the subsequent analysis may be significantly simplified. Thus we assume in what follows that $F \geq \frac{1}{2\sqrt{2}J}$ so that $\beta \geq 0$. Lemma 3.6.1 gives an estimate on the growth of the error in the forecast step. Our aim now is to show that this growth can be controlled by observing Pv discretely in time. It will be required that the time h between observations is sufficiently small.

To ease the notation we introduce three functions that will be used in the proofs of Theorems 3.4.2 and 3.4.6. Namely we define, for $t > 0$,

$$A_1(t) := \frac{16K}{\beta}(e^{\beta t} - 1) + \frac{4R_0^2}{2\beta}(e^{2\beta t} - 1), \tag{3.6.3}$$

$$B_1(t) := \frac{16c^2K^2}{\beta}\left[\frac{e^{\beta t}-e^{-t}}{\beta+1}-(1-e^{-t})\right]+e^{-t}+\frac{4c^2KR_0^2}{2\beta}\left[\frac{e^{2\beta t}-e^{-t}}{2\beta+1}-(1-e^{-t})\right], \tag{3.6.4}$$

and

$$B_2(t) := c^2K\{1-e^{-t}\}. \tag{3.6.5}$$

Here and in what follows c, β and K are as in Property 3.4.2, Lemma 3.6.1 and Proposition 3.3.2. We will use two different norms in $\mathbb{R}^J$ to prove the theorems that follow. In each case, the constant $R_0 > 0$ above quantifies the size of the initial error, measured in the relevant norm for the result at hand.

Proof of Theorem 3.4.5. Define the error $\delta = m - v$. Subtracting equation (3.4.2) from equation (3.4.9) gives

$$\frac{\mathrm{d}\delta}{\mathrm{d}t} + \delta + 2B(v,\delta) + B(\delta,\delta) = 0, \quad t \in (t_k, t_{k+1}), \tag{3.6.6a}$$

$$\delta(t_k) = Q\delta(t_k^-) \tag{3.6.6b}$$

where $\delta(t_{k+1}^-) := \lim_{t\uparrow t_{k+1}} \delta(t)$ as defined in Section 3.4.2.1. Notice that $B_1(0) = 1$ and $B_1'(0) = -1$, so that there is $h^* > 0$ with the property that $B_1(h) \in (0,1)$ for all $h \in (0, h^*]$. Fix any such assimilation time h and denote $\gamma = B_1(h) \in (0,1)$. Let $R_0 := |\delta_0|$. We show by induction that, for every k, $|\delta_k|^2 \leq \gamma^k R_0^2$. We suppose that it is true for k and we prove it for $k+1$.

Taking the inner product of $P\delta$ with the equation (3.6.6) gives

$$\frac{1}{2}\frac{\mathrm{d}|P\delta|^2}{\mathrm{d}t} + |P\delta|^2 + 2\langle B(v,\delta), P\delta\rangle + \langle B(\delta,\delta), P\delta\rangle = 0$$

so that, by Property 3.3.1, item 4,

$$\frac{1}{2}\frac{\mathrm{d}|P\delta|^2}{\mathrm{d}t} + |P\delta|^2 \leq 4|v||\delta||P\delta| + 2|\delta|^2|P\delta|.$$

By the inductive hypothesis we have $|\delta_k|^2 \leq R_0^2$ since $\gamma \in (0,1)$. Shifting the time origin by setting $\tau := t - t_k$ and using Lemma 3.6.1 gives

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}|P\delta|^2}{\mathrm{d}\tau} + |P\delta|^2 &\leq 4K^{\frac{1}{2}}|\delta||P\delta| + 2|\delta_k|e^{\frac{\beta\tau}{2}}|\delta||P\delta| \\ &\leq 4K^{\frac{1}{2}}|\delta||P\delta| + 2R_0e^{\frac{\beta\tau}{2}}|\delta||P\delta|. \end{aligned} \tag{3.6.7}$$

Applying Young's inequality to each term on the right-hand side we obtain

$$\frac{\mathrm{d}|P\delta|^2}{\mathrm{d}\tau} \leq 16K|\delta|^2 + 4R_0^2 e^{\beta\tau}|\delta|^2. \tag{3.6.8}$$

Integrating from 0 to s, where $s \in (0, h)$, gives

$$|P\delta(s)|^2 \leq A_1(s)|\delta_k|^2. \tag{3.6.9}$$

Now again consider the equation (3.6.1) using Property 3.3.1 item 5 to obtain

$$\frac{1}{2}\frac{\mathrm{d}|\delta|^2}{\mathrm{d}\tau} + |\delta|^2 - |\langle B(\delta,\delta), v\rangle| \leq 0.$$

Using Property 3.4.2 and Young's inequality yields

$$\begin{aligned} \frac{1}{2}\frac{\mathrm{d}|\delta|^2}{\mathrm{d}\tau} + |\delta|^2 &\leq c|v||\delta||P\delta| \\ &\leq cK^{\frac{1}{2}}|\delta||P\delta| \\ &\leq \frac{|\delta|^2}{2} + \frac{c^2K}{2}|P\delta|^2. \end{aligned} \tag{3.6.10}$$

Employing the bound (3.6.9) then gives

$$\frac{\mathrm{d}|\delta|^2}{\mathrm{d}\tau} + |\delta|^2 \leq \left(\frac{16c^2K^2}{\beta}(e^{\beta\tau} - 1) + \frac{4c^2KR_0^2}{2\beta}(e^{2\beta\tau} - 1)\right)|\delta_k|^2.$$

Therefore, upon using Gronwall's lemma,

$$|\delta(s)|^2 \leq B_1(s)|\delta_k|^2.$$

It follows that

$$|\delta_{k+1}|^2 \leq \gamma|\delta_k|^2 \leq \gamma^{k+1}R_0^2,$$

and the induction (and hence the proof) is complete. □

Proof of Theorem 3.4.6. We define the error process $\delta(t)$ as follows:

$$\delta(t) = \begin{cases} \delta_k := m_k - v(t) & \text{if } t = t_k \\ \Psi(m_k, t - t_k) - v(t) & \text{if } t \in (t_k, t_{k+1}). \end{cases} \tag{3.6.11}$$

Observe that δ is discontinuous at times t_k which are multiples of h, since

$m_{k+1} \neq \Psi(m_k; h)$. Subtracting (3.4.11) from (3.4.10) we obtain

$$\delta_{k+1} = \delta(t_{k+1}) = \left(\frac{\eta}{1+\eta}P + Q\right) + \frac{1}{1+\eta}\nu_{k+1}. \tag{3.6.12}$$

Let $A_1(\cdot)$, $B_1(\cdot)$ and $B_2(\cdot)$ be as in (3.6.3, 3.6.4, 3.6.5), and set

$$\begin{aligned} M_1(t) :=& \frac{2\eta}{1+\eta}\sqrt{A_1(t)} + \sqrt{B_1(t)}, \\ M_2(t) :=& \frac{2\eta}{1+\eta} + \sqrt{B_2(t)}. \end{aligned}$$

Since $A_1(0) = 0, B_1(0) = 1$, $B_2(0) = 0$ and

$$\frac{d}{dt}\sqrt{B_1(t)}\Big|_{t=0} = -1/2 < 0$$

it is possible to find $h, \eta > 0$ small such that

$$M_2(h) < M_1(h) =: \alpha < 1.$$

Let $R_0 = \|\delta_0\|$. We show by induction that for such h and η, and provided that ϵ is small enough so that

$$\alpha R_0 + 2\epsilon < R_0,$$

we have that $\|\delta_k\| \leq R_0$ for all k. Suppose for induction that it is true for k. Then $|\delta_k| \leq \|\delta_k\| \leq R_0$ and we can apply (after shifting time as before) Lemma 3.6.3 below to obtain that

$$|P\delta(t)| \leq \sqrt{A_1(t)|\delta_k|^2 + |P\delta_k|^2} \leq \sqrt{A_1(t)}|\delta_k| + |P\delta_k|$$

and

$$|\delta(t)| \leq \sqrt{B_1(t)|\delta_k|^2 + B_2(t)|P\delta_k|^2} \leq \sqrt{B_1(t)}|\delta_k| + \sqrt{B_2(t)}|P\delta_k|.$$

Therefore,

$$\begin{aligned} |P\delta_{k+1}| + |\delta_{k+1}| &\leq \left(\frac{2\eta}{1+\eta}\sqrt{A_1(h)} + \sqrt{B_1(h)}\right)|\delta_k| + \left(\frac{2\eta}{1+\eta} + \sqrt{B_2(h)}\right)|P\delta_k| + 2\epsilon \\ &= M_1(h)|\delta_k| + M_2(h)|P\delta_k| + 2\epsilon. \end{aligned}$$

Since $M_2(h) < M_1(h) = \alpha$ we deduce that

$$\|\delta_{k+1}\| \leq \alpha\|\delta_k\| + 2\epsilon,$$

which proves (3.4.12). Furthermore, the induction is complete, since

$$\|\delta_{k+1}\| \leq \alpha\|\delta_k\| + 2\epsilon \leq \alpha R_0 + 2\epsilon \leq R_0.$$

□

Lemma 3.6.2. *Let $v \in \mathcal{B}$. Then, for any δ,*

$$\langle \delta + 2B(v,\delta) + B(\delta,\delta) + \frac{1}{\eta}P\delta, \delta\rangle \geq \Big(1 - \frac{c^2 K\eta}{4}\Big)|\delta|^2.$$

Proof. Use of Property 3.3.1, items 3 and 5, together with Property 3.4.2, shows that

$$\begin{aligned}
\langle \delta + 2B(v,\delta) + B(\delta,\delta) + \frac{1}{\eta}P\delta, \delta\rangle &= |\delta|^2 + 2\langle B(v,\delta),\delta\rangle + \langle B(\delta,\delta),\delta\rangle + \langle\frac{1}{\eta}P\delta,\delta\rangle \\
&= |\delta|^2 - \langle B(\delta,\delta), v\rangle + \langle\frac{1}{\eta}P\delta,\delta\rangle \\
&\geq |\delta|^2 - cK^{\frac{1}{2}}|\delta||P\delta| + \frac{1}{\eta}|P\delta|^2 \\
&\geq |\delta|^2 - \frac{\theta|\delta|^2}{2} - \frac{c^2K|P\delta|^2}{2\theta} + \frac{1}{\eta}|P\delta|^2.
\end{aligned}$$

Now choosing $\theta = \frac{c^2K\eta}{2}$ establishes the claim.

□

Lemma 3.6.3. *In the setting of Theorem 3.4.6, for $t \in [0,h)$ and $R_0 := \|\delta_0\|$ we have*

$$|P\delta(t)|^2 \leq A_1(t)|\delta_0|^2 + |P\delta_0|^2 \tag{3.6.13}$$

and

$$|\delta(t)|^2 \leq B_1(t)|\delta_0|^2 + B_2(t)|P\delta_0|^2, \tag{3.6.14}$$

where the error δ is defined as in (3.6.11) *and A_1, B_1 and B_2 are given by (3.6.3, 3.6.4, 3.6.5).*

Proof. As in equation (3.6.8) we have

$$\frac{\mathrm{d}|P\delta|^2}{\mathrm{d}t} \leq 16K|\delta|^2 + 4R_0^2 e^{\beta t}|\delta|^2.$$

On integrating from 0 to t as before, and noting that now $P\delta_0 \neq 0$ in general, we

obtain

$$|P\delta(t)|^2 \leq \left(\frac{16K}{\beta}\{e^{\beta t} - 1\} + \frac{4R_0^2}{2\beta}\{e^{2\beta t} - 1\} \right) |\delta_0|^2 + |P\delta_0|^2,$$

which proves (3.6.13).

For the second inequality recall the bound (3.6.10)

$$\frac{1}{2}\frac{\mathrm{d}|\delta|^2}{\mathrm{d}t} + |\delta|^2 \leq \frac{|\delta|^2}{2} + \frac{c^2 K}{2}|P\delta|^2,$$

and combine it with (3.6.13) to get

$$\frac{\mathrm{d}|\delta|^2}{\mathrm{d}t} + |\delta|^2 \leq \left(\frac{16c^2K^2}{\beta}\{e^{\beta t} - 1\} + \frac{4c^2KR_0^2}{2\beta}\{e^{2\beta t} - 1\} \right) |\delta_0|^2 + c^2K|P\delta_0|^2.$$

Applying Gronwall's inequality yields (3.6.14). □

Chapter 4

Importance Sampling: Computational Complexity and Intrinsic Dimension

4.1 Introduction

4.1.1 Our Purpose

Our purpose in this chapter is to overview various ways of measuring the computational complexity of importance sampling, to link them to one another through transparent mathematical reasoning, and to create cohesion in the vast published literature on this subject. In addressing these issues we will study importance sampling in a general abstract setting, and then in the particular cases of Bayesian inversion and filtering. These two application settings are particularly important as there are many pressing scientific, technological and societal problems which can be formulated via inversion or filtering. An example of such an inverse problem is the determination of subsurface properties of the Earth from surface measurements; an example of a filtering problem is assimilation of atmospheric measurements into numerical weather forecasts.

The general abstract setting in which we work is as follows. We let μ and π be two probability measures on a measurable space $(\mathcal{X}, \mathcal{F})$ related via the expression

$$\frac{d\mu}{d\pi}(u) := g(u) \Big/ \int_{\mathcal{X}} g(u)\pi(du). \tag{4.1.1}$$

Here, g is the unnormalised *density (or Radon-Nikodym derivative)* of μ with respect to π. Note that the very existence of the density implies that the target is *absolutely*

continuous with respect to the proposal; absolute continuity will play an important role in our subsequent developments of this subject.

Importance sampling is a method for using independent samples from the *proposal* π to approximately compute expectations with respect to the *target* μ. The computational complexity is measured by the number of samples required to control the worst error made when approximating expectations within a class of test functions. Intuitively it is clear that the computational complexity of importance sampling is related to how far the target measure is from the proposal measure. With this in mind, a key quantity in what follows is the second moment, under the proposal, of $d\mu/d\pi$, which throughout the chapter is denoted by ρ. As we observe below, it is simply obtained as $\rho = \pi(g^2)/\pi(g)^2$.

The first application of this setting that we study is the linear inverse problem to determine $u \in \mathcal{X}$ from y where

$$y = Ku + \eta, \quad \eta \sim N(0, \Gamma). \tag{4.1.2}$$

We adopt a Bayesian approach in which we place a prior $u \sim \mathbb{P}_u = N(0, \Sigma)$, assume that η is independent of u, and seek the posterior $u|y \sim \mathbb{P}_{u|y}$. We study importance sampling with $\mathbb{P}_{u|y}$ being the target μ and the prior $\mathbb{P}_u$ being the proposal π.

The second application is the linear filtering problem of sequentially updating the distribution of $v_j \in \mathcal{X}$ given $\{y_i\}_{i=1}^j$ where

$$\begin{aligned} v_{j+1} &= Mv_j + \xi_j, \quad \xi_j \sim N(0, Q), \quad j \geq 0, \\ y_{j+1} &= Hv_{j+1} + \zeta_{j+1}, \quad \zeta_{j+1} \sim N(0, R), \quad j \geq 0. \end{aligned} \tag{4.1.3}$$

We assume that the problem has a Markov structure. We study the approximation of one step of the filtering update by means of particles, building on the study of importance sampling for the linear inverse problem. To this end it is expedient to work on the product space $\mathcal{X} \times \mathcal{X}$, and consider importance sampling for $(v_j, v_{j+1}) \in \mathcal{X} \times \mathcal{X}$. It then transpires that, for two different proposals, which are commonly termed the *standard proposal* and the *optimal proposal*, the complexity of one step of particle filtering may be understood by the study of a linear inverse problem on $\mathcal{X}$; we show this for both proposals, and then use the link to an inverse problem to derive results about the complexity of particle filters based on these two proposals.

For the abstract importance sampling problem we will relate ρ to a number of other natural quantities. These include the *effective sample size* ess, used heuristically in many application domains, and a variety of *distance metrics* between π and μ. Since the existence of a density between target and proposal is central in

this discussion, we will also discuss what happens as this absolute continuity property breaks down. We study this first in *high dimensional problems*, and second in *singular parameter limits* (by which we mean limits in which important parameters defining the problem tend to zero). The motivation for studying high dimensional problems can be appreciated by considering the two examples mentioned at the start of the introduction: inverse problems from the Earth's subsurface, and filtering for numerical weather prediction. In both cases the unknown which we are trying to determine from data is best thought of as a spatially varying field for subsurface properties such as permeability, or atmospheric properties, such as temperature. In practice the field will be discretized and represented as a high dimensional vector, for computational purposes, but for these types of application the state dimension can be of order 10^9. Furthermore as computer power advances there is pressure to resolve more physics, and hence for the state dimension to increase. Thus, it is important to understand infinite dimensional problems, and sequences of approximating finite dimensional problems which approach the infinite dimensional limit. A motivation for studying singular parameter limits arises, for example, from problems in which the noise is small and the relevant log-likelihoods scale inversely with the noise variance. Breakdown of absolute continuity will be related to limits in which the target and proposal become increasingly close to being *mutually singular*.

We will highlight a variety of notions of *intrinsic dimension* that have been introduced in the inverse problem literature; these may differ substantially from the dimensions of the spaces where the unknown u and the data y live. We then go on to show how these intrinsic dimensions relate to the parameter ρ, previously demonstrated to be central to computational complexity. We do so in various limits arising from large dimension of u and y, and/or small observational noise. We also link these concepts to breakdown of absolute continuity. Finally we apply our understanding of linear inverse problems to particle filters, translating the results from one to the other via the correspondence between the two problems, for both standard and optimal proposals, as described above.

It is often claimed that importance sampling suffers from the curse of dimensionality. Whilst there is some empirical truth in this fact, there is a great deal of confusion in the literature about what exactly makes importance sampling hard. In fact such a statement about the role of dimension is vacuous unless "dimension" is defined precisely. Throughout this chapter we use the following convention:

- State space dimension is the dimension of the measurable space where the measures μ and π are defined. We will be mostly interested in the case where the measurable space $\mathcal{X}$ is a separable Hilbert space, in which case the state

space dimension is the cardinality of an orthonormal basis of the space. In the context of inverse problems and filtering, the state space dimension is the dimension of the unknown.

- Data space dimension is the dimension of the space where the data lives.
- Nominal dimension is the minimum of the state space dimension and the data state dimension.
- Intrinsic dimension: we will use two notions of intrinsic dimension for inverse problems, denoted by efd and τ. These combine state/data dimension and small noise parameters. They can be interpreted as a measure of how informative the data is relative to the prior.

Our presentation shows how the intrinsic dimensions are natural when studying computational complexity of importance sampling. Furthermore we relate these quantities to the second moment of the Radon-Nikodym derivative between proposal and target, ρ, which will also be shown to arise naturally in the same context. In studying these quantities, and their inter-relations, we aim to achieve the purpose set out at the start of this subsection. Furthermore, a bibliography subsection, within each section, will link our overarching mathematical framework to the published literature in this area.

4.1.2 Organization of the Chapter and Notation

Section 4.2 describes importance sampling in abstract form. In Sections 4.3 and 4.4 the linear Gaussian inverse problem and the linear Gaussian filtering problem are studied. Our aim is to provide a digestible narrative and hence all proofs are left to an appendix in Section 4.6. Furthermore, as we study the inverse and filtering problems in both finite dimensional Euclidean space and infinite dimensional Hilbert space, there are some technical matters related to Gaussian measures in infinite dimensional spaces that we also detail in the appendix, Subsection 4.6.1, in order not to distract from the narrative flow.

Given a probability measure ν on a measurable space $(\mathcal{X}, \mathcal{F})$ expectations of a measurable function $\phi : \mathcal{X} \to \mathbb{R}$ with respect to ν will be written as both $\nu(\phi)$ and $\mathbb{E}_\nu[\phi]$. When it is clear which measure is being used we may drop the suffix ν and write simply $\mathbb{E}[\phi]$. Similarly, the variance will be written as $\mathrm{Var}_\nu(\phi)$ and again we may drop the suffix when no confusion arises from doing so. All test functions ϕ appearing in this chapter are assumed to be measurable.

We will be interested in sequences of measures indexed by time, by the state space dimension or by a tempering scheme. These are denoted with a subscript, e.g. ν_t, ν_d or ν_i. Anything to do with samples from a measure is denoted with a superscript: N for the number of samples, and n for the indices of the samples. The i-th coordinate of a vector u is denoted by $u(i)$. Thus, $u_t^n(i)$ denotes i-th coordinate of the n-th sample from the measure of interest at time t. Finally, the law of a random variable v will be denoted by $\mathbb{P}_v$.

4.1.3 Literature Review

Some early developments of importance sampling as a method to reduce the variance in Monte Carlo estimation date back to the early 1950's [Kahn and Marshall, 1953], [Kahn, 1955]. In particular the paper [Kahn and Marshall, 1953] demonstrates how to optimally choose the proposal density for given test function ϕ and target density. A modern view of importance sampling in the general framework (4.1.1) is given in [Chopin and Papaspiliopoulos, 2016]. A comprehensive description of Bayesian inverse problems in finite state/data space dimensions can be found in [Kaipio and Somersalo, 2005], and its formulation in infinite dimensional spaces in [Dashti and Stuart, 2016; Lasanen, 2007, 2012a,b; Stuart, 2010]. Text books overviewing the subject of filtering and particle filters include [Del Moral, 2004; Bain and Crisan, 2009], and the article [Crisan and Doucet, 2002] provides a readable introduction to the area. For an up-to-date and in-depth survey of nonlinear filtering see [Crisan and Rozovskii, 2011]. The linear Gaussian inverse problem and the linear Gaussian filtering problem have been extensively studied because they arise naturally in many applications, lead to considerable algorithmic tractability, and provide theoretical insight. For references concerning linear Gaussian inverse problems see [Franklin, 1970; Mandelbaum, 1984; Lehtinen et al., 1989; Kekkonen et al., 2015]. The linear Gaussian filter –the Kalman filter– was introduced in [Kalman, 1960]; see [Lancaster and Rodman, 1995] for further analysis. The inverse problem of determining sub-surface properties of the Earth from surface measurements is discussed in [Oliver et al., 2008], while the filtering problem of assimilating atmospheric measurements for numerical weather prediction is discussed in [Kalnay, 2003].

The key role of ρ, the second moment of the Radon-Nikodym derivative between the target and the proposal, has long been acknowledged [Liu, 1996], [Pitt and Shephard, 1999]. The value of ρ is indeed known to be asymptotically linked to the effective sample size [Kong, 1992], [Kong et al., 1994], [Liu, 1996]. Recent justification for the use of the effective sample size within particle methods is given in [Whiteley et al., 2016]. We will provide a further nonasymptotic justification of

the relevance of ρ through its appearance in error bounds on the error in importance sampling; in this context it is of relevance to highlight the paper [Crisan et al., 1998] which proved non-asymptotic bounds on the error in the importance-sampling based particle filter algorithm. We will also bound the importance sampling error in terms of different notions of distance between the target and the proposal measures, as in [Chen, 2005]; a useful overview of the subject of distances between probability measures is [Gibbs and Su, 2002].

We formulate problems in both finite dimensional and infinite dimensional state spaces. We refer to [Kallenberg, 2002] for a modern presentation of probability appropriate for understanding the material in this chapter. Some of our results are built on the rich area of Gaussian measures in Hilbert space; we include all the required background on this material in the appendix Subsection 4.6.1, and references are included there. However we emphasize that the presentation in the main body of the text is designed to keep technical material to a minimum and to be accessible to readers who are not versed in the theory of probability in infinite dimensional spaces. Absolute continuity of the target with respect to the proposal – or the existence of a density of the target with respect to the proposal – is central to our developments. This concept also plays a pivotal role in the understanding of Markov chain Monte Carlo (MCMC) methods in high and infinite dimensional spaces [Tierney, 1998]. A key idea in MCMC is that breakdown of absolute continuity on sequences of problems of increasing state space dimension is responsible for poor algorithmic performance with respect to increasing dimension; this should be avoided if possible, such as for problems with a well-defined infinite dimensional limit [Cotter et al., 2013]. Similar ideas will come in to play in this chapter.

As well as the breakdown of absolute continuity through increase in dimension, small noise limits can also lead to sequences of proposal/target measures which are increasingly close to mutually singular and for which absolute continuity breaks down. Small noise regimes are of theoretical and computational interest for both inverse problems and filtering. For instance, in inverse problems there is a growing interest in the study of the concentration rate of the posterior in the small observational noise limit, see [Knapik et al., 2011], [Agapiou et al., 2013], [Knapik et al., 2013], [Agapiou et al., 2014], [Ray, 2013], [Vollmer, 2013], [Kekkonen et al., 2015]. In filtering and multiscale diffusions, the analysis and development of improved proposals to deal with small noise limits or simulation of rare events is an active research area [Vanden-Eijnden and Weare, 2012], [Zhang et al., 2013], [Dupuis et al., 2012], [Spiliopoulos, 2013] [Tu et al., 2013].

In order to quantify the computational complexity of a problem, a recur-

rent concept is that of intrinsic dimension. Several notions of intrinsic dimension have been used in different fields, including dimension of learning problems [Bishop, 2006], [Zhang, 2002], [Zhang, 2005], of statistical inverse problems [Lu and Mathé, 2014], of functions in the context of quasi Monte Carlo (QMC) integration in finance applications [Caflisch et al., 1997], [Moskowitz and Caflisch, 1996], [Kuo and Sloan, 2005], and of data assimilation problems [Chorin and Morzfeld, 2013]. The underlying theme is that in many application areas where models are formulated in high dimensional state spaces, there is often a small subspace which captures most of the features of the system. It is the dimension of this subspace that effects the complexity of the problem. In the context of inverse problems the paper [Bengtsson et al., 2008] proposed a notion of intrinsic dimension, which was shown to have a direct connection with the performance of importance sampling. We introduce a further notion of intrinsic dimension for Bayesian inverse problems which agrees with the notion of effective number of parameters used in machine learning and statistics [Bishop, 2006]. We also establish that this notion of dimension and the one in [Bengtsson et al., 2008] are finite, or otherwise, at the same time. Both intrinsic dimensions account for three key features of the complexity of the inverse problem: the nominal dimension (i.e. the minimum of the dimension of the state space and the data), the size of the observational noise and the regularity of the prior relative to the observation noise. Varying the parameters related to these three features may cause a break-down of absolute continuity. The deterioration of importance sampling in large nominal dimensional limits has been widely investigated [Bengtsson et al., 2008], [Bickel et al., 2008], [Snyder et al., 2008], [Snyder et al., 2015], [Snyder, 2011], [Slivinski and Snyder, 2015]. In particular, the key role of the intrinsic dimension, rather than the nominal one, in explaining this deterioration was studied in [Bengtsson et al., 2008]. Here we study the different behaviour of importance sampling as absolute continuity is broken in the three regimes above, and we investigate whether, in all these regimes, the deterioration of importance sampling may be quantified by the various intrinsic dimensions that we introduce.

4.2 Importance Sampling

In Subsection 4.2.1 we define importance sampling and in Subsection 4.2.2 we demonstrate the role of the second moment of the target-proposal density, ρ; we prove two non-asymptotic theorems showing $\mathcal{O}\big((\rho/N)^{\frac{1}{2}}\big)$ convergence rate of importance sampling with respect to the number N of particles. Then in Subsection 4.2.3 we show how ρ relates to the effective sample size ess as often defined by practition-

ers, whilst in Subsection 4.2.4 we link ρ to various distances between probability measures. In Subsection 4.2.5 we highlight the role of the breakdown of absolute continuity in the growth of ρ, as the dimension of the space $\mathcal{X}$ grows. Subsection 4.2.6 follows with a similar discussion relating to singular limits of the density between target and proposal. Subsection 4.2.7 contains a literature review and, in particular, sources for all the material in this section.

4.2.1 General Setting

We consider target μ and proposal π, both probability measures on the measurable space $(\mathcal{X}, \mathcal{F})$, related by (4.1.1). In many statistical applications interest lies in estimating expectations under μ, for a collection of test functions, using samples from π. For a test function $\phi : \mathcal{X} \to \mathbb{R}$ such that $\mu(|\phi|) < \infty$, the identity

$$\mu(\phi) = \frac{\pi(\phi g)}{\pi(g)},$$

leads to the *autonormalized importance sampling* estimator:

$$\begin{aligned} \mu^N(\phi) &:= \frac{\frac{1}{N}\sum_{n=1}^N \phi(u^n) g(u^n)}{\frac{1}{N}\sum_{m=1}^N g(u^m)}, \quad u^n \sim \pi \quad \text{i.i.d.} \\ &= \sum_{n=1}^N w^n \phi(u^n), \qquad w^n := \frac{g(u^n)}{\sum_{m=1}^N g(u^m)}; \end{aligned} \tag{4.2.1}$$

here the w^n's are called the *normalized weights*. As suggested by the notation, it is useful to view (4.2.1) as integrating a function ϕ with respect to the random probability measure $\mu^N := \sum_{n=1}^N w^n \delta_{u^n}$. Under this perspective, importance sampling consists of approximating the target μ by the measure μ^N, which is typically called the *particle approximation of* μ. Note that, while μ^N depends on the proposal π, we suppress this dependence for economy of notation. Our aim is to understand the quality of the approximation μ^N of μ. In particular we would like to know how large to choose N in order to obtain small error. This will quantify the computational complexity of importance sampling.

4.2.2 The Second Moment of the Target-Proposal Density

A fundamental quantity in addressing this issue is ρ, defined by

$$\rho := \frac{\pi(g^2)}{\pi(g)^2}\,. \tag{4.2.2}$$

Thus ρ is the second moment of the Radon-Nikodym derivative of the target with respect to the proposal. The Cauchy-Schwarz inequality shows that $\pi(g)^2 \leq \pi(g^2)$ and hence that $\rho \geq 1$. Our first non-asymptotic result shows that, for bounded test functions ϕ, both the bias and the mean square error (MSE) of the autonormalized importance sampling estimator are $\mathcal{O}\big(N^{-1}\big)$ with constant of proportionality linear in ρ. The proof is in the appendix, Subsubsection 4.6.2.1.

Theorem 4.2.1. *Assume that μ is absolutely continuous with respect to π, with square-integrable density g, that is, $\pi(g^2) < \infty$. The bias and MSE of importance sampling over bounded test functions may be characterized as follows:*

$$\sup_{|\phi|\leq 1} \Big| \mathbb{E}\big[\mu^N(\phi) - \mu(\phi)\big] \Big| \leq \frac{12}{N}\rho\,,$$

and

$$\sup_{|\phi|\leq 1} \mathbb{E}\left[\big(\mu^N(\phi) - \mu(\phi)\big)^2\right] \leq \frac{4}{N}\rho.$$

Remark 4.2.2. *For a bounded test function $|\phi| \leq 1$, we trivially get $|\mu^N(\phi) - \mu(\phi)| \leq 2$; hence the bounds on bias and MSE provided in Theorem 4.2.1 are useful only when they are smaller than 2 and 4, respectively. The result is strongly suggestive that it is necessary to keep ρ/N small in order to obtain good importance sampling approximations. This heuristic dominates the developments in the remainder of the chapter, and in particular our wish to study the behaviour of ρ in various limits.*

It is interesting to contrast Theorem 4.2.1 to a well-known elementary asymptotic result. First, note that

$$\mu^N(\phi) - \mu(\phi) = \frac{N^{-1}\sum_{n=1}^N \frac{g(u^n)}{\pi(g)}\big[\phi(u^n) - \mu(\phi)\big]}{N^{-1}\sum_{n=1}^N \frac{g(u^n)}{\pi(g)}}\,.$$

Therefore, under the condition $\pi\big(g^2\big) < \infty$, and provided additionally that $\pi\big(g^2\phi^2\big) < \infty$, an application of the Slutsky lemmas gives that

$$\sqrt{N}\big(\mu^N(\phi) - \mu(\phi)\big) \Longrightarrow N\left(0, \frac{\pi\big(g^2\overline{\phi}^2\big)}{\pi(g)^2}\right), \quad \text{where } \overline{\phi} := \phi - \mu(\phi)\,. \tag{4.2.3}$$

For bounded $|\phi| \leq 1$, the only condition needed for appealing to the asymptotic result is $\pi\big(g^2\big) < \infty$. Then (4.2.3) gives that, for large N and since $|\overline{\phi}| \leq 2$,

$$\mathbb{E}\left[\big(\mu^N(\phi) - \mu(\phi)\big)^2\right] \lessapprox \frac{4}{N}\rho\,,$$

which is in precise agreement with our non-asymptotic bound.

In comparison with the asymptotic result (4.2.3), our non-asymptotic theorem makes an identical assumption on the importance weights, that is $\pi\big(g^2\big) < \infty$, but stronger assumptions on the test functions. We can obtain non-asymptotic bounds on the MSE and bias for much larger classes of test functions but at the expense of more assumptions on the importance weights. The next theorem addresses the issue of relaxing the class of test functions, whilst still deriving nonasymptotic bounds; the proof can be found in the appendix, Subsubsection 4.6.2.2. To simplify the statement we first introduce the following notation. We write $m_t[h]$ for the t-th *central moment* with respect to π of a function $h : \mathcal{X} \to \mathbb{R}$. That is,

$$m_t[h] := \pi(|h(u) - \pi(h)|^t).$$

We also define, as above, $\overline{\phi} := \phi - \mu(\phi)$.

Theorem 4.2.3. *Suppose that ϕ and g are such that C_{MSE} defined below is finite:*

$$C_{\mathrm{MSE}} := \frac{3}{\pi(g)^2} m_2[\phi g] + \frac{3}{\pi(g)^4} \pi(|\phi g|^{2d})^{\frac{1}{d}} C_{2e}^{\frac{1}{e}} m_{2e}[g]^{\frac{1}{e}} + \frac{3}{\pi(g)^{2(1+\frac{1}{p})}} \pi(|\phi|^{2p})^{\frac{1}{p}} C_{2q(1+\frac{1}{p})}^{\frac{1}{q}} m_{2q(1+\frac{1}{p})}[g]^{\frac{1}{q}}.$$

Then the bias and MSE of importance sampling when applied to approximate $\mu(\phi)$ may be characterized as follows:

$$\Big|\mathbb{E}\big[\mu^N(\phi) - \mu(\phi)\big]\Big| \leq \frac{1}{N}\left(\frac{2}{\pi(g)^2} m_2[g]^{\frac{1}{2}} m_2[\overline{\phi} g]^{\frac{1}{2}} + 2C_{\mathrm{MSE}}^{\frac{1}{2}} \frac{\pi(g^2)^{\frac{1}{2}}}{\pi(g)}\right)$$

and

$$\mathbb{E}\Big[\big(\mu^N(\phi) - \mu(\phi)\big)^2\Big] \leq \frac{1}{N} C_{\mathrm{MSE}}.$$

The constants $C_t > 0, t \geq 2$, satisfy $C_t^{\frac{1}{t}} \leq t - 1$ and the two pairs of parameters d, e, and p, q are conjugate indices.

Remark 4.2.4. *In Bayesian inverse problems $\pi(g) < \infty$ often implies that $\pi\big(g^s\big) < \infty$ for any positive s; we will demonstrate this in a particular case in Section 4.3. In such a case, Theorem 4.2.3 combined with Hölder's inequality shows that importance sampling converges at rate N^{-1} for any test function ϕ satisfying $\pi\big(|\phi|^{2+\epsilon}\big) < \infty$ for some $\epsilon > 0$.*

4.2.3 Effective Sample Size

Many practitioners define the *effective sample size* by the formula

$$\mathsf{ess} := \left(\sum_{n=1}^{N}(w^n)^2\right)^{-1} = \frac{\left(\sum_{n=1}^{N} g(u^n)\right)^2}{\sum_{n=1}^{N} g(u^n)^2} = N\frac{\pi^N_{\mathrm{MC}}(g)^2}{\pi^N_{\mathrm{MC}}\left(g^2\right)},$$

where π^N_{MC} is the empirical Monte Carlo random measure

$$\pi^N_{\mathrm{MC}} := \frac{1}{N}\sum_{n=1}^{N}\delta_{u^n}, \quad u^n \sim \pi.$$

By the Cauchy-Schwarz inequality it follows that $\mathsf{ess} \leq N$. Furthermore, since the weights lie in $[0, 1]$, we have

$$\sum_{n=1}^{N}(w^n)^2 \leq \sum_{n=1}^{N} w^n = 1$$

so that $\mathsf{ess} \geq 1$. These upper and lower bounds may be attained as follows. If all the weights are equal, and hence take value N^{-1}, then $\mathsf{ess} = N$, the optimal situation. On the other hand if exactly k weights take the same value, with the remainder then zero, $\mathsf{ess} = k$; in particular the lower bound of 1 is attained if precisely one weight takes the value 1 and all others are zero.

For large enough N, and provided $\pi\left(g^2\right) < \infty$, the strong law of large numbers gives

$$\mathsf{ess} \approx N/\rho\,.$$

Recalling that $\rho \geq 1$ we see that ρ^{-1} quantifies the proportion of particles that effectively characterize the sample size, in the large particle size asymptotic. Furthermore, by Theorem 4.2.1, we have that, for large N,

$$\sup_{|\phi|\leq 1} \mathbb{E}\left[\left(\mu^N(\phi) - \mu(\phi)\right)^2\right] \lessapprox \frac{4}{\mathsf{ess}}\,.$$

This provides a further justification for the use of ess as an effective sample size, in the large N asymptotic regime.

4.2.4 Probability Metrics

Intuition tells us that importance sampling will perform well when the distance between proposal π and target μ is not too large. Furthermore we have shown the

role of ρ in measuring the rate of convergence of importance sampling. It is hence of interest to explicitly link ρ to distance metrics between π and μ. In fact we consider asymmetric divergences as distance measures; these are not strictly metrics, but certainly represent useful distance measures in many contexts in probability. First consider the χ^2 divergence, which satisfies

$$D_{\chi^2}(\mu\|\pi) := \pi\left(\left[\frac{g}{\pi(g)} - 1\right]^2\right) = \rho - 1\,. \tag{4.2.4}$$

The Kullback-Leibler divergence is given by

$$D_{\mathrm{KL}}(\mu\|\pi) := \pi\left(\frac{g}{\pi(g)}\log\frac{g}{\pi(g)}\right)\,,$$

and may be shown to satisfy

$$\rho \geq e^{D_{\mathrm{KL}}(\mu\|\pi)}\,. \tag{4.2.5}$$

Thus Theorem 4.2.1 suggests that the number of particles required for accurate importance sampling scales exponentially with the Kullback-Leibler divergence between proposal and target and linearly with the χ^2 divergence.

4.2.5 High State Space Dimension and Absolute Continuity

The preceding three subsections have demonstrated how, when the target is absolutely continuous with respect to the proposal, importance sampling converges as the square root of ρ/N. It is thus natural to ask if, and how, this desirable convergence breaks down for sequences of target and proposal measures which become increasingly close to singular. To this end, suppose that the underlying space is the Cartesian product $\mathbb{R}^d$ equipped with the corresponding product σ-algebra, the proposal is a product measure and the un-normalized weight function also has a product form, as follows:

$$\pi_d(du) = \prod_{i=1}^{d}\pi_1(du(i)), \quad \mu_d(du) = \prod_{i=1}^{d}\mu_1(du(i)), \quad g_d(u) = \exp\left\{-\sum_{i=1}^{d}h(u(i))\right\},$$

for probability measures π_1, μ_1 on $\mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}^+$ (and we assume it is not constant to remove the trivial case $\mu_1 = \pi_1$). We index the proposal, target, density and ρ with respect to d since interest here lies in the limiting behaviour as d increases. In the setting of (4.1.1) we now have

$$\mu_d(du) \propto g_d(u)\pi_d(du).$$

By construction g_d has all polynomial moments under π_d and importance sampling for each d has the good properties developed in the previous sections. It is also fairly straightforward to see that μ_∞ and π_∞ are mutually singular when h is not constant: one way to see this is to note that

$$\frac{1}{d}\sum_{i=1}^{d} u(i)$$

has a different almost sure limit under μ_∞ and π_∞. Two measures cannot be absolutely continuous unless they share the same almost sure properties. Therefore μ_∞ is not absolutely continuous with respect to π_∞ and importance sampling is undefined in the limit $d = \infty$. As a consequence we should expect to see a degradation in its performance for large state space dimension d.

To illustrate this degradation, assume that $\pi_1(h^2) < \infty$. Under the product structure (4.2.7), we have $\rho_d = (\rho_1)^d$. Furthermore $\rho_1 > 1$ (since h is not constant). Thus ρ_d grows exponentially with the state space dimension suggesting, when combined with Theorem 4.2.1, that exponentially many particles are required, with respect to dimension, to make importance sampling accurate.

A useful perspective on the preceding, which links to our discussion of the small noise limit in the next subsection, is as follows. By the central limit theorem we have that, for large d,

$$g_d(u) \approx c' \exp\bigl(-\sqrt{d}cz\bigr), \quad z \sim N(0,1), \tag{4.2.6}$$

where $c, c' > 0$ are constants with respect to z; in addition c is independent of dimension d, whilst c' may depend on d. From this it follows that (noting that any constant scaling, such as c', disappears from the definition of ρ_d)

$$\rho_d = \frac{\pi_d(g_d^2)}{\pi_d(g_d)^2} \approx \frac{\mathbb{E}\exp(-2\sqrt{d}cz)}{\bigl(\mathbb{E}\exp(-\sqrt{d}cz)\bigr)^2}, \tag{4.2.7}$$

where, here, $\mathbb{E}$ denotes expectation with respect to $z \sim N(0,1)$. Using the fact that $\mathbb{E}e^{-az} = e^{a^2/2}$ we see that $\rho_d \approx e^{c^2 d}$.

It is important to realise that it is not the product structure *per se* that leads to the collapse, rather the lack of absolute continuity in the limit of infinite state space dimension. Thinking about the role of high dimensions in this way is very instructive in our understanding of high dimensional problems, but is very much related to the setting in which all the coordinates of the problem play a similar role. This does not happen in many application areas. Often there is a

diminishing response of the likelihood to perturbations in growing coordinate index. When this is the case, increasing the state space dimension has only a mild effect in the complexity of the problem, and it is possible to have well-behaved infinite dimensional limits; we will see this perspective in Subsections 4.3.1, 4.3.2 and 4.3.3 for inverse problems, and Subsections 4.4.1, 4.4.2 and 4.4.3 for filtering.

4.2.6 Singular Limits

In the previous subsection we saw an example where for high dimensional state spaces the target and proposal became increasingly close to being mutually singular, resulting in ρ which grows exponentially with the state space dimension. In this subsection we observe that mutual singularity can also occur because of small parameters in the unnormalized density g appearing in (4.1.1), even in problems of fixed dimension; this will lead to ρ which grows algebraically with respect to the small parameter. To understand this situation let $\mathcal{X} = \mathbb{R}$ and consider (4.1.1) in the setting where

$$g(u) = \exp\bigl(-\epsilon^{-1}h(u)\bigr)$$

where $h : \mathbb{R} \to \mathbb{R}^+$. Furthermore assume, for simplicity, that h is twice differentiable and has a unique minimum at $u^\star$, and that $h''(u^\star) > 0$. Assume, in addition, that π has a Lebesgue density with bounded first derivative. Then the Laplace method shows that

$$\mathbb{E}\exp\bigl(-2\epsilon^{-1}h(u)\bigr) \approx \exp\bigl(-2\epsilon^{-1}h(u^\star)\bigr)\sqrt{\frac{2\pi\epsilon}{2h''(u^\star)}}$$

and that

$$\mathbb{E}\exp\bigl(-\epsilon^{-1}h(u)\bigr) \approx \exp\bigl(-\epsilon^{-1}h(u^\star)\bigr)\sqrt{\frac{2\pi\epsilon}{h''(u^\star)}}.$$

It follows that

$$\rho \approx \sqrt{\frac{h''(u^\star)}{4\pi\epsilon}}.$$

Thus Theorem 4.2.1 indicates that the number of particles required for importance sampling to be accurate should grow at least as fast as $\epsilon^{-\frac{1}{2}}$.

4.2.7 Literature Review

In Subsection 4.2.1 we introduced the importance sampling approximation of a target μ using a proposal π, both related by (4.1.1). The resulting particle approximation measure μ^N is random because it is based on samples from π. Hence $\mu^N(\phi)$ is a *random* estimator of $\mu(\phi)$. This estimator is in general biased, and therefore a

reasonable metric for its quality is the MSE

$$\mathbb{E}\Big[\big(\mu^N(\phi) - \mu(\phi)\big)^2\Big],$$

where the expectation is with respect to the randomness in the measure μ^N. We bound the MSE over the class of bounded test functions in Theorem 4.2.1. In fact we may view this theorem as giving a bound on a distance between the measure μ and its approximation μ^N. To this end let ν and μ denote mappings from an underlying probability space (which for us will be that associated with π) into the space of probability measures on $(\mathcal{X}, \mathcal{F})$; in the following, expectation $\mathbb{E}$ is with respect to this underlying probability space. In [Rebeschini and van Handel, 2015] a distance $d(\cdot,\cdot)$ beween such random measures is defined by

$$d(\nu,\mu)^2 = \sup_{|\phi|\leq 1} \mathbb{E}\Big(\big(\nu(\phi) - \mu(\phi)\big)^2\Big), \tag{4.2.8}$$

where the supremum is taken over bounded measurable functions. The paper [Rebeschini and van Handel, 2015] used this distance to study the convergence of particle filters. Note that if the measures are not random the distance reduces to total variation. Using this distance, together with the discussion in Subsection 4.2.4 linking ρ to the χ^2 divergence, we see that Theorem 4.2.1 states that

$$d(\mu^N,\mu)^2 \leq \frac{4}{N}\big(1 + D_{\chi^2}(\mu\|\pi)\big).$$

In Subsection 4.2.4 we also link ρ to the Kullback-Leibler divergence; the bound (4.2.5) can be found in Theorem 4.19 of [Boucheron et al., 2013].

In Subsections 4.2.5 and 4.2.6 we studied how limits in which the target and proposal become closer and closer

As was already noted, this suggests the need to increase the number of particles linearly with $D_{\chi^2}(\mu\|\pi)$ or exponentially with $D_{\mathrm{KL}}(\mu\|\pi)$. Provided that $\log\big(\frac{g(u)}{\pi(g)}\big)$, $u \sim \mu$, is concentrated around its expected value, as often happens in large dimensional and singular limits, it has recently been shown [Chatterjee and Diaconis, 2015] that using a sample size of approximately $\exp\big(D_{\mathrm{KL}}(\mu\|\pi)\big)$ is both necessary and sufficient in order to control the L^1 error of the importance sampling estimator $\mu^N(\phi)$. Theorem 4.2.1 is similar to [Del Moral, 2004, Theorem 7.4.3]. However the later result uses a metric defined over subclasses of bounded functions. The resulting constants in their bounds rely on covering numbers, which are often intractable. In contrast, the constant ρ in Theorem 4.2.1 is more amenable to analysis and has several meaningful interpretations that will be explored in the

remainder of the chapter, including the one resulting in the preceding display. We also refer to [Del Moral, 2013, Section 11.2] for a more recent analysis. The central limit result in equation (4.2.3) shows that for large N the upper bound in Theorem 4.2.1 is sharp. Equation (4.2.3) can be seen as a trivial application of deeper central limit theorems for particle filters, see [Chopin, 2004]. The constants $C_t > 0$, $t \geq 2$ in Theorem 4.2.3 are determined by the Marcinkiewicz-Zygmund inequality [Ren and Liang, 2001]. The proof of Theorem 4.2.3, provided in Subsection 4.6.2.2 of the appendix, follows the approach of [Doukhan and Lang, 2009] for evaluating moments of ratios. Further importance sampling results have been proved within the study of convergence properties of various versions of the particle filter as a numerical method for the approximation of the true filtering/smoothing distribution. These results are often formulated in finite dimensional state spaces, under bounded likelihood assumptions and for bounded test functions, see [Crisan et al., 1998], [Del Moral and Miclo, 2000], [Crisan and Doucet, 2002], [Míguez et al., 2013], [Achutegui et al., 2014]. Generalizations for continuous time filtering can be found in [Bain and Crisan, 2009] and [Han, 2013].

The effective sample size ess, introduced in Subsection 4.2.3, is a standard statistic used to assess and monitor particle approximation errors in importance sampling [Kong, 1992], [Kong et al., 1994]. The effective sample size ess does not depend on any specific test function, but is rather a particular function of the normalized weights which quantifies their variability. So does ρ, and as we show in Subsection 4.2.3 there is an asymptotic connection between both. When interested in assessing the quality of the estimator $\mu^N(\phi)$ for a particular test function, a common diagnosis is the empirical variance of such estimator. In [Chatterjee and Diaconis, 2015], the authors study the limitations of such a diagnosis by showing that in the non-asymptotic regime it fails to capture the distance between the target and the proposal; they also propose a new diagnosis. Our discussion of ess relies on the condition $\pi\left(g^2\right) < \infty$. Intuitively, the particle approximation will be rather poor when this condition is not met. Extreme value theory provides some clues about the asymptotic particle approximation error. First it may be shown that, regardless of whether $\pi\left(g^2\right)$ is finite or not, but simply on the basis that $\pi(g) < \infty$, the largest normalised weight, $w^{(N)}$, will converge to 0 as $N \to \infty$; see for example Section 3 of [Downey and Wright, 2007] for a review of related results. On the other hand, [McLeish and O'Brien, 1982] shows that, for large N,

$$\mathbb{E}\left[\frac{N}{\mathsf{ess}}\right] \approx \int_0^N \gamma S(\gamma) d\gamma,$$

where $S(\gamma)$ is the survival function of the distribution of the un-normalized weights, $\gamma := g(u)$ for $u \sim \pi$. For instance, if the weights have density proportional to γ^{-a-1}, for $1 < a < 2$, then $\pi\big(g^2\big) = \infty$ and, for large enough N and constant C,

$$\mathbb{E}\left[\frac{N}{\mathsf{ess}}\right] \approx C\, N^{-a+2}\,.$$

Thus, in contrast to the situation where $\pi\big(g^2\big) < \infty$, in this setting the effective sample size does not grow linearly with N.

In Subsections 4.2.5 and 4.2.6 we studied how limits in which the target and proposal become closer and closer to being mutually singular (breakdown of absolute continuity) lead to problems for importance sampling. In Subsection 4.2.5 we studied high dimensional problems, using analysis of problems with product structure to enable analytical tractability of the calculations. This use of product structure was pioneered for MCMC methods in [Gelman et al., 1996]. The product structure was then used in a number of recent papers concerning the behaviour of importance sampling in high nominal dimensions, starting with the seminal paper [Bengtsson et al., 2008], and leading on to others such as [Beskos et al., 2014a], [Beskos et al., 2014b], [Bickel et al., 2008], [Snyder et al., 2008], [Snyder, 2011], [Slivinski and Snyder, 2015], and [Snyder et al., 2015].

In [Bengtsson et al., 2008, Section 3.2] it is shown that, using (4.2.6), the maximum normalised importance sampling weight can be approximately written as

$$w^{(N)} \approx \frac{1}{1 + \sum_{n>1} \exp\big\{-\sqrt{d}c(z^{(n)} - z^{(1)})\big\}},$$

where $\{z^n\}_{n=1}^N$ are samples from $N(0,1)$ and the $z^{(n)}$ are the ordered statistics. In [Bickel et al., 2008] a direct but non-trivial calculation shows that if N does not grow exponentially with d, the sum in the denominator converges to 0 in probability and as a result the maximum weight to 1. Of course this means that all other weights are converging to zero, and that the effective sample size is 1. It chimes with the heuristic derived in Subsection 4.2.5 where we show that ρ grows exponentially with d and that choosing N to grow exponentially is thus necessary to keep the upper bound in Theorem 4.2.1 small. The phenomenon is an instance of what is sometimes termed *collapse of importance sampling* in high dimensions. This type of behaviour can be obtained for other classes of targets and proposals; see [Bengtsson et al., 2008], [Snyder et al., 2008].

Within the product setting it may be possible, for some limited classes of problems, to avoid degeneracy of importance sampling-based algorithms for large d

at polynomial cost. The idea is to use *tempering*, that is, to introduce a sequence of intermediate distributions $\{\mu_{d,i}\}_{i=1}^{p}$, with $p \geq 1$ depending on d, that 'bridge' the target and proposal measures

$$\frac{d\mu_d}{d\pi_d}(u) = \prod_{i=0}^{p} \frac{d\mu_{d,i+1}}{d\mu_{d,i}}(u),$$

where we have set $\mu_{d,0} := \pi_d$ and $\mu_{d,p+1} := \mu_d$. The distributions $\{\mu_{d,i}\}_{i=1}^{p+1}$ are targeted sequentially using some form of particle filter. A natural way to define the intermediate distributions is by

$$\frac{d\mu_{i+1}}{d\mu_i}(u) = g_d(u)^{a_i}, \quad 0 \leq i \leq p, \tag{4.2.9}$$

where the temperatures $0 < a_i < 1$ satisfy $\sum_{i=0}^{p} a_i = 1$, and have the effect of 'flattening' the change of measure g_d. The main idea underlying [Beskos et al., 2014a] is that using $p = d$ bridging distributions in $\mathbb{R}^d$ and $a_i = 1/d$ leads to d importance sampling steps with $\rho = \mathcal{O}(1)$. On the other hand, not using tempering leads to one importance sampling step with $\rho = \mathcal{O}(e^d)$. Therefore, *as long as one can guarantee that by solving d problems sequentially the errors do not grow exponentially with d*, tempering is advantageous. In this scenario, in order to avoid degradation of importance sampling without tempering, the number of particles needs to grow exponentially with d; with tempering one can hope to avoid collapse with computational cost $\mathcal{O}(Nd^2)$ under the stated assumption about growth of errors. On a related note, [Frei and Künsch, 2013] proposed a method to combine the ensemble Kalman filter and particle filters. They introduced $p = 1$ bridging distributions, and used an ensemble Kalman filter approximation of $\mu_{d,1}$ to build a weighted particle approximation of μ.

Finally, in Subsection 4.2.6 we use the Laplace method. This is a classical methodology for approximating integrals against near singular integrands, and can be found in many textbooks; see for instance [Bender and Orszag, 1999]. The interested reader may compare the calculation in Subsection 4.2.5, using the Gaussian approximation, with that arising in Subsection 4.2.6, where the small noise limit is studied. At first glance they are similar in form, but the former calculation leads to exponential behaviour in dimension (since it results from different exponents) whilst the latter leads to algebraic behaviour in small noise (since it results from different normalizing constants).

4.3 Importance Sampling and Inverse Problems

The previous section showed that the distance between the proposal and the target is key in understanding the computational complexity of importance sampling and the central role played by ρ. In this section we study the computational complexity of importance sampling applied in the context of Bayesian inverse problems. In doing so we introduce a notion of intrinsic dimension.

The Bayesian approach to inverse problems consists of updating incomplete knowledge concerning a variable u, encoded in a prior probability distribution $\mathbb{P}_u$, based on some noisy observations of u, denoted by y. The updated knowledge is encoded in a posterior probability distribution $\mathbb{P}_{u|y}$. We study importance sampling with target $\mu := \mathbb{P}_{u|y}$ and proposal $\pi := \mathbb{P}_u$. To make the analysis tractable we consider linear Gaussian inverse problems.

In Subsection 4.3.1 we describe the setting of the problem, working in a general Hilbert space, but developing finite dimensional intuition in parallel to aid the reader who is not familiar with the theory of Gaussian measures in Hilbert space; furthermore, we include Subsection 4.6.1 in the appendix which gives background on this theory. Subsection 4.3.2 introduces various notions of "intrinsic dimension" associated with this problem; a key point to appreciate in the sequel is that this dimension can be finite even when the problem is posed in an infinite dimensional Hilbert space.

We highlight that a useful notion of intrinsic dimension for an inverse problem summarizes how much information is contained in the data – relative to the prior – rather than the dimensions of the unknown u (the state space dimension) or the data y (the data space dimension). We show, in Subsection 4.3.3, that when these latter dimensions are infinite then it is crucial that the posterior is absolutely continuous with respect to the prior in order for the intrinsic dimension to be finite; we also link absolute continuity and finite intrinsic dimension with boundedness of the second moment, ρ, of the Radon-Nikodym derivative of posterior with respect to prior. We then investigate, in Subsection 4.3.4, the behaviour of the intrinsic dimension of the inverse problem as the measures μ and π approach mutual singularity; we study both high nominal dimensional limits and small noise limits. We conclude the section with a literature review in Subsection 4.3.5, containing sources for all the material in this section.

4.3.1 General Setting

We study the inverse problem of finding u from y where

$$y = Ku + \eta. \tag{4.3.1}$$

In particular we work in the setting where u is an element of the (potentially infinite dimensional) separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle, \|\cdot\|)$. Two cases will help guide the reader:

Example 4.3.1 (Linear Regression Model)**.** *In the context of the linear regression model, $u \in \mathbb{R}^{d_u}$ is the regression parameter vector, $y \in \mathbb{R}^{d_y}$ is a vector of training outputs and $K \in \mathbb{R}^{d_y \times d_u}$ is the so-called design matrix whose column space is used to construct a linear predictor for the scalar output. In this setting, $d_u, d_y < \infty$, although in modern applications they might be both very large, and the case $d_u \gg d_y$ is the so-called "large p (here d_u) small N (here d_y)" problem.*

Example 4.3.2 (Deconvolution Problem)**.** *In the context of signal deconvolution, $u \in L^2(0,1)$ is a square integrable unknown signal on the unit interval, $K : L^2(0,1) \to L^2(0,1)$ is a convolution operator $Ku(x) = (\phi \star u)(x) = \int_0^1 \phi(x-z)u(z)dz$, and $y = Ku + \eta$ is the noisy observation of the convoluted signal where η is observational noise. The convolution kernel ϕ might be, for example, a Gaussian kernel $\phi(x) = e^{-\delta x^2}$. Note also that discretization of the deconvolution problem will lead to a family of instances of the preceding linear regression model, parametrised by the dimension of the discretization space.*

The infinite dimensional setting does require some technical background, and this is outlined in the first subsection of the appendix. Nevertheless, the reader versed only in finite dimensional Gaussian concepts will readily make sense of the notions of intrinsic dimension described in Subsection 4.3.2 simply by thinking of (potentially infinite dimensional) matrix representations of covariances. In particular, the adjoint, denoted $\cdot^*$, can be thought of as generalization of the concept of transpose, and self-adjoint operators as symmetric matrices. However, to fully appreciate the links made in Subsection 4.3.3, the infinite dimensional setting and the background material from the appendix Subsection 4.6.1 will be helpful.

In equation (4.3.1) the data y is comprised of the image of the unknown u under a linear map K, with added observational noise η. Here K can be formally thought of as being a bounded linear operator in $\mathcal{H}$, which is ill-posed in the sense that if we attempt to invert the data using the (generalized) inverse of K, we get amplification of small errors η in the observation to large errors in the reconstruction

of u. In such situations, we need to use regularization techniques in order to stably reconstruct of the unknown u, from the noisy data y.

We assume Gaussian observation noise $\eta \sim \mathbb{P}_\eta := N(0, \Gamma)$ and adopt a Bayesian approach by putting a prior on the unknown $u \sim \mathbb{P}_u = N(0, \Sigma)$, where $\Gamma : \mathcal{H} \to \mathcal{H}$ and $\Sigma : \mathcal{H} \to \mathcal{H}$ are bounded, self-adjoint, positive-definite linear operators. As discussed in Subsection 4.6.1, if covariance Γ (respectively Σ) is trace class then $\eta \sim \mathbb{P}_\eta$ (respectively $u \sim \mathbb{P}_u$) is almost surely in $\mathcal{H}$. On the other hand, as also discussed in Subsection 4.6.1, when covariance Γ (respectively Σ) is not trace-class we have that $\eta \notin \mathcal{H}$ but $\eta \in \mathcal{Y}$ $\mathbb{P}_\eta$-almost surely (respectively $u \notin \mathcal{H}$ but $u \in \mathcal{X}$ $\mathbb{P}_u$-almost surely) where $\mathcal{Y}$ (respectively $\mathcal{X}$) strictly contains $\mathcal{H}$; indeed $\mathcal{H}$ is compactly embedded into $\mathcal{X}, \mathcal{Y}$.

In this setting the prior $\mathbb{P}_u$ and posterior $\mathbb{P}_{u|y}$ are Gaussian conjugate and $\mathbb{P}_{u|y} = N(m, C)$, with mean and covariance given, under appropriate conditions detailed in the literature review Subsection 4.3.5, by

$$m = \Sigma K^*(K\Sigma K^* + \Gamma)^{-1} y, \tag{4.3.2}$$

$$C = \Sigma - \Sigma K^*(K\Sigma K^* + \Gamma)^{-1} K\Sigma. \tag{4.3.3}$$

The reader wishing to derive these formulae using finite dimensional intuition may note that, using Bayes' rule and completion of the square, the posterior mean and covariance can be expressed via precision matrices as

$$C^{-1} = \Sigma^{-1} + K^*\Gamma^{-1}K, \tag{4.3.4}$$

$$C^{-1}m = K^*\Gamma^{-1}y. \tag{4.3.5}$$

Use of the Schur complement yields (4.3.2).

We tacitly assume that K can be extended to act on elements in $\mathcal{X}$ and that the sum of Ku and η makes sense in $\mathcal{Y}$. In the setting outlined above we assume that the prior acts as a regularization for the inversion of the data y. This is encoded in the following assumption on the relationship between the operators K, Σ and Γ.

Assumption 4.3.3. *Define $S = \Gamma^{-\frac{1}{2}} K \Sigma^{\frac{1}{2}}$, $A = S^*S$ and assume that A, viewed as a linear operator in $\mathcal{H}$, is bounded. Furthermore, assume that the spectrum of A consists of a countable number of eigenvalues, sorted without loss of generality in a non-increasing way*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_j \geq \cdots \geq 0.$$

In Section 4.3.5 we give an intuitive explanation for the centrality of A and S, and discuss the role of the assumption in the context of inverse problems.

4.3.2 Intrinsic Dimension

The operator A defined in Assumption 4.3.3 plays an important role in what follows because it measures the size of the difference between the prior and posterior covariances Σ and C. The developments in Section 4.2 indicate that a key measure determining the computational complexity of importance sampling is the distance between the target (here the posterior) and the proposal (here the prior). In the Gaussian setting considered in this section the differences between posterior and prior covariances will contribute to this distance and we now develop this idea. Note, however, that we say nothing here about the differences between prior and posterior means.

We illustrate the ideas in finite state/data space dimensions in the first instance, a setting in which we have the following result, proved in the appendix, Subsubsection 4.6.3.1. For extensions to Hilbert spaces, see the discussion in the literature review Subection 4.3.5.

Proposition 4.3.4. *In the finite dimensional setting, and under the assumption that both Σ and C are invertible,*

$$\mathrm{Tr}\big((C^{-1} - \Sigma^{-1})\Sigma\big) = \mathrm{Tr}(A), \quad \mathrm{Tr}\big((\Sigma - C)\Sigma^{-1}\big) = \mathrm{Tr}\big((I + A)^{-1}A\big).$$

Thus the traces of A and of $(I + A)^{-1}A$ measure the relative differences between the posterior and prior precision and covariance operators, respectively, relative to their prior values. For this reason they provide useful measures of the computational compexity of importance sampling, motivating the following definitions:

$$\tau := \mathrm{Tr}(A), \qquad \mathsf{efd} := \mathrm{Tr}\big((I + A)^{-1}A\big)\,. \tag{4.3.6}$$

Note that the trace calculates the sum of the eigenvalues and is well-defined, although may be infinite, in the Hilbert space setting. We refer to efd as effective dimension; both τ and efd are measures of the intrinsic dimension of the inverse problem at hand. Remaining for the moment in the finite dimensional setting, we have the next result. The proof is given in the appendix, Subsubsection 4.6.3.1:

Proposition 4.3.5. *Let S and A be defined as in Assumption 4.3.3, and consider the finite dimensional setting.*

1. *The matrices $\Gamma^{1/2}S(I + A)^{-1}S^*\Gamma^{-1/2} \in \mathbb{R}^{d_y \times d_y}$, $S(I + A)^{-1}S^* \in \mathbb{R}^{d_y \times d_y}$ and $(I + A)^{-1}A \in \mathbb{R}^{d_u \times d_u}$ have the same non-zero eigenvalues and hence the same trace.*

2. *If $\lambda_i > 0$ is a non-zero eigenvalue of A then these three matrices have corresponding eigenvalue $\lambda_i(1+\lambda_i)^{-1} < 1$, and*

$$\mathsf{efd} = \sum_i \frac{\lambda_i}{1+\lambda_i} \leq d = \min\{d_u, d_y\}\,.$$

Here, recall, $d = \min\{d_u, d_y\}$ is referred to as the nominal dimension of the problem. Part 2. of the preceding result demonstrates the connection between efd and the physical dimensions of the unknown and observation spaces, whilst part 1. demonstrates the equivalence between the traces of a variety of operators, all of which are used in the literature; this is discussed in greater detail in the literature review of Subsection 4.3.5. In the Hilbert space setting, recall, the intrinsic dimensions efd and τ can be infinite. It is important to note, however, that this cannot happen if the rank of K is finite. That is, the intrinsic dimension efd is finite whenever the unknown u or the data y live in a finite dimensional subspace of $\mathcal{H}$. The following result, proved in Subsubsection 4.6.3.1 of the appendix, relates efd and τ. It shows in particular that they are finite, or otherwise, at the same time. It holds in the infinite dimensional setting.

Lemma 4.3.6. *Let Assumption 4.3.3 hold. Then A is trace class if and only if $(I+A)^{-1}A$ is trace class, and the following inequalities hold*

$$\frac{1}{\|I+A\|}\mathrm{Tr}(A) \leq \mathrm{Tr}\left((I+A)^{-1}A\right) \leq \mathrm{Tr}(A).$$

As a consequence

$$\frac{1}{\|I+A\|}\tau \leq \mathsf{efd} \leq \tau. \tag{4.3.7}$$

We are now ready to study the performance of importance sampling with posterior as target and prior as proposal. In Subsection 4.3.3 we identify conditions under which we can guarantee that ρ in Theorem 4.2.1 is finite and absolute continuity holds. In Subsection 4.3.4 we then study the growth of ρ as mutual singularity is approached in different regimes. The intrinsic dimensions τ and efd will be woven into these developments.

4.3.3 Absolute Continuity

In the finite dimensional setting, when both covariance matrices Σ and Γ are strictly positive-definite, the Gaussian proposal and target distributions have densities with respect to the Lebesgue measure. They are hence mutually absolutely continuous

and it is hence straightforward to find the Radon-Nikodym derivative of the target with respect to the proposal by taking the ratio of the respective Lebesgue densities once the posterior is identified via Bayes' theorem; this gives:

$$\frac{d\mu}{d\pi}(u) = \frac{d\mathbb{P}_{u|y}}{d\mathbb{P}_u}(u;y) \propto \exp\left(-\frac{1}{2}u^*K^*\Gamma^{-1}Ku + u^*K^*\Gamma^{-1}y\right) =: g(u;y). \quad (4.3.8)$$

Direct calculation shows that, for $d_u, d_y < \infty$ and Γ invertible, the ratio ρ defined in (4.2.2) is finite, and indeed that g admits all polynomial moments, all of which are positive. In this subsection we study ρ in the Hilbert space setting. In general there is no guarantee that the posterior is absolutely continuous with respect to the prior; when it is not, g, and hence ρ, are not defined. We thus seek conditions under which such absolute continuity may be established.

To this end, we define the likelihood measure $y|u \sim \mathbb{P}_{y|u} := N(Ku, \Gamma)$, and the joint distribution of (u, y) under the model $\nu(du, dy) := \mathbb{P}_{y|u}(dy|u)\mathbb{P}_u(du)$, recalling that $\mathbb{P}_u = N(0, \Sigma)$. We also define the marginal distribution of the data under the joint distribution, $\nu_y(dy) = \mathbb{P}_y(dy)$. We have the following result, proved in Subsubsection 4.6.3.2 of the appendix:

Theorem 4.3.7. *Let Assumption 4.3.3 hold and let $\mu = \mathbb{P}_{u|y}$ and $\pi = \mathbb{P}_u$. The following are equivalent:*

i) $\mathsf{efd} < \infty$;

ii) $\tau < \infty$;

iii) $\Gamma^{-1/2}Ku \in \mathcal{H}$, *$\pi$-almost surely;*

iv) for ν_y-almost all y, the posterior μ is well defined as a measure in $\mathcal{X}$ and is absolutely continuous with respect to the prior with

$$\frac{d\mu}{d\pi}(u) \propto \exp\left(-\frac{1}{2}\left\|\Gamma^{-1/2}Ku\right\|^2 + \frac{1}{2}\langle\Gamma^{-1/2}y, \Gamma^{-1/2}Ku\rangle\right) =: g(u;y), \quad (4.3.9)$$

where $0 < \pi\big(g(\cdot;y)\big) < \infty$.

Remark 4.3.8. *Due to the exponential structure of g, we have that assertion (iv) of the last theorem is immediately equivalent to g being ν-almost surely positive and finite and for ν_y-almost all y the second moment of the target-proposal density is finite:*

$$\rho = \frac{\pi\left(g(\cdot;y)^2\right)}{\pi\big(g(\cdot;y)\big)^2} < \infty.$$

Note that item (iii) can also be interpreted as quantifying the dimension of the problem, since it is a requirement on the regularity of the forward image of the unknown, relative to the noise; such regularity condition typically relates to smoothness of the underlying field, and thus to intrinsic dimension, as we show here.

We have established something very interesting: there are meaningful notions of intrinsic dimension for inverse problems formulated in infinite state/data state dimensions and, when the intrinsic dimension is finite, importance sampling may be possible as there is absolute continuity; moreover, in such situation ρ is finite. Thus, under any of the equivalent conditions i)-iv), Theorem 4.2.1 can be used to provide bounds on the effective sample size ess, defined in Subsection 4.2.3; indeed the effective sample size is then proportional to N.

It is now of interest to understand how ρ, and the intrinsic dimensions τ and efd, depend on various parameters arising in the problem, such as small noise or the dimension of finite dimensional approximations of the inverse problem. Such questions are studied in the next subsection.

4.3.4 Singular Limits

The parameter ρ is a complicated nonlinear function of the eigenvalues of A and the data y. However, there are some situations in which we can lower bound ρ in terms of the intrinsic dimensions τ, efd and the size of the eigenvalues of A. We present two classes of examples of this type. The first is a simple but insightful example in which the eigenvalues cluster into a finite dimensional set of large eigenvalues and a set of small remaining eigenvalues. The second involves asymptotic considerations in a simultaneously diagonalizable setting.

4.3.4.1 Spectral Jump

Consider the setting where u and y both live in finite dimensional spaces of dimensions d_u and d_y respectively. Suppose that A has eigenvalues $\{\lambda_i\}_{i=1}^{d_u}$ with $\lambda_i = C \gg 1$ for $1 \leq i \leq k$, and $\lambda_i \ll 1$ for $k+1 \leq i \leq d_u$; indeed we assume that

$$\sum_{i=k+1}^{d_u} \lambda_i \ll 1.$$

Then $\tau(A) \approx Ck$, whilst the effective dimension satisfies $\mathsf{efd} \approx k$. Using the identity

$$2D_{\mathrm{KL}}(\mathbb{P}_{u|y} \| \mathbb{P}_u) = \log\Big(\det(I+A)\Big) - \mathrm{Tr}\Big((I+A)^{-1}A\Big) + m^* \Sigma^{-1} m\,.$$

and studying the asymptotics for fixed m, with k and C large, we obtain

$$D_{\mathrm{KL}}(\mathbb{P}_{u|y}||\mathbb{P}_u) \approx \frac{\mathsf{efd}}{2}\log(C)\,.$$

Therefore, using (4.2.5),

$$\rho \gtrapprox C^{\frac{\mathsf{efd}}{2}}\,.$$

This suggests that ρ grows exponentially with the *number* of large eigenvalues, whereas it has an algebraic dependence on the *size* of the eigenvalues. Theorem 4.2.1 then suggests that the number of particles required for accurate importance sampling will grow exponentially with the number of large eigenvalues, and algebraically with the size of the eigenvalues. A similar distinction may be found by comparing the behaviour of ρ in large state space dimension in Subsection 4.2.5 (exponential) and with respect to small scaling parameter in Subsection 4.2.6 (algebraic).

4.3.4.2 Spectral Cascade

We now introduce a three-parameter family of inverse problems, defined through the eigenvalues of A. These three parameters represent the regularity of the prior and the forward map, the size of the observational noise, and the number of positive eigenvalues of A, which corresponds to the nominal dimension. We are interested in investigating the performance of importance sampling, as quantified by ρ, in different regimes for these parameters. We work in the framework of Assumption 4.3.3, and under the following additional assumption:

Assumption 4.3.9. *Within the framework of Assumption 4.3.3, we assume that* $\Gamma = \gamma I$ *and that* A *has eigenvalues* $\left\{\frac{j^{-\beta}}{\gamma}\right\}_{j=1}^{\infty}$ *with* $\gamma > 0$, *and* $\beta \geq 0$. *We consider a truncated sequence of problems with* $A(\beta, \gamma, d)$, *with eigenvalues* $\left\{\frac{j^{-\beta}}{\gamma}\right\}_{j=1}^{d}$, $d \in \mathbb{N} \cup \{\infty\}$. *Finally, we assume that the data is generated from a fixed underlying infinite dimensional truth* $u^\dagger$,

$$y = Ku^\dagger + \eta, \quad Ku^\dagger \in \mathcal{H},$$

and for the truncated problems the data is given by projecting y *onto the first* d *eigenfunctions of* A.

Note that d in the previous assumption is the data space dimension, which agrees here with the nominal dimension. The setting of the previous assumption arises, for example, when d is finite, from discretizing the data of an inverse problem formulated in an infinite dimensional state space. Provided that the forward map K

and the prior covariance Σ commute, our analysis extends to the case where both the unknown and the data are discretized in the common eigenbasis. In all these cases, interest lies in understanding how the complexity of importance sampling depends on the level of the discretizations. The parameter γ may arise as an observational noise scaling, and it is hence of interest to study the complexity of importance sampling when γ is small. And finally the parameter β reflects regularity of the problem, as determined by the prior and noise covariances, and the forward map; critical phase transitions occur in computational complexity as this parameter is varied, as we will show.

The intrinsic dimensions $\tau = \tau(\beta, \gamma, d)$ and $\mathsf{efd} = \mathsf{efd}(\beta, \gamma, d)$ read

$$\tau = \frac{1}{\gamma} \sum_{j=1}^{d} j^{-\beta}, \qquad \mathsf{efd} = \sum_{j=1}^{d} \frac{j^{-\beta}}{\gamma + j^{-\beta}}. \tag{4.3.10}$$

Table 4.1 shows the scalings of the effective dimensions efd and τ with the model parameters. It also shows how ρ behaves under these scalings and hence gives, by Theorem 4.2.1, an indication of the number of particles required for accurate importance sampling in a given regime. In all the scaling limits where ρ grows to infinity the posterior and prior are approaching mutual singularity; we can then apply Theorem 4.2.1 to get an indication of how importance sampling deteriorates in these limits.

Note that by Theorem 4.3.7 we have $\tau(\beta, \gamma, d) < \infty$ if and only if $\mathsf{efd}(\beta, \gamma, d) < \infty$. It is clear from (4.3.10) that $\tau = \infty$ if and only if $\{d = \infty, \beta \leq 1\}$. By Theorem 4.3.7 again, this implies, in particular, that absolute continuity is lost in the limit as $d \to \infty$ when $\beta \leq 1$, and as $\beta \searrow 1$ when $d = \infty$. Absolute continuity is also lost in the limit $\gamma \to 0$, in which the posterior is fully concentrated around the data (at least in those directions in which the data live). In this limit we always have $\tau = \infty$, whereas $\mathsf{efd} < \infty$ in the case where $d < \infty$ and $\mathsf{efd} = \infty$ when $d = \infty$. Note that in the limit $\gamma = 0$ Assumption 4.3.3 does not hold, which explains why τ and efd are not finite simultaneously. Indeed, as was noted before, efd is always bounded by the nominal dimension d irrespective of the size γ of the noise.

Some important remarks on Table 4.1 are:

- ρ grows *algebraically* in the small noise limit ($\gamma \to 0$) if the nominal dimension d is finite.

- ρ grows *exponentially* in τ or efd as the nominal dimension grows ($d \to \infty$), or as the prior becomes rougher ($\beta \searrow 1$).

Regime	Parameters	efd	τ	ρ
Small noise	$\gamma \to 0$, $d < \infty$	d	γ^{-1}	$\gamma^{-d/2}$
	$\gamma \to 0$, $d = \infty$, $\beta > 1$	$\gamma^{-1/\beta}$	γ^{-1}	$\gamma^{-\frac{\epsilon\beta}{2}(\gamma^{-1/\beta-\epsilon})}$
Large d	$d \to \infty$, $\beta < 1$	$d^{1-\beta}$	$d^{1-\beta}$	$\exp(d^{1-\beta})$
Small noise and large d	$\gamma = d^{-\alpha}$, $d \to \infty, \beta > 1, \alpha > \beta$	d	d^{α}	$d^{(\alpha-\beta)d}$
	$\gamma = d^{-\alpha}$, $d \to \infty, \beta > 1, \alpha < \beta$	$d^{\alpha/\beta}$	d^{α}	$d^{\epsilon d^{\alpha/\beta-\epsilon}}$
	$\gamma = d^{-\alpha}$, $d \to \infty, \beta < 1, \alpha > \beta$	d	$d^{1+\alpha-\beta}$	$d^{(\alpha-\beta)d}$
	$\gamma = d^{-\alpha}$, $d \to \infty, \beta < 1, \alpha < \beta$	$d^{1+\alpha-\beta}$	$d^{1+\alpha-\beta}$	$d^{\epsilon d^{\alpha/\beta-\epsilon}}$
Regularity	$d = \infty, \beta \searrow 1$	$\frac{1}{\beta-1}$	$\frac{1}{\beta-1}$	$\exp(\frac{1}{\beta-1})$

Table 4.1: The third and fourth columns show the scaling of the intrinsic dimensions with model parameters. The fourth one gives a lower bound on the growth of ρ, suggesting that the number of particles should be increased *at least* as indicated by this column in terms of the model parameters. This lower bound holds for all realizations of the data y when $\gamma \to 0$, and in probability for those regimes where γ is fixed. ϵ can be chosen arbitrarily small.

- ρ grows *factorially* in the small noise limit ($\gamma \to 0$) if $d = \infty$, and in the joint limit $\gamma = d^{-\alpha}, d \to \infty$. The exponent in the rates relates naturally to efd.

The scalings of τ and efd can be readily deduced by comparing the sums defining τ and efd with integrals. The analysis of the sensitivity of ρ to the model parameters relies on an explicit expression for this quantity. The details are in the appendix, Subsubsection 4.6.3.3.

4.3.5 Literature Review

Some more examples of linear inverse problems in both finite and infinite dimensions include the Radon Inversion used for X-ray imaging, the determination of the initial temperature from later measurements and the inversion of the Laplace transform. Many case studies as well as more elaborate nonlinear inverse problems can be found for example in [Kaipio and Somersalo, 2005], [Stuart, 2010] which adopt a Bayesian approach to their solution, and [Engl et al., 1996], [Mueller and Siltanen, 2012] which adopt a classical approach. The Bayesian approach we undertake, in the example of linear regression (Example 4.3.1) becomes the Gaussian conjugate Bayesian analysis of linear regression models, as in [Lindley and Smith, 1972].

Formulae (4.3.4), (4.3.5) for the mean and covariance expressed via precisions in the finite dimensional setting may be found in [Lindley and Smith, 1972]. In fact sense can be given to these formulae in the infinite dimensional setting as well; see [Agapiou et al., 2013, Section 5]. Formulae (4.3.2), (4.3.3) in the infinite dimensional setting are derived in [Mandelbaum, 1984], [Lehtinen et al., 1989]; in the specific

case of inverting for the initial condition in the heat equation they were derived in [Franklin, 1970]. The appendix, Subsection 4.6.1, has a discussion of Gaussian measures in Hilbert spaces and contains further background references.

As mentioned above, we tacitly assume that K can be extended to act on elements in $\mathcal{X}$ and that the sum of Ku and η makes sense in $\mathcal{Y}$. This assumption holds trivially if the three operators K, Σ, Γ are simultaneously diagonalizable. It also holds in non-diagonal settings, in which it is possible to link the domains of powers of the three operators by appropriate embeddings; for some examples see [Agapiou et al., 2013, Section 7].

The assumption that the spectrum of A introduced in Assumption 4.3.3 consists of a countable number of eigenvalues, means that the operator A can be thought of as an infinitely large diagonal matrix. It holds if A is compact [Lax, 2002, Theorem 3, Chapter 28], but is in fact more general since it covers, for example, the non-compact case $A = I$.

In the finite dimensional setting the assumption that A is bounded holds automatically if the noise covariance is invertible. The centrality of $S = \Gamma^{-\frac{1}{2}} K \Sigma^{\frac{1}{2}}$ may then be understood as follows. Under the prior and noise models we may write $u = \Sigma^{\frac{1}{2}} u_0$ and $\eta = \Gamma^{\frac{1}{2}} \eta_0$ where u_0 and η_0 are independent centred Gaussians with identity covariance operators (white noises). Under the assumption that Γ is invertible we then find that we may write (4.3.1), for $y_0 = \Gamma^{-\frac{1}{2}} y$, as

$$y_0 = S u_0 + \eta_0. \tag{4.3.11}$$

Thus all results may be derived for this inverse problem, and translated back to the original setting. The role of S, and hence A, is thus clear in the finite dimensional setting. This intuition carries over to infinite dimensions.

We note here that the inverse problem

$$y_0 = w_0 + \eta_0 \tag{4.3.12}$$

with η_0 a white noise and $w_0 \sim N(0, SS^*)$ is equivalent to (4.3.11), but formulated in terms of unknown $w_0 = Au_0$, rather than unknown u_0. In this picture the key operator is SS^* rather than $A = S^*S$. Note that by Lemma 4.6.5 $\mathrm{Tr}(S^*S) = \mathrm{Tr}(SS^*)$. Furthermore, if S is compact the operators SS^* and S^*S have the same nonzero eigenvalues [Engl et al., 1996, Section 2.2], thus $\mathrm{Tr}((I + SS^*)^{-1} SS^*) = \mathrm{Tr}((I + S^*S)^{-1} S^*S)$. The last equality holds even if S is non-compact, since then Lemma 4.6.5 together with Lemma 4.3.6 imply that both sides are infinite. Combining, we see that the intrinsic dimension (τ or efd) is the same regardless of whether we

view w_0 or u_0 as the unknown. In particular, the assumption that A is bounded is equivalent to assuming that the operators S, S^* or SS^* are bounded [Lax, 2002, Theorem 14, Chapter 19]. For the equivalent formulation (4.3.12), the posterior mean equation (4.3.2) is

$$m = SS^*(SS^* + I)^{-1}y.$$

If SS^* is compact, that is, if its nonzero eigenvalues λ_i go to 0, then m is a regularized approximation of w_0, since the components of the data corresponding to small eigenvalues λ_i are shrunk towards zero. On the other hand, if SS^* is unbounded, that is, if its nonzero eigenvalues λ_i go to infinity, then there is no regularization and high frequency components in the data remain almost unaffected by SS^* in m. Therefore, the case SS^* is bounded is the borderline case for having that the prior has a regularizing effect in the inversion of the data.

In Subsection 4.3.2 we study notions of dimension for Bayesian inverse problems. In the Bayesian setting, the prior infuses information and correlations on the components of the unknown u, reducing the number of parameters that are estimated. In the context of Bayesian or penalized likelihood frameworks, this has led to the notion of *effective number of parameters*, defined as

$$\mathrm{Tr}\Big(\Gamma^{1/2}S(I + S^*S)^{-1}S^*\Gamma^{-1/2}\Big).$$

This quantity agrees with efd by Proposition 4.3.5 and has been used extensively in Statistics and Machine Learning, see for example [Spiegelhalter et al., 2002], and Section 3.5.3 of [Bishop, 2006] and references therein. One motivation for this definition is based on a Bayesian version of the "hat matrix", see for example [Spiegelhalter et al., 2002]. However, in this article we provide a different motivation that is more relevant to our aims. Moreover, rather than as an effective number of parameters, we interpret efd as the effective dimension of the Bayesian linear model. Similar forms of effective dimension have been used for learning problems in [Zhang, 2002], [Zhang, 2005], [Caponnetto and De Vito, 2007] and for statistical inverse problems in [Lu and Mathé, 2014]. In all of these contexts the size of the operator A quantifies how informative the data is; see the discussion below. The paper [Bickel et al., 2008] introduced the notion of $\tau = \mathrm{Tr}(A)$ as an effective dimension for importance sampling within linear inverse problems and filtering. In that paper several transformations of the inverse problem are performed before doing the analysis. We undo these transformations. The role of τ in the performance of the Ensemble Kalman filter had been previously studied in [Furrer and Bengtsson, 2007].

The operator A has played an important role in the study of linear inverse

problems. First, it has been used for obtaining posterior contraction rates in the small noise limit, see the operator B^*B in [Lin et al., 2015], [Agapiou and Mathé, 2014]. Its use was motivated by techniques for analyzing classical regularization methods, in particular regularization in Hilbert scales see [Engl et al., 1996, Chapter 8]. Furthermore, its eigenvalues and eigendirections can be used to determine (optimal) low-rank approximations of the posterior covariance [Bui-Thanh et al., 2013], [Spantini et al., 2015, Theorem 2.3]. The analogue of A in nonlinear Bayesian inverse problems is the so-called prior-preconditioned data-misfit Hessian, which has been used in [Martin et al., 2012] to design Metropolis Hastings proposals.

Proposition 4.3.5 shows that efd is at most as large as the nominal dimension, in finite dimensional settings. The difference between both is a measure of the effect the prior has on the inference relative to the maximum likelihood solution. This difference increases as the size of Σ increases, or as the correlation among the vectors that form the columns of K increases, while the difference decreases as the size of Γ decreases or as the correlations in Γ increase. Note also that in finite dimensional settings, Proposition 4.3.4 shows that efd quantifies how much change there is in going from the posterior to the prior, measured in terms of change in the covariance, in units of the prior; and τ plays a similar role expressed in terms of change in the precisions, again in units of the prior. By the cyclic property of the trace, Lemma 4.6.5(ii), and by Proposition 4.3.4, τ and efd may also be characterized as follows:

$$\begin{aligned}\tau &= \mathrm{Tr}\big((C^{-1}-\Sigma^{-1})\Sigma\big) = \mathrm{Tr}\big((\Sigma - C)C^{-1}\big),\\ \mathsf{efd} &= \mathrm{Tr}\big((\Sigma - C)\Sigma^{-1}\big) = \mathrm{Tr}\big((C^{-1}-\Sigma^{-1})C\big).\end{aligned}$$

Thus we may also view efd as measuring the change in the precision, measured in units given by the posterior precision; whilst τ measures the change in the covariance, measured in units given by the posterior covariance.

Note that Proposition 4.3.4 also holds in the general Hilbert space setting, provided formula (4.3.4) for the posterior precision operator can be justified; see Remark 4.6.6 in the appendix. The above alternative identities for τ and efd can also be justified in those settings, using analogous techniques. We hence have that the interpretations of τ and efd discussed in the previous paragraph, carry over to such infinite dimensional settings.

In many applications, the unknown $u \in \mathbb{R}^{d_u}$ and often the data $y \in \mathbb{R}^{d_y}$ correspond to discretizations of continuum functions living in Hilbert spaces. The canonical illustration arises from discretizting Example 4.3.2 to obtain Example 4.3.1. In such situations the three matrices K, Γ, Σ defining the Bayesian inverse

problem also correspond to discretizations of infinite dimensional linear operators. It is of interest to understand the performance of importance sampling as the discretization level increases in order to decide how to distribute the available budget between using more particles or investing in higher discretization levels. A deep analysis of importance sampling in the large d limit can be found in [Bengtsson et al., 2008]. The authors show that, if $\beta \leq 1$ and $d \to \infty$, the maximum importance sampling weight converges to 1 in probability, unless the number of particles grows super-exponentially with, essentially, $\tau(d)$. Here we show that $\rho(d)$ grows exponentially with $\tau(d)$ (and $\mathsf{efd}(d)$), which together with Theorem 4.2.1 suggests also the need to increase the number of samples exponentially with dimension.

It is straightforward to check that since $Ku^\dagger \in \mathcal{H}$, the probability measure of the data in Assumption 4.3.9 is equivalent to the marginal probability measure of the data under the model, $\nu_y(dy)$. Hence for data of the form of Assumption 4.3.9, Theorem 4.3.7 implies that the posterior is absolutely continuous with respect to the prior, almost surely with respect to the noise distribution.

The deviance information criterion introduced in [Spiegelhalter et al., 2002], is based on a notion of effective number of parameters that generalises the one we discuss in this chapter to more general Bayesian hierarchical models.

In the context of inverse problems, by (4.3.9), the tempered un-normalized likelihood $g(u;y)^a$ takes the form

$$g(u;y)^a = \exp\left(-\frac{a}{2\gamma}\left\|\Gamma^{-1/2}Ku\right\|^2 + \frac{a}{\gamma}\langle\Gamma^{-1/2}y, \Gamma^{-1/2}Ku\rangle\right). \tag{4.3.13}$$

This corresponds to the likelihood of our standard inverse problem, but where Γ is replaced by Γ/a and hence A in Assumption 4.3.3 is scaled by a. In particular, in the context of an inverse problem in the Euclidean space $\mathbb{R}^d$, if $a = \frac{1}{d}$ and $A(d)$ is a discretization of an operator A with eigenvalues bounded by $\lambda_{\max}$ we easily deduce that the tempered problem has intrinsic dimensions $\mathsf{efd}, \tau \leq \lambda_{\max}$, bounded independently of d. Applying this sequentially then leads to the sequence of measures $\mu_{d,i}$ as explained at the end of Subsection 4.2.7. We remark that under the tempering approach d of these problems with bounded effective dimension would need to be solved sequentially; a careful study of the propagation of errors of such sequential scheme would be necessary to understand its complexity, but is beyond the scope of our work. In practice this issue can be ameliorated by including approriate mixing kernels, invariant with respect to $\mu_{d,i}$ for each i, as demonstrated in [Kantas et al., 2014].

4.4 Importance Sampling and Filtering

In Section 4.2 we introduced importance sampling, and studied its computational complexity. We highlighted the role of the density of the target with respect to the proposal. We also studied the behaviour of importance sampling when approaching loss of absolute continuity between target and proposal. In particular we studied the effect of various singular limits (large nominal dimension, small parameters) in this breakdown. Section 4.3 studied these issues for Bayesian linear inverse problems. Here we study them for the filtering problem, using the relationship between Bayesian inversion and filtering outlined in the introductory section, and detailed here. In Subsection 4.4.1 we set-up the problem and derive a link between importance sampling based particle filters and the inverse problem. In Subsections 4.4.2 and 4.4.3 respectively we use this connection to study the intrinsic dimension of filtering, and the connection to absolute continuity between proposal and target, and in doing so make comparisons between the standard and optimal proposals. Subsection 4.4.4 contains some explicit computations which enable comparison of the complexity of the two proposals in various singular limits relating to high dimension or small observational noise. We conclude with the literature review Subsection 4.4.5 which overviews the sources for the material herein.

The component of particle filtering which we analyze in this section is only that related to sequential importance sampling; we do not discuss the interaction between the simulated particles which arises via resampling schemes. Such interaction would not typically be very relevant in the two time-unit dynamical systems we study here, but would be necessary to get reasonable numerical schemes when assimilating data over many time units. We comment further on this, and the choice of the assimilation problem we study, in the literature review.

4.4.1 General Setting

We simplify the notation by setting $j = 0$ in (4.1.3) to obtain

$$\begin{aligned} v_1 &= Mv_0 + \xi, \quad v_0 \sim N(0, P), \quad \xi \sim N(0, Q), \\ y_1 &= Hv_1 + \zeta, \quad \zeta \sim N(0, R). \end{aligned} \tag{4.4.1}$$

Note that we have also imposed a Gaussian assumption on v_0. Because of the Markov assumption on the dynamics for $\{v_j\}$, we have that v_0 and ξ are independent. As in Section 4.3 we set-up the problem in a separable Hilbert space $\mathcal{H}$, although the reader versed only in finite dimensional Gaussian measures should have no trouble following the developments, simply by thinking of the covariance operators as

(possibly infinite) matrices. We assume throughout that the covariance operators $P, Q, R : \mathcal{H} \to \mathcal{H}$ are bounded, self-adjoint, positive linear operators, but not necessarily trace-class (see the discussion on this trace-class issue in Section 4.3). We also assume that the operators $M, H : \mathcal{H} \to \mathcal{H}$ that describe, respectively, the unconditioned signal dynamics and the observation operator, can be extended to larger spaces if necessary; see the appendix Subsection 4.6.1 for further details on these technical issues.

Our goal in this section is to study the complexity of importance sampling within the context of both the standard and optimal proposals for particle filtering. For both these proposals we show that there is an inverse problem embedded within the particle filtering method, and compute the proposal covariance, the observation operator and the observational noise covariance. We may then use the material from the previous section, concerning inverse problems, to make direct conclusions about the complexity of importance sampling for particle filters.

The aim of one step of filtering may be expressed as sampling from the target $\mathbb{P}_{v_1,v_0|y_1}$. Particle filters do this by importance sampling, with this measure on the product space $\mathcal{X} \times \mathcal{X}$ as the target. We wish to compare two ways of doing this, one by using the proposal distribution $\mathbb{P}_{v_1|v_0}\mathbb{P}_{v_0}$ and the second by using as proposal distribution $\mathbb{P}_{v_1|v_0,y_1}\mathbb{P}_{v_0}$. The first is known as the *standard proposal*, and the second as the *optimal proposal.* We now connect each of these proposals to a different inverse problem.

4.4.1.1 Standard Proposal

For the standard proposal we note that, using Bayes' theorem, conditioning, and that the observation y_1 does not depend on v_0 explicitly,

$$\begin{aligned}\mathbb{P}_{v_1,v_0|y_1} &\propto \mathbb{P}_{y_1|v_1,v_0}\mathbb{P}_{v_1,v_0} \\ &= \mathbb{P}_{y_1|v_1,v_0}\mathbb{P}_{v_1|v_0}\mathbb{P}_{v_0} \\ &= \mathbb{P}_{y_1|v_1}\mathbb{P}_{v_1|v_0}\mathbb{P}_{v_0}.\end{aligned}$$

Thus the density of the target $\mathbb{P}_{v_1,v_0|y_1}$ with respect to the proposal $\mathbb{P}_{v_1|v_0}\mathbb{P}_{v_0}$ is proportional to $\mathbb{P}_{y_1|v_1}$. Although this density concerns a proposal on the joint space of (v_0, v_1), since it involves only v_1 we may consider the related inverse problem of finding v_1, given y_1, and ignore v_0.

In this picture filtering via the standard proposal proceeds as follows:

$$\mathbb{P}_{v_0} \mapsto \mathbb{P}_{v_1} \mapsto \mathbb{P}_{v_1|y_1}.$$

Here the first step involves propagation of proability measures under the dynamics. This provides the proposal $\pi = \mathbb{P}_{v_1}$ used for importance sampling to determine the target $\mu = \mathbb{P}_{v_1|y_1}$. The situation is illustrated in the upper branch of Figure 4.1. Since

$$\mathbb{E}(v_1 v_1^*) = \mathbb{E}(Mv_0 + \xi)(Mv_0 + \xi)^*,$$

and v_0 and ξ are independent under the Markov assumption, the proposal distribution is readily seen to be a centred Gaussian with covariance $\Sigma = MPM^* + Q$. The observation operator is $K = H$ and the noise covariance $\Gamma = R$. We have established a direct connection between the particle filter, with standard proposal, and the inverse problem of the previous section. We will use this connection to study the complexity of the particle filter, with standard proposal, in what follows.

4.4.1.2 Optimal Proposal

For the optimal proposal we note that, by conditioning on v_0,

$$\begin{aligned}\mathbb{P}_{v_1,v_0|y_1} &= \mathbb{P}_{v_1|v_0,y_1}\mathbb{P}_{v_0|y_1} \\ &= \mathbb{P}_{v_1|v_0,y_1}\mathbb{P}_{v_0}\frac{\mathbb{P}_{v_0|y_1}}{\mathbb{P}_{v_0}}.\end{aligned}$$

Thus the density of the target $\mathbb{P}_{v_1,v_0|y_1}$ with respect to the proposal $\mathbb{P}_{v_1|v_0,y_1}\mathbb{P}_{v_0}$ is the same as the density of $\mathbb{P}_{v_0|y_1}$ with respect to $\mathbb{P}_{v_0}$. As a consequence, although this density concerns a proposal on the joint space of (v_0, v_1), it is equivalent to an inverse problem involving only v_0. We may thus consider the related inverse problem of finding v_0 given y_1, and ignore v_1.

In this picture filtering via the optimal proposal proceeds as follows:

$$\mathbb{P}_{v_0} \mapsto \mathbb{P}_{v_0|y_1} \mapsto \mathbb{P}_{v_1|y_1}.$$

Here the first step involves importance sampling with proposal $\pi = \mathbb{P}_{v_0}$ and target $\mu = \mathbb{P}_{v_0|y_1}$. This target measure is then propagated under the conditioned dynamics to find $\mathbb{P}_{v_1|y_1}$; the underlying assumption of the optimal proposal is that $\mathbb{P}_{v_1|v_0,y_1}$ can be sampled so that this conditioned dynamics can be implemented particle by particle. The situation is illustrated in the lower branch of Figure 4.1. Since

$$y_1 = HMv_0 + H\xi + \zeta$$

the proposal distribution is readily seen to be a centred Gaussian with covariance $\Sigma = P$, the observation operator $K = HM$ and the noise covariance given by

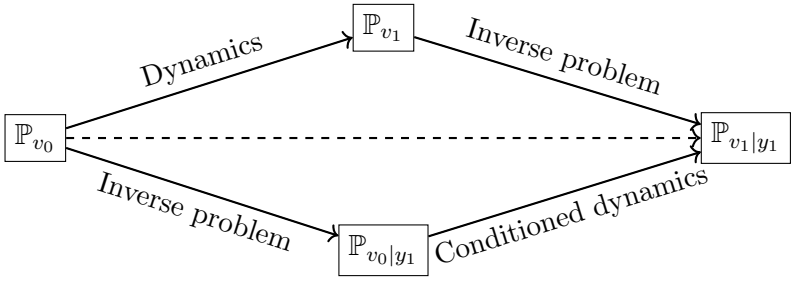


Figure 4.1: Filtering step decomposed in two different ways. The upper path first pushes forward the measure $\mathbb{P}_{v_0}$ using the signal dynamics, and then incorporates the observation y_1. The lower path assimilates the observation y_1 first, and then propagates the conditioned measure using the signal dynamics. The standard proposal corresponds to the upper decomposition and the optimal one to the lower decomposition.

the covariance of $H\xi + \zeta$, namely $\Gamma = HQH^* + R$. Again we have established a direct connection between the particle filter, with optimal proposal, and the inverse problem of the previous section. We will use this connection to study the complexity of the particle filter, with optimal proposal, in what follows.

A key assumption of the optimal proposal is the second step: the ability to sample from the conditioned dynamics $\mathbb{P}_{v_1|v_0,y_1}$ and we make a few comments on this before returning to our main purpose, namely to study complexity of particle filtering via the connection to an inverse problem. The first comment is to note that since we are in a purely Gaussian setting, this conditioned dynamics is itself determined by a Gaussian and so may in principle be performed in a straightforward fashion. In fact the conditioned dynamics remains Gaussian even if the forward model Mv_0 is replaced by a nonlinear map $f(v_0)$, so that the optimal proposal has wider applicability than might at first be appreciated. Secondly we comment that the Gaussian arising in the conditioned dynamics has mean m and variance Ξ given by the formulae

$$\begin{aligned}\Xi &= Q - QH^*(HQH^* + R)^{-1}HQ,\\ m &= Mv_0 + QH^*(HQH^* + R)^{-1}(y_1 - HMv_0).\end{aligned}$$

It is a tacit assumption in what follows that the operators defining the filtering problem are such that $\Xi : \mathcal{H} \to \mathcal{H}$ is well-defined and that $m \in \mathcal{H}$ is well-defined. More can be said about these points, but doing so will add further technicalities without contributing to the main goals of this chapter.

	Standard Proposal	Optimal proposal
Proposal	$\mathbb{P}_{v_0}(dv_0)\mathbb{P}_{v_1\|v_0}(dv_1)$	$\mathbb{P}_{v_0}(dv_0)\mathbb{P}_{v_1\|v_0,y_1}(dv_1)$
BIP	$y_1 = Hv_1 + \eta_{st}$	$y_1 = HMv_0 + \eta_{op}$
Prior Cov.	$MPM^* + Q$	P
Data Cov.	R	$R + HQH^*$
$\log g(u; y_1)$	$-\frac{1}{2}\|Hv_1\|_R^2 + \langle y_1, Hv_1\rangle_R$	$-\frac{1}{2}\|HMv_0\|_{R+HQH^*}^2 + \langle y_1, HMv_0\rangle_{R+HQH^*}$

Table 4.2

4.4.2 Intrinsic Dimension

Using the inverse problems that arise for the standard proposal and for the optimal proposal, and employing them within the definition of A from Assumption 4.3.3, we find the two operators A arising for these two different proposals:

$$A := A_{st} := (MPM^* + Q)^{1/2} H^* R^{-1} H (MPM^* + Q)^{1/2}$$

for the standard proposal, and

$$A := A_{op} := P^{\frac{1}{2}} M^* H^* (R + HQH^*)^{-1} HMP^{1/2}$$

for the optimal proposal. Again here it is assumed that these operators are bounded in $\mathcal{H}$:

Assumption 4.4.1. *The operators A_{st} and A_{op}, viewed as linear operators in $\mathcal{H}$, are bounded. Furthermore, assume that the spectra of both A_{st} and A_{op} consist of a countable number of eigenvalues.*

Using these definitions of A_{st} and A_{op} we may define, from (4.3.6), the intrinsic dimensions τ_{st}, efd_{st} for the standard proposal and τ_{op}, efd_{op} for the optimal one in the following way

$$\tau_{st} = \mathrm{Tr}(A_{st}), \qquad \mathsf{efd}_{st} = \mathrm{Tr}\big((I + A_{st})^{-1} A_{st}\big)$$

and

$$\tau_{op} = \mathrm{Tr}(A_{op}), \qquad \mathsf{efd}_{op} = \mathrm{Tr}\big((I + A_{op})^{-1} A_{op}\big).$$

4.4.3 Absolute Continuity

The following two theorems are a straightforward application of Theorem 4.3.7, using the connections between filtering and inverse problems made above. The contents

of the two theorems are summarized in Table 4.2.

Theorem 4.4.2. *Consider one-step of particle filtering for* (4.4.1). *Let* $\mu = \mathbb{P}_{v_1|y_1}$ *and* $\pi = \mathbb{P}_{v_1} = N(0, Q + MPM^*)$. *Then the following are equivalent:*

i) $\mathsf{efd}_{st} < \infty$;

ii) $\tau_{st} < \infty$;

iii) $R^{-1/2}Hv_1 \in \mathcal{H}$, π*-almost surely;*

iv) for ν_y*-almost all* y, *the target distribution* μ *is well defined as a measure in* $\mathcal{X}$ *and is absolutely continuous with respect to the proposal with*

$$\frac{d\mu}{d\pi}(v_1) \propto \exp\left(-\frac{1}{2}\left\|R^{-1/2}Hv_1\right\|^2 + \frac{1}{2}\langle R^{-1/2}y_1, R^{-1/2}Hv_1\rangle\right) =: g_{st}(v_1; y_1), \tag{4.4.2}$$

where $0 < \pi\big(g_{st}(\cdot; y_1)\big) < \infty$.

Theorem 4.4.3. *Consider one-step of particle filtering for* (4.4.1). *Let* $\mu = \mathbb{P}_{v_0|y_1}$ *and* $\pi = \mathbb{P}_{v_0} = N(0, Q)$. *Then, for* $R_{op} = R + HQH^*$, *the following are equivalent:*

i) $\mathsf{efd}_{op} < \infty$;

ii) $\tau_{op} < \infty$;

iii) $R_{op}^{-1/2}HMv_0 \in \mathcal{H}$, π*-almost surely;*

iv) for ν_y*-almost all* y, *the target distribution* μ *is well defined as a measure in* $\mathcal{X}$ *and is absolutely continuous with respect to the proposal with*

$$\frac{d\mu}{d\pi}(v_0) \propto \exp\left(-\frac{1}{2}\left\|R_{op}^{-1/2}HMv_0\right\|^2 + \frac{1}{2}\langle R_{op}^{-1/2}y_1, R_{op}^{-1/2}HMv_0\rangle\right) =: g_{op}(v_0; y_1), \tag{4.4.3}$$

where $0 < \pi\big(g_{op}(\cdot; y_1)\big) < \infty$.

Remark 4.4.4. *Because of the exponential structure of* g_{st} *and* g_{op}, *the assertion* (iv) *in the preceding two theorems is equivalent to* g_{st} *and* g_{op} *being* ν*-almost surely positive and finite and for almost all* y_1 *the second moment of the target-proposal density is finite. This second moment is given, for the standard and optimal proposals, by*

$$\rho_{st} = \frac{\pi\left(g_{st}(\cdot; y)^2\right)}{\pi\big(g_{st}(\cdot; y)\big)^2} < \infty$$

and

$$\rho_{op} = \frac{\pi\left(g_{op}(\cdot; y)^2\right)}{\pi\big(g_{op}(\cdot; y)\big)^2} < \infty$$

respectively. The relative sizes of ρ_{st} and ρ_{op} determine the relative efficiency of the standard and optimal proposal versions of filtering.

The following theorem shows that there is loss of absolute continuity for the standard proposal whenever there is for the optimal one. The result is formulated in terms of the intrinsic dimension τ, and we show that $\tau_{op} = \infty$ implies $\tau_{st} = \infty$. By Theorem 4.3.7, this implies the result concerning absolute continuity. Recalling that poor behaviour of importance sampling is intimately related to such breakdown, this suggests that the optimal proposal is always at least as good as the standard one. The following theorem also gives a condition on the operators H, Q and R under which collapse for both proposals occurs at the same time, irrespective of the regularity of the operators M and P. Roughly speaking this simultaneous collapse result states that if R is large compared to Q then absolute continuity for both proposals is equivalent; and hence collapse of importance sampling happens under one proposal if and only if it happens under the other. Intuitively the advantages of the optimal proposal stem from the noise in the dynamics; they disappear completely if the dynamics is deterministic. The theorem quantifies this idea. Finally, an example demonstrates that there are situations where τ_{op} is finite, so that optimal proposal based importance sampling works well for finite dimensional approximations of an infinite dimensional problem, whilst τ_{st} is infinite, so that standard proposal based importance sampling works poorly for finite dimensional approximations. The proof of the theorem is given in the appendix, Subsection 4.6.4.

Theorem 4.4.5. *Suppose that Assumption 4.4.1 holds. Then,*

$$\tau_{op} \leq \tau_{st}. \tag{4.4.4}$$

Moreover, if $\mathrm{Tr}\Big(HQH^*R^{-1}\Big) < \infty$, *then*

$$\tau_{st} < \infty \iff \tau_{op} < \infty.$$

We remark that, under additional simplifying assumptions, we can obtain bounds of the form (4.4.4) for efd and ρ. We chose to formulate the result in terms of τ since we can prove the bound (4.4.4) in full generality. Moreover, by Theorem 4.3.7 the bound in terms of τ suffices in order to understand the different collapse properties of both proposals.

The following example demonstrates that it is possible that $\tau_{op} < \infty$ while $\tau_{st} = \infty$; in this situation filtering via the optimal proposal is well-defined, whilst using the standard proposal it is not. Loosely speaking, this happens if y_1 provides more information on v_1 than v_0.

Example 4.4.6. *Suppose that*

$$H = Q = R = M = I, \qquad \mathrm{Tr}(P) < \infty.$$

Then, it is straightforward from the definitions that $A_{st} = P + I$ and $A_{op} = P/2$. In an infinite dimensional Hilbert the identity operator has infinite trace, $\mathrm{Tr}(I) = \infty$, and so

$$\tau_{st} = \mathrm{Tr}(A_{st}) = \mathrm{Tr}(P + I) = \infty, \qquad \tau_{op} = \mathrm{Tr}(A_{op}) = \mathrm{Tr}(P/2) < \infty.$$

We have thus established an example of a filtering model for which $\tau_{st} = \infty$ and $\tau_{op} < \infty$. We note that by Theorem 4.4.5, any such example satisfies the condition $\mathrm{Tr}(HQH^*R^{-1}) = \infty$. When this condition is met, automatically $\tau_{st} = \infty$ (see the proof of the Theorem 4.4.5 in the appendix, Subsection 4.6.4). However, τ_{op} can still be finite. Indeed, within the proof of that theorem we show that the inequality

$$\tau_{op} \leq \mathrm{Tr}(R^{-1}HMPM^*H^*)$$

always holds. The right-hand side may be finite provided that the eigenvalues of P decay fast enough. A simple example of this situation is where HM is a bounded operator and all the relevant operators have eigenvalues. In this case the Rayleigh-Courant-Fisher theorem – see the appendix, Subsection 4.6.3 for a reference – guarantees that the eigenvalues of $HMPM^*H^*$ can be bounded in terms of those of P. Again by the Rayleigh-Courant-Fisher theorem, since we are always assuming that the covariance R is bounded, it is possible to bound the eigenvalues of $R^{-1}HMPM^*H^*$ in terms of those of $HMPM^*H^*$. This provides a wider range of examples where $\tau_{st} = \infty$ while $\tau_{op} < \infty$.

4.4.4 Singular Limits

We are interested in the computational complexity of particle filtering. As stated in Remark 4.4.4 the values of the second moment of the target-proposal density, ρ_{st} and ρ_{op}, characterize the performance of particle filtering using impotance sampling with the standard and optimal proposals respectively. By comparing the values of ρ_{st} and ρ_{op} we can ascertain situations in which the optimal proposal has significant

Regime	Param.	$\text{eig}(A_{st})$	$\text{eig}(A_{op})$	$\text{eig}(P_\infty)$	ρ_{st}	ρ_{op}
Small obs. noise	$r \to 0$	r^{-1}	r	r	$r^{-d/2}$	1
	$r = q \to 0$	1	1	$r(=q)$	1	1
Large d	$d \to \infty$	1	1	1	$\exp(d)$	$\exp(d)$

Table 4.3: Scalings of the standard and optimal proposals in small noise and large d regimes for one filter step initialized from stationarity ($P = P_\infty$). This table and the one below should be interpreted in the same way as Table 4.1.

Regime	Param.	$\text{eig}(A_{st})$	$\text{eig}(A_{op})$	ρ_{st}	ρ_{op}
Small obs. noise	$r \to 0$	r^{-1}	1	$r^{-d/2}$	1
	$r = q \to 0$	r^{-1}	r^{-1}	$r^{-d/2}$	$r^{-d/2}$
Large d	$d \to \infty$	1	1	$\exp(d)$	$\exp(d)$

Table 4.4: Scalings of the standard and optimal proposals in small noise and large d regimes for one filter step initialized from $P = pI$.

advantage over the standard proposal. We also recall, from Section 4.3, the role of the intrinsic dimensions in determining the scaling of the second moment of the target-proposal density.

The following example will illustrate a number of interesting phenomena in this regard. In the setting of fixed finite state/data state dimension it will illustrate how the scalings of the various covariances entering the problem effect computational complexity. In the setting of increasing nominal dimension d, when the limiting target is singular with respect to the proposal, it will illustrate how computational complexity scales with d. And finally we will contrast the complexity of the filters in two differing initialization scenarios: (i) from an arbitrary initial covariance P, and from a steady state covariance P_∞. Such a steady state covariance is a fixed point of the covariance update map for the Kalman filter defined by (4.1.3).

Example 4.4.7. *Suppose that $M = H = I \in \mathbb{R}^{d\times d}$, and $R = rI$, $Q = qI$, with $r, q > 0$. A simple calculation shows that the steady state covariance is given by*

$$P_\infty = \frac{\sqrt{q^2 + 4qr} - q}{2} I,$$

and that the operators A_{st} and A_{op} when $P = P_\infty$ are

$$A_{st} = \frac{\sqrt{q^2 + 4qr} + q}{2r} I, \qquad A_{op} = \frac{\sqrt{q^2 + 4qr} - q}{2(q + r)} I.$$

Note that A_{st} and A_{op} are a function of q/r, whereas P_∞ is not.

If the filtering step is initialized outside stationarity at $P = pI$, with $p > 0$,

then

$$A_{st} = \frac{p+q}{r} I, \qquad A_{op} = \frac{p}{q+r} I.$$

Both the size and number of the eigenvalues of A_{op}/A_{st} play a role in determining the size of ρ, the second moment of the target-proposal variance. It is thus interesting to study how ρ scales in both the small observational noise regime $r \ll 1$ and the high dimensional regime $d \gg 1$. The results are summarized in Tables 4.3 and 4.4. Some conclusions from these tables are:

- *The standard proposal degenerates at an algebraic rate as $r \to 0$, for fixed dimenson d, for both initializations of P.*
- *The optimal proposal is not sensitive to the small observation limit $r \to 0$ if the size of the signal noise, q, is fixed. If started outside stationarity, the optimal proposal degenerates algebraically if $q \propto r \to 0$. However, even in this situation the optimal proposal scales well if initialized in the stationary regime.*
- *In this example the limiting problem with $d = \infty$ has infinite intrinsic dimension for both proposals, because the target and the proposal are mutually singular. As a result, ρ grows exponentially in the large d limit.*
- *Example 4.4.6 suggests that there are cases where ρ_{st} grows exponentially in the large dimensional limit $d \to \infty$ but ρ_{op} converges to a finite value. This may happen if $\mathrm{Tr}\left(HQH^*R^{-1}\right) < \infty$, but the prior covariance P is sufficiently smooth.*

4.4.5 Literature Review

In Subsection 4.4.1 we follow [Bengtsson et al., 2008], [Bickel et al., 2008], [Snyder et al., 2008], [Snyder, 2011], [Slivinski and Snyder, 2015], [Snyder et al., 2015] and consider one step of the filtering model (4.1.3). There are two main motivations for studying one step of the filter. Firstly, if keeping the filter error small is prohibitively costly for one step, then there is no hope that an online particle filter will be successful [Bengtsson et al., 2008]. Secondly, it can provide insight for filters initialized close to stationarity [Chorin and Morzfeld, 2013]. As in [Snyder, 2011], [Slivinski and Snyder, 2015], [Snyder et al., 2015] we cast the analysis of importance sampling in joint space and consider as target $\mu := \mathbb{P}_{u|y_1}$, with $u := (v_0, v_1)$ and with the standard and optimal proposals defined in Subsection 4.4.1.

In general nonlinear, non-Gaussian problems the optimal proposal is usually not implementable, since it is not possible to evaluate the corresponding weights,

or to sample from the distribution $\mathbb{P}_{v_1|v_0,y_1}$. However, the optimal proposal is implementable in our framework (see for example [Doucet et al., 2000]) and understanding its behaviour is important in order to build and analyze improved and computable proposals which are informed by the data [Tu et al., 2013], [Goodman et al., 2015], [van Leeuwen, 2010]. It is worth making the point that the so-called "optimal proposal" is really only locally optimal. In particular, this choice is optimal in minimizing the variance of the weights at the given step given that all previous proposals have been already chosen. This choice does not minimize the Monte Carlo variance for some time horizon for some family of test functions. A different optimality criterion is obtained by trying to simultaneously minimize the variance of weights at times $t \leq r \leq t+m$, for some $m \geq 1$, or minimize some function of these variances, say their sum or their maximum. Such look ahead procedures might not be feasible in practice. Surprisingly, examples exist where the standard proposal leads to smaller variance of weights some steps ahead relative to the locally optimally tuned particle fiter; see for example Section 3 of [Johansen and Doucet, 2008], and the discussion in [Chopin and Papaspiliopoulos, 2016, Chapter 10]. Still, such examples are quite contrived and experience suggests that local adaptation is useful in practice.

Similarly as for inverse problems, the values of ρ_{st} and ρ_{op} determine the performance of importance sampling for the filtering model with the standard and optimal proposals. These depend in a nonlinear fashion on the eigenvalues of A_{st} and A_{op}. In Subsection 4.4.3 we show that the conditions of collapse for the standard and optimal proposals (found in [Snyder, 2011] and [Bickel et al., 2008], respectively) correspond to any of the equivalent conditions of finite dimension or finite ρ described in Theorems 4.4.2 and 4.4.3.

In Subsection 4.4.4 we study singular limits in the framework of [Chorin and Morzfeld, 2013]. Thus, we consider a diagonal filtering setting in the Euclidean space $\mathbb{R}^d$, and assume that all coordinates of the problem play the same role, which corresponds to the extreme case $\beta = 0$ in Subsection 4.3.4. The paper [Chorin and Morzfeld, 2013] introduced a notion of effective dimension for detectable and stabilizable linear Gaussian data assimilation problems as the Frobenius norm of the steady state covariance of the filtering distribution. It is well known that the detectability and stabilizability conditions ensure the existence of such steady state covariance [Lancaster and Rodman, 1995]. This notion of dimension quantifies the success of data assimilation in having reduced uncertainty on the unknown once the data has been assimilated. Therefore the definition of dimension given in [Chorin and Morzfeld, 2013] is at odds with both τ and efd : it does not quantify how much

is learned from the data in one step, but instead how concentrated the filtering distribution is in the time asymptotic regime when the filter is in steady state. Our calculations demonstrate differences which can occur in the computational complexity of filtering, depending on whether it is initialized in this statistical steady state, or at an arbitrary point.

4.5 Conclusions

The main motivation for this article is the study of computational complexity of importance sampling, and in particular provision of a framework which unifies the multitude of publications with bearing on this question. We study inverse problems and particle filters in Bayesian models that involve high and infinite state space and data dimensions.

Our study has required revisiting the fundamental structure of importance sampling on general state spaces. We have derived non-asymptotic concentration inequalities for the particle approximation error and related what turns out to be the key parameter of performance, the second moment of the density between the target and proposal, to many different importance sampling input and output quantities.

As a reasonable compromise between mathematical tractability and practical relevance we have focused on Bayesian linear models for regression and statistical inversion of ill-posed inverse problems. We have studied the efficiency of sampling-based posterior inference in these contexts carried out by importance sampling using the prior as proposal. We have demonstrated that performance is controlled by an intrinsic dimension, as opposed to the state space or data dimensions, and we have discussed and related two different measures of this dimension. It is important to emphasise that the intrinsic dimension is really a measure of relative strength between the prior and the likelihood in forming the posterior, as opposed to a measure of "degrees of freedom" in the prior. In other words, infinite-dimensional Bayesian linear models with finite intrinsic dimension are not identified with models for which the prior for the unknown is concentrated on a finite-dimensional manifold of the infinite-dimensional state space.

A similar consideration of balancing tractability and practical relevance has dictated the choice not to study interacting particles typically used for filtering, but rather to focus on one-step filtering using importance sampling. For such problems we introduce appropriate notions of intrinsic dimension and compare the relative merits of popular alternative schemes.

The most pressing topic for future research stemming from this article is the

development of concrete recommendations for algorithmic design within classes of Bayesian models used in practice. Within the model structure we have studied here, practically relevant and important extensions include models with non-Gaussian priors on the unknown, nonlinear operators that link the unknown to the data, and unknown hyperparameters involved in the model specification. Linearisation of a nonlinear model around some reasonable value for the unknown (e.g. the posterior mean) is one way to extend our measures of intrinsic dimension in such frameworks. We can expect the subject area to see considerable development in the coming decade.

4.6 Appendix

4.6.1 Gaussian Measures in Hilbert Space

In Section 4.3 we study Bayesian inverse problems in the Hilbert space setting. This enables us to talk about infinite dimensional limits of sequences of high dimensional inverse problems and is hence useful when studying the complexity of importance sampling in high dimensions. Here we provide some background on Gaussian measures in Hilbert space. We start by describing how to construct a random draw from a Gaussian measure on an infinite dimensional separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle, \|\cdot\|)$. Let $\mathcal{C} : \mathcal{H} \to \mathcal{H}$ be a self-adjoint, positive-definite and trace class operator. It then holds that $\mathcal{C}$ has a countable set of eigenvalues $\{\kappa_j\}_{j\in\mathbb{N}}$, with corresponding normalized eigenfunctions $\{e_j\}_{j\in\mathbb{N}}$ which form a complete orthonormal basis in $\mathcal{H}$.

Example 4.6.1. *We use as a running example the case where $\mathcal{H}$ is the space of square integrable real-valued functions on the unit interval, $\mathcal{H} = L^2(0,1)$ and where the Gaussian measure of interest is a unit centred Brownian bridge on the interval $(0,1)$. Then $m = 0$ and $\mathcal{C}$ is the inverse of the negative Laplacian on $(0,1)$ with homogeneous Dirichlet boundary conditions. The eigenfunctions and eigenvalues of $\mathcal{C}$ are given by*

$$e_j(t) = \sqrt{2}\sin(j\pi t), \quad \kappa_j = (j\pi)^{-2}.$$

The eigenvalues are summable and hence the operator $\mathcal{C}$ is trace class. For further details see [Stuart, 2010].

For any $m \in \mathcal{H}$, we can write a draw $x \sim N(m, \mathcal{C})$ as

$$x = m + \sum_{j=1}^{\infty} \sqrt{\kappa_j}\zeta_j e_j,$$

where ζ_j are independent standard normal random variables in $\mathbb{R}$; this is the Karhunen-Loeve expansion [Adler, 1990, Chapter III.3]. The trace class assumption on the operator $\mathcal{C}$, ensures that $x \in \mathcal{H}$ with probability 1, see Lemma 4.6.2 in Subsection 4.6.3. The particular rate of decay of the eigenvalues $\{\kappa_j\}$ determines the almost sure regularity properties of x. The idea is that the quicker the decay, the smoother x is, in a sense which depends on the basis $\{e_j\}$. For example if $\{e_j\}$ is the Fourier basis, which is the case if $\mathcal{C}$ is a function of the Laplacian on a tours, then a quicker decay of the eigenvalues of $\mathcal{C}$ means a higher Hölder and Sobolev regularity (see [Stuart, 2010, Lemmas 6.25 & 6.27] and [Dashti and Stuart, 2016, Section 2.4]). For the Brownian bridge Example 4.6.1 above, draws are almost surely in spaces of both Hölder and Sobolev regularity upto (but not including) one half.

The above considerations suggest that we can work entirely in the "frequency" domain, namely the space of coefficients of the element of $\mathcal{H}$ in the eigenbasis of the covariance, the sequence space ℓ^2. Indeed, we can identify the Gaussian measure $N(m, \mathcal{C})$ with the independent product measure $\bigotimes_{j=1}^{\infty} N(m_j, \kappa_j)$, where $m_j = \langle m, e_j \rangle$. Using this identification, we can define a sequence of Gaussian measures in $\mathbb{R}^d$ which converge to $N(m, \mathcal{C})$ as $d \to \infty$, by truncating the product measure to the first d terms. Even though in $\mathbb{R}^d$ any two Gaussian measures with strictly positive covariances are absolutely continuous with respect to each other (that is, equivalent as measures), in the infinite-dimensional limit two Gaussian measures can be mutually singular, and indeed are unless very stringent conditions are satisfied.

For $N(m, \mathcal{C})$ in $\mathcal{H}$, we define its Cameron-Martin space E as the domain of $\mathcal{C}^{-1/2}$, which can be characterized as the space of all the shifts in the mean which result in an equivalent Gaussian measure. Since $\mathcal{C}$ is a trace class operator, its inverse (hence also its square root) is an unbounded operator, therefore E is a compact subset of $\mathcal{H}$. In fact E has zero measure under $N(0, \mathcal{C})$. For example, if $\mathcal{C}$ is given by the Brownian bridge Example 4.6.1, then the Cameron-Martin space E is the Sobolev space of functions which vanish on the boundary and whose first derivative is in $\mathcal{H}$; as mentioned above, draws from this measure only have upto half a derivative in the Sobolev sense. The equivalence or singularity of two Gaussian measures with different covariance operators and different means depends on the compatibility of both their means and covariances, as expressed in the three conditions of the Feldman-Hajek theorem. For more details on the equivalence and singularity of Gaussian measures see [Da Prato and Zabczyk, 1992].

The Karhunen-Loeve expansion makes sense even if $\mathcal{C}$ is not trace class, in which case it defines a Gaussian measure in a space $\mathcal{X} \supset \mathcal{H}$ with a modified covariance operator which *is* trace class. Indeed, let $D : \mathcal{H} \to \mathcal{H}$ be any injective

bounded self-adjoint operator such that: a) D is diagonalizable in $\{e_j\}_{j\in\mathbb{N}}$, with (positive) eigenvalues $\{d_j\}_{j\in\mathbb{N}}$; b) the operator $D\mathcal{C}D$ is trace class, that is, $\{\kappa_j d_j^2\}_{j\in\mathbb{N}}$ is summable. Define the weighted inner product $\left\langle \cdot, \cdot \right\rangle_{D^{-2}} := \left\langle D\cdot, D\ \cdot \right\rangle$, the weighted norm $\|\cdot\|_{D^{-2}} = \|D\cdot\|$ and the space

$$\mathcal{X} := \overline{\text{span}\{e_j : j \in \mathbb{N}\}}^{\|\cdot\|_{D^{-2}}}.$$

Then the functions $\psi_j = d_j^{-1} e_j,\ j \in \mathbb{N}$, form a complete orthonormal basis in the Hilbert space $(\mathcal{X}, \left\langle \cdot, \cdot \right\rangle_{D^{-2}}, \|\cdot\|_{D^{-2}})$. The Karhunen-Loeve expansion can then be written as

$$x = m + \sum_{j=1}^{\infty} \sqrt{\kappa_j}\zeta_j e_j = m + \sum_{j=1}^{\infty} \sqrt{\kappa_j} d_j \zeta_j \psi_j,$$

so that we can view x as drawn from the Gaussian measure $N(m, D\mathcal{C}D)$ in $\mathcal{X}$, where $D\mathcal{C}D$ is trace class by assumption. For example, the case $\mathcal{H} = L^2(0,1)$ and $\mathcal{C} = I$, corresponding to Gaussian white noise for functions on the interval $(0,1)$, can be made sense of in negative Sobolev-Hilbert spaces with $-1/2 - \epsilon$ derivatives, for any $\epsilon > 0$. Finally, we stress that absolute continuity in general and the Cameron-Martin space in particular, are concepts which are independent of the space in which we make sense of the measure. In the Gaussian white noise example, we hence have that the Cameron-Martin space is $E = \mathcal{H}$.

The following lemma is similar to numerous results concerning Gaussian measures in function spaces. Because the precise form which we use is not in the literature, we provide a direct proof.

Lemma 4.6.2. *Let $\mathcal{X}$ be a separable Hilbert space with orthonormal basis $\{\varphi_j\}_{j\in\mathbb{N}}$. Define the Gaussian measure γ through the Karhunen-Loeve expansion*

$$\gamma := \mathcal{L}\Big(\sum_{j=1}^{\infty} \sqrt{\lambda_j}\xi_j\varphi_j\Big),$$

where λ_j is a sequence of positive numbers and where ξ_j are i.i.d. standard normal. Then draws from γ are in $\mathcal{X}$ almost surely if and only if $\sum_{j=1}^{\infty}\lambda_j < \infty$.

Proof. If $\sum_{j=1}^{\infty}\lambda_j < \infty$, then

$$\mathbb{E}_\gamma \|x\|_{\mathcal{X}}^2 = \mathbb{E}\sum_{j=1}^{\infty}\lambda_j\xi_j^2 = \sum_{j=1}^{\infty}\lambda_j < \infty,$$

hence $x \sim \gamma$ is in $\mathcal{X}$ almost surely.

For the converse, suppose that $x \sim \gamma$ is in $\mathcal{X}$ almost surely. Then

$$\|x\|_{\mathcal{X}}^2 = \sum_{j=1}^{\infty} \lambda_j \xi_j^2 < \infty, \quad \text{a.s.}$$

Note that this implies that $\lambda_j \to 0$, and so in particular $\lambda_\infty := \sup_j \lambda_j < \infty$.

By [Kallenberg, 2002, Theorem 3.17], since $\sqrt{\lambda_j}\xi_j \sim N(0, \lambda_j)$ are independent and symmetric random variables, we get that

$$\sum_{j=1}^{\infty} \mathbb{E}[\lambda_j \xi_j^2 \wedge 1] < \infty.$$

A change of variable gives

$$\begin{aligned}\mathbb{E}[\lambda_j \xi_j^2 \wedge 1] &\geq \frac{2}{\sqrt{2\pi\lambda_j}} \int_0^1 y^2 e^{-\frac{y^2}{2\lambda_j}} dy \\ &= \frac{2\lambda_j^{\frac{3}{2}}}{\sqrt{2\pi\lambda_j}} \int_0^{1/\sqrt{\lambda_j}} z^2 e^{-\frac{z^2}{2}} dz = \frac{2\lambda_j}{\sqrt{2\pi}} \int_0^{1/\sqrt{\lambda_j}} z^2 e^{-\frac{z^2}{2}} dz.\end{aligned}$$

Thus, for every $j \in \mathbb{N}$,

$$\mathbb{E}[\lambda_j \xi_j^2 \wedge 1] \geq \frac{2\lambda_j}{\sqrt{2\pi}} \int_0^{1/\sqrt{\lambda_\infty}} z^2 e^{-\frac{z^2}{2}} dz.$$

Since the left hand side is summable, we conclude that

$$\sum_{j=1}^{\infty} \lambda_j < \infty.$$

□

4.6.2 Proofs Section 4.2

Throughout we denote by π_{MC}^N the empirical random measure

$$\pi_{\mathrm{MC}}^N := \frac{1}{N} \sum_{n=1}^{N} \delta_{u^n}, \quad u^n \sim \pi.$$

We recall that μ^N denotes the particle approximation of μ based on sampling from the proposal π.

4.6.2.1 Proof of Theorem 4.2.1

Proof of Theorem 4.2.1. For the bias we write

$$\begin{aligned}\mu^N(\phi) - \mu(\phi) &= \frac{1}{\pi^N_{\mathrm{MC}}(g)}\pi^N_{\mathrm{MC}}(\phi g) - \mu(\phi) \\ &= \frac{1}{\pi^N_{\mathrm{MC}}(g)}\pi^N_{\mathrm{MC}}\Big((\phi - \mu(\phi))g\Big).\end{aligned}$$

Then, letting $\overline{\phi} := \phi - \mu(\phi)$ and noting that

$$\pi(\overline{\phi}g) = 0$$

we can rewrite

$$\mu^N(\phi) - \mu(\phi) = \frac{1}{\pi^N_{\mathrm{MC}}(g)}\Big(\pi^N_{\mathrm{MC}}(\overline{\phi}g) - \pi(\overline{\phi}g)\Big).$$

The first of the terms in brackets is an unbiased estimator of the second one, and so

$$\begin{aligned}\mathbb{E}\big[\mu^N(\phi) - \mu(\phi)\big] &= \mathbb{E}\left[\Big(\frac{1}{\pi^N_{\mathrm{MC}}(g)} - \frac{1}{\pi(g)}\Big)\Big(\pi^N_{\mathrm{MC}}(\overline{\phi}g) - \pi(\overline{\phi}g)\Big)\right] \\ &= \mathbb{E}\left[\frac{1}{\pi^N_{\mathrm{MC}}(g)\pi(g)}\Big(\pi(g) - \pi^N_{\mathrm{MC}}(g)\Big)\Big(\pi^N_{\mathrm{MC}}(\overline{\phi}g) - \pi(\overline{\phi}g)\Big)\right].\end{aligned}$$

Therefore,

$$\begin{aligned}&\Big|\mathbb{E}\big[\mu^N(\phi) - \mu(\phi)\big]\Big| \\ &\quad\le \Big|\mathbb{E}\Big[\big(\mu^N(\phi) - \mu(\phi)\big)1_{\{2\pi^N_{\mathrm{MC}}(g) > \pi(g)\}}\Big]\Big| + \Big|\mathbb{E}\Big[\big(\mu^N(\phi) - \mu(\phi)\big)1_{\{2\pi^N_{\mathrm{MC}}(g) \le \pi(g)\}}\Big]\Big| \\ &\quad\le \frac{2}{\pi(g)^2}\mathbb{E}\Big[\big|\pi(g) - \pi^N_{\mathrm{MC}}(g)\big|\big|\pi^N_{\mathrm{MC}}(\overline{\phi}g) - \pi(\overline{\phi}g)\big|\Big] + 2\mathbb{P}\Big(2\pi^N_{\mathrm{MC}}(g) \le \pi(g)\Big) \\ &\quad\le \frac{2}{\pi(g)^2}\frac{1}{\sqrt{N}}\pi\big(g^2\big)^{1/2}\frac{2}{\sqrt{N}}\pi\big(g^2\big)^{1/2} + 2\mathbb{P}\Big(2\pi^N_{\mathrm{MC}}(g) \le \pi(g)\Big),\end{aligned}$$

where in the second and third inequality we used that $|\phi| \le 1$. Now note that

$$\mathbb{P}\Big(2\pi^N_{\mathrm{MC}}(g) \le \pi(g)\Big) = \mathbb{P}\Big(2(\pi^N_{\mathrm{MC}}(g) - \pi(g)) \le -\pi(g)\Big) \le \mathbb{P}\Big(2|\pi^N_{\mathrm{MC}}(g) - \pi(g)| \ge \pi(g)\Big).$$

By the Markov inequality $\mathbb{P}\Big(2\pi^N_{\mathrm{MC}}(g) \le \pi(g)\Big) \le \frac{4}{N}\frac{\pi(g^2)}{\pi(g)^2}$, and so

$$\sup_{|\phi|\le 1}\Big|\mathbb{E}\big[\mu^N(\phi) - \mu(\phi)\big]\Big| \le \frac{12}{N}\frac{\pi(g^2)}{\pi(g)^2}.$$

This completes the proof of the result for the bias. For the MSE

$$\begin{aligned}\mu^N(\phi) - \mu(\phi) &= \frac{1}{\pi^N_{\mathrm{MC}}(g)}\pi^N_{\mathrm{MC}}(\phi g) - \frac{1}{\pi(g)}\pi(\phi g) \\ &= \left(\frac{1}{\pi^N_{\mathrm{MC}}(g)} - \frac{1}{\pi(g)}\right)\pi^N_{\mathrm{MC}}(\phi g) - \frac{1}{\pi(g)}\Big(\pi(\phi g) - \pi^N_{\mathrm{MC}}(\phi g)\Big) \\ &= \frac{1}{\pi(g)}\Big(\pi(g) - \pi^N_{\mathrm{MC}}(g)\Big)\mu^N(\phi) - \frac{1}{\pi(g)}\Big(\pi(\phi g) - \pi^N_{\mathrm{MC}}(\phi g)\Big), \quad (4.6.1)\end{aligned}$$

and so using the inequality $(a+b)^2 \leq 2(a^2+b^2)$ we obtain

$$\big(\mu^N(\phi) - \mu(\phi)\big)^2 \leq \frac{2}{\pi(g)^2}\left\{\Big(\pi(g) - \pi^N_{\mathrm{MC}}(g)\Big)^2\mu^N(\phi)^2 + \Big(\pi(\phi g) - \pi^N_{\mathrm{MC}}(\phi g)\Big)^2\right\}.$$

Therefore, for $|\phi| \leq 1$,

$$\begin{aligned}\mathbb{E}\left[\big(\mu^N(\phi) - \mu(\phi)\big)^2\right] &\leq \frac{2}{\pi(g)^2}\left\{\mathbb{E}\left[\Big(\pi(g) - \pi^N_{\mathrm{MC}}(g)\Big)^2\right] + \mathbb{E}\left[\Big(\pi(\phi g) - \pi^N_{\mathrm{MC}}(\phi g)\Big)^2\right]\right\} \\ &= \frac{2}{\pi(g)^2}\Big\{\mathrm{Var}_\pi\big(\pi^N_{\mathrm{MC}}(g)\big) + \mathrm{Var}_\pi\big(\pi^N_{\mathrm{MC}}(\phi g)\big)\Big\} \\ &\leq \frac{2}{N\pi(g)^2}\Big\{\pi\big(g^2\big) + \pi\big(\phi^2 g^2\big)\Big\} \\ &\leq \frac{4}{N}\frac{\pi\big(g^2\big)}{\pi(g)^2},\end{aligned}$$

and the proof is complete. □

Remark 4.6.3. *The constant* 12 *for the bias can be somewhat reduced by using in the proof the indicator* $1_{\{a\pi^N_{\mathrm{MC}}(g)\leq\pi(g)\}}$ *instead of* $1_{\{2\pi^N_{\mathrm{MC}}(g)\leq\pi(g)\}}$ *and optimizing over* $a > 0$. *Doing this yields the constant* $C \approx 10.42$ *rather than* $C = 12$.

4.6.2.2 Proof of Theorem 4.2.3

The proof of the MSE part of Theorem 4.2.3 uses the approach of [Doukhan and Lang, 2009] for calculating moments of ratios of estimators. The proof of the bias part is very similar to the proof of the bias part of Theorem 4.2.1.

In order to estimate the MSE, we use [Doukhan and Lang, 2009, Lemma 2] which in our setting becomes:

Lemma 4.6.4. *For $0 < \theta < 1$, it holds*

$$\big|\mu^N(\phi) - \mu(\phi)\big| \leq \frac{|\pi^N_{\mathrm{MC}}(\phi g) - \pi(\phi g)|}{\pi(g)} + \frac{|\pi^N_{\mathrm{MC}}(\phi g)|}{\pi(g)^2}|\pi^N_{\mathrm{MC}}(g) - \pi(g)| + \max_{1\leq n\leq N}|\phi(u^n)|\frac{|\pi^N_{\mathrm{MC}}(g) - \pi(g)|^{1+\theta}}{\pi(g)^{1+\theta}}.$$

The main novelty of the above lemma compared to the bounds we used in the proof of Theorem 4.2.1, is not the bound on ϕ using the maximum, but rather the introduction of $\theta \in (0, 1)$. This will be apparent in the proof of Theorem 4.2.3 below.

We also repeatedly use Hölder's inequality in the form

$$\mathbb{E}\big[|uv|^s\big] \leq \mathbb{E}\big[|u|^{sa}\big]^{\frac{1}{a}}\mathbb{E}\big[|v|^{sb}\big]^{\frac{1}{b}},$$

for any $s > 0$ and for $a, b > 1$ such that $\frac{1}{a} + \frac{1}{b} = 1$, as well as the Marcinkiewicz-Zygmund inequality [Ren and Liang, 2001], which for centered i.i.d. random variables X_n gives

$$\mathbb{E}\left[\Big|\sum_{n=1}^N X_n\Big|^t\right] \leq C_t N^{\frac{t}{2}}\mathbb{E}\big[|X_1|^t\big], \quad \forall t \geq 2.$$

There are known bounds on the constants, namely $C_t^{\frac{1}{t}} \leq t - 1$, [Ren and Liang, 2001]. We apply this inequality in several occasions with $X_n = h(u^n) - \pi(h)$ for different functions h, in which case we get

$$\mathbb{E}\Big[\big|\pi^N_{\mathrm{MC}}(h) - \pi(h)\big|^t\Big] \leq C_t\mathbb{E}\Big[\big|h(u^1) - \pi(h)\big|^t\Big]N^{-\frac{t}{2}}, \quad \forall t \geq 2. \tag{4.6.2}$$

We are now ready to prove Theorem 4.2.3.

Proof of Theorem 4.2.3. We first prove the MSE part. By Lemma 4.6.4 we have that

$$\mathbb{E}\Big[\big(\mu^N(\phi) - \mu(\phi)\big)^2\Big] \leq 3A_1 + 3A_2 + 3A_3,$$

where A_1, A_2, A_3 correspond to the second moments of the three terms respectively.

1. For the first term we have

$$A_1 = \frac{1}{\pi(g)^2}\mathbb{E}\left[\Big(\pi^N_{\mathrm{MC}}(\phi g) - \pi(\phi g)\Big)^2\right] \leq \frac{1}{\pi(g)^2}\mathbb{E}\left[\Big(\phi(u^1)g(u^1) - \pi(\phi g)\Big)^2\right]N^{-1}.$$

2. For the second term, Hölder's inequality gives

$$\begin{aligned} A_2 &= \frac{1}{\pi(g)^4}\mathbb{E}\Big[\big|\pi^N_{\mathrm{MC}}(\phi g)\big(\pi^N_{\mathrm{MC}}(g)-\pi(g)\big)\big|^2\Big] \\ &\leq \frac{1}{\pi(g)^4}\mathbb{E}\Big[\big|\pi^N_{\mathrm{MC}}(\phi g)\big|^{2d}\Big]^{\frac{1}{d}}\mathbb{E}\Big[\big|\pi^N_{\mathrm{MC}}(g)-\pi(g)\big|^{2e}\Big]^{\frac{1}{e}}, \end{aligned}$$

where $\frac{1}{d}+\frac{1}{e}=1$. Use of the triangle inequality yields

$$\begin{aligned} \mathbb{E}\Big[\big|\pi^N_{\mathrm{MC}}(\phi g)\big|^{2d}\Big]^{\frac{1}{d}} &= \frac{1}{N^2}\mathbb{E}\Big[\Big|\sum_{n=1}^N \phi(u^n)g(u^n)\Big|^{2d}\Big]^{\frac{1}{d}} \\ &\leq \pi\big(|\phi g|^{2d}\big)^{\frac{1}{d}}. \end{aligned}$$

Combining with (4.6.2) (note that $t=2e>2$) we get

$$A_2 \leq \frac{1}{\pi(g)^4}\pi\big(|\phi g|^{2d}\big)^{\frac{1}{d}}C_{2e}^{\frac{1}{e}}\mathbb{E}\Big[\big|g(u_1)-\pi(g)\big|^{2e}\Big]^{\frac{1}{e}}N^{-1}.$$

3. By Hölder we have

$$\begin{aligned} A_3 &= \frac{1}{\pi(g)^{2(1+\theta)}}\mathbb{E}\left[\max_{1\leq n\leq N}|\phi(u^n)|^2\big|\pi(g)-\pi^N_{\mathrm{MC}}(g)\big|^{2(1+\theta)}\right] \\ &\leq \frac{1}{\pi(g)^{2(1+\theta)}}\mathbb{E}\left[\max_{1\leq n\leq N}\big|\phi(u^n)\big|^{2p}\right]^{\frac{1}{p}}\mathbb{E}\left[\big|\pi(g)-\pi^N_{\mathrm{MC}}(g)\big|^{2q(1+\theta)}\right]^{\frac{1}{q}}, \end{aligned}$$

where $\frac{1}{p}+\frac{1}{q}=1$. Note that

$$\mathbb{E}\left[\max_{1\leq n\leq N}\big|\phi(u^n)\big|^{2p}\right]^{\frac{1}{p}} \leq \mathbb{E}\left[\sum_{n=1}^N\big|\phi(u^n)\big|^{2p}\right]^{\frac{1}{p}} = N^{\frac{1}{p}}\pi\big(|\phi|^{2p}\big)^{\frac{1}{p}}.$$

Combining with (4.6.2), with $t_\theta = 2q(1+\theta)>2$, we get

$$A_3 \leq \frac{1}{\pi(g)^{2(1+\theta)}}N^{\frac{1}{p}}\pi\big(|\phi|^{2p}\big)^{\frac{1}{p}}C_{t_\theta}^{\frac{1}{q}}\mathbb{E}\Big[\big|g(u^1)-\pi(g)\big|^{t_\theta}\Big]^{\frac{1}{q}}N^{-1-\theta}.$$

Now choosing $\theta=\frac{1}{p}\in(0,1)$ gives the desired order of convergence

$$A_3 \leq \frac{1}{\pi(g)^{2(1+\frac{1}{p})}}\pi(|\phi|^{2p})^{\frac{1}{p}}C_{2q(1+\frac{1}{p})}^{\frac{1}{q}}\mathbb{E}\Big[\big|g-\pi(g)\big|^{2q(1+\frac{1}{p})}\Big]^{\frac{1}{q}}N^{-1}.$$

This completes the proof of the MSE part. For the bias, as in the proof of

Theorem 4.2.1 we have

$$\begin{aligned}
&\Big|\mathbb{E}\big[\mu^N(\phi) - \mu(\phi)\big]\Big| \\
&\quad \leq \frac{2}{\pi(g)^2}\mathbb{E}\Big[\Big|\pi(g) - \pi^N_{\mathrm{MC}}(g)\Big|\Big|\pi^N_{\mathrm{MC}}(\overline{\phi}g) - \pi(\overline{\phi}g)\Big|\Big] + \Big|\mathbb{E}\Big[\big(\mu^N(\phi) - \mu(\phi)\big)1_{\{2\pi^N_{\mathrm{MC}}(g)\leq\pi(g)\}}\Big]\Big|,
\end{aligned}$$

where $\overline{\phi} = \phi - \mu(\phi)$. Using the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}
&\Big|\mathbb{E}\big[\mu^N(\phi) - \mu(\phi)\big]\Big| \\
&\quad \leq \frac{2}{\pi(g)^2}\mathbb{E}\Big[\big|\pi(g) - \pi^N_{\mathrm{MC}}(g)\big|^2\Big]^{\frac{1}{2}}\mathbb{E}\Big[\big|\pi^N_{\mathrm{MC}}(\overline{\phi}g) - \pi(\overline{\phi}g)\big|^2\Big]^{\frac{1}{2}} \\
&\qquad\qquad + \mathbb{E}\Big[\big(\mu^N(\phi) - \mu(\phi)\big)^2\Big]^{\frac{1}{2}}\mathbb{P}\Big(2\pi^N_{\mathrm{MC}}(g) \leq \pi(g)\Big)^{\frac{1}{2}} \\
&\quad \leq \frac{2}{\pi(g)^2}\frac{1}{N}\mathbb{E}\Big[\big|g(u^1) - \pi(g)\big|^2\Big]^{\frac{1}{2}}\mathbb{E}\Big[\big|\overline{\phi}(u^1)g(u^1) - \pi(\overline{\phi}g)\big|^2\Big]^{\frac{1}{2}} + \frac{C_{\mathrm{MSE}}^{\frac{1}{2}}}{N^{\frac{1}{2}}}\frac{2}{N^{\frac{1}{2}}}\frac{\pi(g^2)^{\frac{1}{2}}}{\pi(g)},
\end{aligned}$$

where to bound the probability of $2\pi^N_{\mathrm{MC}}(g) \leq \pi(g)$ we use the Markov inequality similarly as in the analogous part of the proof of Theorem 4.2.1. □

4.6.3 Proofs Section 4.3

We next state a lemma collecting several useful properties of the trace of linear operators. A compact linear operator T is said to belong in the trace class family, if its singular values $\{\sigma_i\}_{i=1}^{\infty}$ are summable. In this case we write $\mathrm{Tr}(T) = \sum_{i=1}^{\infty}\sigma_i$, while for notational convenience we define the trace even for non-trace class operators, with infinite value. T is said to belong in the Hilbert-Schmidt family, if its singular values are square summable (equivalently if T^*T is Hilbert-Schmidt).

Lemma 4.6.5. *Let T be an operator on a Hilbert space $\mathcal{H}$. Suppose for the next three items that T is trace class. Then*

i) $\mathrm{Tr}(T^*) = \overline{\mathrm{Tr}(T)}$. *In particular, if the eigenvalues of T are real then* $\mathrm{Tr}(T^*) = \mathrm{Tr}(T)$;

ii) for any bounded operator B in $\mathcal{H}$, $\mathrm{Tr}(TB) = \mathrm{Tr}(BT)$. *This assertion also holds if T and B are Hilbert-Schmidt;*

iii) for any bounded operator B in $\mathcal{H}$, $\mathrm{Tr}(TB) = \mathrm{Tr}(BT) \leq \|B\|\,\mathrm{Tr}(T)$.

For any bounded linear operator T, it holds that

iv) $\mathrm{Tr}(T^*T) = \mathrm{Tr}(TT^*)$,

where if T (equivalently T^*) is not Hilbert-Schmidt, we define the trace to be $+\infty$.

If T is a linear operator and P is bounded and positive definite, such that TP^{-1} (equivalently $P^{-\frac{1}{2}}TP^{-\frac{1}{2}}$ or $P^{-1}T$) is bounded, it holds that

v) $\mathrm{Tr}(TP) = \mathrm{Tr}(P^{\frac{1}{2}}TP^{\frac{1}{2}}) = \mathrm{Tr}(PT)$,

where as in (iv) we allow infinite values of the trace.

Finally, suppose that D_1 is positive definite and D_2 is positive semi definite, and that T is self adjoint and bounded in $\mathcal{H}$. Furthermore, assume that $D_1^{-1}T$ and $(D_1 + D_2)^{-1}T$ have eigenvalues. Then

vi) $\mathrm{Tr}(D_1^{-1}T) \geq \mathrm{Tr}\big((D_1 + D_2)^{-1}T\big)$.

Proof. The proofs of parts (i)-(iii) can be found in [Lax, 2002, Section 30.2], while (iv) is an exercise in [Lax, 2002, Section 30.8]. Part (v) can be shown using the infinite-dimensional analogue of matrix similarity, see [Apostol et al., 1982, Section 2]. In particular, if we multiply TP to the left by $P^{1/2}$ and to the right by $P^{-1/2}$, we do not change its eigenvalues hence neither its trace, so $\mathrm{Tr}(TP) = \mathrm{Tr}(P^{\frac{1}{2}}TP^{\frac{1}{2}})$. Similarly, if we multiply TP to the left by P and to the right by P^{-1}, we get $\mathrm{Tr}(TP) = \mathrm{Tr}(PT)$. Part (vi) follows from the stronger fact that the ordered eigenvalues of $D_1^{-1}T$ are one by one bounded by the ordered eigenvalues of $(D_1+D_2)^{-1}T$. This in turn can be established using that the eigenvalues of these operators are determined by the generalized eigenvalue problem $Tv = \lambda D_1 v$ and $Tv = \lambda(D_1 + D_2)v$, with associated Rayleigh quotients

$$\frac{\langle x, Tx\rangle}{\langle x, D_1 x\rangle} \geq \frac{\langle x, Tx\rangle}{\langle x, (D_1 + D_2)x\rangle}, \tag{4.6.3}$$

and an application of the Rayleigh-Courant-Fisher theorem (see [Lax, 2002] and [Reed and Simon, 1978]). □

4.6.3.1 Proofs of subsection 4.3.2

Proof of Proposition 4.3.4. Under the given assumptions, expression (4.3.4) for C^{-1} is well-defined and gives

$$\Sigma^{\frac{1}{2}}C^{-1}\Sigma^{\frac{1}{2}} = I + A. \tag{4.6.4}$$

Thus

$$\begin{aligned}\mathrm{Tr}(A) &= \mathrm{Tr}(C^{\frac{1}{2}}C^{-1}\Sigma^{\frac{1}{2}} - I)\\ &= \mathrm{Tr}\big(C^{\frac{1}{2}}(C^{-1} - \Sigma^{-1})\Sigma^{\frac{1}{2}}\big)\\ &= \mathrm{Tr}\big((C^{-1} - \Sigma^{-1})\Sigma\big),\end{aligned}$$

where the last equality is justified using the cyclic property of the trace, Lemma 4.6.5(ii). For the second identity, since $(I + A)^{-1}A = I - (I + A)^{-1}$, we have again by (4.6.4)

$$\begin{aligned}\mathrm{Tr}((I + A)^{-1}A) &= \mathrm{Tr}\Big(I - (I + A)^{-1}\Big)\\ &= \mathrm{Tr}\Big(I - \Sigma^{-1/2}C\Sigma^{-1/2}\Big)\\ &= \mathrm{Tr}\Big(\Sigma^{-1/2}(\Sigma - C)\Sigma^{-1/2}\Big)\\ &= \mathrm{Tr}\Big((\Sigma - C)\Sigma^{-1}\Big),\end{aligned}$$

where the last equality is again justified via the cyclic property of the trace. □

Remark 4.6.6. *Proposition 4.3.4 also holds in the general separable Hilbert space setting, provided that formula* (4.3.4) *for the precision operator of the posterior is justified, see [Agapiou et al., 2013, Section 5]. Indeed, the proofs of the two identities are almost identical to the finite dimensional case, the only difference being in the justification of the last equalities in the two sequences of equalities above. In this case the two trace-commutativity equalities have to be justified using Lemma 4.6.5(v) rather than Lemma 4.6.5(ii). In the first case, Lemma 4.6.5(v) can be applied, since* $A = \Sigma^{\frac{1}{2}}(C^{-1} - \Sigma^{-1})\Sigma^{\frac{1}{2}}$ *is bounded by Assumption 4.3.3, and* Σ *is assumed to be positive definite and bounded. In the second case, Lemma 4.6.5(v) can be applied, since by Assumption 4.3.3 the operator* $(I + A)^{-1}A$ *is bounded, and* Σ *is bounded and positive definite.*

Proof of Proposition 4.3.5.

1. We have that (v_i, μ_i) is an eigenvector/value pair of the first matrix if and only if $(\Gamma^{-1/2}v_i, \mu_i)$ is of the second. It is also immediate that (v_i, μ_i) is a pair for the second if and only if (S^*v_i, μ_i) is for $A(I + A)^{-1}$. However, it is also easy to check that $A(I + A)^{-1} = (I + A)^{-1}A$.

2. In view of the above, note that (v_i, μ_i) is a pair for $(I + A)^{-1}A$ if an only if $(v_i, \mu_i/(1 - \mu_i))$ is for A. Hence, if λ_i is an eigenvalue of A, $\lambda_i/(1 + \lambda_i)$ is one for the other matrices. Given that this is always less or equal to 1 and

the efd is a trace of either $d_y \times d_y$ or $d_u \times d_u$ matrices, the inequality follows immediately.

□

Proof of Lemma 4.3.6. If A is trace class then it is compact and since it is also self-adjoint and nonnegative it can be shown (for example using the spectral representation of A) that $\left\|(I+A)^{-1}\right\| \leq 1$. Then Lemma 4.6.5(iii) implies that

$$\mathrm{Tr}\big((I+A)^{-1}A\big) \leq \mathrm{Tr}(A).$$

Assume now that $(I+A)^{-1}A$ is trace class. Then A is too since it is the product of the bounded operator $I+A$ and the trace class operator $(I+A)^{-1}A$, see again Lemma 4.6.5(iii). In particular,

$$\mathrm{Tr}(A) \leq \|I+A\|\mathrm{Tr}\big((I+A)^{-1}A\big).$$

□

4.6.3.2 Proofs of subsection 4.3.3

Proof of Theorem 4.3.7. i) $\Leftrightarrow$ *ii)* is immediate from Lemma 4.3.6.
ii) $\Leftrightarrow$ *iii)* It holds that $\Gamma^{-\frac{1}{2}}Ku \sim N(0, \Gamma^{-\frac{1}{2}}K\Sigma K^*\Gamma^{-\frac{1}{2}})$ since $\Gamma^{-\frac{1}{2}}Ku$ is a linear transformation of the Gaussian $u \sim \mathbb{P}_u = N(0,\Sigma)$. By Lemma 4.6.2 and since A has eigenvalues, we hence have that $\Gamma^{-\frac{1}{2}}Ku \in \mathcal{H}$ if and only if $\mathrm{Tr}(\Gamma^{-\frac{1}{2}}K\Sigma K^*\Gamma^{-\frac{1}{2}}) < \infty$.

iii) $\Rightarrow$ *iv)* According to the discussion in Subsection 4.6.1 on the absolute continuity of two Gaussian measures with the same covariance but different means, the Gaussian likelihood measure $\mathbb{P}_{y|u} = N(Ku, \Gamma)$ and the Gaussian noise measure $\mathbb{P}_\eta = N(0,\Gamma)$ are equivalent if and only if $\Gamma^{-\frac{1}{2}}Ku \in \mathcal{H}$. Under *iii)*, we hence have that $\mathbb{P}_{y|u}$ and $\mathbb{P}_\eta$ are equivalent for π-almost all u and under the Cameron-Martin formula [Da Prato and Zabczyk, 1992] for π-almost all u we have

$$\frac{d\mathbb{P}_{y|u}}{d\mathbb{P}_\eta}(y) = \exp\left(-\frac{1}{2}\left\|\Gamma^{-1/2}Ku\right\|^2 + \left\langle\Gamma^{-1/2}y, \Gamma^{-1/2}Ku\right\rangle\right) =: g(u;y).$$

Defining the measure $\nu_0(u,y) := \pi(u) \times \mathbb{P}_\eta(y)$ in $\mathcal{X} \times \mathcal{Y}$, we then immediately have that

$$\frac{d\nu}{d\nu_0}(u,y) = g(u;y),$$

where ν is the joint distribution of (u,y) under the model $y = Ku + \eta$ with u and

η independent Gaussians $N(0,\Sigma)$ and $N(0,\Gamma)$ respectively.

We next show that $\pi(g(\cdot;y)) > 0$ for $\mathbb{P}_\eta$-almost all y, which will in turn enable us to use a standard conditioning result to get that the posterior is well defined and absolutely continuous with respect to the prior. Indeed, it suffices to show that $g(u;y) > 0$ ν_0-almost surely. Fix $u \sim \pi$. Then, as a function of $y \sim \mathbb{P}_\eta$ the negative exponent of g is distributed as $N(\frac{1}{2}\|\Gamma^{-\frac{1}{2}}Ku\|^2, \|\Gamma^{-\frac{1}{2}}Ku\|^2)$ where $\|\Gamma^{-\frac{1}{2}}Ku\|^2 < \infty$ with π probability 1. Therefore, for ν_0-almost all (u,y) the exponent is finite and thus g is ν_0-almost surely positive implying that $\pi(g(\cdot;y)) > 0$ for $\mathbb{P}_\eta$-almost all y. Noticing that the equivalence of ν and ν_0 implies the equivalence of the marginal distribution of the data under the model, ν_y, with the noise distribution $\mathbb{P}_\eta$, we get that $\pi(g(\cdot;y)) > 0$ for ν_y-almost all y. Hence, we can apply Lemma 5.3 of [Hairer et al., 2007], to get that the posterior measure $\mathbb{P}_{u|y}(\cdot) = \nu(\cdot|y)$ exists ν_y-almost surely and is given by

$$\frac{d\mu}{d\pi}(u) = \frac{1}{\pi(g)} \exp\left(-\frac{1}{2\gamma}\left\|\Gamma^{-1/2}Ku\right\|^2 + \frac{1}{\gamma}\langle\Gamma^{-1/2}y, \Gamma^{-1/2}Ku\rangle\right).$$

Finally, we note that since $\frac{d\nu}{d\nu_0} = g$, we have that $\int_{\mathcal{X}\times\mathcal{Y}} g\, d\nu_0(u,y) = 1$. Thus the Fubini-Tonelli theorem implies that $\pi(g(\cdot;y)) < \infty$ for $\mathbb{P}_\eta$-almost all y and hence also for ν_y-almost all y.

iv) $\Rightarrow$ *ii)* Under *iv)* we have that the posterior measure μ which, as discussed in Subsection 4.3.1, is Gaussian with mean and covariance given by (4.3.2) and (4.3.3), is y-almost surely absolutely continuous with respect to the prior $\pi = N(0,\Sigma)$. By the Feldman-Hajek theorem [Da Prato and Zabczyk, 1992], we hence have that y-almost surely the posterior mean lives in the common Cameron-Martin space of the two measures. This common Cameron-Martin space is the image space of $\Sigma^{\frac{1}{2}}$ in $\mathcal{H}$. Thus we deduce that $w := \Sigma^{-\frac{1}{2}}\Sigma K^*(K\Sigma K^* + \Gamma)^{-1}y \in \mathcal{H}$ almost surely. We next observe that, under ν, $\Gamma^{-\frac{1}{2}}y \sim N(0, SS^* + I)$. Furthermore

$$w = S^*(SS^* + I)^{-1}\Gamma^{-\frac{1}{2}}y,$$

thus under ν, $w \sim N(0, S^*(SS^* + I)^{-1}S)$ where S is defined in Assumption 4.3.3. Using Lemma 4.6.2, we thus get that *iv)* implies that $S^*(SS^* + I)^{-1}S$ is trace class. Using Lemma 4.6.5(iv) with $T = (SS^* + I)^{-\frac{1}{2}}S$, we then also get that $(SS^* + I)^{-\frac{1}{2}}SS^*(SS^* + I)^{-\frac{1}{2}}$ is trace class. Since $(SS^* + I)^{\frac{1}{2}}$ is bounded, using Lemma 4.6.5(iii) twice we get that SS^* is trace class. Finally, again using Lemma 4.6.5(iv) we get that S^*S is trace class, thus ii) holds. □

4.6.3.3 Proofs of subsection 4.3.4

The scalings of τ and efd can be readily deduced by comparing the sums defining τ and efd with integrals:

$$\tau(\beta,\gamma,d) \approx \frac{1}{\gamma}\int_1^d \frac{1}{x^\beta}\,dx, \qquad \mathsf{efd} \approx \int_1^d \frac{1}{1+\gamma x^\beta} = \gamma^{-1/\beta}\int_\gamma^{d\gamma^{1/\beta}} \frac{1}{1+y^\beta}\,dy.$$

Our analysis of the sensitivity of $\rho = \rho(\beta,\gamma,d)$ to the model parameters relies in the following expression for ρ, which is valid unless the effective dimension is infinite, i.e. unless $d=\infty$, $\beta \le 1$.

In the next result, and in the analysis that follows, we ease the notation by using subscripts to denote the coordinate of a vector. Thus we write, for instance, y_j rather than $y(j)$.

Lemma 4.6.7. *Under Assumption 4.3.9*

$$\rho = \rho(\beta,\gamma,d) := \prod_{j=1}^{d} \frac{\frac{j^{-\beta}}{\gamma}+1}{\sqrt{2\frac{j^{-\beta}}{\gamma}+1}} \exp\left(\sum_{j=1}^{d}\left(\frac{2}{2+\gamma j^\beta} - \frac{1}{1+\gamma j^\beta}\right)\frac{y_j^2}{\gamma}\right), \tag{4.6.5}$$

which is finite for ν_y-almost all y.

Proof of Lemma 4.6.7. We rewrite the expectation with respect to π as an expectation with respect to the law of Ku as follows. Note that here u_j is a dummy integration variable, which represents the j-th corrdinate of Ku, rather than that

of u. Precisely, we have:

$$\begin{aligned}
\pi\big(g(\cdot,y)\big) &= \int_{\mathcal{X}} g(u,y)d\pi(u) \\
&= \int_{\mathcal{X}} \exp\left(-\frac{1}{2\gamma}\sum_{j=1}^{\infty}u_j^2 + \frac{1}{\gamma}\sum_{j=1}^{d}y_ju_j\right) d\left(\bigotimes_{j=1}^{d} N\left(0, j^{-\beta}\right)(u_j)\right) \\
&= \prod_{j=1}^{d}\int_{\mathbb{R}} \exp\left(-\frac{1}{2\gamma}u_j^2 + \frac{1}{\gamma}y_ju_j\right)\frac{\exp\left(-\frac{j^\beta u_j^2}{2}\right)}{\sqrt{2\pi j^{-\beta}}}du_j \\
&= \prod_{j=1}^{d}\frac{1}{\sqrt{2\pi j^{-\beta}}}\int_{\mathbb{R}} \exp\left(-(\gamma^{-1}+j^\beta)\frac{u_j^2}{2} + \frac{1}{\gamma}y_ju_j\right)du_j \\
&= \prod_{j=1}^{d}\frac{\exp\left(\frac{\gamma^{-2}y_j^2}{2(\gamma^{-1}+j^\beta)}\right)}{\sqrt{2\pi j^{-\beta}}}\int_{\mathbb{R}} \exp\left(-(\gamma^{-1}+j^\beta)\frac{\left(u_j - \frac{\gamma^{-1}y_j}{\gamma^{-1}+j^\beta}\right)^2}{2}\right)du_j \\
&= \prod_{j=1}^{d}\sqrt{\frac{j^\beta}{\gamma^{-1}+j^\beta}}\exp\left(\frac{\gamma^{-2}y_j^2}{2(\gamma^{-1}+j^\beta)}\right) \\
&= \prod_{j=1}^{d}\sqrt{\frac{\gamma j^\beta}{1+\gamma j^\beta}}\exp\left(\frac{\gamma^{-1}y_j^2}{2(1+\gamma j^\beta)}\right).
\end{aligned}$$

Thus,

$$\pi\big(g(\cdot,y)\big)^2 = \prod_{j=1}^{d}\frac{\gamma j^\beta}{1+\gamma j^\beta}\exp\left(\frac{\gamma^{-1}y_j^2}{1+\gamma j^\beta}\right)$$

and

$$\pi\big(g(\cdot,y)^2\big) = \prod_{j=1}^{d}\sqrt{\frac{\gamma j^\beta}{2+\gamma j^\beta}}\exp\left(\frac{2\gamma^{-1}y_j^2}{2+\gamma j^\beta}\right),$$

Taking the corresponding ratio gives the expression for ρ. □

Analysis of scalings of ρ. Here we show how to obtain the scalings in Table 4.1. Taking logarithms in (4.6.5)

$$\log(\rho) = \sum_{j=1}^{d}\log\left(\frac{\frac{j^{-\beta}}{\gamma}+1}{\sqrt{2\frac{j^{-\beta}}{\gamma}+1}}\right) + \sum_{j=1}^{d}\left(\frac{2}{2+\gamma j^\beta} - \frac{1}{1+\gamma j^\beta}\right)\gamma^{-1}y_j^2. \tag{4.6.6}$$

Note that every term of both sums is positive. In the small noise regimes the first sum dominates, whereas in the large d, $\beta \searrow 1$ the second does. We show here how to find the scaling of $\gamma \to 0$ when $d = \infty$.

We have that

$$\log(\rho) \geq \sum_{j=1}^{\infty} \log\left(\frac{\frac{j^{-\beta}}{\gamma}+1}{\sqrt{2\frac{j^{-\beta}}{\gamma}+1}}\right)$$
$$\approx \int_1^{f(\gamma)} \log\left(\frac{\frac{x^{-\beta}}{\gamma}+1}{\sqrt{2\frac{x^{-\beta}}{\gamma}+1}}\right) dx + \int_{f(\gamma)}^{\infty} \log\left(\frac{\frac{x^{-\beta}}{\gamma}+1}{\sqrt{2\frac{x^{-\beta}}{\gamma}+1}}\right) dx$$

where $f(\gamma)$ is a function of γ that we are free to choose. Choosing $f(\gamma) = \gamma^{-1/\beta-\epsilon}$ (ϵ small) the first integral dominates the second one and, for small γ, $\log(\rho) \geq \gamma^{-1/\beta-\epsilon}\log(\gamma^{-\epsilon\beta/2})$ from where the result in Table 4.1 follows. The joint large d, small γ scalings can be established similarly.

When the second sum in (4.6.6) dominates, the scalings hold in probability. To illustrate this, we study here how to derive the large d limit with $\beta < 1$. Without loss of generality we can assume in what follows that each y_j is centered, i.e. $y_j \sim N(0,\gamma)$ instead of $y_j \sim N\big((Ku)_j^{\dagger},\gamma\big)$. This is justified since, for any $c > 0$,

$$\mathbb{P}(y_j^2 \geq c) = \mathbb{P}(|y_j| \geq c^{1/2}) \geq \mathbb{P}(|y_j - (Ku)_j^{\dagger}| \geq c^{1/2}).$$

Neglecting the first sum in (4.6.6), which can be shown to be of lower order in d, we get

$$\sum_{j=1}^{d}\left(\frac{2}{2+\gamma j^{\beta}} - \frac{1}{1+\gamma j^{\beta}}\right)\gamma^{-1}y_j^2 = S(y,d).$$

Using that $\mathbb{E}y_j^2 = \gamma$,

$$\mathbb{E}\log(\rho) \geq \sum_{j=1}^{d}\left(\frac{2}{2+\gamma j^{\beta}} - \frac{1}{1+\gamma j^{\beta}}\right)$$
$$\approx \int_1^{d}\left(\frac{2}{2+\gamma x^{\beta}} - \frac{1}{1+\gamma x^{\beta}}\right) dx \approx d^{1-\beta} =: m(d).$$

Also, since $\mathrm{Var}(y_j^2) = 3\gamma^2$,

$$\mathrm{Var}\log(\rho) \geq \sum_{j=1}^{d}\left(\frac{2}{2+\gamma j^{\beta}} - \frac{1}{1+\gamma j^{\beta}}\right)^2\gamma^2$$
$$\approx \int_1^{d}\left(\frac{2}{2+\gamma x^{\beta}} - \frac{1}{1+\gamma x^{\beta}}\right)^2 dx \approx d^{1-2\beta} =: c(d).$$

Thus we have

$$\begin{aligned}
\mathbb{P}\Big(\log(\rho) \geq m(d)/2\Big) &\geq \mathbb{P}\Big(S(y,d) \geq m(d)/2\Big) \\
&\geq \mathbb{P}\Big(S(y,d) \geq \mathbb{E}S(y,d)/2\Big) \\
&\geq \mathbb{P}\Big(|S(y,d) - \mathbb{E}S(y,d)| \leq \mathbb{E}S(y,d)/2\Big) \\
&= 1 - \mathbb{P}\Big(|S(y,d) - \mathbb{E}S(y,d)| \geq \mathbb{E}S(y,d)/2\Big) \\
&\geq 1 - \mathbb{P}\Big(|S(y,d) - \mathbb{E}S(y,d)| \geq m(d)/2\Big) \\
&\geq 1 - 4\frac{c(d)}{m(d)^2} \to 1.
\end{aligned}$$

□

4.6.4 Proofs Section 4.4

The following lemma will be used in the proof of Theorem 4.4.5. It justifies the use of the cyclic property in calculating certain traces in the infinite dimensional setting.

Lemma 4.6.8. *Suppose that $A = S^*S$, where $S = \Gamma^{-1/2}K\Sigma^{1/2}$ as in Assumption 4.3.3 is bounded. Then*

$$\tau = \mathrm{Tr}(A) = \mathrm{Tr}(\Gamma^{-1}K\Sigma K^*).$$

Therefore, using the equivalence in Table 4.2 we have that τ_{st} and τ_{op} admit the following equivalent expressions:

$$\tau_{st} = \mathrm{Tr}\big(R^{-1}H(MPM^* + Q)H^*\big) \tag{4.6.7}$$

and

$$\tau_{op} = \mathrm{Tr}\big((R + HQH^*)^{-1}HMPM^*H^*\big). \tag{4.6.8}$$

Proof. Using Lemma 4.6.5(iv) we have that $\tau = \mathrm{Tr}(S^*S) = \mathrm{Tr}(SS^*)$. Now note that $SS^* = \Gamma^{-1/2}K\Sigma K^*\Gamma^{-1/2}$ is bounded since A is, and that $\Gamma^{1/2}$ is also bounded, hence we can use Lemma 4.6.5(v) to get the desired result. □

Proof of Theorem 4.4.5. Using the previous lemma,

$$\begin{aligned}\tau_{st} &= \mathrm{Tr}\Big(R^{-1}HMPM^*H^*\Big) + \mathrm{Tr}\Big(R^{-1}HQH^*\Big)\\ &\geq \mathrm{Tr}\Big(R^{-1}HMPM^*H^*\Big)\\ &\geq \mathrm{Tr}\Big((R+HQH^*)^{-1}HMPM^*H^*\Big) = \tau_{op},\end{aligned}$$

where the first inequality holds because R is positive-definite and HQH^* is positive semi definite, and the second one follows from Lemma 4.6.5(vi).

If $\mathrm{Tr}(HQH^*R^{-1}) < \infty$ then there is $c > 0$ such that, for all x, $\|HQH^*x\| \leq c\|Rx\|$. Hence applying again Lemma 4.6.5(vi) for both directions of the equivalence, we obtain that

$$\begin{aligned}\tau_{op} = \mathrm{Tr}\Big((R+HQH^*)^{-1}HMPM^*H^*\Big) < \infty &\iff \mathrm{Tr}\Big(R^{-1}HMPM^*H^*\Big) < \infty\\ &\iff \tau_{st} < \infty.\end{aligned}$$

□

Bibliography

H. D. I. Abarbanel. *Predicting the Future: Completing Models of Observed Complex Systems*. Springer. Series: Understanding Complex Systems, 2013.

K. Achutegui, D. Crisan, J. Miguez, and G. Rios. A simple scheme for the parallelization of particle filters and its application to the tracking of complex stochastic systems. *arXiv preprint arXiv:1407.8071*, 2014.

M. Ades and P. J. Van Leeuwen. An exploration of the equivalent weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, 139(672):820–840, 2013.

R. J. Adler. An introduction to continuity, extrema, and related topics for general Gaussian processes. *Lecture Notes-Monograph Series*, 1990.

S. Agapiou and P. Mathé. Preconditioning the prior to overcome saturation in Bayesian inverse problems. *arXiv preprint arXiv:1409.6496*, 2014.

S. Agapiou, S. Larsson, and A. M. Stuart. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Processes and their Applications*, 123(10):3828–3860, 2013.

S. Agapiou, A. M. Stuart, and Y-X Zhang. Bayesian posterior contraction rates for linear severely ill-posed inverse problems. *Journal of Inverse and Ill-posed Problems*, 22(3):297–321, 2014.

S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. M. Stuart. Importance sampling: Computational complexity and intrinsic dimension. *arXiv preprint arXiv:1511.06196*, 2015.

C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 72(3):269–342, 2010.

C. Apostol, D. A. Herrero, and D. Voiculescu. The closure of the similarity orbit of a Hilbert space operator. *Bulletin of the American Mathematical Society*, 6(3): 421–426, 1982.

A. Azouani, E. Olson, and E. S. Titi. Continuous data assimilation using general interpolant observables. *Journal of Nonlinear Science*, 24:277–304, 2014.

A. Bain and D. Crisan. *Fundamentals of Stochastic Filtering*, volume 3. Springer, 2009.

H. T. Banks and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems*. Springer Science & Business Media, 2012.

C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers I*. Springer Science & Business Media, 1999.

G. Benettin, L. Galgani, and J.M. Strelcyn. Kolmogorov entropy and numerical experiments. *Physical Review A*, 14:2338–2345, Dec 1976. doi: 10.1103/PhysRevA.14.2338.

T. Bengtsson, P. Bickel, B. Li, et al. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In *Probability and statistics: Essays in honor of David A. Freedman*, pages 316–334. Institute of Mathematical Statistics, 2008.

A. Bennett. *Inverse Modeling of the Ocean and Atmosphere*. Cambridge University Press, 2003.

J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. Wiley Series in Probability and Statistics, 1994.

A. Beskos, D. Crisan, and A. Jasra. On the stability of sequential Monte Carlo methods in high dimensions. *The Annals of Applied Probability*, 24(4):1396–1445, 2014a.

A. Beskos, D. Crisan, A. Jasra, K. Kamatani, and Y. Zhou. A stable particle filter in high-dimensions. *arXiv preprint arXiv:1412.3501*, 2014b.

A. Beskos, A. Jasra, E. A. Muzaffer, and A. M. Stuart. Sequential Monte Carlo methods for Bayesian elliptic inverse problems. *Statistics and Computing*, 25(4): 727–737, 2015.

P. Bickel, B. Li, T. Bengtsson, et al. Sharp failure rates for the bootstrap particle filter in high dimensions. In *Pushing the limits of contemporary statistics: Contributions in honor of Jayanta K. Ghosh*, pages 318–329. Institute of Mathematical Statistics, 2008.

C. M. Bishop. *Pattern recognition and machine learning*. Springer New York, 2006.

D. Bloemker, K. J. H. Law, A. M. Stuart, and K. Zygalakis. Accuracy and stability of the continuous-time 3DVAR filter for the Navier-Stokes equation. *Nonlinearity*, 26:2193–2219, 2013. doi: 10.1088/0951-7715/26/8/2193.

D. Bloemker, K. J. H. Law, A. M. Stuart, and K. C. Zygalakis. Accuracy and stability of the continuous-time 3DVAR filter for the navier-stokes equation. *Nonlinearity*, 2014.

S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. ISBN 978-0-19-953525-5. doi: 10.1093/acprof:oso/9780199535255.001.0001.

C. E. A. Brett, K. F. Lam, K. J. H. Law, D. S. McCormick, M. R. Scott, and A. M. Stuart. Accuracy and stability of filters for dissipative pdes. *Physica D: Nonlinear Phenomena*, 2013.

A. Budhiraja. Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter. In *Annales de l'IHP Probabilités et statistiques*, volume 39, pages 919–941, 2003.

T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler. A computational framework for infinite-dimensional Bayesian inverse problems part i: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, 35(6):A2494–A2523, 2013.

R. E. Caflisch, W. J. Morokoff, and A. B. Owen. *Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension*. Department of Mathematics, University of California, Los Angeles, 1997.

A. Caponnetto and E. De Vito. Optimal rates for the regularized least-squares algorithm. *Foundations of Computational Mathematics*, 7(3):331–368, 2007.

O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models*. Springer Heidelberg, 2009.

F. Cérou. Long time behavior for some dynamical noise free nonlinear filtering problems. *SIAM Journal on Control and Optimization*, 38(4):1086–1101, 2000.

S. Chatterjee and P. Diaconis. The sample size required in importance sampling. *arXiv preprint arXiv:1511.01437*, 2015.

Y. Chen. Another look at rejection sampling through importance sampling. *Statistics & Probability Letters*, 72(4):277–283, 2005.

N. Chopin. A sequential particle filter method for static models. *Biometrika*, 89(3): 539–552, 2002.

N. Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Annals of Statistics*, pages 2385–2411, 2004.

N. Chopin and O. Papaspiliopoulos. *A concise introduction to sequential Monte Carlo*. 2016.

N. Chopin, P. E. Jacob, and O. Papaspiliopoulos. SMC2: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 75(3):397–426, 2013.

A. J. Chorin and M. Morzfeld. Conditions for successful data assimilation. *Journal of Geophysical Research: Atmospheres*, 118(20):11–522, 2013.

I. Chueshov. A squeezing property and its applications to a description of long-time behaviour in the three-dimensional viscous primitive equations. *Proceedings of the Royal Society of Edinburgh, Section A*, 144(04):711–729, 2014.

N. Chustagulprom, S. Reich, and M. Reinhardt. A hybrid ensemble transform filter for high dimensional dynamical systems. *arXiv preprint arXiv:1509.06669*, 2015.

P. Constantin and C. Foias. Navier-Stokes equations. *Chicago Lectures in Mathematics, University of Chicago, Chicago/London*, 1988.

S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, 28 (3):424–446, 2013.

D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. *Signal Processing, IEEE Transactions on*, 50(3):736–746, 2002.

D. Crisan and K. Heine. Stability of the discrete time filter in terms of the tails of noise distributions. *Journal of the London Mathematical Society*, 78(2):441–458, 2008.

D. Crisan and B. Rozovskii. *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2011.

D. Crisan, P. Del Moral, and T. Lyons. Discrete filtering using branching and interacting particle systems. *Université de Toulouse. Laboratoire de Statistique et Probabilités*, 1998.

G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge University Press, 1992.

M. Dashti and A. M. Stuart. Uncertainty quantification and weak approximation of an elliptic inverse problem. *SIAM Journal on Numerical Analysis*, 49:2524–2542, 2011.

M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. *Handbook of Uncertainty Quantification*, 2016.

M. Dashti, S. Harris, and A. M. Stuart. Besov priors for bayesian inverse problems. *Inverse Problems and Imaging*, 6:183–200, 2012.

M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. Map estimators and posterior consistency in bayesian nonparametric inverse problems. *Inverse Problems*, 29: 095017, 2013.

P. Del Moral. *Feynman-Kac Formulae*. Springer, 2004.

P. Del Moral. *Mean Field Simulation for Monte Carlo Integration*. CRC Press, 2013.

P. Del Moral and L. Miclo. *Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering*. Springer, 2000.

F. X. Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, 38(2):97–110, 1986.

R. Douc, E. Moulines, and Y. Ritov. Forgetting of the initial condition for the filter in general state-space hidden markov chain: a coupling approach. *Electronic Journal of Probability*, 14:27–49, 2009.

A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, 12(656-704):3, 2009.

A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

P. Doukhan and G. Lang. Evaluation for moments of a ratio with application to regression estimation. *Bernoulli*, 15(4):1259–1286, 2009.

P. J. Downey and P. E. Wright. The ratio of the extreme to the sum in a random sequence. *Extremes*, 10(4):249–266, 2007. doi: 10.1007/s10687-007-0044-0.

M. M. Dunlop and A. M. Stuart. The bayesian formulation of eit: Analysis and algorithms. *arXiv preprint arXiv:1508.04106*, 2015.

P. Dupuis, K. Spiliopoulos, and H. Wang. Importance sampling for multiscale diffusions. *Multiscale Modeling & Simulation*, 10(1):1–27, 2012.

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

G. Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003.

J. N. Franklin. Well-posed stochastic extensions of ill-posed linear problems. *Journal of Mathematical Analysis and Applications*, 31(3):682–716, 1970.

M. Frei and H. R. Künsch. Bridging the ensemble kalman and particle filters. *Biometrika*, 100(4):781–800, 2013.

R. Furrer and T. Bengtsson. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, 98(2):227–255, 2007.

A. Gelman, G. O. Roberts, and W. Gilks. Efficient Metropolis jumping hules. *Bayesian statistics*, 5(599-608):42, 1996.

R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: a Spectral Approach*. Courier Corporation, 2003.

A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.

J. Goodman, K. K. Lin, and M. Morzfeld. Small-noise analysis and symmetrization of implicit monte carlo samplers. *Communications on Pure and Applied Mathematics*, 2015.

N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *Radar and Signal Processing, IEE Proceedings F*, volume 140, pages 107–113. IET, 1993.

G. A. Gottwald and A. J. Majda. A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks. *Nonlinear Processes in Geophysics*, 20(5):705–712, 2013.

M. Hairer, A. M. Stuart, J. Voss, et al. Analysis of spdes arising in path sampling part ii: the nonlinear case. *The Annals of Applied Probability*, 17(5/6):1657–1706, 2007.

T. M. Hamill and C. Snyder. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Monthly Weather Review*, 128(8):2905–2919, 2000.

W. Han. *On the Numerical Solution of the Filtering Problem*. PhD thesis, Ph. D. Thesis. Department of Mathematics, Imperial College London, 2013.

J. Harlim and A. J. Majda. Catastrophic filter divergence in filtering nonlinear dissipative systems. *Communications in Mathematical Sciences*, 8(1):27–43, 2010.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

K. Hayden, E. Olson, and E. S. Titi. Discrete data assimilation in the Lorenz and 2d Navier-Stokes equations. *Physica D: Nonlinear Phenomena*, 240:1416–1425, 2011.

M. A. Iglesias, K. J. H. Law, and A. M. Stuart. Ensemble Kalman methods for inverse problems. *Inverse Problems*, 29(4):045001, 2013.

M. A. Iglesias, Y. Lu, and A. M. Stuart. A bayesian level set method for geometric inverse problems. *arXiv preprint arXiv:1504.00313*, 2015.

A. H. Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007.

A. M. Johansen and A. Doucet. A note on auxiliary particle filters. *Statistics & Probability Letters*, 78(12):1498–1504, 2008. doi: 10.1016/j.spl.2008.01.032.

S. J. Julier and J. K. Uhlmann. New extension of the Kalman filter to nonlinear systems. In *AeroSense'97*, pages 182–193. International Society for Optics and Photonics, 1997.

H. Kahn. *Use of different Monte Carlo sampling techniques*. Rand Corporation, 1955.

H. Kahn and A. W. Marshall. Methods of reducing sample size in Monte Carlo computations. *Journal of the Operations Research Society of America*, 1(5):263–278, 1953.

J. P. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160. Springer, 2005.

O. Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer Science, 2nd edition edition, 2002.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2003.

N. Kantas, A. Beskos, and A. Jasra. Sequential Monte Carlo methods for high-dimensional inverse problems: a case study for the Navier–Stokes equations. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):464–489, 2014.

H. Kekkonen, M. Lassas, and Siltanen S. Posterior consistency and convergence rates for bayesian inversion with hypoelliptic operators. *arXiv preprint arXiv:1507.01772*, 2015.

D. T. B. Kelly, K. J. H. Law, and A. M. Stuart. Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. *Nonlinearity*, 27, 2014.

D. T. B. Kelly, A. J. Majda, and X. T. Tong. Concrete ensemble kalman filters with rigorous catastrophic filter divergence. *Proceedings of the National Academy of Sciences*, 112(34):10589–10594, 2015.

M. L. Kleptsyna and A. Y. Veretennikov. On discrete time ergodic filters with wrong initial data. *Probability Theory and Related Fields*, 141(3-4):411–444, 2008.

B. T. Knapik, A. W. van Der Vaart, and J. H. van Zanten. Bayesian inverse problems with Gaussian priors. *The Annals of Statistics*, 39(5):2626–2657, 2011.

B. T. Knapik, A. W. van der Vaart, and J. H. van Zanten. Bayesian recovery of the initial condition for the heat equation. *Communications in Statistics-Theory and Methods*, 42(7):1294–1313, 2013.

A. Kong. A note on importance sampling using standardized weights. *University of Chicago, Dept. of Statistics, Tech. Rep*, 348, 1992.

A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288, 1994.

M. Kostuk. *Synchronization and statistical methods for the data assimilation of HVc neuron models*. PhD thesis, University of California, San Diego, 2012.

H. Kunita. Asymptotic behavior of the nonlinear filtering errors of Markov processes. *Journal of Multivariate Analysis*, 1(4):365–393, 1971.

F. Y. Kuo and I. H. Sloan. Lifting the curse of dimensionality. *Notices of the AMS*, 52(11):1320–1328, 2005.

H. J. Kushner and G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*, volume 35. Springer, 2003.

P. Lancaster and L. Rodman. *Algebraic Riccati Equations*. Oxford University Press, 1995.

S. Lasanen. Measurements and infinite-dimensional statistical inverse theory. *Proceedings in Applied Mathematics and Mechanics*, 7(1):1080101–1080102, 2007.

S. Lasanen. Non-Gaussian statistical inverse problems. Part I: Posterior distributions. *Inverse Problems and Imaging*, 6(2):215–266, 2012a.

S. Lasanen. Non-Gaussian statistical inverse problems. Part II: Posterior convergence for approximated unknowns. *Inverse Problems and Imaging*, 6(2):267–287, 2012b.

K. J. H. Law and A. M. Stuart. Evaluating data assimilation algorithms. *Monthly Weather Review*, 140:3757–3782, 2012.

K. J. H. Law, A. Shukla, and A. M. Stuart. Analysis of the 3DVAR Filter for the Partially Observed Lorenz '63 Model. *Discrete and Continuous Dynamical Systems A*, 34:1061–1078, 2014.

K. J. H. Law, A. M. Stuart, and K. C. Zygalakis. *Data Assimilation: A Mathematical Introduction.* Springer, 2015.

K. J. H. Law, D. Sanz-Alonso, A. Shukla, and A. M. Stuart. Filter accuracy for the Lorenz 96 model: fixed versus adaptive observation operators. *Physica D: Nonlinear Phenomena*, 325:1–13, 2016.

P. D. Lax. *Functional analysis.* Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], New York, 2002. ISBN 0-471-55604-1.

M. S. Lehtinen, L. Paivarinta, and E. Somersalo. Linear inverse problems for generalised random variables. *Inverse Problems*, 5(4):599, 1989.

J. Lei and P. Bickel. On convergence of recursive Monte Carlo filters in non-compact state spaces. *Statistica Sinica*, 23:429–450, 2013.

K. Lin, S. Lu, and P. Mathé. Oracle-type posterior contraction rates in Bayesian inverse problems. *Inverse Problems & Imaging*, 9(3), 2015.

D. V. Lindley and A. F. M. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, pages 1–41, 1972.

J. S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119, 1996.

J. S. Liu. *Monte Carlo Strategies in Scientific Computing.* Springer Science & Business Media, 2008.

A. C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986.

E. N. Lorenz and K. A. Emanuel. Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, 55: 399–414, 1998.

S. Lu and P. Mathé. Discrepancy based model selection in statistical inverse problems. *Journal of Complexity*, 30(3):290–308, 2014.

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

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

A. Mandelbaum. Linear estimators and measurable linear transformations on a Hilbert space. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(3):385–397, 1984.

J. Martin, L. C. Wilcox, C. Burstedde, and O. Ghattas. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487, 2012.

D. L. McLeish and G. L. O'Brien. The expected ratio of the sum of squares to the square of the sum. *The Annals of Probability*, 10(4):1019–1028, 1982.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

J. Míguez, D. Crisan, and P. M. Djurić. On the convergence of two sequential Monte Carlo methods for maximum a posteriori sequence estimation and stochastic global optimization. *Statistics and Computing*, 23(1):91–107, 2013.

B. Moskowitz and R. E. Caflisch. Smoothness and dimension reduction in quasi-Monte Carlo methods. *Mathematical and Computer Modelling*, 23(8):37–54, 1996.

J. L. Mueller and S. Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications*, volume 10. SIAM, 2012.

D. Oliver, A. Reynolds, and N. Liu. *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge University Press, 2008.

E. Olson and E. S. Titi. Determining modes for continuous data assimilation in 2d turbulence. *Journal of Statistical Physics*, 113:799–840, 2003.

E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, 56(5):415–428, 2004.

D. F. Parrish and J. C. Derber. The national meteorological center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120(8):1747–1763, 1992.

M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Sssociation*, 94(446):590–599, 1999.

K. Ray. Bayesian inverse problems with non-conjugate priors. *Electronic Journal of Statistics*, 7:2516–2549, 2013.

P. Rebeschini and R. van Handel. Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25(5):2809–2866, 2015.

M. Reed and B. Simon. *Analysis of Operators, Vol. IV of Methods of Modern Mathematical Physics*. New York, Academic Press, 1978.

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

Y-F Ren and H-Y Liang. On the best constant in Marcinkiewicz–Zygmund inequality. *Statistics & Probability Letters*, 53(3):227–233, 2001.

G. R. Richter. An inverse problem for the steady state diffusion equation. *SIAM Journal on Applied Mathematics*, 41(2):210–221, 1981.

J. C. Robinson. *Infinite-dimensional Dynamical Systems: an Introduction to Dissipative Parabolic PDEs and the Theory of Global Attractors*, volume 28. Cambridge University Press, 2001.

W. Rudin. *Real and Complex Analysis*. Tata McGraw-Hill Education, 1987.

D. Sanz-Alonso and A. M. Stuart. Long-time asymptotics of the filtering distribution for partially observed chaotic dynamical systems. *SIAM/ASA Journal on Uncertainty Quantification*, 3:1200–1220, 2015. doi: 10.1137/140997336.

S. Särkkä. *Bayesian Filtering and Smoothing*, volume 3. Cambridge University Press, 2013.

C. Schillings and A. M. Stuart. Analysis of the ensemble Kalman filter for inverse problems. *arXiv preprint arXiv:1602.02020v1*, 2016.

L. Slivinski and C. Snyder. Exploring practical estimates of the ensemble size necessary for particle filters. *Monthly Weather Review*, (2015), 2015.

C. Snyder. Particle filters, the "optimal" proposal and high-dimensional systems. In *Proceedings of the ECMWF Seminar on Data Assimilation for Atmosphere and Ocean*, 2011.

C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136(12):4629–4640, 2008.

C. Snyder, T. Bengtsson, and M. Morzfeld. Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143(11):4750, 2015.

E. Somersalo, M. Cheney, and D. Isaacson. Existence and uniqueness for electrode models for electric current computed tomography. *SIAM Journal on Applied Mathematics*, 52(4):1023–1040, 1992.

E. D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, volume 6. Springer, 1998.

A. Spantini, A. Solonen, T. Cui, J. Martin, L. Tenorio, and Y. Marzouk. Optimal low-rank approximations of Bayesian linear inverse problems. *SIAM Journal on Scientific Computing*, 37(6):A2451–A2487, 2015.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 64(4):583–639, 2002. doi: 10.1111/1467-9868.00353.

K. Spiliopoulos. Large deviations and importance sampling for systems of slow-fast motion. *Applied Mathematics & Optimization*, 67(1):123–161, 2013.

A. S. Stordal, H. A. Karlsen, G. Nævdal, H. J. Skaug, and B. Vallès. Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Computational Geosciences*, 15(2):293–305, 2011.

A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

T. Tarn and Y. Rasis. Observers for nonlinear stochastic systems. *Automatic Control, IEEE Transactions*, 21(4):441–488, 1976.

R. Temam. *Navier-Stokes Equations and Nonlinear Functional Analysis*, volume 66. SIAM, 1995.

R. Temam. *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, volume 68 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1997.

F. E. Thau. Observing the state of non-linear dynamic systems. *International Journal of Control*, 17(3):471–479, 1973.

L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *The Annals of Applied Probability*, 8(1):1–9, 1998.

X. T. Tong and R. Van Handel. Conditional ergodicity in infinite dimension. *The Annals of Probability*, 42(6):2243–2313, 2014.

X. T. Tong, A. J. Majda, and D. T. B. Kelly. Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation. *arXiv preprint arXiv:1507.08319*, 2015.

A. Trevisan and F. Uboldi. Assimilation of standard and targeted observations within the unstable subspace of the observation analysis forecast cycle system. *Journal of the Atmospheric Sciences*, 61(1):103–113, 2004.

X. Tu, M. Morzfeld, J. Wilkening, and A. J. Chorin. Implicit sampling for an elliptic inverse problem in underground hydrodynamics. *arXiv preprint arXiv:1308.4640*, 2013.

R. Van Handel. *Filtering, stability, and robustness*. PhD thesis, California Institute of Technology, 2006.

P. J. van Leeuwen. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653): 1991–1999, 2010.

E. Vanden-Eijnden and J. Weare. Data assimilation in the low noise, accurate observation regime with application to the Kuroshio current. *Monthly Weather Review*, 141:1, 2012.

P. Vidoni. Exponential family state space models based on a conjugate latent process. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61(1):213–221, 1999.

S. J. Vollmer. Posterior consistency for Bayesian inverse problems through stability and regression results. *Inverse Problems*, 29(12):125011, 2013.

N. Whiteley, A. Lee, and K. Heine. On the role of interaction in sequential Monte Carlo algorithms. *Bernoulli*, 22(1):494–529, 2016.

D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.

D. Xiu. *Numerical Methods for Stochastic Computations: a Spectral Method Approach*. Princeton University Press, 2010.

F. Zhang, M. Zhang, and J. A. Hansen. Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Advances in Atmospheric Sciences*, 26 (1):1–8, 2009.

T. Zhang. Effective dimension and generalization of kernel learning. In *Advances in Neural Information Processing Systems*, pages 454–461, 2002.

T. Zhang. Learning bounds for kernel regression using effective data dimensionality. *Neural Computation*, 17(9):2077–2098, 2005.

W. Zhang, C. Hartmann, M. Weber, and C. Schütte. Importance sampling in path space for diffusion processes. *Multiscale Modeling & Simulation*, 2013.