Sequential Data Assimilation in High Dimensional Nonlinear Systems

Andreas S. Stordal

Dissertation for the degree of Philosophiae Doctor (PhD)

Department of Mathematics
University of Bergen

August 2011

Abstract

Since its introduction in 1994, the ensemble Kalman filter (EnKF) has gained a lot of attention as a tool for sequential data assimilation in many scientific areas. Due to its computationally fast and easy to implement algorithm its popularity has increased vastly over the last decades especially in the fields of oceanography and petroleum engineering. Although EnKF has been successfully applied to many real world problems it has a major drawback from a statistical point of view. The algorithm only converge to the optimal solution if the system under consideration is linear and all random variables describing the system are Gaussian.

There exist sequential Monte Carlo methods (SMC) with correct asymptotic properties, but both numerical and theoretical studies have shown that the number of samples must increase exponentially with the dimension of the model to avoid a collapse of the algorithm. For large scale geophysical systems, such as petroleum reservoirs or ocean models, each sample requires the solution of a system of partial differential equations on a large grid. The computational burden of solving these equations using numerical schemes naturally puts an upper limit on the number of samples we can use in practice. Hence these sequential Monte Carlo methods are not applicable, at least in their simplest form, in large scale geophysical models.

This thesis explores the possibility of bridging the gap between EnKF and one of the asymptotically correct SMC methods, known as particle filters, by extending already known theory on Gaussian mixture filters. In addition a sensitivity analysis is carried out for a new type of data in reservoir models.

A new approximative filter is developed by introducing an additional parameter in the standard Gaussian mixture filter. The adaptive Gaussian mixture filter (AGM) consists of two parameters and by choosing these differently the filter may run as EnKF, a particle filter, or something in between. The method is tested on the Lorenz96 model for comparison with EnKF and Gaussian mixture filters. Further comparison with EnKF is made after running AGM on a 2D two-phase and a 3D three-phase petroleum reservoir.

We generalize AGM and compute error bounds and asymptotic properties using classical approximation techniques before we explore the effects of estimating the first and second order moments locally in Kalman type filters. By local estimation we mean local in value and not in space. Two different approaches are suggested and applied to the chaotic Lorenz63 model and a 1D reservoir model. Finally the sensitivity of reservoir parameters to a new type of data, nanosensor observations, are calculated. Both analytical and numerical results are provided. A simulation experiment is included from which we can compute the resolution of the estimated parameters numerically.

Acknowledgements

I would first and foremost like to thank my supervisors (in alphabetic order): Dr. Hans A. Karlsen, Dr. Geir Nævdal, Dr. Dean S. Oliver and Dr. Hans J. Skaug for their excellent guidance and their contributions to both the research content and the formulation of the papers that are included in this thesis. I would also like to thank them for sharing their vast knowledge in statstics, mathematics and reservoir engineering with me over the last three years.

A thanks also to my two other coauthors, Dr. Randi Valestrand and Dr. Brice Vallès, my collegues at IRIS and my fellow PhD students at the University of Bergen.

A special thanks to Dr. Dean Oliver and his PhD students at the University of Oklahoma for many great times and scientific discussion during my time as a visitor from February to August 2010.

During my time as a PhD student I have attended two summer schools on data assimilation in Romania and would like to express my gratitude and thanks to Dr. Remus Hanea, TU Delft and TNO for arranging and organizing them in the best possible way.

I would also like to thank Dr. Chris Farmer and his PhD student Lian Duan at the University of Oxford for their hospitality and all the scientific and non-scientific discussions we had during my stay as a visitor in November 2010.

Finally I would like to thank my family, friends, the IRIS hiking group and my teammates at F.K. Hald for all the joyful times we have had during my time as a PhD student and also for helping me take my mind off the thesis at times.

List of papers

- **Paper A:** Andreas S. Stordal, Hans A. Karlsen, Geir Nævdal, Hans J. Skaug and Brice Vallès, *Bridging the Ensemble Kalman filter and particle filters: The adaptive Gaussian mixture filter*, Computational Geosciences **15**, 293-305, 2011.
- **Paper B:** Andreas S. Stordal, Hans A. Karlsen, Geir Nævdal, Hans J. Skaug and Randi Valestrand, *Comparing the adaptive Gaussian mixture filter with the ensemble Kalman filter on synthetic reservoir models*, resubmitted to Computational Geosciences.
- **Paper C:** Andreas S. Stordal and Hans A. Karlsen, *A generalization of the adaptive Gaussian mixture filter with error bounds and large sample asymptotics*, manuscript.
- **Paper D:** Andreas S.Stordal, Hans A. Karlsen, Geir Nævdal, Dean S. Oliver and Hans J. Skaug, *Filtering with local Kalman gain*, resubmitted to Physica D.
- **Paper E:** Andreas S. Stordal and Dean S. Oliver, *Characterization of permeability and porosity from Nanosensor Observations*, Advances in Water Resources, **34**, 946-956, 2011.

Contents

Chapter 1

Introduction

The main focus in this thesis is on data assimilation in high dimensional dynamic systems. Data assimilation differs a bit from classical filter problems where the underlying model is assumed to be a known stochastic process and the goal is to find the best estimate of the dynamic states of the hidden process at each time where a measurement is taken. The measurements are usually a nonlinear mapping of the states with an additional stochastic error term. In data assimilation for large scale models such as atmospheric models and petroleum reservoirs, the underlying system is deterministic and the goal is to estimate either the states itself, parameters governing the forward equations of the model or both. In this thesis however, we approach the data assimilation problem in the same manner as the filter problem with two exceptions which have become standard in the scientific community [1]. The first is that we include parameters in the unknown state vector, which is usually not the case in filter problems. Secondly we include simulated measurements in the state vector so that the measurement operator in the augmented state space model is a binary linear operator. This however, does not imply a linear relationship between the measurements and the states we want to estimate. Further, we formulate the filter problem in discrete time. We always assume that the measurements are taken at discrete times, as a nonlinear mapping of the state vector at the current time with an additional stochastic error term. Although geophysical models are usually continuous in time, the only computational difference between a discrete model and a continuous model in this set up is the numerical errors when solving the forward equations numerically (if no analytical solution is available) from one measurement time to the next (the transition from $t-1$ to t). All examples in this thesis are synthetic, that is the states and observations are generated from a reference model using the same discretization and numerical solvers as in the filter algorithms, hence we may restrict ourselves to discrete time systems.

Even for simple low dimensional dynamic models the optimal statistical solution to the filter problem, given by the probability distribution of the hidden states conditioned on all measurements, is analytically tractable only in a few special cases. Due to the complexity, the number of unknown states (or parameters) and the computational cost of running the forward model in large scale geophysical models even suboptimal solutions (approximative solutions which converge to the optimal solution as the number of iterations increase) are out of reach. Hence alternative methods without optimal asymptotic behavior is sought. Over the last two decades a sampling version of the classical Kalman filter, known as the ensemble Kalman filter (EnKF) [15], has developed into one of the most popular algorithms for data assimilation, especially if we

restrict ourselves to sequential algorithms. The EnKF, as the Kalman filter, is a linear minimization algorithm but contrary to the standard Kalman filter, which is optimal in linear Gaussian systems, EnKF is only optimal asymptotically in linear Gaussian systems [25] as it uses low rank approximations of all matrices involved in the Kalman update. Our focus is on nonlinear systems so when applying Kalman type filters a linearization of the system is then required. Instead of linearizing the model based on a gradient approach as in the extended Kalman filter (EKF) [2] the model is linearized by the ensemble itself in the EnKF algorithm.

There have already been several attempts of combining ideas from EnKF with particle filters such as [30] where they compute the importance weights in a SMC with a Gaussian importance function obtained from EnKF or [26] where they use a moment matching idea to update the samples while using importance sampling to estimate posterior moments. For a review on SMC methods in geophysical models see [43]. The primary goal of this work is to improve upon the EnKF algorithm by combining it with algorithms in the class of Gaussian mixture filters [10]. We try to clarify the relationship of EnKF to particle filters and improve upon the methodology by bringing in ideas from statistical literature, such as local estimation of covariance matrices. In order to explain the Gaussian mixture filters we start by formulating the filter problem, its conceptual solution and asymptotic optimal approximation techniques.

1.1 The filter problem

Let $X_t(\omega), t \geq 0$ and $Y_t(\omega), t > 0$ be two $\mathcal{F}$ measurable stochastic processes defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. The filter problem consists of estimating the probability distribution of the states $X_t \in \mathcal{X}$ of a hidden dynamic system given the partial information $\mathcal{F}_Y = \sigma(Y_1, \ldots, Y_t)$ from a sequence of measurements $Y_{1:t} \triangleq (Y_1, \ldots, Y_t)$, $Y_j \in \mathcal{Y}$, $j = 1 \ldots t$. The systems under consideration may consist of both dynamic variables $U_t = (U_t^1, \ldots, U_t^{n_u})'$ and static parameters $\theta = (\theta_1, \ldots, \theta_{n_\theta})'$. For notational purposes we include both the dynamic variables and static parameters into the state vector, $X_t = (\theta, U_t)'$. Throughout the thesis we assume that all random variables are elements of $L^2(\Omega, \mathcal{F}, \mathcal{P})$. We consider systems of the form

$$X_t = \mathcal{M}_t(X_{t-1}, \xi_t) \tag{1.1}$$

$$Y_t = \mathcal{H}_t(X_t, \varepsilon_t), \tag{1.2}$$

with initial condition X_0 and where $\{\xi_t, t \geq 1\}$ and $\{\varepsilon_t, t \geq 1\}$ are noise sequences of mutually independent variables which are also independent of X_0. We will always assume that given $X_t = x$ the law of Y_t admits a known conditional probability density $p_{Y_t|X_t}(y_t|x_t)$. For fixed $Y_t = y_t$ we refer to $p_{Y_t|X_t}(y_t|x_t)$ as the likelihood function at time t. We assume that the observations are taken at discrete time steps either as direct measurements or as an integral of the measurement process between two time steps if Y_t is a continuous process. (Then $\mathcal{H}_t$ would depend on the state at time $s \in (t-1, t]$. If the model is discrete the model operator $\mathcal{M}_t$ is the transition from one time step to the next. If the system is continuous $\mathcal{M}_t$ is the integral operator of a deterministic or stochastic differential equation. The numerical algorithms discussed in this thesis are the same for both the discrete and continuous model scenario if the measurement noise is additive. That is we only need to know X_t to evaluate the density for Y_t. All the

case studies in this thesis are synthetic, that is numerical errors when solving $\mathcal{M}_t$ and $\mathcal{H}_t$ (in the continuous case) are non existing and these types of errors are not an issue in this thesis. In other words we assume that we can sample correctly from (1.1). We also consider the prior knowledge of the parameters and initial state of the system to be described in terms of a known probability density function which we denote as the prior. From (1.1) it follows that $\{X_t\}$ is Markov and in most of what follows we assume that the transition function is a conditional probability density. We are now ready to formulate the conceptual solution of the filter problem. For an overview see e.g. [12] or [37].

1.1.1 Conceptual solution of the filter problem

Assume that a prior density $p(x_0)$ exists and that at each time step $t > 0$ the conditional densities $p(x_t|x_{t-1})$ and $p(y_t|x_t)$ exists, where $p(\cdot)$ denotes the probability density function with the argument of the function indicating the random variables under consideration as long as there is no danger of confusion, i.e. $p(x_t|y_t) \triangleq p_{X_t|Y_t}(x_t|y_t)$. From a statistical point of view, the optimal solution to the filter problem described in the previous section is given by the conditional density of X_t given all measurements up to and including time t, $p(x_t|y_{1:t})$, denoted as the posterior density. Bayes' theorem allows us to recursively define the posterior density in terms of the densities described above for $t > 1$.

$$\text{forecast:} \quad p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})\,dx_{t-1} \tag{1.3}$$

$$\text{update:} \quad p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{\int p(y_t|x_t)p(x_t|y_{1:t-1})\,dx_t}. \tag{1.4}$$

For $t = 0$ the solution is $p(x_0)$ and for $t = 1$ the forecast is given by $p(x_1|x_0)$. To analytically solve these equations in general requires the evaluation of complex high dimensional integrals which is impossible in the general case. However there are certain cases where it is possible to obtain analytical solutions to the filter problem.

- If the system is linear and Gaussian the functional recursions satisfies the Kalman filter equations [21].
- If the state space is discrete valued with a finite number of states the optimal solution is obtained using a grid based method [37].
- For a certain class of nonlinear problems analyzed by [6] and [11], it is possible to formulate exact analytical solutions.

The Kalman filter is an important part of this thesis and we therefore give a quick overview. Assume that in addition to a Gaussian prior the model is described by linear operators and additive Gaussian noise

$$\begin{aligned} X_t &= \mathbb{F}_t X_{t-1} + \mathbb{B}_t \xi_t \\ Y_t &= \mathbb{H}_t X_t + \mathbb{R}_t \varepsilon_t, \end{aligned} \tag{1.5}$$

where ξ_t and ε_t are standardized Gaussian vectors, i.e, with zero mean, no correlations and with unit marginal variance. In this scenario the solution of the filter problem is

a Gaussian density with parameters given by the Kalman recursions. Starting with μ_0 and $\mathbb{P}_0$ from the prior density and defining

$$\begin{aligned}\mu_t &= \mathbf{E}[X_t|y_{1:t}],\\ \mathbb{S}_t &= \mathrm{Cov}(X_t|y_{1:t-1}),\\ \mathbb{P}_t &= \mathrm{Cov}(X_t|y_{1:t}).\end{aligned}$$

The parameters in the Gaussian posterior density evolves as follows

$$\begin{aligned}\mu_t &= \mathbb{F}_t\mu_{t-1} + \mathbb{S}_t\mathbb{H}_t'(\mathbb{H}_t\mathbb{S}_t\mathbb{H}_t' + \mathbb{R}_t\mathbb{R}_t')^{-1}(Y_t - \mathbb{H}_t\mathbb{F}_t\mu_{t-1}),\\ \mathbb{P}_t &= (\mathbb{I} - \mathbb{S}_t\mathbb{H}_t'(\mathbb{H}_t\mathbb{S}_t\mathbb{H}_t')^{-1}\mathbb{H}_t)\mathbb{S}_t,\end{aligned} \tag{1.6}$$

where

$$\mathbb{S}_t = \mathbb{F}_t\mathbb{P}_{t-1}\mathbb{F}_t' + \mathbb{B}_t\mathbb{B}_t'. \tag{1.7}$$

and

$$\mathbb{S}_t\mathbb{H}_t'(\mathbb{H}_t\mathbb{S}_t\mathbb{H}_t' + \mathbb{R}_t\mathbb{R}_t')^{-1} = \mathrm{Cov}(X_t, Y_t)[\mathrm{Cov}(Y_t, Y_t)]^{-1} \tag{1.8}$$

is referred to as the Kalman gain matrix.

We started this introduction by claiming that the posterior density is the optimal solution to the filter problem and recursively updating the posterior using Bayes' theorem coincides with the Kalman filter when we have linearity and Gaussianity. In order to verify these claims it is useful to study the filter problem in Hilbert space where it is natural to define the optimal solution as the element $\hat{x}$ in the subspace spanned by the measurements minimizing $||x-\hat{x}||$.

1.1.2 Filtering as a projection

Consider again the probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Let $X \in L^2(\Omega, \mathcal{F}, \mathcal{P})$. Denote by $\mathcal{F}_Y \subset \mathcal{F}$ the σ-algebra generated by the random variables $Y_{1:t} \in L^2(\Omega, \mathcal{F}, \mathcal{P})$. We assume that all variables are zero mean without loss of generality. We denote the covariance and cross-covariance matrices by $\mathbb{C}_{X,X}$, $\mathbb{C}_{Y,Y}$ and $\mathbb{C}_{X,Y}$ respectively. Let $\mathcal{L}_1$ be the subspace of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathcal{P})$ defined by

$$U_1 = \{Y \in L^2(\Omega, \mathcal{F}, \mathcal{P}); \text{ such that } Y \text{ is } \mathcal{F}_Y \text{ measurable}\}.$$

Then there exists a $\mathcal{P}$ unique element in $\mathcal{L}_1$ with the minimum distance to X (i.e. the best estimate) which is the orthogonal projection of X, denoted $P_{\mathcal{L}_1}(X)$, into $\mathcal{L}_1$. $P_{\mathcal{L}_1}$ is $\mathcal{F}_Y$ measurable and we have the identity

$$P_{\mathcal{L}_1}(X) = \mathbf{E}[X|\mathcal{F}_Y].$$

For a proof see e.g.[35]

Further let $\mathcal{L}_2$ be the subspace of $L^2(\Omega, \mathcal{F}, \mathcal{P})$ defined by the linear span of $Y = Y_{1:t}$. Again we may define an orthogonal projection of X into $\mathcal{L}_2$ which is the point in $\mathcal{L}_2$ closest to X

$$P_{\mathcal{L}_2}X = \min \mathbf{E}[X - \mathbb{A}Y]^2,$$

where the minimum is given by $\mathbb{A} = \mathbb{C}_{X,Y}\mathbb{C}_{Y,Y}^{-1}$. If $Y = \mathbb{H}X + \mathbb{R}\varepsilon$ where ε is zero mean Gaussian with identity covariance matrix then $\mathbb{C}_{X,Y} = \mathbb{C}_{X,X}\mathbb{H}'$, $\mathbb{C}_{Y,Y} = \mathbb{H}\mathbb{C}_{X,X}\mathbb{H}' + \mathbb{R}\mathbb{R}'$.

Further if X is Gaussian then $P_{\mathcal{L}_1}$ and $P_{\mathcal{L}_2}$ coincides and we recognize $\mathbb{A}$ from (1.8) as the Kalman gain matrix. From the above we conclude that

$$\mathbf{E}[X - \mathbf{E}[X|\mathcal{F}_Y]]^2 \leq \mathbf{E}[X - \mathbb{A}Y]^2,$$

with equality in the linear Gaussian case.

Now that we have defined the optimal solution to the filter problem we turn our focus on how to solve it in practice.

1.2 Sampling the posterior

In general, as discussed in the previous section, it is impossible to obtain analytical solutions to the filter problem. We are interested in quantities such as $P(|X_t| > C|Y_{1:t})$ or in general $\mathbf{E}[v(X_t)|Y_{1:t}]$ which can be written as

$$\int v(x)p(x_t|y_{1:t})\,dx_t \tag{1.9}$$

for an appropriate function v. If $v = 1_A$ for some Borel set A then (1.9) is equal to $P(X_t \in A|Y_{1:t} = y_{1:t})$. However, not knowing $p(x_t|y_{1:t})$ analytically we are forced to use approximative solutions of (1.9). As the convergence of standard numerical integration techniques depend on the dimension of the integrand they are not suitable to solve the filter problem. However, we see that the integral given by (1.9) is the expectation of a stochastic variable with density $p(x_t|y_{1:t})$ where $y_{1:t}$ is fixed. Thus we get by the strong law of large numbers that

$$\frac{1}{N}\sum_{i=1}^{N} v(X_t^i) \xrightarrow{a.s.} \int v(x)p(x_t|y_{1:t})\,dx, \tag{1.10}$$

as $N \to \infty$ if $\{X_t^i\}_{i=1}^N$ is a random sample from $p(x_t|y_{1:t})$. Here a.s. denotes almost sure convergence. Moreover, the order of convergence rate of (1.10) is equal to the square root of N for any dimension of the integral. However, the actual convergence also contains a constant that is clearly dependent on the dimension. There exists several algorithms that samples correctly from the posterior density (e.g. Langevin sampling [4] and Markov chain Monte Carlo [38]) however, the focus in this thesis is on the sequential Monte Carlo methods as these are more suitable for online estimation and do not require re-staring the algorithm from time zero when a new observation arrive. These sequential Monte Carlo methods are often referred to as particle filters and originate from sequential importance sampling (SIS). In order to understand SIS and particle filters it is useful to start with a description of ordinary importance sampling (IS).

1.2.1 Importance sampling

Let $X \in \mathcal{X}$ be a random vector with probability density function $p(x)$ and let $v : \mathcal{X} \to \mathbb{R}$ be an integrable function w.r.t.the density p. Assume that we want to evaluate

$$I(v) = \int v(x)p(x)\,dx.$$

Very often, as in the filter problem, it is not possible to sample directly from $p(x)$. However, let g be any density dominating p. Then we can write

$$I(v) = \int \frac{v(x)p(x)}{g(x)} g(x)\,dx.$$

If $\{X^i\}_{i=1}^N$ is a sample from g then by the strong law of large numbers

$$I_N(v) = \frac{1}{N}\sum_{i=1}^{N} \frac{v(X^i)p(X^i)}{g(X^i)} \xrightarrow{a.s.} \mathbf{E}_g\left[\frac{v(X)p(X)}{g(X)}\right] = I(v),$$

as $N \to \infty$. An important situation is when $p(x)$ is known only up to a normalizing constant,i.e, $p = C\widetilde{p}$ where $\widetilde{p}$ is known. In that case we use a weight function $w = \widetilde{p}/g$ and since any density integrates to one

$$I(v) = \frac{\int v(x)Cw(x)\,g(x)\,dx}{\int Cw(x)\,g(x)\,dx} = \frac{\int v(x)w(x)\,g(x)\,dx}{\int w(x)\,g(x)\,dx} = \frac{\mathbf{E}_g[v(X)w(X)]}{\mathbf{E}_g[w(X)]}.$$

Then, although biased (ratio of two estimators), the estimator

$$I_N(v) = \frac{\sum_{i=1}^N v(X^i)w(X^i)}{\sum_{j=1}^N w(X^j)}, \tag{1.11}$$

where $\{X^j\}_{j=1}^N$ is a random sample from g, converges almost surely to $I(v)$ [17]. This is exactly the scenario we have in the filter problem in (1.3). The joint posterior density satisfies

$$p(x_{0:t}|y_{1:t}) \propto p(y_t|x_t)p(x_t|x_{t-1})p(x_{0:t-1}|y_{1:t-1}),$$

by iterating we get

$$p(x_{0:t}|y_{1:t}) \propto p(x_0)\prod_{k=1}^{t} p(y_k|x_k)p(x_k|x_{k-1}), \tag{1.12}$$

and we see that all ingredients are assumed to be known and the problem is the normalizing constant. A suboptimal, but convenient choice for the importance function g is

$$g(x_{0:t}|y_{1:t}) = p(x_{0:t}) = p(x_0)\prod_{k=1}^{t} p(x_k|x_{k-1}), \tag{1.13}$$

since we have assumed that each transition density is known and that we know how to sample from it. The importance weights are then given by

$$w(x_{0:t}) = \frac{p(x_{0:t}|y_{1:t})}{g(x_{0:t}|y_{1:t})} \propto \frac{p(x_0)\prod_{k=1}^t p(y_k|x_k)p(x_k|x_{k-1})}{p(x_0)\prod_{k=1}^t p(x_k|x_{k-1})} = \prod_{k=1}^{t} p(y_k|x_k). \tag{1.14}$$

From equation (1.11) any question regarding the joint posterior distribution can be answered using estimators of the form

$$I_N(v) = \frac{\sum_{i=1}^N v(X_{0:t}^i)\prod_{k=1}^t p(y_k|X_k^i)}{\sum_{j=1}^N \prod_{k=1}^t p(y_k|X_k^j)},$$

where $X^i_{0:t}$ is a sample from (1.13) provided that N is sufficiently large. Usually we are only interested in the marginal posterior distribution $p(x_t|y_{1:t})$, then one may simply omit the first t components in each of the random vectors $X^i_{0:t}$.

There are in general two reasons why importance sampling as described above is not applied to the filter problem.

- The measurements often arrive sequentially in time and we would like to update the posterior density each time a new measurement arrives. It is then important to be able to start with the sample from the posterior density at the previous time step and not having to sample the entire vector $x_{0:t}$. This is also one of the main reasons why Markov chain Monte Carlo is not suitable for sequential sampling.
- As time increases the variance of the weights increase. From (1.14) we see that each of the components of $X^i_{1:t}$ has to statistically match the data at the given time to avoid having negligible weight. Being a random vector, this is very unlikely as time increases. The result is that most of the vectors have negligible weights, and more important, due to the normalization of the weights the ones that do have significant weights do not necessarily statistically match the data well, they just match the data better than the other random vectors. Instead we would like to have an algorithm where the weights at time t only depends on the statistical match between y_t and the sample $\{X^i_t\}_{i=1}^N$.

The first problem can be solved by selecting the importance function in such a way that both sampling and evaluation of the importance weights may be done sequentially (Note that this is already the case in (1.13)). This is known as sequential importance sampling (SIS). The second problem is solved by introducing a bootstrapping step in the SIS algorithm, this is known as sequential importance resampling (SIR) and is often referred to as the standard particle filter.

1.2.2 Sequential importance sampling and sequential importance resampling

Consider again the importance sampling setup where one has to construct a weighted sample from a chosen importance function g. As mentioned in the previous section we would like select the importance function in such a way that we may sample and evaluate the weights sequentially in time. For the filter problem this can be achieved if we choose an importance function of the form

$$g_t(x_{0:t}|y_{1:t}) = g_{t|t-1}(x_t|x_{0:t-1},y_{1:t})g_{t-1}(x_{0:t-1}|y_{1:t-1}). \tag{1.15}$$

We now see that a sample from $g_t(x_{0:t}|y_{1:t})$ is obtained by sampling from $g_{t|t-1}(x_t|X^i_{0:t-1},y_{1:t})$ given that one has a sample $X^i_{0:t-1}$ from $g_{t-1}(x_{0:t-1}|y_{1:t-1})$. The weights can also be updated sequentially as

$$\begin{aligned} W^i_t = w(X^i_{0:t}) &= \frac{p(X^i_{0:t}|y_{1:t})}{g_t(X^i_{0:t}|y_{1:t})} \\ &\propto \frac{p(y_t|X^i_t)p(X^i_t|X^i_{t-1})p(X^i_{0:t-1}|y_{1:t-1})}{g_{t|t-1}(X^i_t|X^i_{0:t-1},y_{1:t})g_{t-1}(X^i_{0:t-1}|y_{1:t-1})} \\ &= \frac{p(y_t|X^i_t)p(X^i_t|X^i_{t-1})}{g_{t|t-1}(X^i_t|X^i_{0:t-1},y_{1:t})}W^i_{t-1}. \end{aligned} \tag{1.16}$$

Further if we are only interested in the marginal posterior distribution at each time step it is convenient to select the importance function in such a way that the identity $g_{t|t-1}(x_t|x_{0:t-1},y_{1:t}) = g_{t|t-1}(x_t|x_{t-1},y_t)$ holds for each t. Then it is sufficient to only store $(X^i_{t-1},W^i_{t-1},y_t)$ in order to sample X^i_t and evaluate W^i_t. From (1.13) and (1.14) (1.15) and (1.16) are satisfied if we choose

$$\begin{aligned} &g_{t|t-1}(x_t|x_{0:t-1},y_{1:t}) = p(x_t|x_{t-1}) \\ &,g_{t-1}(x_{0:t-1}|y_{1:t-1}) = p(x_0)\prod_{k=1}^{t-1} p(x_k|x_{k-1}), \\ &W^i_t \propto W^i_{t-1} p(y_t|X^i_t). \end{aligned}$$

From (1.14) we see that each weight decays rapidly with time and after the weights are normalized the difference between the largest and smallest weight increase rapidly, hence the number of samples actually contributing to the estimator $I_N(v)$ decreases with time. A measure of how many samples that actually contributes to the estimate was introduced in [22] as the estimated effective ensemble size (omitting the time index)

$$\hat{N}_{\text{eff}} = \frac{(\sum_{j=1}^N \bar{W}^j)^2}{\sum_{i=1}^N (\bar{W}^i)^2} = \frac{1}{\sum_{i=1}^N (W^i)^2}, \tag{1.17}$$

where we now have defined $\bar{W}$ as the unnormalized weights and W as the normalized weights. To avoid a degeneracy of the weights [18] reintroduced a resampling step of the SIS algorithm, hence the name sequential importance resampling. The idea is to replace the weighted estimator

$$I_N(v) = \sum_{i=1}^N v(X^i)W^i$$

with the unweighted estimator

$$\hat{I}_n(v) = N^{-1}\sum_{i=1}^N v(X^i)N^i,$$

where N^i is a random number such that $\mathbf{E}[N^i] = NW^i$. The simplest (but not optimal) way of doing this is by drawing $M = \{N^i\}_{i=1}^N$ from the multinomial distribution $(N,p_1,\ldots,p_N) = (N,W^1,\ldots,W^N)$ where However, this is equivalent to sample N i.i.d. variables $\{\hat{X}^i\}_{i=1}^N$ from the weighted empirical distribution

$$\sum_{i=1}^N W^i \delta_{X^i}(x),$$

where δ is the Dirac-delta measure. A rule of thumb is to perform resampling if $\hat{N}_{\text{eff}}$ is below some predetermined threshold value [13]. There are several ways to reduce the variance of the resampling step or to increase the diversity of the re-sampled particles, but they will not be addressed here. For an overview see [12] or [24].

The estimates $\hat{I}_N(v)$ produced by the SIR algorithm has the correct asymptotic behavior assuming some regularity conditions on the model process and the actual v. This means that $\hat{I}_N(v)$ is strongly consistent and converge almost surely to the correct value $I(v)$. Moreover the normalized error, $\sqrt{N}\left(\hat{I}_N(v) - I(v)\right)$ is asymptotically normal with zero mean. For more details and theoretical results on SIR filters see e.g. [24, 32]

1.3 Curse of dimensionality and the ensemble Kalman filter

Although the SMC methods presented in the previous section have nice theoretical properties their applications are somewhat limited. If the complexity or the dimension of the phenomena being studied increase, the particle filter methods collapse in the sense that all but a few particles have significant weight at a given time step if the sample size is limited. Note that this is different from the weight degeneracy described in the previous section. When high dimensional or very complex models are studied such as petroleum reservoir or oceanographic and climate models, the computational cost of sampling forward in time increases as it demands solving partial differential equations numerically over a large domain. This naturally puts on upper limit on the number of samples, N, in the Monte Carlo algorithms. Intuitively the weight collapse occur as the solution space is extremely small compared to the space that we sample from. Mathematically it can be shown under some restrictions that the largest of the importance weights converges to one in probability as the dimension of the problem goes to infinity unless the number of samples grow exponentially with the dimension [7, 41]. The computational cost of forward sampling in large scale system force us to look for more robust sampling algorithms with the expense of no longer having asymptotic optimality.

1.3.1 The ensemble Kalman filter

Although particle filters are included in the geophysical literature [43], the most popular SMC method for application in high dimensional geophysical systems is the Ensemble Kalman filter (EnKF) [15]. Several improved versions of EnKF have been successfully applied to the history matching problem in petroleum [1] and in oceanography [9]. Usually in these applications, it is assumed that the measurement error is additive and Gaussian with zero mean and a known covariance matrix $\mathbb{R}_t$ so that for each t, $p(y_t|x_t)$ is a known Gaussian density w.r.t. y_t. Also it is common to include the simulated measurements in the state vector giving a linear relationship between the observation and the state vector with additive Gaussian noise, $Y_t = \mathbb{H}_t X_t + \varepsilon_t$, where $\mathbb{H}_t$ is a matrix of zeros and ones that selects the simulated measurements from the state vector.

From a statistical point of view EnKF is best described as a sequential Monte Carlo method with an additional Gaussian assumption. The EnKF evolves by sampling from the Markov transitions $p(x_t|x_{t-1})$. The joint density of the state vector is then assumed to be Gaussian and is approximated by the Gaussian density with the forecast sample mean $\bar{X}_t^f$ and sample covariance $\mathbb{S}_t^f$ as parameters. According to *Bayes*' theorem the posterior density is again a Gaussian density with parameters μ_t and $\mathbb{\Sigma}_t$ given by the Kalman equations

$$\begin{aligned} \mu_t &= \bar{X}_t^f + \mathbb{K}_t(y_t - \mathbb{H}_t \bar{X}_t^f) \\ \mathbb{K}_t &= \mathbb{S}_t^f \mathbb{H}_t'(\mathbb{H}_t \mathbb{S}_t^f \mathbb{H}_t' + \mathbb{R}_t)^{-1} \\ \mathbb{\Sigma}_t &= (\mathbb{I} - \mathbb{K}_t \mathbb{H}_t)\mathbb{S}_t^f. \end{aligned} \tag{1.18}$$

Again these parameters are estimated by a sample, where a sample $\{X_t^{i,u}\}_{i=1}^N$ from the posterior is obtained from the prior sample by

$$X_t^{i,u} = X_t^{i,f} + \mathbb{K}_t(y_T - (\mathbb{H}_t X_t^{i,f} + \varepsilon_t^i)), \quad i = 1, \ldots, N \tag{1.19}$$

where $\mathbb{K}_t$ is the same as in (1.18) and ε_t^i is a zero mean Gaussian random variable with covariance matrix $\mathbb{R}_t$. The parameters μ_t and Σ_t are estimated by the posterior sample and we no longer have the identities in (1.18), however these identities can be obtained by using a mean correction and updating the sample using a square root transformation [39] instead of (1.19). Note that under the Gaussian assumption, this sampling strategy corresponds to using the posterior density as importance function, that is all the weights are equal to N^{-1} which is equivalent to an unweighted sample. In other words, the curse of dimensionality as described above does not apply in the same manner to EnKF which makes it applicable to large geophysical models. However, there are other dimensionality issues regarding EnKF and such as the impact of spurious correlations and inbreeding of the sample. They are both consequences of the fact that we use sample covariance matrices in the Kalman Gain when the sample is updated at each time step. Each ensemble member is now a function of the other ensemble members and we no longer have an independent sample which results in under estimation of the posterior uncertainty. Also, as the number of ensemble members is low compared to the dimension of $\mathbb{S}^f$, the entries s_{ij}^f may be very unreliable especially the ones that are theoretically zero. This means that the sample update differs from the theoretical update (the one obtained using the theoretical covariance matrix). There exists several techniques we can apply to reduce the effect of spurious correlations e.g. [3, 16, 33, 45].

Another drawback, which is the major concern in this thesis, is that the asymptotic solution obtained by EnKF is only optimal in linear Gaussian systems [25]. The main bias for EnKF enters through the linear update of the non-observed entries in the state vector which may have a strong nonlinear relationship with the measurements.

As the SIR filter is asymptotically optimal but collapses in high dimensions and EnKF may suffer from severe bias, intuitively it might be reasonable to believe that the optimal filter for high dimensional systems lies somewhere in between.

1.3.2 Improved importance functions and Gaussian mixture filters

The aim of this section is to describe a class of filters, known as Gaussian mixture filters (GM), equipped with properties from both SIR and EnKF. The starting point however, is improving the importance function in the SIR algorithm. As in EnKF, we work with augmented state vectors and we therefore restrict ourselves to systems with a linear measurement operator. If we in addition assume that the model and measurement errors are additive with known Gaussian densities, it is possible to improve the choice of importance function. Conditionally on $X_{t-1} = x_{t-1}$ and $Y_t = y_t$, the optimal importance function (for any system) is $p(x_t|x_{t-1}, y_t)$ [14]. For the system described above, with model error covariance $\mathbb{Q}_t$ and measurement error covariance $\mathbb{R}_t$ we have

$$\begin{aligned} &p(x_t|x_{t-1}, y_t) = \Phi(\mu_t, \mathbb{P}_t), \\ &\mu_t = \mathcal{M}_t(x_{t-1}) + \mathbb{Q}_t\mathbb{H}_t(\mathbb{H}_t\mathbb{Q}_t\mathbb{H}_t' + \mathbb{R}_t)^{-1}(y_t - \mathbb{H}_t\mathcal{M}_t(x_{t-1})), \end{aligned} \tag{1.20}$$

where Φ denotes a Gaussian density. After sampling X_t^i from (1.20) the importance weights are updated as

$$w_t^i \propto w_{t-1}^i \Phi(y_t - \mathbb{H}_t\mathcal{M}_t(X_{t-1}^i), \mathbb{H}_t\mathbb{Q}_t\mathbb{H}_t' + \mathbb{R}_t),\ i = 1, \ldots, N.$$

Mathematically we see that the extra covariance term in the likelihood function, $\mathbb{H}_t\mathbb{Q}_t\mathbb{H}_t'$, results in a heavier tail weight function and thus the weights are less spread. Intuitively,

the weights are less spread as we use information about y_t when we sample X_t, thus we expect the sample to all match the data quite well or at least better than if we sample randomly from $p(x_t|x_{t-1})$.

In many geophysical applications the model error is unknown and very often not accounted for in the forward modeling. If we represent these unknown error quantities with a Gaussian random variable, it is not necessarily a good idea to add noise from this density to each sample but instead sample from the forward model and then represent the uncertainty with a Gaussian density. In this setting a natural representation of the forecast distribution would be a weighted sum of Gaussian densities (a Gaussian mixture) where the mean of each density is represented by one of the ensemble members/particles. The covariance matrix can be chosen arbitrarily, but should somehow reflect the uncertainty in the model to make the approximation reasonable. Note that it is these covariance matrices that controls the linear update of each ensemble member. If we let $\mathbb{P}$ be the covariance matrix of the Gaussian density and by defining $X_t^{i,f} = \mathcal{M}_t(X_{t-1}^{i,u})$, the updated particles in the GM filter are given by

$$X_t^{i,u} = X_t^{i,f} + \mathbb{P}_t\mathbb{H}_t(\mathbb{H}_t\mathbb{P}_t\mathbb{H}_t' + \mathbb{R}_t)^{-1}(y_t - \mathbb{H}_t X_t^{i,f}),\ i = 1,\dots,N, \tag{1.21}$$

and the weights are given by

$$W_t^i \propto W_{t-1}^i \Phi(y_t - \mathbb{H}_t X_t^{i,f}, \mathbb{H}_t\mathbb{P}_t\mathbb{H}_t' + \mathbb{R}_t),\ i = 1,\dots,N. \tag{1.22}$$

The Gaussian mixture (GM) filters can also be derived from nonparametric density estimation theory [40] by approximating the prior density at each time step with a Gaussian kernel estimator. For more details on GM filters see e.g. [8, 19, 23].

Although the GM filters combine ideas from EnKF and SIR, the asymptotic theory of the weights as the dimension increases are still valid. Hence the GM also suffer from the curse of dimensionality, at least in its basic form. From a practical point of view this means that $\mathbb{P}_t$ has to be large in some sense to avoid a collapse of the weights. A typical choice for $\mathbb{P}_t$ [19] is

$$\mathbb{P}_t = h^2\mathbb{S}_t^f,$$

that is $\mathbb{P}_t$ is proportional to the sample covariance of the forecast ensemble.

The first objective of this thesis is to construct a new class of Gaussian mixture filters with reduced bias. As mentioned above the size of the covariance matrix in each Gaussian kernel decides the impact of the linear update, increasing the size of each matrix increases the bias of the estimates if the posterior distribution is not Gaussian. At the same time, the covariance matrix of the likelihood function is increased by $\mathbb{H}_t\mathbb{P}_t\mathbb{H}_t'$ thus increasing the size of the matrices reduces the chance of a weight collapse and makes the filter more robust. To avoid these contradictions we introduce a new parameter, influencing the weights independent of the linear update. In the new setting it is possible to reduce the size of the linear update and at the same time avoid a weight collapse. The details are given in paper A.

1.3.3 Summary of paper A

In paper A we discuss the approximation of the optimal SIR filter using Gaussian mixtures with covariance matrices of the form

$$\mathbb{P}_t = h^2\mathbb{S}^f \tag{1.23}$$

where $\mathbb{S}^f$ is the sample covariance of the augmented forecast state vector $(X_t^{1,f},\dots,X_t^{N,f})$ and $h \in [0,1]$. It is necessary that $h > 0$ in order to update the sample, but in many applications we are forced to have $h \approx 1$ in order to avoid a weight collapse and the bias of the filter is almost as severe as EnKF. In paper A we argue that the correct thing to do is to introduce a weight reduction parameter independent of the linear update. Thus we introduce the weight interpolation

$$W_{\text{new}} = \alpha W_{\text{old}} + (1-\alpha)N^{-1}. \tag{1.24}$$

We also adapt α at each assimilation step as (remembering (1.17))

$$\alpha = \hat{N}_{\text{eff}} N^{-1},$$

based on bias variance tradeoff arguments. The resulting filter, denoted the adaptive Gaussian mixture filter (AGM), can behave as a SIR filter, a GM filter and EnKF depending on the values of the two parameters. A simulation study of the Lorenz40 model [28] where we compute the mean squared error (MSE) and the Kullback-Leibler divergence (KL) confirms that the bias of the filter is reduced compared to both GM and EnKF by introducing the new parameter.

1.3.4 Summary of paper B

One of the main areas of application for EnKF is reservoir characterization of petroleum reservoirs. It is therefore natural to apply AGM to such type of problems and compare the performance with EnKF. In paper B we compare the AGM with the standard EnKF on two synthetic petroleum reservoir models. The first is a 2D model with large initial uncertainty and a Gaussian prior with a spherical variogram. The second is the well known 3D model Punq-S3 [5]. While the initial ensemble in the 2D case are randomly generated from the prior, the initial ensemble in the Punq model are conditioned to well data.

Comparing different filter solutions on large scale models is not a simple task for several reasons. First, we have no good approximation of the true posterior distribution. Secondly, we don't know the distribution of the measures calculated (usually L_2 norms), except for the linear Gaussian case, but in this case the true posterior could be well approximated. In many geophysical applications, the only measure considered is the data mismatch. That is the (root) mean squared error between the observations and the filter solution. A rule of thumb is that the lower history match, the better the filter. Although a reasonable assumption, one has to understand that the N filtered samples are supposed to represent an independent sample from the posterior distribution and not N "best" estimates. A low data mismatch is often connected to under estimation of the posterior uncertainty. For petroleum reservoirs, the main interest in the filter problem is the geophysical parameters such as permeability and porosity. These quantities are usually assumed to be constant in time, hence we are able to compare the solution at the final time step with the prior density to see how much the samples have changed. If we have a sample from the true posterior, then it should not contradict the prior distribution. This sometimes happens when EnKF is applied, that is the posterior sample has a geostatistical structure that is far from the prior geostatistical structure. If we assume that there exists a unique porosity and permeability vector $\hat{\theta}$ minimizing the sum

of inner products

$$\sum_{t=1}^{T}(Y_t - \mathcal{H}_t(\theta))' R_t^{-1}(Y_t - \mathcal{H}_t(\theta)), \tag{1.25}$$

then the sample $\theta_i = \hat{\theta}, i = 1,\ldots,N$ has lower history match than any other sample, however, it is not a good solution to the filter problem as it represents the posterior distribution as a single point.

For Gaussian priors (which is the case in the two synthetic models) we may evaluate the Euclidean inner product

$$(\theta(s)_i^u - \theta_p(s))' \mathbb{P}^{-1}(\theta(s)_i^u - \theta_p(s)), \tag{1.26}$$

where $\theta(s)_i^u$ is filtered ensemble member i, $\theta(s)_p$ is the prior mean, $\mathbb{P}$ is the prior covariance, and s is the spatial variable. Minimizing (1.26) is equivalent to maximizing the prior likelihood for each ensemble member, thus we are able to compare how well each filtered ensemble member match fits the prior density. For large dimensional systems the posterior mean is usually far from the prior mean and the values of (1.26) is not that informative (remember the curse of dimensionality problem). Perhaps a more interesting measure is the deviation between the initial sample and the filtered sample

$$(\theta(s)_i^u - \theta_i^p(s))' \mathbb{P}^{-1}(\theta(s)_i^u - \theta_p^i(s)). \tag{1.27}$$

If small changes have been made to the prior update then we expect that the samples are more consistent with the prior distribution which they initially were sampled from. We compute (1.27) in for the posterior ensemble in Punq-S3 where no resampling is performed.

For the 2D model, where resampling is performed, the empirical variogram formula for each ensemble member

$$\frac{1}{N_h}\sum_{i,j\in N_h}(X(s_i) - X(s_j))^2, \tag{1.28}$$

where N_h is the number of ensemble members such that $\| s_i - s_j \| = h$, is computed and the mean over the ensemble is used to measure the geostatistical properties of the posterior ensemble. As mentioned above the theoretical posterior variogram is not the same as the prior, but as the posterior process is no longer stationary (1.28) is no longer an estimate of the variogram. However, it does say something about the geostatistical properties within each sample and should reflect the geostatistical properties of the true field which is generated using the prior theoretical variogram.

For the Punq-S3 model, the consistency of the filters are compared, in terms of the empirical cdf for the total oil production, using twenty different initial ensembles as previous studies have shown inconsistent results with EnKF using different initial ensembles [27].

Over all, the data mismatch for both filters are comparable, however the samples from AGM are closer to the prior samples and, produces estimated fields that are closer to the true field in root means squared error and the estimated cdfs of the total cumulativ oil production had a lower spread than the cdfs estimated from the EnKF sample.

The AGM was developed in paper A and tested on large scale models in paper B. In paper C we generalize it and study the asymptotic properties with some simplifications.

1.3.5 Summary of paper C

A generalization of AGM is the topic of paper C. We consider a d dimensional Markov system with transition kernels $Q_p(x,x')$, $p=1,\ldots,n$ with Q_p being the transition form time t_{p-1} to t_p. The measurements arrive sequentially in time at discrete times t_p, $p=1,\ldots,n$ and we assume that conditional on $X_p=x'$ the measurement Y_p has a known bounded density $g_p(x')$. Given the posterior density $\hat{\eta}_{p-1}$ at time $p-1$, the posterior expectation of a function f at time p is then proportional to

$$\hat{\eta}_{p-1}(K_p f(x)),$$

where we have defined the kernels K_p by

$$K_p f(x) = \int f(x')g_p(x')Q_p(x,x')\,dx'.$$

Our general interpolated kernel filter is defined through an approximating of K_p given by

$$\widehat{K}_p(x) \stackrel{\text{def}}{=} \alpha g_p(x') \int G_{p,h}(x'-u)Q_p(x,u)\,du + (1-\alpha)g_p(x') \int G_{p,h}(x'-u)\theta_{p,h}(u)Q_p(x,u)\,du, \tag{1.29}$$

where $G_{p,h}(x)=h^{-d}G_p(x/h)$, G_p is a symmetric kernel. $\theta_{p,h}(x)$ is a function such that for $h=0$ we have $\theta_{p,h}(x)g_p(x)$ is a bounded function for all h. By sampling $\{X_p^i\}_{i=1}^N$ from Q_p we get the particle approximation of (1.29) by

$$\sum_{i=1}^{N} \left(\alpha g_p(x_p)G_{p,h}(x_p-X_p^i) + (1-\alpha)g_p(x_p)\theta_{p,h}(X_p^i)G_{p,h}(x_p-X_p^i)\right).$$

We find that by appropriate functions $G_{p,h}$ and $\theta_{p,h}$ together with some regularity assumptions the filter reduces to AGM. However, we also show that there is a one to one correspondence between the adaptive weights in paper A and the new weights. Using methods developed by [31] error bounds for the asymptotic distribution of the new filter are provided along with a central limit theorem for the particular choice $\alpha=1-h^2$ and $h=h(N)=N^{-1/4}$ using a general SMC theorem [32].

1.3.6 Local Kalman gain

In paper C the kernel G_p depends on a parameter h, however it is not necessary that the kernel depends on h through (1.23) as in paper A and B. In fact, it is easy to construct an example where this particular choice is not a good one. Consider the simple model

$$Y = X^2 + \varepsilon,$$

where X and ε are independent standard Gaussian variables. Although dependent we see that

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X,X^2+\varepsilon) = \mathrm{Cov}(X,X^2) = \mathbf{E}X^3 = 0,$$

since X is Gaussian. This means that we have $\mathbf{E}[h^2\mathbb{S}]=0$ so that we expect GM, AGM and EnKF to do no linear update of the sample. This situation occurs since the covariance is a measure of global linear dependence between variables. A natural extension is then to measure a local linear dependence between the variables and hence estimate

the Kalman gain locally. Viewing the sample correlation as a linear approximation of the measurement operator, it is clear that a local approximation is better than a global approximation. Both parametric and non-parametric estimation of a local covariance matrix is discussed and tested on small dynamic systems in paper D.

1.3.7 Summary of paper D

In paper D we discuss the the failure of EnKF when the posterior distribution is multimodal. A local parametric Kalman filter is proposed where a the Kalman gain is evaluated locally at each ensemble member using a local Gaussian correlation [20]. The new filter is applied to a simple toy problem with bimodal posterior distribution [44]. Due to the computational cost of the filter, two nonparametric approaches are discussed. The first is a k-cluster mean algorithm [29] and the second is a kernel estimation approach using ideas from kernel regression [34]. It is shown that the kernel approach approximates an ensemble based EKF where the Jacobian of $\mathcal{H}_t$ is used to calculate the cross covariance between the states and measurements. The new methods are compared with EnKF and SIR in two different models, one chaotic model with dynamic states being filtered and a petroleum reservoir type model where we estimate geological parameters. For both models the true posterior is multimodal and is approximated with SIR using a large number of particles. The new methods proposed show great potential in capturing several (if not all) modes with as few as 100 particles. The standard SIR collapse in both cases with this few particles. As expected, EnKF fails to capture the multiple modes, and approximates the posterior with a single mode.

1.4 Resolution, sensitivities and the Fréchet derivative

So far we have only studied estimation techniques, however, an important task in complex models with many unknown parameters is to determine the information content in the data. Taking measurements can often be time consuming and/or expensive and it is therefore important to understand what the different measurements tell us about the underlying system. Sensitivities and resolutions are important tools in order to estimate the information content of the data and we will give a brief overview.

Assume that we study a linear model $y = \mathbb{G}m$ with prior mean and covariance $(m_p, \mathbb{C}_p)$ and observations y_o with error covariance $\mathbb{C}_y$. The least square estimate, m_{est}, minimizing

$$(m-m_p)'\mathbb{C}_p^{-1}(m-m_p)+(y_o-\mathbb{G}m)'\mathbb{C}_y^{-1}(y_o-\mathbb{G}m), \tag{1.30}$$

is given by

$$m_{est} = m_p + \mathbb{C}_p\mathbb{G}'(\mathbb{G}\mathbb{C}_m\mathbb{G}'+\mathbb{C}_y)^{-1}(y_o-\mathbb{G}m_p), \tag{1.31}$$

which is the mode of the posterior distribution if m the prior density and the likelihood are assumed Gaussian. The resolution of the estimate [42] is defined by

$$m_{est} - m_p = \mathbb{R}(m_{true} - m_p). \tag{1.32}$$

With an exact measurement $y_o = y = \mathbb{G}m_{true}$, then $\mathbb{R}$ is well defined in the linear case by

$$\mathbb{R} = \mathbb{C}_p\mathbb{G}'(\mathbb{G}\mathbb{C}_p\mathbb{G}+\mathbb{C}_y)^{-1})\mathbb{G}. \tag{1.33}$$

From (1.31) we see that if $\mathbb{R}$ is the identity matrix, we would have a perfect estimate. The result remain approximately valid for mildly nonlinear problems where $\mathbb{G}$ is a linearization of the model using the sensitivity matrix [36].

The posterior uncertainty is related to the resolution matrix using (1.32)

$$\mathbb{C}_{est} = (\mathbb{I} - \mathbb{R})\mathbb{C}_p, \tag{1.34}$$

so the resolution matrix $\mathbb{R}$ can be calculated without linearizing the model from the alternative formulation

$$\mathbb{R} = \mathbb{I} - \mathbb{C}_{est}\mathbb{C}_p^{-1}, \tag{1.35}$$

where $\mathbb{C}_{est}$ is the model uncertainty obtained from the solution (e.g. an MCMC run). From (1.35) we see that in the Kalman filter the resolution matrix is given by $\mathbb{KH}$.

For a continuous problem where the model $m = m(x)$ is a function in $L^2([a,b])$, the resolution is defined by

$$m_{est}(x) - m_p(x) = \int_a^b R(\xi)(m_{true}(\xi) - m_p(\xi)), \tag{1.36}$$

and we would have a perfect estimate if $R(\xi) = \delta_x(\xi)$, where δ is the Dirac measure.

In order to estimate the resolution it is useful to compute the sensitivity coefficients for a discretized problem, or the Fréchet derivative for a continuous problem. Both the sensitivities and the Fréchet derivative describe the change of model output from a small change in the input.

Definition 1 *Let $f : E \to Y$ be a mapping from an open set E in a normed linear space X into a normed linear space Y. Let $x \in E$. If there exists a bounded linear operator $\mathbb{G} : X \to Y$ such that*

$$\lim_{\delta \to 0} \frac{\| f(x+\delta) - f(x) - \mathbb{G}\delta \|}{\| \delta \|} = 0, \tag{1.37}$$

then f is said to be Fréchet differentiable at x.

In $L^2([a,b])$ the Fréchet derivative of a functional f at $m \in L^2([a,b])$(Parker 1994) is the integral kernel G such that

$$f(m+\delta m) = f(m) + (G, \delta m) + o(\| \delta m \|), \tag{1.38}$$

where $(G, \delta m)$ is the inner product $\int_a^b G(m(x))\delta m(x)\,dx$. From (1.38) we see that the Fréchet derivative or sensitivity can be obtained from a perturbation analysis.

1.4.1 Summary of paper E

In paper E we investigate the information content in nanosensors with limited functionality injected in a petroleum reservoir and transported with the fluid to provide information about spatial properties (Ullo, Chapman).

The model we study is a 1 dimensional core with incompressible flow. For this simple model we are able to derive analytical expressions of the Fréchet derivative of the location and pressure of a nanosensor with respect to permeability and porosity. To determine the resolution we compute linearized estimates of the sensitivity for the discretized model with unknown sensor location and for the scenario with known sensor location. The preliminary results is that the location of the sensor is sensitive to the

porosity between its location and the injection location, while it is sensitive to the effective permeability over the entire interval. The pressure on the sensor is sensitive to the permeability and porosity in the region between the particle and the injector, however, if the location is known the pressure is not sensitive to porosity.

The resolution of model estimates are computed in a numerical example with varying precision of the measurements where the reference log permeability and porosity are samples from a Gaussian prior. Not surprisingly, the model resolution is poor beyond the range traversed by the particle.

In the next numerical example the measurements consists of the flow rate at the inlet and censored pressure values from the particles. We assume that the nanosensors are encapsulated to protect them from the environment until a predetermined time is reached with a certain error. After a given time, 16 threshold levels are exposed to the reservoir conditions and level j will burst if the pressure exceeds some predetermined pressure value. The posterior pdf is approximated using a Markov chain Monte Carlo approach and from these results we also compute the resolution of the estimates. The conclusion is much the same as for the linearized example, the posterior uncertainty is smallest at the inlet and increases beyond the range traversed by the particles.

Bibliography

[1] AANONSEN, S. I., NÆVDAL, G., OLIVER, D. S., REYNOLDS, A. C., AND VALLÈS, B. Ensemble Kalman filter in reservoir engineering — a review. *SPE Journal 14*, 3 (2009), 393–412. 1, 1.3.1

[2] ANDERSON, B. D., AND MOORE, J. B. *Optimal filtering*. Dover, 2005. 1

[3] ANDERSON, J. L. Exploring the need for localization in ensemble data assimilation using hierarchical ensemble filter. *Physica D 230* (2007), 99–111. 1.3.1

[4] APTE, A., HAIRER, M., STUART, A. M., AND VOSS, J. Sampling the posterior: an approach to non-Gaussian data assimilation. *Physica D 230* (2007), 50–64. 1.2

[5] BARKER, J. W., CUYPERS, M., AND HOLDEN, L. Quantifying uncertainty in production forecasts: Another look at the PUNQ-S3 problem. *SPE Journal 6* (2001), 433–441. 1.3.4

[6] BENES, V. Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics* (1981), 65–92. 1.1.1

[7] BENGTSSON, T., BICKEL, P., AND LI, B. Curse-of-dimensionality revisited: Collapse of particle filter in very large scale systems. *Probability and Statistic 2* (2008), 316–334. 1.3

[8] BENGTSSON, T., SNYDER, C., AND NYCHKA, D. Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research—Atmospheres 108*, D24 (2003), ARTN 8775. 1.3.2

[9] BERTINO, L., EVENSEN, G., AND WACKERNAGEL, H. Sequential data assimilation techniques in oceanography. *International Statistical Review 71* (2007), 223–241, doi: 10.1111/j.1751-5823.2003.tb00194.x. 1.3.1

[10] CHEN, R., AND LIU, J. S. Mixture Kalman filters. *Journal of the Royal Statistical Society Series B-Statistical Methodology 60* (2000), 493–508. 1

[11] DAUM, F. Exact finite-dimensional nonlinear filters. *IEEE Transactions on Automatic Control 31* (1986), 616–622, doi: 10.1109/TAC.1986.1104344. 1.1.1

[12] DOUCET, A., DE FREITAS, N., AND GORDON, N., Eds. *Sequential Monte-Carlo methods in practice*. Springer-Verlag, New York, 2001. 1.1, 1.2.2

[13] DOUCET, A., DE FREITAS, N., AND GORDON, N. *Sequential Monte Carlo methods in practice*. Springer Verlag, 2001. 1.2.2

[14] DOUCET, A., GODSILL, S., AND ANDRIEU, C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing 10* (2000), 197–208. 1.3.2

[15] EVENSEN, G. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2007. 1, 1.3.1

[16] FURRER, R., AND BENGTSSON, T. Estimation of high-dimensional prior and posterior covariance matrices in kalman filter variants. *Journal of Multivariate Analysis* (2005). 1.3.1

[17] GEWEKE, J. Bayesian inference in econometric models using monte carlo integration. *Econometrica 24* (1989), 1317–1399. 1.2.1

[18] GORDON, N. *Bayesian methods for tracking*. PhD thesis, University of London, 1993. 1.2.2

[19] HOTEIT, I., PHAM, D.-T., TRIANTAFYLLOU, G., AND KORRES, G. A New approximative solution of the optimal nonlinear filter for Data Assimilation in Meteorology and Oceanography. *Monthly Weather Review 136* (2008), 317–334. 1.3.2

[20] HUFTHAMMER, K. O. *Local Gaussian Correlation*. PhD thesis, University of Bergen, 2008. 1.3.7

[21] KALMAN, R. E. Contributions to the theory of optimal control. 1.1.1

[22] KONG, A., LIU, J., AND WONG, W. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association 89*, 425 (1994), 278–288. 1.2.2

[23] KOTECHA, J. H., AND DJURIĆ, P. M. Gaussian sum particle filtering. *IEEE 51*, 10 (October 2003), 2602–2612. 1.3.2

[24] KÜNSCH, H. R. Recursive Monte Carlo filters: algorithms and theoretical analysis. *The Annals of Statistics 33*, 5 (2005), 1983–2021. 1.2.2

[25] LE GLAND, F., MONBET, V., AND DUC TRAN, V. Large sample asymptotics for the ensemble Kalman filter. INRIA report, 2009. 1, 1.3.1

[26] LEI, J., AND BICKEL, P. Ensemble filtering for high dimensional non-linear state space models. *Preprint* (2009). 1

[27] LORENTZEN, R. J., NÆVDAL, G., VALLÈS, B., BERG, A. M., AND GRIMSTAD, A.-A. Analysis of the ensemble Kalman filter for estimation of permeability and porosity in reservoir models. In *SPE Annual Technical Conference and Exhibition* (Dallas, Texas, 9-12 October 2005). Paper SPE96375. 1.3.4

[28] LORENZ, E. N., AND EMANUEL, K. A. Optimal sites for supplementary weather observations: Simulations with a small model. *J. Atmos. Sci. 55* (1998), 399–414. 1.3.3

[29] MACQUEEN, J. Some methods for classification and analysis of multivariate observation. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability* (1967), pp. 281–297. 1.3.7

[30] MANDEL, J., AND BEEZLEY, J. D. An ensemble Kalman-particle predictor-corrector filter for non-Gaussian data assimilation. In *ICCS 2009: Proceedings of the 9th International Conference on Computational Science* (Berlin, Heidelberg, 2009), Springer-Verlag, pp. 470–478. 1

[31] MORAL, P. D., AND JACOD, J. *Interacting particle filtering with discrete observations*. Springer, 2001, pp. 43–75. 1.3.5

[32] MORAL, P. D., AND JACOD, J. The Monte-Carlo method for filtering with discrete time observations: Central limit theorems. *The Fields Institute Communications, Numerical Methods and Stochastics, Ed. T.J. Lyons, T.S. Salisbury, American Mathematical Society* (2002). 1.2.2, 1.3.5

[33] MYRSETH, I., AND OMRE, H. Hierarchical ensemble kalman filter. *Spe Journal 15* (2010), doi: 10.2118/125851-PA. 1.3.1

[34] NADARAYA, E. On estimating regression. *Theory of Probability and its Applications 59*, 5 (1964), 749–757. 1.3.7

[35] ØKSENDAL, B. *Stochastic Differential Equations, Sixth edition*. Springer, 2007. 1.1.2

[36] OLIVER, D. S., REYNOLDS, A. C., AND LIU, N. *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge, 2008. 1.4

[37] RISTIC, B., ARULAMPALAM, S., AND GORDON, N. *Beyond the Kalman filter*. Artech House, 2004, pp. 45–47. 1.1, 1.1.1

[38] ROBERT, C. P., AND CASELLA, G. *Monte Carlo Statistical Methods*. Springer, 2004. 1.2

[39] SAKOV, P., AND OKE, P. R. Implications of the form of the ensemble transformation in the ensemble square root filters. *Mon. Weather Rev.* (2008), 1042–1053. 1.3.1

[40] SILVERMAN, B. W. *Density estimation for statistics and data analysis*. Chapman and Hall, 1986. 1.3.2

[41] SNYDER, C., BENGTSSON, T., BICKEL, P., AND ANDERSON, J. Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev. 136* (2008), 4629–4640. 1.3

[42] TARANTOLA, A. *Inverse Problem Theory: Methods for Data Fit-ting and Model Parameter Estimation*. 1987. 1.4

[43] VAN LEEUWEN, P. J. Particle filtering in geophysical systems. *Monthly Weather Review 137*, 12 (2009), 4089–4114. 1, 1.3.1

[44] ZAFARI, M., AND REYNOLDS, A. C. Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter. *SPE Journal 12*, 3 (2007), 382–391, doi: 10.2118/95750-PA. 1.3.7

[45] ZHANG, Y., AND OLIVER, D. S. Improving the ensemble estimate of the Kalman gain by bootstrap sampling. *Mathematical Geosciences 42*, 3 (2010), 327–345, doi: 10.1007/s11004-010-9267-8. 1.3.1