A guided sequential Monte Carlo method for the assimilation of data into stochastic dynamical systems*

Sebastian Reich

November 13, 2012

Abstract

Assimilation of measurements into stochastic dynamical systems is challenging due to the generally non-Gaussian behavior of the underlying probability density functions. While sequential Monte Carlo methods have emerged as a methodology for tackling assimilation problems under rather general circumstances, those methods suffer from the curse of dimensionality. At the same time ensemble transform filters, such as the ensemble Kalman filter, have emerged as attractive alternatives to sequential Monte Carlo methods since they also work for high dimensional problems. Typical ensemble transform filters are however based on rather crude approximations to the involved probability density functions and are therefore of limited accuracy. For that reason there have been a number of recent attempts to combine sequential Monte Carlo methods with ensemble transform techniques, so called, guided sequential Monte Carlo methods. In this paper, we first put ensemble transform filters in the context of coupling and optimal transportation and secondly propose a new guided sequential Monte Carlo method based on combining approximate couplings with importance sampling. The effect of various filtering strategies is demonstrated for a simple Brownian dynamics model.

1 Introduction

We consider random dynamical systems (Gardiner, 2004) induced by the iteration

$$X_{n+1} = \Psi(X_n) + \Xi_n, \qquad n \geq 0, \tag{1}$$

under the assumption that $X_0 : \Omega \to \mathbb{R}^N$ is a multivariate random variable over some sample space Ω with given probability density function (PDF) $\pi_0(x)$, $\Xi_n : \Omega \to \mathbb{R}^N$ are independent and identically distributed Gaussian random variables with mean zero and covariance matrix $Q \in \mathbb{R}^{N\times N}$, i.e. $\Xi_n \sim \mathrm{N}(0, Q)$, and $\Psi : \mathbb{R}^N \to \mathbb{R}^N$ is an appropriate map. The associated Chapman-Kolmogorov equation (Gardiner, 2004) for the marginal PDFs $\pi_n(x)$, i.e. $X_n \sim \pi_n$, is given by

$$\begin{aligned}\pi_{n+1}(x) &= \int_{\mathbb{R}^N} \frac{1}{(2\pi)^{N/2}|Q|^{1/2}} \exp\left(-\frac{1}{2}(x-\Psi(x'))^T Q^{-1}(x-\Psi(x'))\right) \pi_n(x')\, dx' \\ &= \int_{\mathbb{R}^N} \pi(x|x')\, \pi_n(x')\, dx' \end{aligned} \tag{2}$$

*Universität Potsdam, Institut für Mathematik, Am Neuen Palais 10, D-14469 Potsdam, Germany

with Markov transition kernel

$$\pi(x|x') = \frac{1}{(2\pi)^{N/2}|Q|^{1/2}} \exp\left(-\frac{1}{2}(x-\Psi(x'))^T Q^{-1}(x-\Psi(x'))\right).$$

Here $|Q|$ denotes the determinant of Q.

In this paper, we will assume that (1) arises from the discretization of an underlying stochastic differential equation (SDE) by the Euler-Maruyama method (Kloeden and Platen, 1992) with step-size $\Delta t > 0$, i.e.

$$X_{n+1} = X_n + \Delta t f(X_n) + \sqrt{\Delta t} Z_n, \qquad t_{n+1} = t_n + \Delta t,$$

and the marginal PDFs will be denoted by $\pi_X(x, t_n) = \pi_n(x)$. To avoid confusion we will also use the notation $X(t_n) = X_n$ from now on. Hence $\Psi(X(t_n)) = X(t_n) + \Delta t f(X(t_n))$ is the drift term and $\Xi(t_n) = \sqrt{\Delta t} Z_n$ describes diffusion with $Z_n \sim \mathrm{N}(0, \gamma I)$ and diffusion coefficient $\gamma > 0$. The zero diffusion limit $\gamma = 0$, i.e.

$$X(t_{n+1}) = \Psi(X(t_n)) = X(t_n) + \Delta t f(X(t_n)),$$

gives formally rise to the Chapman-Kolmogorov equation

$$\pi_X(x, t_{n+1}) = \int_{\mathbb{R}^N} \delta(x - \Psi(x'))\, \pi_X(x', t_n)\, dx',$$

which is equivalent to

$$\pi_X(\Psi(x), t_{n+1})\, |D\Psi(x)| = \pi_X(x, t_n), \tag{3}$$

where $D\Psi(x) \in \mathbb{R}^{N\times N}$ denotes the Jacobian matrix of partial derivatives of $\Psi(x)$ and $\delta(\cdot)$ the Dirac delta function. We emphasize that we are not interested in the $\Delta t \to 0$ limit in this paper and will consider $\Delta t > 0$ as a given, fixed quantity. We emphasize furthermore that the algorithms considered in this paper do not depend on this assumption and are applicable to general intermittent data assimilation problems.

We assume the availability of partial observations $y_{\mathrm{obs}}(j\Delta t_{\mathrm{obs}})$ of the stochastic process generated by (1) in discrete time intervals Δt_{obs} and for $j \geq 1$. We also assume that $\Delta t_{\mathrm{obs}} = L\Delta t$, i.e. measurements are taken every $L \geq 1$ time-steps. The forward model for the observational process is assumed to be linear, i.e.

$$Y = HX + \Theta \tag{4}$$

with forward operator $H : \mathbb{R}^N \to \mathbb{R}^K$, measurement error $\Theta \sim \mathrm{N}(0, R)$, and measurement error covariance matrix $R \in \mathbb{R}^{K\times K}$. It is assumed that measurement errors at different instances in time are independent and identically distributed. The associated likelihood function is denoted by $\pi_Y(y|x)$.

The task of intermittent data assimilation is to determine the conditional PDFs $\pi_X(x, t_n|\mathrm{Y}_k)$ of the random variable $X(t_n)$ at $t_n = n\Delta t$ given collected measurements

$$\mathrm{Y}_k = (y_{\mathrm{obs}}(\mathrm{t}_1)^T, y_{\mathrm{obs}}(\mathrm{t}_2)^T, \cdots, y_{\mathrm{obs}}(\mathrm{t}_k)^T)^T$$

at observation times $\mathrm{t}_j = j\Delta t_{\mathrm{obs}}$, $j = 1, \ldots, k$. We will consider the two cases $t_n = \mathrm{t}_k$ (filtering) and $t_n > \mathrm{t}_k$ (prediction). See Jazwinski (1970) for an excellent introduction to stochastic processes and filtering.

Let us first consider the pure prediction problem with no observations ($k = 0$). A Monte Carlo simulation of (1) would proceed as follows. First one finds M independent realizations $x_i(t_0)$ from the initial PDF $\pi_X(x, t_0)$. Secondly, each realization is updated recursively and independently according to

$$x_i(t_{n+1}) = \Psi(x_i(t_n)) + \xi_i(t_n), \qquad n \geq 0, \tag{5}$$

where $\xi_i(t_n)$ are independent realizations from the normal distribution $\mathrm{N}(0, Q)$ with $Q = \gamma \Delta t I$. Under appropriate conditions on the marginal PDFs $\pi_X(x, t_n)$ it can be shown that the empirical distribution

$$\pi_X^{\mathrm{em}}(x, t_n) = \frac{1}{M} \sum_{i=1}^{M} \delta(x - x_i(t_n))$$

converges weakly to $\pi_X(x, t_n)$ as $M \to \infty$ for any fixed $n > 0$.

Monte Carlo simulation approaches for the pure prediction problem have been extended to the combined filtering-prediction problem and have given rise to a broad range of sequential Monte Carlo methods (Doucet et al., 2001; Bain and Crisan, 2009). The essential idea is to augment the realizations $x_i(t_n)$ by weights $w_i(t_n) > 0$ subject to

$$\sum_i w_i(t_n) = 1$$

and to adjust the weights such that they reflect the importance of the samples $x_i(t_n)$ relative to the available measurements Y_k while the samples $x_i(t_n)$ continue to follow the stochastic dynamics (1). A common problem of this basic importance sampling approach is a degeneracy of weights which requires resampling techniques. Straightforward resampling can be achieved by eliminating samples with small weights and duplication of those with large weights. Practical experience shows that the just described combined importance sampling-resampling approach of sequential Monte Carlo methods does not work well for high dimensional problems unless the number of samples M is increased at a rate which scales exponentially in the phase space dimension N (Bengtsson et al., 2008). At the same time, the ensemble Kalman filter (EnKF) (Evensen, 2006) has emerged as a robust alternative to sequential Monte Carlo methods applicable to high dimensional problems. The EnKF relies on Gaussian approximations to the marginal PDFs $\pi_X(x, t_n|\mathrm{Y}_k)$ and can be shown to be statistically inconsistent in the limit $M \to 0$ (contrary to sequential Monte Carlo methods) (Lei and Bickel, 2011). Hence the EnKF is only applicable to problems with an unimodal and nearly Gaussian behavior of the underlying PDFs.

Recently, increased efforts have been made to turn sequential Monte Carlo methods viable for large-scale problems. All these efforts have in common that one tries to dynamically steer samples $x_i(t_n)$ to regions of high probability in $\pi_X(x, t_n|\mathrm{Y}_k)$ and hence to maintain nearly uniform weights $w_i(t_n)$ without the need for frequent resampling. We will call these methods guided sequential Monte Carlo (GSMC) methods. Particular instances of GSMC methods have been discussed, for example, by Leeuwen (2010); Chorin et al. (2010); Morzfeld et al. (2012); Bocquet et al. (2010). An alternative line of research is focused on improvements of the EnKF. We mention the rank histogram filter (RHF) of Anderson (2010) and the moment corrected EnKFs of Lei and Bickel (2011).

In this paper, we combine the coupling/transportation perspective on filtering with sequential Monte Carlo methods in order to propose a novel GSMC method. The outline of the paper

is as follows. We will first review Bayes' theorem and its connection to coupling of random variables in Section 2. This will provide us with an abstract Monte Carlo methodology for the combined filtering-prediction problem. Finding exact couplings is impossible in most practical cases which suggest to combine available couplings, such as EnKFs (Evensen, 2006), with sequential Monte Carlo methods. Algorithmic details will be given in Section 3 while a numerical demonstration is provided in Section 4.

2 Bayes' theorem, filtering and coupling of random variables

Recall that we have assumed that observations are taken in intervals of Δt_{obs} which satisfy $\Delta t_{\text{obs}} = \Delta t L$ for an appropriate integer $L \geq 1$. In this context it is helpful to generalize the Chapman-Kolmogorov equation (2) to its L-fold recursive application, i.e.

$$\begin{aligned}\pi_X(x, t_n + \Delta t_{\text{obs}}) &= \int_{\mathbb{R}^N} \cdots \int_{\mathbb{R}^N} \pi(x|x')\, \pi(x'|x'') \cdots \pi(x^{(L-1)}|x^{(L)})\, \pi_X(x^{(L)}, t_n)\, dx' \cdots dx^{(L)} \\ &= \int_{\mathbb{R}^N} \pi_L(x|\tilde{x})\, \pi_X(\tilde{x}, t_n)\, d\tilde{x}.\end{aligned}$$

At the level of PDFs, the sequential data assimilation problem can now be stated as follows: For $j = 0, 1, \ldots$ alternate between

(i) Prediction:

$$\pi_X(x, \mathrm{t}_{j+1}|\mathrm{Y}_j) = \int_{\mathbb{R}^N} \pi_L(x|x')\, \pi_X(x', \mathrm{t}_j|\mathrm{Y}_j)\, dx', \tag{6}$$

(ii) Filtering:

$$\pi_X(x, \mathrm{t}_{j+1}|\mathrm{Y}_{j+1}) = \frac{\pi_Y(y_{\text{obs}}(\mathrm{t}_{j+1})|x)\, \pi_X(x, \mathrm{t}_{j+1}|\mathrm{Y}_j)}{\int_{\mathbb{R}^N} \pi_Y(y_{\text{obs}}(\mathrm{t}_{j+1})|x)\, \pi_X(x, \mathrm{t}_{j+1}|\mathrm{Y}_j)\, dx} \tag{7}$$

with likelihood function

$$\pi_Y(y|x) = \frac{1}{(2\pi)^{K/2}|R|^{1/2}} \exp\left(-\frac{1}{2}(y - Hx)^T R^{-1}(y - Hx)\right)$$

from the forward model (4).

Recall that $\pi_X(x, 0|\mathrm{Y}_0)$ is equal to the given PDF $\pi_X(x, t_0)$ of the initial random variable $X(t_0)$.

We now summarize a few Monte Carlo approaches for sequential data assimilation. We have already discussed a Monte Carlo approach to simulating the Chapman-Kolmogorov equation (2). We formally extend this approach to the data assimilation problem (6)-(7). We introduce the notation $(x_i^f(\mathrm{t}_j), w_i^f(\mathrm{t}_j))$, $i = 1, \ldots, M$, to denote M weighted samples from the forecast (or predicted) distribution $\pi_X(x, \mathrm{t}_j|\mathrm{Y}_{j-1})$ and, correspondingly, $(x_i^a(\mathrm{t}_j), w_i^a(\mathrm{t}_j))$, $i = 1, \ldots, M$, to denote weighted samples from the analysed (or filtered) distribution $\pi_X(x, \mathrm{t}_j|\mathrm{Y}_j)$ at time $\mathrm{t}_j = j\Delta t_{\text{obs}}$. It follows that expectation values $\bar{g}$ of a function $g : \mathbb{R}^N \to \mathbb{R}$ can be approximated according to

$$\bar{g}_M^f = \sum_{i=1}^M w_i^f(\mathrm{t}_j)\, g(x_i^f(\mathrm{t}_j)) \approx \int_{\mathbb{R}^N} g(x)\, \pi_X(x, \mathrm{t}_j|\mathrm{Y}_{j-1})\, dx$$

and

$$\bar{g}_M^a = \sum_{i=1}^M w_i^a(\mathrm{t}_j)\, g(x_i^a(\mathrm{t}_j)) \approx \int_{\mathbb{R}^N} g(x)\, \pi_X(x, \mathrm{t}_j|\mathrm{Y}_j)\, dx,$$

respectively.

The basic sequential Monte Carlo method is based on the following importance sampling approach (Doucet et al., 2001; Bain and Crisan, 2009): For $j = 0, 1, \ldots$ alternate between

(i) Prediction:

$$x_i^f(\mathrm{t}_{j+1}) \sim \pi_L(\cdot|x_i^a(\mathrm{t}_j)), \qquad w_i^f(\mathrm{t}_{j+1}) = w_i^a(\mathrm{t}_j), \tag{8}$$

(ii) Filtering:

$$x_i^a(\mathrm{t}_{j+1}) = x_i^f(\mathrm{t}_{j+1}), \qquad w_i^a(\mathrm{t}_{j+1}) \propto w_i^f(\mathrm{t}_{j+1})\, \pi_Y(y_{\mathrm{obs}}(\mathrm{t}_{j+1})|x_i^f(\mathrm{t}_{j+1})), \tag{9}$$

where the constant of proportionality is chosen such that

$$\sum_{i=1}^M w_i^a(\mathrm{t}_{j+1}) = 1.$$

Due to a possible degeneracy of weights, it is necessary to perform resampling either after each filtering step or whenever an appropriate criterion on the distribution of weights is satisfied. Residual resampling is one of the popular resampling methods (Künsch, 2005; Bain and Crisan, 2009).

We now summarize an alternative approach which leads to constant weights $w_i^f = w_i^a = 1/M$. The basic idea is that of coupling the prior and posterior distributions (Reich, 2011; Moselhy and Marzouk, 2011; Cotter and Reich, 2013). In order to explain this idea in more detail, we simplify the notation in (7) and use the shorthands

$$\pi_X^{\mathrm{prior}}(x) = \pi_X(x, \mathrm{t}_{j+1}|\mathrm{Y}_j), \qquad \pi_X^{\mathrm{post}}(x) = \pi_X(x, \mathrm{t}_{j+1}|\mathrm{Y}_{j+1})$$

for the prior and posterior distributions at t_{j+1}, respectively. A coupling between these two distributions is defined by a joint PDF $\pi_{XZ}(x, z)$ (or more generally by a joint measure μ_{XZ} on $\mathbb{R}^N \times \mathbb{R}^N$) such that

$$\pi_X^{\mathrm{prior}}(x) = \int_{\mathbb{R}^N} \pi_{XZ}(x, z)\, dz$$

and

$$\pi_X^{\mathrm{post}}(z) = \int_{\mathbb{R}^N} \pi_{XZ}(x, z)\, dx,$$

respectively. Given a coupling, one can replace (9) by the following Monte Carlo approach:

(ii) Filtering:

$$x_i^a(\mathrm{t}_{j+1}) \sim \pi_Z(\cdot|x_i^f(\mathrm{t}_{j+1})) := \frac{\pi_{XZ}(x_i^f(\mathrm{t}_{j+1}), \cdot)}{\pi_X^{\mathrm{prior}}(x_i^f(\mathrm{t}_{j+1}))}, \qquad w_i^a(\mathrm{t}_{j+1}) = w_i^f(\mathrm{t}_{j+1}). \tag{10}$$

Hence the filtering step has now the same structural form as (8). The important difference is that the Markov transition kernel $\pi_L(x|x')$ is determined explicitly by the model (1) while such a transition kernel needs to be constructed using a coupling in case of the Bayesian filtering step (7) and depends on the observed $y_{\text{obs}}(\mathrm{t}_{j+1})$. In the literature, this is sometimes called the McKean approach to filtering (Del Moral, 2004).

There is a further difficulty in that the prior and posterior PDFs are not explicitly available in the context of sequential Monte Carlo methods and that approximations $\bar{\pi}_X^{\text{prior}}$ and $\bar{\pi}_X^{\text{post}}$, respectively, need to be estimated from the available samples x_i^f, their weights w_i^f and likelihoods $\pi_Y(y_{\text{obs}}|x_i^f)$ either via parametric or non-parametric statistics (Wand and Jones, 1995; Hastie et al., 2009).

We now continue with a simple demonstration of the concept of coupling by means of two univariate Gaussian random variables $X \sim \mathrm{N}(\bar{x}, \sigma_{xx}^2)$ and $Z \sim \mathrm{N}(\bar{z}, \sigma_{zz}^2)$. Since the marginal are given a priori, a joint Gaussian has to be of the form $\mathrm{N}(m, P)$ with mean $m = (\bar{x}, \bar{z})^T \in \mathbb{R}^2$ and covariance matrix

$$P = \begin{pmatrix} \sigma_{xx}^2 & \sigma_{xz}^2 \\ \sigma_{zx}^2 & \sigma_{zz}^2 \end{pmatrix} \in \mathbb{R}^{2\times 2},$$

where the only free parameter $\sigma_{xz}^2 = \sigma_{zx}^2$ has to satisfy

$$\sigma_{xx}^2\sigma_{zz}^2 - \sigma_{xz}^4 > 0$$

to make P symmetric positive definite. Setting $\sigma_{xz} = 0$ implies independence of X and Z and is equivalent to defining a coupling via the product PDF

$$\pi_{XZ}(x, z) = \pi_X^{\text{prior}}(x)\pi_X^{\text{post}}(z)$$

in the general case. We now consider $\sigma_{xz}^2 > 0$ and recall that (10) requires the conditional PDF $\pi_Z(z|x)$ which in this example is characterized by the conditional mean

$$z = \bar{z} + \frac{\sigma_{xz}^2}{\sigma_{xx}^2}\,(x - \bar{x})$$

and the variance

$$\sigma^2 = \sigma_{zz}^2 - \sigma_{zx}^2\sigma_{xx}^{-2}\sigma_{xz}^2.$$

If one sets $\sigma_{xz}^2 = \sqrt{\sigma_{xx}^2\sigma_{zz}^2}$, then σ^2 becomes zero and one obtains a deterministic coupling of the two random variables via

$$Z = \bar{z} + \frac{\sigma_{zz}}{\sigma_{xx}}(X - \bar{x}),$$

which amounts to the well-known transformation of univariate Gaussians under a linear function.

Inspired by this example we will search for general deterministic couplings $Z = T(X)$ with associated joint probability measure

$$\mu_{XZ}(dx, dz) = \delta(z - T(x))\,\pi_X^{\text{prior}}(x)\,dxdz$$

from which it follows via margenalization that T has to satisfy

$$\pi_X^{\text{post}}(T(x))\,|DT(x)| = \pi_X^{\text{prior}}(x) \tag{11}$$

(compare (3)). Once a deterministic coupling has been found, (10) can be replaced by

(ii) Filtering:

$$x_i^a(\mathrm{t}_{j+1}) = T_{j+1}(x_i^f(\mathrm{t}_{j+1})), \qquad w_i^a(\mathrm{t}_{j+1}) = w_i^f(\mathrm{t}_{j+1}) \tag{12}$$

where T_{j+1} is a transport map, satisfying (11) at t_{j+1}, given the prior $\pi_X(x, \mathrm{t}_{j+1}|\mathrm{Y}_j)$ and the measurement $y_{\mathrm{obs}}(\mathrm{t}_{j+1})$.

Optimality in the sense of Monge-Kantorovitch (Villani, 2009) is defined by

$$\mu_{XZ}^* = \arg \inf_{\mu_{XZ} \in \Pi} \int_{\mathbb{R}^N \times \mathbb{R}^N} \|x - z\|^2 \mu_{XZ}(dx, dz),$$

where the infimum runs over the set of all couplings μ_{XZ}, denoted by Π, with marginals π_X^{prior} and π_X^{post}. We first note that finding the optimal coupling between empirical measures

$$\mu_X^{\mathrm{prior}}(dx) = \frac{1}{M} \sum_{i=1}^M \delta(x - x_i^f) dx$$

and

$$\mu_X^{\mathrm{post}}(dz) = \sum_{i=1}^M w_i \delta(z - x_i^f) dz$$

leads to a linear programming problem (Cotter and Reich, 2013; Reich, 2012b). Another key result of optimal transportation states that the optimal coupling is induced by a transport map $T(x)$ for sufficiently regular prior PDFs π_X^{prior} (Villani, 2009). Furthermore, the transport map satisfies $T(x) = \nabla_x \psi(x)$ for an appropriate potential $\psi : \mathbb{R}^N \to \mathbb{R}$. It follows from (11) that the potential ψ has to satisfy the nonlinear elliptic PDE

$$\pi_X^{\mathrm{post}}(\nabla_x \psi(x)) \, |D \nabla_x \psi(x)| = \pi_X^{\mathrm{prior}}(x). \tag{13}$$

For univariate prior and posterior random variables with cumulative probability distribution functions

$$F_{\mathrm{prior}}(x) = \int_{-\infty}^x \pi_X^{\mathrm{prior}}(x') \, dx'$$

and $F_{\mathrm{post}}(x)$, respectively, a transport map is easily found via

$$z = T(x) = F_{\mathrm{post}}^{-1}(F_{\mathrm{prior}}(x)). \tag{14}$$

It becomes however computationally infeasible to solve (13) for ψ in case the dimension N of phase space is large and/or π_X^{prior} is non-Gaussian. Being forced to give up the idea of strict optimality, one can resort to an idea of Moser (1965) (see also Villani (2009)) to find *a* deterministic coupling. Moser suggested to utilize a dynamic embedding of the form

$$\frac{dx}{ds} = -\frac{1}{\pi_s(x)} \nabla_x \phi(x), \tag{15}$$

with linearly interpolated PDFs

$$\pi_s = (1 - s)\pi_X^{\mathrm{prior}} + s\pi_X^{\mathrm{post}}, \qquad s \in [0, 1],$$

and the potential ϕ determined by the Poisson equation

$$\nabla_x \cdot (\nabla_x \phi) = -\pi_X^{\mathrm{prior}} + \pi_X^{\mathrm{post}}.$$

The desired transport map T is defined as the time-one flow map of the ODE (15). The embedding technique of Moser has been applied and refined for the Bayesian filtering step in sequential data assimilation by Reich (2011). It has been demonstrated by Reich (2012a) that explicit solutions to the embedding technique (15) can be found in case $\pi_X^{\rm prior}$ is a multivariate Gaussian or a mixture of multivariate Gaussians.

Furthermore, the popular family of EnKFs (Evensen, 2006) can be viewed as providing couplings under the assumption that the prior distribution is approximated at t_{j+1} by a multivariate Gaussian with mean $\bar{x}^f$ and covariance matrix P^f. For example, the EnKF with perturbed observations leads to a non-deterministic (non-optimal) coupling, which gives rise to

$$x_i^a(\mathrm{t}_{j+1}) \sim \pi_Z(\cdot|x_i^f(\mathrm{t}_{j+1})) = \mathrm{N}(x_i^f(\mathrm{t}_{j+1}) - K(Hx_i^f(\mathrm{t}_{j+1}) - \mathrm{y}_{\rm obs}), KRK^T)$$

in (10) with Kalman gain matrix

$$K = P^f H^T (HP^f H^T + R)^{-1}.$$

See also the discussion in Cotter and Reich (2013) on an optimal coupling for EnKFs based on the work of Olkin and Pukelsheim (1982).

We will now utilize available couplings for Bayesian inference in order to derive GSMC methods for more general classes of prior and posterior distributions.

3 A guided sequential Monte Carlo (GSMC) method

We assume that, at initial time $\mathrm{t}_0 = 0$, an ensemble of M independent samples $x_i(0) \in \mathbb{R}^N$ from the given PDF $\pi_X(x, 0)$ is being generated. Each sample is given an initial weight of $w_i^f(0) = 1/M$.

In between observations, the ensemble is propagated under the dynamical model (5). The initial conditions for each simulation interval are provided by the analysed ensemble members $x_i^a(\mathrm{t}_j)$ from the most current data assimilation step. The model predictions at the next observation point t_{j+1} are denoted by $x_i^f(\mathrm{t}_{j+1})$. The weights $w_i(\mathrm{t}_j)$ do not change during model simulations. Observed values $y(\mathrm{t}_j) \in \mathbb{R}^K$ are assimilated in time intervals of $\Delta t_{\rm obs}$ using the forward model (4).

An essential ingredient of the proposed GSMC method is to find an appropriate estimate $\bar{\pi}_X^{\rm prior}(x, \mathrm{t}_{j+1})$ of the prior PDF from the weighted samples $(x_i^f(\mathrm{t}_{j+1}), w_i^f(\mathrm{t}_{j+1}))$, $i = 1, \ldots, M$. For example, the prior distribution at time t_{j+1} can be approximated using a Gaussian mixture or a Gaussian kernel density estimator (Wand and Jones, 1995) of the form

$$\bar{\pi}_X^{\rm prior}(x, \mathrm{t}_{j+1}) = \sum_{i=1}^M \frac{w_i^f}{(2\pi h)^{N/2}|P^f|^{1/2}} \exp\left(-\frac{1}{2h}(x - x_i^f)^T (P^f)^{-1}(x - x_i^f)\right),$$

where $x_i^f = x_i^f(\mathrm{t}_{j+1})$, $w_i^f = w_i^f(\mathrm{t}_{j+1})$, P^f denotes the empirical covariance matrix of the forecast ensemble, and $1 \geq h > 0$ is the bandwidth of the estimator. From now on we will drop the time argument and assume that all relevant quantities are computed for $t = \mathrm{t}_{j+1}$ unless indicated otherwise.

Our GSMC approach relies on an appropriate transport map $\hat{T} : \mathbb{R}^N \to \mathbb{R}^N$, which will depend on both the forecast ensemble $\{(x_i^f, w_i^f)\}_{i=1}^M$ and the measured $y_{\rm obs}$. It should be chosen such that the transformed posterior distribution $\tilde{\pi}_X^{\rm post}$, defined according to (11) by

$$\tilde{\pi}_X^{\rm post}(\hat{T}(x))|D\hat{T}(x)| = \bar{\pi}_X^{\rm prior}(x),$$

is close to the desired posterior distribution

$$\bar{\pi}_X^{\text{post}}(x) := \pi_X(x|y_{\text{obs}}) \propto \pi_Y(y_{\text{obs}}|x)\bar{\pi}_X^{\text{prior}}(x).$$

Following the idea of importance sampling, we now define the analysed ensemble by $x_i^a = \hat{T}(x_i^f)$ with updated weights

$$\begin{aligned} w_i^a &\propto w_i^f \frac{\pi_Y(y_{\text{obs}}|x_i^a)\,\bar{\pi}_X^{\text{prior}}(x_i^a)}{\tilde{\pi}_X^{\text{post}}(x_i^a)} \\ &= w_i^f\, \pi_Y(y_{\text{obs}}|x_i^a)\,|D\hat{T}(x_i^f)|\,\frac{\bar{\pi}_X^{\text{prior}}(x_i^a)}{\bar{\pi}_X^{\text{prior}}(x_i^f)}, \end{aligned}$$

where the normalization constant is chosen such that $\sum_i w_i^a = 1$.

Furthermore, we will assume that the transport map $\hat{T}$ couples a prior $\hat{\pi}_X^{\text{prior}}$ and its associated posterior PDF $\hat{\pi}_X^{\text{post}}$ exactly. Such transport maps exists for Gaussian prior PDFs as well as Gaussian mixture prior PDFs. The coupling property implies that

$$|D\hat{T}(x)| = \frac{\hat{\pi}_X^{\text{prior}}(x)}{\hat{\pi}_X^{\text{post}}(\hat{T}(x))}$$

and, furthermore, since

$$\hat{\pi}_X^{\text{post}}(\hat{T}(x)) \propto \pi_Y(y_{\text{obs}}|\hat{T}(x))\,\hat{\pi}_X^{\text{prior}}(\hat{T}(x))$$

we may conclude that

$$|D\hat{T}(x)| \propto \frac{\hat{\pi}_X^{\text{prior}}(x)}{\pi_Y(y_{\text{obs}}|\hat{T}(x))\,\hat{\pi}_X^{\text{prior}}(\hat{T}(x))}.$$

Combining all these results leads to the modified filtering step:

(ii) Filtering:

$$x_i^a = \hat{T}(x_i^f), \qquad w_i^a \propto w_i^f\, \frac{\hat{\pi}_X^{\text{prior}}(x_i^f)}{\hat{\pi}_X^{\text{prior}}(x_i^a)}\, \frac{\bar{\pi}_X^{\text{prior}}(x_i^a)}{\bar{\pi}_X^{\text{prior}}(x_i^f)}, \tag{16}$$

where the time argument t_{j+1} has been dropped for notational convenience. Here $\bar{\pi}_X^{\text{prior}}(x)$ denotes an estimate of the true underlying prior and $\hat{\pi}_X^{\text{prior}}(x)$ denotes a PDF which allows for the computation of a transport map $\hat{T}(x)$.

The PDF $\hat{\pi}_X^{\text{prior}}(x)$ should be chosen such that particle weights remain nearly uniform. If the accumulated weights become however strongly non-uniform, particles can be resampled by the same techniques as being employed for traditional sequential Monte Carlo methods.

We mention that the proposed GSMC algorithm can be extended to the case that an exact coupling cannot be found for the given likelihood $\pi_Y(y|x)$ and a simplified likelihood $\hat{\pi}_Y(y|x)$ needs be employed. In that case one would use

$$w_i^a \propto w_i^f\, \frac{\pi_Y(y_{\text{obs}}|x_i^a)}{\hat{\pi}_Y(y_{\text{obs}}|x_i^a)}\, \frac{\hat{\pi}_X^{\text{prior}}(x_i^f)}{\hat{\pi}_X^{\text{prior}}(x_i^a)}\, \frac{\pi_X^{\text{prior}}(x_i^a)}{\pi_X^{\text{prior}}(x_i^f)}$$

in (16).

We next provide a simple numerical demonstration of the proposed GSMC method using one-dimensional Brownian dynamics.

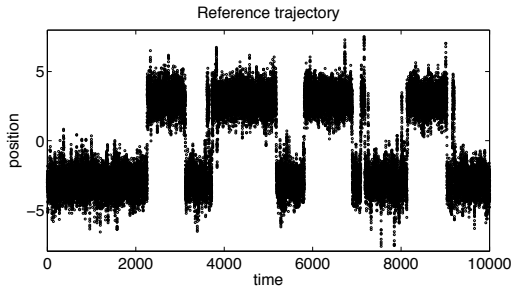

Figure 1: Shown is the reference solution from which observations are generated by adding Gaussian noise with mean zero and variance $R = 36$.

4 Brownian dynamics under a double well potential

We consider one-dimensional Brownian dynamics

$$dx = -V'(x)\,dt + dW(t)$$

with potential

$$V(x) = \cos(x) + \frac{3}{4}\left(\frac{x}{6}\right)^4$$

and standard Brownian motion $W(t)$ as a test example. The stochastic equations are solved numerically by the Euler-Mayuama method, i.e.

$$x_{n+1} = x_n - \Delta t V'(x_n) + \sqrt{\Delta t} Z_n$$

with step-size $\Delta t = 0.1$ and $Z_n \sim \mathrm{N}(0, 1)$.

The measurement equation is

$$Y = X + \sqrt{R}\Xi$$

with $\Xi \sim \mathrm{N}(0, 1)$ and $R = 36$ is the variance of the measurement error. Actual measurements are obtained from a reference trajectory (see Figure 1) of our Brownian dynamics model and added measurement noise. Ensemble sizes vary between $M = 20, 50$ and 100. Measurements are taken every 10 units of time (i.e. $\Delta t_{\mathrm{obs}} = 10\Delta t$) and a total of 1,000 assimilation steps are performed.

We compute a highly accurate reference solution by solving the Fokker-Planck equation

$$\frac{\partial \pi_X}{\partial t} = \frac{1}{2}\frac{\partial}{\partial x}\left(\pi_X^* \frac{\partial \pi_X}{\partial x}\right)$$

with canonical PDF

$$\pi_X^*(x) \propto \exp(-V(x))$$

over a computational grid with mesh-size $\Delta x = 1/16$. The results are used to approximate the prediction step (6) directly on the level of PDFs. The grid approximations to $\pi_X(x, \mathrm{t}_{j+1}|\mathrm{Y}_j)$ are then used to find grid approximation to the analysed PDFs $\pi_X(x, \mathrm{t}_{j+1}|\mathrm{Y}_{j+1})$ using (7). These density approximations are finally used to approximate the time-evolved means $\bar{x}(t_n)$ which are taken as a reference for the ensemble-based filter algorithms.

Given a set of particles $x_i^f \in \mathbb{R}$ with weights $w_i > 0$ we compute the unweighted ensemble mean

$$\bar{x}^f = \frac{1}{M} \sum_{i=1}^{M} x_i^f$$

and the corresponding covariance matrix

$$P^f = \frac{1}{M-1} \sum_{i=1}^{M} (x_i^f - \bar{x}^f)^2,$$

which implies a Gaussian prior

$$\hat{\pi}_X^{\text{prior}}(x) = \frac{1}{(2\pi P^f)^{1/2}} \exp\left(-\frac{1}{2P^f}(x - \bar{x}^f)^2 \right)$$

for the application of the ensemble square root filter as a proposal map $\hat{T}(x)$.

For the implementation of the GSMC approach we assume furthermore that particles are given an index $\alpha_i \in \{-1, 1\}$ indicating whether they belong to the left or the right, respectively, potential well, i.e. $\alpha_i = +1$ if $x_i^f > 0$ and $\alpha_i = -1$ if $x_i^f < 0$. We then approximate the prior distribution $\pi_X(x, \mathrm{t}_{j+1}|\mathrm{Y}_j)$ by

$$\bar{\pi}_X^{\text{prior}}(x) = \frac{\gamma_{-1}}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \bar{x}_{-1})^2 \right) + \frac{\gamma_{+1}}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \bar{x}_{+1})^2 \right)$$

with

$$\gamma_j = \sum_{i=1}^{M} \delta_{j,\alpha_i} w_i, \qquad j \in \{-1, +1\},$$

where δ_{jk} denotes the Kronecker delta,

$$\bar{x}_j = \sum_{i=1}^{M} \delta_{j,\alpha_i} w_i x_i^f$$

and fixed standard deviation $\sigma = \sqrt{2}$. Clearly, a more sophisticated Gaussian mixture model could be fitted to the available ensembles using the expectation-maximization (EM) algorithm (Dempster et al., 1977).

We also implemented the RHF of Anderson (2010) using piecewise linear ensemble based approximations to the prior and posterior cumulative distribution functions F_{prior}, F_{post} and subsequent construction of a transport map for (12) using formula (14).

Numerical results are presented in Table 1, where the root mean square (RMS) error is defined as

$$\text{RMS error} = \sqrt{\frac{1}{J} \sum_{j=1}^{J} (\bar{x}^a(\mathrm{t}_j) - \bar{x}^{\text{ref}}(\mathrm{t}_j))^2},$$

with $\bar{x}^a(\mathrm{t}_j)$ denoting the analysed ensemble average from the filter at time t_j and $\bar{x}^{\text{ref}}(\mathrm{t}_j)$ its numerical approximation from the Fokker-Planck approach.

The RHF with a transport map based on cumulative distribution functions yields the most accurate filter results. In fact, the RHF converges to the analytic filtering solution as $M \to$

Table 1: RMS errors for ensemble means obtained from an ensemble square root filter (EnKF), the ensemble Gaussian mixture filter (EGMF), the guided sequential Monte Carlo method (GSMC), and the rank histogram filter (RHF) compared to the expected value computed by a Fokker-Planck discretization with error variance $R = 36$ and ensemble sizes of $M = 20, 50, 100$ particles/ensemble members.

	EnKF	EGMF	GSMC	RHF
$M = 20$	1.1590	0.7683	1.0200	0.6551
$M = 50$	1.0701	0.5127	0.7172	0.3717
$M = 100$	1.0477	0.4033	0.6534	0.2691

∞. The second best result is obtained for the ensemble Gaussian mixture filter (EGMF) of Reich (2012a) for which a transport map is constructed using a binary Gaussian mixture approximation for the distributions in x using the EM algorithm. We also find that the GSMC approach yields an improvement over the EnKF while not delivering results as accurate as those from the RHF and the EGMF. Note that the GSMC implementation with $\hat{T}$ based on a ensemble square root filter can be viewed as an inexpensive post-processing step to the associated EnKF algorithm.

5 Conclusions

From a mathematical perspective the coupling and optimal transportation approach to Bayesian data assimilation offers attractive opportunities. Practical implementations are limited by the fact that optimal transport maps are difficult to compute numerically. It has, however, been demonstrated for a wide range of problems that rather crude approximations to the coupling/transport problem, such as the EnKF, can lead to surprisingly robust data assimilation algorithms. In this paper, we have followed a recent trend of combining crude ensemble transform approximations with sequential Monte Carlo methods. More specifically, we have proposed an importance sampling approach for post-processing existing ensemble transform filter formulations. Such an approach should be useful whenever the underlying ensemble transform filter is capable of tracking regions of high posterior probability while being of limited statistical accuracy. Under those circumstances, postprocessing the data should lead to improved statistics.

References

J.L. Anderson. A non-Gaussian ensemble filter update for data assimilation. *Monthly Weather Review*, 138:4186–4198, 2010.

A. Bain and D. Crisan. *Fundamentals of stochastic filtering*, volume 60 of *Stochastic modelling and applied probability*. Springer-Verlag, New-York, 2009.

Th. Bengtsson, P. Bickel, and B. Li. Curse-of-dimensionality revisited: Collapse of the particle

filter in very large scale systems. In D. Nolan and T. Speed, editors, *Probability and Statistics: Essays in Honor of David A. Freedman*, pages 316–334. Institute of Mathematical Statistics, Beachwood, Ohio, USA, 2008.

M. Bocquet, C.A. Pires, and L. Wu. Beyond Gaussian statistical modeling in geophysical data assimilaition. *Mon. Wea. Rev.*, 138:2997–3022, 2010.

A.J. Chorin, M. Morzfeld, and X. Tu. Implicit filters for data assimilation. *Comm. Appl. Math. Comp. Sc.*, 5:221–240, 2010.

C.J. Cotter and S. Reich. Ensemble filter techniques for intermittent data assimilation - a survey. In H.W. Engel et al., editor, *Radon Series on Computational and Applied Mathematics*, page in press. De Gruyter, Boston, 2013.

P. del Moral. *Feynman-Kac formulae: Genealogical and interacting particle systems with applications*. Springer-Verlag, New York, 2004.

A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Soc.*, 39B:1–38, 1977.

A. Doucet, N. de Freitas, and N. Gordon (eds.). *Sequential Monte Carlo methods in practice*. Springer-Verlag, Berlin Heidelberg New York, 2001.

G. Evensen. *Data assimilation. The ensemble Kalman filter*. Springer-Verlag, New York, 2006.

C.W. Gardiner. *Handbook on stochastic methods*. Springer-Verlag, 3rd edition, 2004.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer-Verlag, New York, 2nd edition, 2009.

A.H. Jazwinski. *Stochastic processes and filtering theory*. Academic Press, New York, 1970.

P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer-Verlag, Berlin Heidelberg New York, 1992.

H.R. Künsch. Rekursive Monte Carlo filter: Algorithms and theoretical analysis. *Ann. Statist.*, pages 1983–2021, 2005.

P.J. Van Leeuwen. Nonlinear data assimilation in the geosciences: an extremely efficient particle filter. *Q.J.R. Meteorolog. Soc.*, 136:1991–1996, 2010.

J. Lei and P. Bickel. A moment matching ensemble filter for nonlinear and non-Gaussian data assimilation. *Mon. Weath. Rev.*, 139:3964–3973, 2011.

M. Morzfeld, X. Tu, E. Atkins, and A.J. Chorin. A random map implementation of implicit filters. *J. Comput. Phys.*, 231:2049–2066, 2012.

T.A. El Moselhy and Y.M. Marzouk. Bayesian inference with optimal maps. *submitted*, 2011.

J. Moser. On the volume elements on a manifold. *Trans. Amer. Math. Soc.*, 120:286–294, 1965.

I. Olkin and F. Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257–263, 1982.

S. Reich. A dynamical systems framework for intermittent data assimilation. *BIT Numer Math*, 51:235–249, 2011.

S. Reich. A Gaussian mixture ensemble transform filter. *Q. J. R. Meterolog. Soc.*, 138:222–233, 2012a.

S. Reich. A non-parametric ensemble transform method for Bayesian inference. Technical Report arXiv:1210.0376, Potsdam University, 2012b.

C. Villani. *Optimal transportation: Old and new*. Springer-Verlag, Berlin Heidelberg, 2009.

M.P. Wand and M.C. Jones. *Kernel smoothing*. Chapmann and Hall, London, 1995.