Fast Approximate Data Assimilation for High-Dimensional Problems

Xiao Lin, Gabriel Terejanu*

Department of Computer Science and Engineering, University of South Carolina

Abstract
Currently, real-time data assimilation techniques are overwhelmed by data volume, velocity and increasing complexity of computational models. In this paper, we propose a novel data assimilation approach which only requires a small number of samples and can be applied to high-dimensional systems. This approach is based on linear latent variable models and leverages machinery to achieve fast implementation. It does not require computing the high-dimensional sample covariance matrix, which provides significant computational speed-up. Since it is performed without calculating likelihood function, it can be applied to data assimilation problems in which likelihood is intractable. In addition, model error can be absorbed implicitly and reflected in the data assimilation result. Two numerical experiments are conducted, and the proposed approach shows promising results.

Keywords: Latent variable model, Nonlinear filtering, Probabilistic principal component analysis

1. Introduction
For a linear system, under the assumption of Gaussian probability distributions, the problem of estimating the states of the system has an exact closed-form solution given by the Kalman Filter (KF) (Kalman 1960). If the probability distributions are non-Gaussian or the system is nonlinear, in general no closed-form solutions are available. Different assumptions and approximations have been made for quasi-optimal solutions maintaining both accuracy and tractability. Among the best understood and most frequently cited nonlinear filters are the Extended Kalman Filter (EKF), the Ensemble Kalman Filter (EnKF), the Unscented Kalman Filter (UKF), and the Particle Filter (PF).

The EKF is historically the first, and still the most widely adopted approach to nonlinear state estimation problems. It is based on the assumption that over small time increments, nonlinear system dynamics can be accurately modeled by a first-order Taylor series expansion (Crassidis and Junkins 2004). However, when the process model and observation model are highly non-linear, EKF may give extremely poor performance (Julier and Uhlmann 2004). The UKF, which is a derivative-free alternative to EKF, bypasses differentiation by using a deterministic sampling approach (Julier and Uhlmann 2004; Wan and van der Merwe 2001). The EnKF is a reduced rank sampling filter which propagates the states through the system nonlinearities and updates a relatively small ensemble of samples from which an assumed Gaussian distribution which captures the main characteristics in the uncertainty is estimated (Evensen 2003). Due to its simplicity in both theory and implementation, EnKF has been widely used, especially in meteorology where high dimensional data assimilation is performed.

The PF uses a sampling approach, estimating the posterior probability distribution, including its higher order moments, by propagating and updating a number of particles, without the assumption of Gaussian statistics (Ristic et al. 2004). Particle filters have mostly been applied to low dimensional problems. Snyder et al. (Snyder et al. 2008a) points out that particle filter has the tendency to collapse and for a good estimation, it is required that the number of samples scales exponentially with the problem size. Nakano et al. (Nakano et al. 2007) proposed a merging particle filter (MPF) to avoid the collapse problem and the performance of MPF, PF and EnKF are compared by applying them to a 40-dimensional nonlinear dynamic system. The results show that as the number of samples increases, MPF...
converges faster than PF and after a certain number of samples they will both outperform EnKF. However, for a relatively small number of samples, for example  \( N < 256 \) which is tested in their experiments, EnKF works much better than MPF and PF.

Stordal et al. (Stordal et al., 2011) bridged EnKF and PF through an adaptive Gaussian mixture filter. Basically, EnKF is equivalent to the Gaussian mixture filter which takes sample covariance as covariance for each Gaussian kernel and has uniform weights. When model error is added, PF can also be viewed as the Gaussian mixture filter whose Gaussian kernel has zero covariance. Stordal et al. (Stordal et al., 2011) also pointed out increasing the variance of model noise in PF can force the weights to be more uniform so as to avoid filter collapse. In real applications, model error is an inevitable issue in data assimilation. The estimation of quantity of interest can be biased and overconfident without considering model error. Often, general formulations that exploit the structure of physics-based models by using internal discrepancies to capture structural errors yield intractable likelihood functions. In fact, in many cases, the distribution of observation noise is unknown, and the observation model is viewed as a black box which outputs observation data given the input states. These problems make typical PF, EnKF unsuitable as they rely on computation of likelihood functions.

In this paper, we propose a novel data assimilation method which only requires a small number of samples even in high dimensional systems. The proposed method is based on linear latent variable models. Samples are first mapped into latent space which has much lower dimensionality. Then update is performed in the latent space and posterior samples are mapped back to the original space. Compared with EnKF, this method doesn’t require Gaussian distribution assumption and is able to capture higher order statistics. Moreover, it can absorb model error and reflect the model error in the uncertainty of posterior distribution. This is useful since there is always a discrepancy between real world data and mathematical models. In addition, the method is performed without calculating likelihood of observation, thus it can be applied to data assimilation problems in which likelihood is intractable. Other filtering methods mentioned above all require direct computation of likelihood functions. In this paper, the proposed method is assessed on two nonlinear dynamic systems and also compared with EnKF in various scenarios. Our method shows promising results.

The rest of paper is organized as follows. In section 2, the proposed method is detailed. The method is not restricted to a particular type of linear latent models and several options are provided. In section 3, two numerical experiments are performed to assess its predictability, which includes comparisons with EnKF in several different scenarios.

2. Methodology

Consider the following general parameterized nonlinear dynamical system perturbed by process noise \( w_k \), measurement noise \( v_k \) and uncertain initial conditions. Given a set of measurements, the goal is to infer the state of the system and parameters to improve model predictions and as a result decision making under uncertainty.

\[
\begin{align*}
x_{k+1} &= f(x_k, \theta, w_k) \\
d_k &= h(x_k, \theta, v_k) \\
x_0 &\sim p(x_0)
\end{align*}
\]

(1) (2) (3)

Given a set of observation \( D_k = \{d_i|1 \leq i \leq k\} \), Bayesian filtering is the problem of finding the joint probability density function (pdf) of the states \( x_k \) and parameters \( \theta \) conditioned on all the observations up to and including current time \( t_k \), \( p(x_k, \theta | D_k) \).

The posterior pdf, \( p(x_k, \theta | D_k) \), may be obtained recursively using Bayes’ rule:

\[
p(x_k, \theta | D_k) = \frac{p(d_k | x_k, \theta, D_{k-1})p(x_k, \theta | D_{k-1})}{p(d_k | D_{k-1})}
\]

(4)

In Eq (4), \( p(d_k | x_k, \theta, D_{k-1}) \) is the likelihood function, \( p(d_k | D_{k-1}) \) is the normalization constant known as the evidence and the prior pdf \( p(x_k, \theta | D_{k-1}) \) is obtained as follows:

\[
p(x_k, \theta | D_{k-1}) = \int p(x_k | x_{k-1}, \theta)p(x_{k-1}, \theta | D_{k-1})dx_{k-1}
\]

(5)
where \( p(x_{k-1}, \theta | D_{k-1}) \) is posterior pdf at time \( t_{k-1} \) and \( p(x_k | x_{k-1}, \theta) \) can be obtained through Eq. (1).

If \( f(\cdot) \) and \( h(\cdot) \) are both linear models and all the probability distribution involved are Gaussian, then Bayesian filtering can be easily solved with Kalman filter. Otherwise, methods based on Monte Carlo samples are commonly adopted. However, current data assimilation methods suffer from the curse of dimensionality and struggle to deal with intractable likelihood functions. In general nonlinear dynamical systems, the measurement noise might not be explicitly presented in Eq. (2), it can be embedded into the observation model with unknown distribution, which makes it impossible to create likelihood function for EnKF and particle filter.

In this paper, we tackle Bayesian filtering from a different angle. Instead of relying on Eq. (2) to calculate likelihood, we operate on the joint distribution of \( x_k, \theta \) and \( d_k \) directly. For simplicity, in the followings, all the quantities of interest (QoIs) at any time \( k \) will be denoted by \( q_k \).

\[
q_k = [x_k, \theta]^T
\]  

(6)

Furthermore, let \( q_{k+1|k} \) denote \( q \) which is obtained through Eq. (1) after observing \( D_k \) but before \( d_{k+1} \). Thus \( p(x_{k+1}, \theta | D_k) \) can be written as \( p(q_{k+1|k}) \). Similarly, let \( d_{k+1|k} \) denote \( d \) which is obtained by propagating \( q_{k+1|k} \) through Eq. (2).

Assume there is a joint distribution \( p(q_{k+1|k}, d_{k+1|k}) \), then posterior pdf of \( q_{k+1|k} \) given \( d_{k+1} \) is the conditional distribution \( p(q_{k+1|k} | d_{k+1}) \) which is denoted by \( q_{k+1|k+1} \). Since there are no analytical forms for these distributions, they have to be estimated with Monte Carlo samples. This means that at any time \( k \) the uncertainty in the QoIs is described by a set of \( N \) samples.

\[
q_{k|k} \sim p(q_{k|k}), \quad \text{for } i = 1, \ldots, N
\]  

(7)

These samples are propagated through dynamics in Eq. (1) to obtain predictive samples at time \( k + 1, \{q_{k+1|k+1}^{(i)}\}_{i=1}^{N} \). Then \( \{q_{k+1|k+1}^{(i)}\}_{i=1}^{N} \) are propagated through Eq. (2) to obtain \( \{d_{k+1|k+1}^{(i)}\}_{i=1}^{N} \). Given the measurement \( d_{k+1} \), the goal is to obtain \( N \) samples, \( \{q_{k+1|k+1}^{(i)}\}_{i=1}^{N} \), from the posterior pdf of the QoIs, \( p(q_{k+1|k+1}) \). The whole flow is shown in Fig. 1. The input of our method are \( \{q_{k+1|k+1}^{(i)}\}_{i=1}^{N}, \{d_{k+1|k+1}^{(i)}\}_{i=1}^{N} \) and \( d_{k+1} \), and the expected output is \( \{q_{k+1|k+1}^{(i)}\}_{i=1}^{N} \).

![Bayesian update approximation (Ensemble LVM)](image)

Figure 1: Work flow of the proposed Bayesian filtering method.

If \( \{q_{k+1|k+1}, d_{k+1}^{(i)}\} \) follows multivariate Gaussian distribution, \( p(q_{k+1} | d_{k+1}) \) can be easily obtained \cite{Eaton1983}. If it is not Gaussian, we can still approximate it with a Gaussian distribution, and obtain conditional distribution. But in this case, the conditional distribution is likely to be inaccurate. As is shown in Fig. 2, a single Gaussian distribution is unable to capture multimodal property of samples and may lead to unreliable estimation of posterior \( p(q_{k+1} | d_{k+1}) \).
Figure 2: Fig.(a) shows Monte Carlo samples of $p(q_{k+1|k}, d_{k+1|k})$. The goal is to calculate conditional distribution given $\tilde{d}_{k+1}$. Fig.(b) shows Gaussian approximation of $p(q_{k+1|k}, d_{k+1|k})$ and consequential conditional distribution.

We can also estimate the probability density $p(q_{k+1|k}, d_{k+1|k})$ from Monte Carlo samples. However, if only a small number of samples are available and $[q_{k+1|k}, d_{k+1|k}]$ is in high dimensional space, then probability density estimation (PDE) will suffer badly from the curse of dimensionality. In this paper, we construct $p(q_{k+1|k}, d_{k+1|k})$ based on linear latent variable models and perform Bayesian inference in latent space.

2.1. Linear Latent Variable Model

The linear latent variable model (LLVM) is shown in Eq. (8). $W$ is a coefficient matrix, $z$ is the latent variable, $\mu$ is bias and $\eta$ is the noise. Assume $q_{k+1|k}$ and $d_{k+1|k}$ have a dimensionality of $H_q$ and $H_d$ respectively. Also denote the dimensionality of the joint space as $H$, then $H = H_q + H_d$. Let $M$ be the dimensionality of latent variable $z$. $M$ is supposed to be much smaller than $H$. The grounds for latent model is that data points usually lie close to a manifold of much lower dimensionality than that of the original data space (Bishop, 2006).

$$\begin{align*}
[q_{k+1|k}, d_{k+1|k}]^T &= Wz + \mu + \eta \\
z &\sim N(0, I_{M \times M}) \\
\eta &\sim N(0, \Psi)
\end{align*}$$ (8)

Parameters $W$, $\mu$ and $\Psi$ can be obtained with samples $\{q_{k+1|k}, d_{k+1|k}\}_{i=1...N}$ via maximum likelihood estimation.

$$W, \mu, \Psi = \arg \max \hat{p}(\{q_{k+1|k}, d_{k+1|k}\}_{i=1...N}|W, \mu, \Psi)$$

We may restrict the structure of the parameters, such that the latent space will preserve certain property from the original data set. This will lead to different linear latent variable models, which will be discussed later in the paper.

From Eq. (8), a density estimator for $p(q_{k+1|k}, d_{k+1|k})$ can be easily obtained as is shown in Eq. (9).

$$\begin{align*}
\hat{p}(q_{k+1|k}, d_{k+1|k}) &= N(\mu, \Sigma) \\
\Sigma &= WW^T + \Psi
\end{align*}$$ (9)

However, the pdf $p(q_{k+1|k}, d_{k+1|k})$ in Eq. (9) is still a single Gaussian distribution which might not be sufficient to describe original samples.
2.2. Ensemble Linear Latent Variable Model

Here, we introduce an ensemble linear latent variable model (EnLLVM) to replace the single Gaussian approximation. EnLLVM will lead to a Gaussian mixture which is supposed to give a better description of original samples. EnLLVM is built in the following way. For each sample \( \{q_{k+1|i}, d_{k+1|i}\} \) of \( p(q_{k+1|i}, d_{k+1|i}) \), we project it into the latent space, and there is a corresponding latent variable \( z_i \) which follows Gaussian distribution:

\[
z_i \sim N(E[z_i(q_{k+1|i}, d_{k+1|i})], \text{Cov}[z_i(q_{k+1|i}, d_{k+1|i})])
\]

This process is illustrated in Fig. 3. The black dots denote original samples and are projected into the latent space. For each sample, a corresponding latent variable is updated, which is denoted by a normal distribution in red. Back to original space, we will get a Gaussian representation for each sample. This is denoted by a blue circle. Then \( p(q_{k+1|i}, d_{k+1|i}) \) can be reconstructed by combining all the Gaussian representations. This will give a Gaussian mixture shown in Eq. (10). Since the model is linear Gaussian, \( p(z_i|q_{k+1|i}, d_{k+1|i}) \) can be obtained according to the property of conditional Gaussian distribution (Bishop, 2006).

\[
\tilde{p}_{eq}(q_{k+1|i}, d_{k+1|i}) = \frac{1}{N} \sum_{i=1}^{N} N(\mu_i, \Sigma_i) \tag{10}
\]

\[
\mu_i = WE[z_i(q_{k+1|i}, d_{k+1|i})] + \mu \tag{11}
\]

\[
\Sigma_i = WCov[z_i(q_{k+1|i}, d_{k+1|i})]W^T + \Psi \tag{12}
\]

where the mean and covariance of the latent variable are given by

\[
E[z_i(q_{k+1|i}, d_{k+1|i})] = \text{Cov}[z_i(q_{k+1|i}, d_{k+1|i})]W^T(q_{k+1|i} - \mu) \tag{13}
\]

\[
\text{Cov}[z_i(q_{k+1|i}, d_{k+1|i})] = (I_{M\times M} + WW^T)^{-1} \tag{14}
\]

We can also obtain \( E[z_i(q_{k+1|i}, d_{k+1|i})] \) and \( \text{Cov}[z_i(q_{k+1|i}, d_{k+1|i})] \) according to the prediction step of Kalman filter, which will give the following result

\[
E[z_i(q_{k+1|i}, d_{k+1|i})] = W^T \Sigma^{-1}(q_{k+1|i} - \mu) \tag{15}
\]

\[
\text{Cov}[z_i(q_{k+1|i}, d_{k+1|i})] = I_{M\times M} - W^T \Sigma^{-1}W \tag{16}
\]

Here, \( \Sigma \) is the covariance of prior distribution \( p(q_{k+1|i}, d_{k+1|i}) \), which is given in Eq. (10). These two presentations are equivalent. However, the computational complexity can be different. As we can see, both representations involve inversion of a \( H \) by \( H \) matrix which will cost \( O(H^3) \). The \( H \) by \( H \) matrix is \( \Psi \) in the first representation, and \( \Sigma = WW^T + \Psi \) in the second one. If \( \Psi \) is a diagonal matrix, this inversion will only cost \( O(H) \). Although a \( M \) by \( M \) matrix also needs to be inverted, \( M \) is supposed to be much smaller than \( H \), thus \( O(M^3) \) compared with \( O(H^3) \) is a big computational saving. In fact, by restricting the structure of \( \Psi \), we can have different latent linear models. This will be detailed in Sec. 2.3.

2.3. Property of EnLLVM

Although LLVM is naturally obtained from the latent model, it has a single Gaussian distribution which is unable to capture multimodality. To solve this problem, EnLLVM is developed. In this section, we will investigate the property of EnLLVM and compare it with LLVM.

Before making any statement, let’s first examine what \( \mu \) is in the latent model. All three parameters \( W, \mu, \Psi \) can be obtained by maximizing the likelihood function shown in Eq. (9). The corresponding log likelihood function is given in Eq. (17).

\[
\ln \tilde{p}(q_{k+1|i}, d_{k+1|i}) | w_{i=1}^{N} W, \mu, \Psi) = \sum_{i=1}^{N} \ln p(q_{k+1|i}, d_{k+1|i}) | W, \mu, \Psi \tag{17}
\]

\[
= - \frac{NM}{2} \ln(2\pi) - \frac{N}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^{N} (||q_{k+1|i} - \mu||^2 - \mu)^T \Sigma^{-1}(||q_{k+1|i} - \mu||^2 - \mu) \tag{18}
\]
Take the derivative of Eq. (19) with respect to $\mu$, we have
\[
d\ln \hat{p}(\{q_{k+1|k}, d_{k+1|k}\}_{i=1}^{N} | W, \mu, \Psi) = \frac{1}{2} \sum_{i=1}^{N} \Sigma^{-1} ([q_{k+1|k}, d_{k+1|k}] - \mu) \tag{20}
\]
Set Eq. (20) to 0, we can obtain $\mu$ which is the mean of $\{q_{k+1|k}, d_{k+1|k}\}_{i=1}^{N}$.

This means no matter what latent space is, $\mu$ is always the mean of original samples and thus different latent space is only determined by $W$ and $\Psi$. With this fact, we can get the lemma below.

**Lemma 1.** EnLLVM provides the same estimate for the mean and covariance of the samples as LLVM, while it captures higher order statistics as compared with LLVM.

As shown in Eq. (9), the mean and covariance of LLVM is $\mu$ and $\Sigma$ respectively. So we need to prove the following equations are true.

\[
E_{en}[q_{k+1|k}, d_{k+1|k}] = \mu \tag{22}
\]
\[
\text{Cov}_{en}[q_{k+1|k}, d_{k+1|k}] = \Sigma \tag{23}
\]

**Proof.** The mean and covariance of the Gaussian mixture in Eq. (10) are calculated using the following relations:

\[
E_{en}[q_{k+1|k}, d_{k+1|k}] = \frac{1}{N} \sum_{i=1}^{N} \mu_i \tag{24}
\]
\[
= \frac{1}{N} \sum_{i=1}^{N} WW^T \Sigma^{-1} ([q_{k+1|k}, d_{k+1|k}]^T - \mu) + \mu \tag{25}
\]
\[
= WW^T \Sigma^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} [q_{k+1|k}, d_{k+1|k}]^T - \mu \right) + \mu \tag{26}
\]
\[
= \mu \tag{27}
\]
Eq. (26) uses the fact that $\mu$ equals the mean of original samples. The covariance of EnLLVM can be decomposed

Figure 3: Gaussian mixture approximation given by EnLLVM.
into two parts as is shown below.

\[
\text{Cov}_{\eta}[\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k}] = \frac{1}{N} \sum_{i=1}^{N} \left[ \Sigma_i + (\mu_i - \mu)(\mu_i - \mu)^T \right]
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \Sigma_i + \frac{1}{N} \sum_{i=1}^{N} (\mu_i - \mu)(\mu_i - \mu)^T
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \Sigma_i + \frac{1}{N} \sum_{i=1}^{N} (\mu_i - \mu)(\mu_i - \mu)^T
\]

where

\[
T_1 = W\left(I_{d \times d} - W^T \Sigma^{-1} W\right)W^T + \Psi
\]

\[
= \Sigma - WW^T \Sigma^{-1} WW^T
\]

and

\[
T_2 = WW^T \Sigma^{-1} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( [\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k}]^T - \mu \right) \left( [\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k}]^T - \mu \right)^T \right] \Sigma^{-1} WW^T
\]

\[
\approx \Sigma
\]

Substituting Eq. 29 and Eq. 30 back into Eq. 28 we obtain that the covariance of the EnLLVM approximation coincides with LLVM covariance.

\[
\text{Cov}_{\eta}[\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k}] = \Sigma.
\]

So far, we have shown that EnLLVM provides the same estimate for the mean and covariance of samples as LLVM. As we know, LLVM cannot capture bimodality, since it gives a single Gaussian approximation. While EnLLVM is a mixture of Gaussian models, it is supposed to have better performance in capturing higher order statistics. This is shown in Fig. 4. The solid blue circle indicates bimodality of original samples. The dashed blue circles are components of the Gaussian mixture. Once observation data \( \hat{d}_{k+1} \) is given, the conditional pdf \( p(\mathbf{q}_{k+1|k}|\hat{d}_{k+1}) \) can be obtained by updating each Gaussian components with \( \hat{d}_{k+1} \) and the bimodality can be preserved. The details of updating process will be discussed in Sec. 2.5.

2.4. Options of LLVM

As is discussed in Sec. 2.3, \( \mu \) is the mean of samples for all possible latent spaces. A particular latent model only depends on \( W \) and \( \Psi \). By restricting the structure of \( W \) and \( \Psi \), the latent space can preserve certain property from original samples. To set notation, we decompose Eq. (8) into two separate latent models as is shown in Eq. (32) and (33).

\[
\mathbf{q}_{k+1|k} = W_d z + \mu_q + \eta_q
\]

\[
\mathbf{d}_{k+1|k} = W_d z + \mu_d + \eta_d
\]

\[
z \sim N(0, I_{d \times d})
\]

\[
\eta_q \sim N(0, \Psi_q)
\]

\[
\eta_d \sim N(0, \Psi_d)
\]

Now we have two separate latent models for \( \mathbf{q}_{k+1|k} \) and \( \mathbf{d}_{k+1|k} \) with a common latent variable \( z \). Parameters \( W_q, W_d, \mu_q, \mu_d, \eta_q \) and \( \eta_d \) can be obtained by comparing with Eq. (8). By doing this, we have made the assumption that the noise \( \eta_q \) and \( \eta_d \) are uncorrelated. This is a restriction on the structure of noise \( \eta \) and it is the case for commonly used linear latent models which we will discussed in this paper. Fig. 5 gives another interpretation of LLVM in our problem. Instead of taking it as the projection from \( [\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1|k}] \) to \( z \), we can view it from a generative viewpoint in which a sample of \( \mathbf{q}_{k+1|k} \) is obtained by sampling latent variable \( z \) and noise \( \eta_q \) then sampling the Gaussian representation conditioned on \( z \) and \( \eta_q \). A sample of \( \mathbf{d}_{k+1|k} \) is obtained in a similar way. Note that \( \mathbf{q}_{k+1|k} \) and \( \mathbf{d}_{k+1|k} \) share the same \( z \). This process is shown with red color. Here we introduce three commonly used linear latent models.
Probabilistic Principal component analysis (PPCA). PCA is concerned with finding orthogonal directions in lower dimensional space, on which the projections have the largest variance. It is assumed that these projections in lower dimensional space can reserve most information from the original data set, thus can be used for dimensionality reduction and feature extraction. Probabilistic principal component analysis (PPCA) is the probabilistic interpretation of PCA. It can be viewed as a maximum likelihood solution of the probabilistic latent variable model given in Eq. (8). We have mentioned that by restricting the structure of the covariance matrix $\Psi$, we can obtain different latent space. For PPCA, all off-diagonal values in $\Psi$ are zero and values on the diagonal are the same.

$$\text{blkdiag}(\Psi_q, \Psi_d) = \sigma^2 I_{D \times D}$$

As is discussed before, $\mu$ is the mean of samples no matter what latent space is. The other two parameters $W$ and $\sigma^2$ can be also be obtained through maximum likelihood method. Tipping et al. (Tipping and Bishop, 1999) gave the exact closed-form solution. However, if the dimensionality of the original space is high, which is one of the main focus of this paper, an iterative expectation maximization (EM) procedure can be performed to lower the computational cost. The EM algorithm can also be implemented incrementally. Samples are processed in an incremental and asynchronous way, which can be advantageous if both $N$ and $H$ are large. For the implementation of EM procedure, one can refer to (Bishop, 2006).

Factor analysis. Factor analysis is closely related to PPCA. The only difference in the latent model is that the diagonal of $\Psi$ can have different values, while for PPCA, all the diagonal values are the same.

$$\text{blkdiag}(\Psi_q, \Psi_d) = \text{diag}(\sigma^2_1 \ldots \sigma^2_D)$$

Unlike PPCA, there is no closed-form maximum likelihood solution for $W$. However, it can be done through EM method (Bishop, 2006).

Probabilistic Canonical Correlation Analysis (PCCA). Given two random vectors $q$ and $d$, CCA is concerned with finding two directions on which the projections of $q$ and $d$ have maximum correlation. Assuming $q$ has dimension $H_q$ and $d$ has dimension $H_d$, the total number of canonical correlations is $\min\{H_q, H_d\}$. For PCCA, parameters $W_q$, $W_d$, $\mu_q$, $\mu_d$, $\eta_q$, and $\eta_d$ can also be obtained by maximizing the likelihood function. Bach et al. (Bach and Jordan, 2005) gives the closed-form solution as well as EM algorithm. Note, unlike PPCA and factor analysis, $\eta_q$ and $\eta_d$ have off-diagonal values.
2.5. Bayesian Update of EnLLVM

Once observation data $\tilde{d}_{k+1}$ is obtained, the latent variable $z$ for each component of the Gaussian mixture can be updated though Eq. (33). Then posterior $p(\hat{q}_{k+1|k+1})$ can be estimated by sampling from the updated Gaussian mixture $\hat{p}_m(q_{k+1|k+1})$. The Gaussian mixture is updated in following way. First, for each Gaussian component, latent variable $z_i$ is updated through Eq. (33). This update is performed in the same way as Gaussian component is initialized by each sample. The mean and covariance matrix of $z_i$ given $\tilde{d}_{k+1}$ are as follows

$$E[z_i|\tilde{d}_{k+1}] = Cov[z_i|\tilde{d}_{k+1}](W^T\Psi_d^{-1}(d_{k+1} - \mu_d) + Cov[z|d_{k+1}](\mu'_{q_i} - \mu_d)^{-1}E[z|q_{k+1}, d_{k+1}])$$

(34)

$$Cov[z_i|\tilde{d}_{k+1}] = Cov[z_i|d_{k+1}](W^T\Psi_d^{-1}W_d)^{-1}$$

(35)

where $E[z|q_{k+1}, d_{k+1}]$ and $Cov[z|q_{k+1}, d_{k+1}]$ are given in Eq. (13) and Eq. (14) respectively. Second, the weight of each Gaussian component is calculated, as $d_{k+1}$ has different likelihood for each component. This update procedure shares the idea as particle filter in which each particle is assigned with a weight according to its likelihood. However, the likelihoods in EnLLVM and particle filter are different in essence. In particle filter, the likelihood of observation data is calculated through observation model that is Eq. (2). While in EnLLVM, the likelihood isn’t directly related to the observation model. It is calculated through Eq. (33) which is a latent model that we created. Thus EnLLVM can be applied to the data assimilation problem in which typical likelihood is intractable. After $z$ and weight are updated for each Gaussian component, $\hat{p}_m(q_{k+1|k+1})$ can be obtained as is shown in Eq. (36) from which posterior samples can be generated.

$$\hat{p}_m(q_{k+1|k+1}) = \sum_{i=1}^{N} w_i N(\mu_i, \Sigma_i)$$

(36)

$$\mu_i = W_q E[z_i|q_{k+1}] + \mu_q$$

(37)

$$\Sigma_i = W_q Cov[z_i|q_{k+1}]W_q^T + \Psi_q$$

(38)

One may notice that, since we already have Gaussian representation for each sample $[q_{i,k+1}, d_{i,k+1}]$, given observation data $\tilde{d}_{k+1}$, $p(q_{k+1|k+1})$ can be obtained directly by conditioning in the original joint space. But here we perform the update through latent model. This is due to the potential computational saving.
Lemma 2. Consider latent models shown in Eq. (32) and Eq. (33) with \( z \sim N(\mu_z, \Psi_z) \). Direct conditioning in the original space and update through latent models give equivalent posterior distribution \( p(\mathbf{q}_{k+1|k+1}) \).

Proof. The joint distribution of \( \mathbf{q}_{k|k+1} \) and \( \mathbf{d}_{k|k+1} \) is shown in Eq. (39).

\[
q_{k|k+1}, d_{k|k+1} \sim N(W\mu_z + \mu, WW^T + \Psi)
\]

where the mean and covariance matrix can be decomposed as follows

\[
W\mu_z + \mu = \begin{bmatrix} W_\mu \mu_z + \mu_d \\ W_\mu \mu_d + \mu_d \end{bmatrix} \tag{39}
\]

\[
WW^T + \Psi = \begin{bmatrix} W_\Psi W_\Psi^T + \Psi_q & W_\Psi W_d^T \\ W_d W_\Psi^T + \Psi_d \end{bmatrix} \tag{40}
\]

Given \( \mathbf{d}_{k+1} \), the conditional distribution of \( \mathbf{q}_{k+1|k+1} \) is also a Gaussian distribution of which the mean and covariance matrix are given below

\[
E[q_{k+1|k+1}] = W_\mu \mu_z + \mu + W_\Psi W_d^T(W_d W_\Psi W_d^T + \Psi_d)^{-1}(\mathbf{d}_{k+1} - W_\mu \mu_z + \mu_d) \tag{41}
\]

\[
\text{Cov}[q_{k+1|k+1}] = W_\Psi W_\Psi^T + \Psi_q - W_\Psi W_d^T(W_d W_\Psi W_d^T + \Psi_d)^{-1}W_d W_\Psi^T \tag{42}
\]

If we update \( \mathbf{q}_{k+1|k+1} \) through linear latent model, the mean and covariance matrix of latent variable \( z \) are already given in Eq. (34) and Eq. (35). Since the goal here is to show the equivalence with Eq. (41) and Eq. (41), we will use the other representation given by Kalman filter, which is shown below

\[
E[z|\mathbf{d}_{k+1}] = \mathbf{\Psi}_z + \mathbf{\Psi}_z W_d^T(W_d \mathbf{\Psi}_z W_d^T + \mathbf{\Psi}_d)^{-1}(\mathbf{d}_{k+1} - W_\mu \mu_z + \mu_d) \tag{43}
\]

\[
\text{Cov}[z|\mathbf{d}_{k+1}] = (I_{M	imes M} - \mathbf{\Psi}_z W_d^T(W_d \mathbf{\Psi}_z W_d^T + \mathbf{\Psi}_d)^{-1}W_d)\mathbf{\Psi}_z \tag{44}
\]

Then

\[
E[q_{k+1|k+1}] = W_\mu E[z|\mathbf{d}_{k+1}] + \mu_q \tag{45}
\]

\[
\text{Cov}[q_{k+1|k+1}] = W_\Psi E[z|\mathbf{d}_{k+1}]W_d^T + \Psi_q \tag{46}
\]

\[
\text{Cov}[q_{k+1|k+1}] = W_\Psi E[z|\mathbf{d}_{k+1}]W_d^T + \Psi_d \tag{47}
\]

As we can see, both ways give the equivalent update. However, conditioning in the data space requires the inversion of \( H_d \) by \( H_d \) matrix which is inevitable As is discussed before, if \( \Psi_q \) is a diagonal matrix, inversion of \( \Psi_q \) in Eq. (34) and Eq. (35) only costs \( O(H_d) \). Although an inversion of \( M \) by \( M \) matrix is added and can not be avoided, the dimensionality of latent space is usually much less than the observation space. Thus less computational cost is expected by updating through latent model.

2.6. Noise Inflation

After Bayesian update, each Gaussian component is assigned with a weight which is calculated according to the corresponding likelihood. This may arise a degeneracy problem, that is, the weight is concentrated on very few Gaussian components. In this case, posterior samples are likely to be generated from the same component many times, which will fail to describe the real posterior distribution. Once the generated samples are propagated to the next time points and same operation is performed, less and less diversity will be kept. Another issue is to deal with model error. The existence of model error indicates that the observation data \( \mathbf{d}_{k+1} \) may not be generated from Eq. (1) and Eq. (2), which further leads to the possibility that \( [\mathbf{d}_{k+1}, \mathbf{q}_{k+1|k+1}] \) may be far away from the joint distribution \( p(\mathbf{q}_{k+1|k}, \mathbf{d}_{k+1}) \) described by the current latent model.

To avoid component degeneracy and accommodate for the model error, noise inflation is performed. We introduce a new hyperparameter \( \eta \) in the distribution of the noise \( \eta \)

\[
\eta_d \sim N(0, \alpha\Psi_d) \tag{49}
\]
This hyperparameter is the key in performing data assimilation in the presence of model error. Initially, after solving the LLVM, the Gaussian representation is sufficient to capture the predictive samples. However, since our models always depart from reality due to modeling approximations, there is a discrepancy between our model predictions \( \mathbf{d}_{k+1} \) and the actual measurements \( \tilde{\mathbf{d}}_{k+1} \). The hyperparameter \( \alpha \) will be automatically tuned to accommodate for model discrepancy, it is determined in the same way as getting the parameters of the LLVM. Namely, \( \alpha \) is given by the maximizing the likelihood function provided by Eq. (33).

\[
\alpha = \arg \max N(\tilde{\mathbf{d}}_{k+1}; \mu, W_d W_d^T + \alpha \Psi_d)
\]

There is a closed form solution to Eq. (33). To make it more general, let’s introduce a lemma first.

**Lemma 3.** For fixed observation data \( \tilde{\mathbf{d}} \) and mean \( \mu \), the maximum value of \( N(\tilde{\mathbf{d}}; \mu, \Sigma) \) is achieved when \( \Sigma^* = (\tilde{\mathbf{d}} - \mu)(\tilde{\mathbf{d}} - \mu)^T \).

**Proof.** The log likelihood is given in Eq. (50).

\[
\ln N(\tilde{\mathbf{d}}; \mu, \Sigma) = -\frac{k}{2}\log 2\pi - \frac{1}{2}\log |\Sigma| - \frac{1}{2}(\tilde{\mathbf{d}} - \mu)^T \Sigma^{-1}(\tilde{\mathbf{d}} - \mu) \tag{50}
\]

where \( k \) is dimensionality of the Gaussian distribution. Take the derivative of \( \ln N(\tilde{\mathbf{d}}; \mu, \Sigma) \) with respect to \( \Sigma \), we have

\[
\frac{\partial \ln N(\tilde{\mathbf{d}}; \mu, \Sigma)}{\partial \Sigma} = -\frac{1}{2} \left( \frac{\partial |\Sigma|}{\partial \Sigma} + \frac{\partial (\tilde{\mathbf{d}} - \mu)^T \Sigma^{-1}(\tilde{\mathbf{d}} - \mu)}{\Sigma} \right) \tag{51}
\]

\[
= -\frac{1}{2}(\Sigma^{-1} - \Sigma^{-1}(\tilde{\mathbf{d}} - \mu)(\tilde{\mathbf{d}} - \mu)^T \Sigma^{-1}) \tag{52}
\]

Set Eq. (52) to 0, we obtain \( \Sigma^* \) that maximize the log likelihood.

\[
\Sigma^* = (\tilde{\mathbf{d}} - \mu)(\tilde{\mathbf{d}} - \mu)^T \tag{53}
\]

In our case, the covariance matrix is \( W_d W_d^T + \alpha \Psi_d \), thus \( \alpha^* \) can be obtained by solving the following equation:

\[
\alpha^* I = \Psi_d \Psi_d^{-1} \tag{54}
\]

\[
\Psi_d = (\mathbf{d}_{k+1} - \mu_d)(\mathbf{d}_{k+1} - \mu_d)^T - W_d W_d^T \tag{55}
\]

Usually, the exact solution of Eq. (54) does not exist, as \( \Psi_d \Psi_d^{-1} \) may have nonzero off-diagonal values and different diagonal values. However, we can always find approximate solution by minimizing the Frobenius norm difference between \( \alpha^* I_k \) and \( \Psi_d \Psi_d^{-1} \), which leads to the following solution:

\[
\alpha^* = \frac{1}{k} \text{tr}(\Psi_d \Psi_d^{-1}) \tag{56}
\]

If \( \alpha^* > 1 \), the noise in Eq. (33) will be inflated, which will lead to larger uncertainty in the latent variable. This uncertainty will eventually be reflected in the uncertainty of QoIs through Eq. (32). This means the model error will be absorbed in the uncertainty of QoIs, which provides a way to investigate model error and develop models.

### 3. Numerical experiments

In this section, the predictability of EnLLVM will be assessed by working on two chaotic systems. The first example will show the capability of capturing bimodal distribution. In the second example, EnLLVM and EnKF will be applied to a high dimensional system, and their performance will be compared on several different scenarios.
3.1. Example 1 - Lorenz 63

The Lorenz 63 is a three dimensional system and was used to study bimodal distribution in [Terejanu et al. 2011].

\[
\begin{align*}
\frac{dx_1}{dt} &= -cx_1 + cx_2 \\
\frac{dx_2}{dt} &= -x_1x_3 + rx_1 - y_2 \\
\frac{dx_3}{dt} &= x_1x_2 - bx_3
\end{align*}
\]

Here \( c = 10, b = \frac{8}{3} \) and \( r = 28 \). The discrete measurement model is given in Eq. (60)

\[
d_k = \sqrt{x_1(t_k)^2 + x_2(t_k)^2 + x_3(t_k)^2} + \nu_k, \quad \nu_k \sim N(0, 1)
\]

The total simulation time is 4 and the system is discretized with a time step \( \Delta t = 0.1 \). The measurement is obtained and used to update states every 4 time steps. The initial state pdf is set to

\[
p(x(t_0)) \sim 0.5N([-0.2, -0.2, 8]^T, \sqrt{0.35I_3}) + 0.5N([0.2, 0.2, 8]^T, \sqrt{0.35I_3})
\]

Due to the square form of the measurement model, and bimodal nature of the forecast pdf, the measurements do not offer sufficient information to choose one mode of the conditional pdf, thus the state estimate maintain its bimodal nature.

The joint space of states and observation have a dimensionality of four, and in this simulation, we use two components for PPCA and 30 samples for Bayesian update. Fig. 6 shows the posterior distribution of states after each update. The red lines indicate the truth value. As we can see, EnPPCA is doing well in capturing the truth value. Besides, it can also capture the bimodal distribution of \( x_1 \) and \( x_2 \).

![Figure 6: Posterior distribution of states after each update.](image-url)
3.2. Example 2 - Lorenz 96

In this example, EnPPCA and EnKF will be applied to track the states of Lorenz 96. Lorenz 96 is a 40 dimensional nonlinear system and commonly used for studying data assimilation. The system is given in Eq. \[(61)\].

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

where \(x_0 = x_{40}, x_{-1} = x_{39}, \) and \(x_{41} = x_1\). The system is discretized with a time step of \(\Delta t = 0.001\). The prior is given by \(N(0, 1)\) and the observation model provides an incomplete observation of the state of the system at every \(\Delta t = 0.1\). EnKF and EnPPCA will be compared in different scenarios. Besides the linear measurement model, two different nonlinear measurement models are also used. Another important scenario is the presence of model error. As is discussed earlier, EnPPCA has the property to absorb potential model error, and reflects it in the posterior distribution of QoI. Here, four different models are used to generate samples and they only differ in the constant term.

| Data generation process: |
|--------------------------|
| \(M1: \frac{dx_j(t)}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 9\) |
| \(M2: \frac{dx_j(t)}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 10\) |
| \(M3: \frac{dx_j(t)}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 11\) |
| \(M4: \frac{dx_j(t)}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 12\) |
| True model: \(\frac{dx_j(t)}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + 8\) |
| Measurement models: \(d_j(t) = x_{2j-1}(t) + v_j(t), \quad v_j(t) \sim N(0, I_{20})\) |
| Nonlinear measurement model: \((I): d_j(t) = x_{2j-1}(t)x_{2j}(t) + v_j(t), \quad v_j(t) \sim N(0, I_{20})\) |
| \((II): d_j(t) = x_{2j-1}(t)^2 + v_j(t), \quad v_j(t) \sim N(0, I_{20})\) |

The joint space between the QoIs (states of the system) and the observable has dimensionality of 60. For this study, 5 components are used for EnPPCA, which means the dimensionality of the latent space is 5. Here, we use 30 samples to track the states of the system. Root mean square error (RMSE) is used as the metric to measure the predictability of EnPPCA and EnKF. RMSE is calculated between the mean of posterior samples and the true state value.

Table.\[\text{1}\] shows the RMSE statistics of 100 trials. As we can see, if no model error exists, the performance of EnKF is comparable with or better than EnPPCA in average RMSE for all three measurement models. However, if we look at predictive distribution given by EnKF, which is shown in Fig.\[\text{7}\], the 95% credible interval of posterior samples often fails to capture the true state. This is due to the limited number of samples which underestimate the covariance matrix of the states. Fig.\[\text{8}\] shows predictive trajectories provided by EnPPCA. As we can see, the truth nearly always fall in the 95% credible interval.

Once model error is introduced, the RMSE increases for both EnKF and EnPPCA. However, EnPPCA has a smaller average RMSE than EnKF. And RMSE of EnKF increases significantly with the increasing model error, since there is no procedure to deal with model error. Fig.\[\text{9}\] Fig.\[\text{10}\] show the predictive trajectories of EnKF and EnPPCA when model M4 is used to generate samples. In this case, the model error is large and we can see that, 95% credible interval provided by EnKF can hardly capture the true state value. While for EnPPCA, it is still able to capture the truth.
Table 1: EnKF vs EnPPCA: RMSE statistics of 100 trials

| Model Error                      | EnKF |          |          | EnPPCA |          |          |
|----------------------------------|------|----------|----------|--------|----------|----------|
|                                  | Mean | STD      | Mean     | STD    | Mean     | STD      |
| Linear Measurement Model         |      |          |          |        |          |          |
| No err.                          | 2.71 | 0.61     | 2.92     | 0.23   |          |          |
| M1: 9                            | 3.71 | 0.63     | 2.99     | 0.23   |          |          |
| M2: 10                           | 4.63 | 0.71     | 3.07     | 0.22   |          |          |
| M3: 11                           | 5.31 | 0.81     | 3.18     | 0.21   |          |          |
| M4: 12                           | 5.96 | 0.84     | 3.27     | 0.20   |          |          |
| Nonlinear Measurement Model (I)  |      |          |          |        |          |          |
| No err.                          | 1.87 | 0.53     | 2.85     | 0.26   |          |          |
| M1: 9                            | 3.3  | 0.63     | 2.91     | 0.22   |          |          |
| M2: 10                           | 4.37 | 0.47     | 3.02     | 0.21   |          |          |
| M3: 11                           | 5.01 | 0.44     | 3.17     | 0.17   |          |          |
| M4: 12                           | 5.51 | 0.49     | 3.32     | 0.17   |          |          |
| Nonlinear Measurement Model (II) |      |          |          |        |          |          |
| No err.                          | 2.87 | 0.76     | 2.84     | 0.24   |          |          |
| M1: 9                            | 3.78 | 0.75     | 2.94     | 0.22   |          |          |
| M2: 10                           | 4.78 | 0.56     | 3.03     | 0.20   |          |          |
| M3: 11                           | 5.43 | 0.53     | 3.16     | 0.19   |          |          |
| M4: 12                           | 6.17 | 0.55     | 3.37     | 0.17   |          |          |

Figure 7: Results of data assimilation by EnFK: tracking the state of a 40 dimensional Lorenz system using only 30 samples and sparse incomplete observation model.
Figure 8: Results of data assimilation by PPCA: tracking the state of a 40 dimensional Lorenz system using only 30 samples and sparse incomplete observation model.

Figure 9: Results of data assimilation by EnKF: tracking the state of a 40 dimensional Lorenz system using only 30 samples and sparse incomplete linear observation model with model error M4.
4. Summary

In this paper, a novel data assimilation method EnLLVM is proposed. EnLLVM is an ensemble method which is based on linear latent variable models. In EnLLVM, the joint distribution of QoIs and observation is approximated by a Gaussian mixture which is constructed by projecting each sample into the latent space. Then Bayesian update of QoIs is performed by first updating the latent variable of each Gaussian component and then map it to the original space. By choosing proper linear latent models, e.g., PPCA or FA, one can avoid inverting big matrix in original space. In stead, only inversion of matrix which has the same dimensionality as the latent space is required. This can be a big saving in complexity as the dimensionality of the latent space is supposed to be much lower than that of the original space. EnLLVM only requires a small number of samples. Compared with EnKF, it doesn’t require Gaussian distribution assumption and is able to capture higher order statistics. Since Bayesian update is performed in the latent space without calculating likelihood through observation model, EnLLVM can be applied to data assimilation problems where likelihood function is intractable. In addition, the introduction of noise inflation procedure not only prevent potential degeneracy, but also enable EnLLVM to accommodate for model error which will be absorbed in the inflated noise and finally reflected in the uncertainty of QoIs.

Since EnLLVM is a general framework without specifying the latent model, We can try different latent models which preserve different kinds of properties from the origin samples. In this paper, the performance of EnPPCA is assessed by applying to two nonlinear dynamic systems Lorenz63 and Lorenz96. The first example shows the capability of EnPPCA to capture bimodal distribution. In the second example, we compared EnPPCA with EnKF on the same system in a bunch of different scenarios including different nonlinear observation models and different model error. Results show that EnPPCA gives consistent better performance.

EnLLVM is still an approximate method. Although linear latent model facilitates mapping between original space and latent space, the Gaussian presentation it provides is too simple to give an exact description of the probability distribution, even if ensemble method is used. Since samples are projected a much lower latent space, the loss of information is unavoidable. Can we apply the same method without dimension reduction? The answer is no. This is because if there is no dimension reduction, the covariance of the noise in the latent model will be zero. Look
back to Eq. (10), \( \Psi = 0 \) will lead to \( \Sigma = WW^T \). This will also further lead to a zero covariance for the latent variable:

\[
\begin{align*}
\text{Cov}[z|q_{k+1|k}, d_{k+1|k}] &= I_{M \times M} - W^T \Sigma^{-1} W \\
&= I_{M \times M} - W^T (WW^T)^{-1} W \\
&= I_{M \times M} - (W^T W^{-1})(W^{-1} W) \\
&= 0
\end{align*}
\]

This uses the fact that if \( WW^T \) is invertible, then the rows of \( W \) have to be linearly independent. Since \( W \) is a square matrix, it is invertable. The mean of the latent variable will become

\[
\begin{align*}
E[z|q_{k+1|k}, d_{k+1|k}] &= W^T \Sigma^{-1} [(q_{k+1|k}, d_{k+1|k})^T - \mu] \\
&= W^T (WW^T)^{-1} [(q_{k+1|k}, d_{k+1|k})^T - \mu]
\end{align*}
\]

Replace the mean and covariance of the latent variable in Eq. (11) and Eq. (12), we will get the Gaussian representation for sample \( [q_{k+1|k}, d_{k+1|k}] \):

\[
\begin{align*}
\mu_i &= WE[z|q_{k+1|k}, d_{k+1|k}] + \mu \\
&= [q_{k+1|k}, d_{k+1|k}]^T \\
\Sigma_i &= WCov[z|q_{k+1|k}, d_{k+1|k}]W + \Psi \\
&= 0
\end{align*}
\]

As we can see, the Gaussian representation is the sample itself and no further update can apply. However, this is by no means a disadvantage of EnLLVM, since the fast implementation is achieved by mapping into the latent space which is supposed to have much lower dimensionality.
5. References

References

Bach, F. R., Jordan, M. I., 2005. A probabilistic interpretation of canonical correlation analysis. Tech. Rep. Technical Report 688, Department of Statistics, University of California, Berkeley.

Bishop, C. M., Aug. 2006. Pattern Recognition and Machine Learning. Springer Verlag.

Crassidis, J., Junkins, J., 2004. Optimal Estimation of Dynamic Systems. CRC Press.

Eaton, M. L., 1983. Multivariate Statistics: a Vector Space Approach. John Wiley and Sons.

Evensen, G., 2003. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. Ocean Dynamics 53, 343–367.

Julier, S., Uhlmann, J., 2004. Unscented Filtering and Nonlinear Estimation. Proceedings of the IEEE 92 (3), 401–422.

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME–Journal of Basic Engineering 82, 35–45, series D.

Nakano, S., Ueno, G., Higuchi, T., 2007. Merging particle filter for sequential data assimilation. Nonlinear Processes in Geophysics 14 (4), 395–408.

Ristic, B., Arulampalam, S., Gordon, N., 2004. Beyond the Kalman Filter. Particle Filters for Tracking Applications. Artech House.

Snyder, C., Bengtsson, T., Bickel, P., Anderson, J., 2008. Obstacles to high-dimensional particle filtering. Monthly Weather Review 136 (12), 4629–4640.

Stordal, A. S., Karlsen, H. A., Nøvdal, G., Skaug, H. J., Vallès, B., Mar 2011. Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. Computational Geosciences 15 (2), 293–305.

Terejanu, G., Singla, P., Singh, T., Scott, P. D., 2011. Adaptive gaussian sum filter for nonlinear bayesian estimation. IEEE Transactions on Automatic Control 56 (9), 2151–2156.

Tipping, M. E., Bishop, C. M., 1999. Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61 (3), 611–622.

Wan, E., van der Merwe, R., 2001. The Unscented Kalman Filter. Wiley Publishing.

URL citeeey.iifi.unizh.ch/van01unscented.html