Acta Numerica (2019), pp. 635–711 Published by Cambridge University Press.

doi:10.1017/S0962492919000011 Printed in the United Kingdom

Data assimilation: The Schrödinger perspective

Sebastian Reich
Institute of Mathematics,
University of Potsdam,
D-14476 Potsdam, Germany
and
Department of Mathematics and Statistics,
University of Reading,
Reading RG6 6AX, UK
E-mail: sebastian.reich@uni-potsdam.de

Data assimilation addresses the general problem of how to combine model-based predictions with partial and noisy observations of the process in an optimal manner. This survey focuses on sequential data assimilation techniques using probabilistic particle-based algorithms. In addition to surveying recent developments for discrete- and continuous-time data assimilation, both in terms of mathematical foundations and algorithmic implementations, we also provide a unifying framework from the perspective of coupling of measures, and Schrödinger's boundary value problem for stochastic processes in particular.

CONTENTS

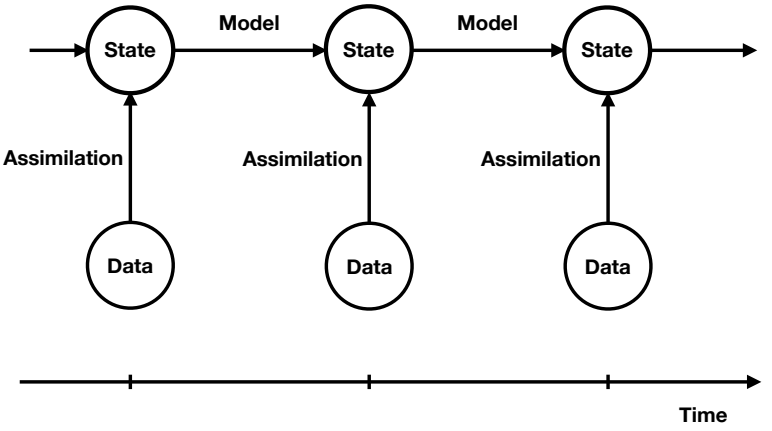


Figure 1.1. Schematic illustration of sequential data assimilation, where model states are propagated forward in time under a given model dynamics and adjusted whenever data become available at discrete instances in time. In this paper, we look at a single transition from a given model state conditioned on all the previous and current data to the next instance in time, and its adjustment under the assimilation of the new data then becoming available.

1. Introduction

This survey focuses on sequential data assimilation techniques for state and parameter estimation in the context of discrete- and continuous-time stochastic diffusion processes. See Figure 1.1. The field itself is well established (Evensen 2006, Särkkä 2013, Law, Stuart and Zygalakis 2015, Reich and Cotter 2015, Asch, Bocquet and Nodet 2017), but is also undergoing continuous development due to new challenges arising from emerging application areas such as medicine, traffic control, biology, cognitive sciences and geosciences.

Data assimilation is typically formulated within a Bayesian framework in order to combine partial and noisy observations with model predictions and their uncertainties with the goal of adjusting model states and model parameters in an optimal manner. In the case of linear systems and Gaussian distributions, this task leads to the celebrated Kalman filter (Särkkä 2013) which even today forms the basis of a number of popular data assimilation schemes and which has given rise to the widely used ensemble Kalman filter (Evensen 2006). Contrary to standard sequential Monte Carlo methods (Doucet, de Freitas and Gordon 2001, Bain and Crisan 2008), the ensemble Kalman filter does not provide a consistent approximation to the sequential filtering problem, while being applicable to very high-dimensional problems. This and other advances have widened the scope of sequential data assimilation and have led to an avalanche of new methods in recent years.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

In this review we will focus on probabilistic methods (in contrast to data assimilation techniques based on optimization, such as 3DVar and 4DVar (Evensen 2006, Law *et al.* 2015)) in the form of sequential particle methods. The essential challenge of sequential particle methods is to convert a sample of M particles from a filtering distribution at time t_k into M samples from the filtering distribution at time t_{k+1} without having access to the full filtering distributions. It will also often be the case in practical applications that the sample size will be small to moderate in comparison to the number of variables we need to estimate.

Sequential particle methods can be viewed as a special instance of interacting particle systems (del Moral 2004). We will view such interacting particle systems in this review from the perspective of approximating a certain boundary value problem in the space of probability measures, where the boundary conditions are provided by the underlying stochastic process, the data and Bayes' theorem. This point of view leads naturally to optimal transportation (Villani 2003, Reich and Cotter 2015) and, more importantly for this review, to Schrödinger's problem (Föllmer and Gantert 1997, Leonard 2014, Chen, Georgiou and Pavon 2014), as formulated first by Erwin Schrödinger in the form of a boundary value problem for Brownian motion (Schrödinger 1931).

This paper has been written with the intention of presenting a unifying framework for sequential data assimilation using coupling of measure arguments provided through optimal transportation and Schrödinger's problem. We will also summarize novel algorithmic developments that were inspired by this perspective. Both discrete- and continuous-time processes and data sets will be covered. While the primary focus is on state estimation, the presented material can be extended to combined state and parameter estimation. See Remark 2.2 below.

Remark 1.1. We will primary refer to the methods considered in the survey as particle or ensemble methods instead of the widely used notion of sequential Monte Carlo methods. We will also use the notions of particles, samples and ensemble members synonymously. Since the ensemble size, M, is generally assumed to be small to moderate relative to the number of variables of interest, we will focus on robust but generally biased particle methods.

1.1. Overall organization of the paper

This survey consists of four main parts. We start Section 2 by recalling key mathematical concepts of sequential data assimilation when the data become available at discrete instances in time. Here the underlying dynamic models can be either continuous (*i.e.* generated by a stochastic differential

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

equation) or discrete-in-time. Our initial review of the problem will lead to the identification of three different scenarios of performing sequential data assimilation, which we denote by (A), (B) and (C). While the first two scenarios are linked to the classical importance resampling and optimal proposal densities for particle filtering (Doucet *et al.* 2001), scenario (C) builds upon an intrinsic connection to a certain boundary value problem in the space of joint probability measures first considered by Erwin Schrödinger (1931).

After this initial review, the remaining parts of Section 2 provide more mathematical details on prediction in Section 2.1, filtering and smoothing in Section 2.2, and the Schrödinger approach to sequential data assimilation in Section 2.3. The modification of a given Markov transition kernel via a twisting function will arise as a crucial mathematical construction and will be introduced in Sections 1.2 and 2.1. The next major part of the paper, Section 3, is devoted to numerical implementations of prediction, filtering and smoothing, and the Schrödinger approach as relevant to scenarios (A)–(C) introduced earlier in Section 2. More specifically, this part will cover the ensemble Kalman filter and its extensions to the more general class of linear ensemble transform filters as well as the numerical implementation of the Schrödinger approach to sequential data assimilation using the Sinkhorn algorithm (Sinkhorn 1967, Peyre and Cuturi 2018). Discrete-time stochastic systems with additive Gaussian model errors and stochastic differential equations with constant diffusion coefficient serve as illustrating examples throughout both Sections 2 and 3.

Sections 2 and 3 are followed by two sections on the assimilation of data that arrive continuously in time. In Section 4 we will distinguish between data that are smooth as a function of time and data which have been perturbed by Brownian motion. In both cases, we will demonstrate that the data assimilation problem can be reformulated in terms of so-called mean-field equations, which produce the correct conditional marginal distributions in the state variables. In particular, in Section 4.2 we discuss the feedback particle filter of Yang, Mehta and Meyn (2013) in some detail. The final section of this review, Section 5, covers numerical approximations to these mean-field equations in the form of interacting particle systems. More specifically, the continuous-time ensemble Kalman–Bucy and numerical implementations of the feedback particle filter will be covered in detail. It will be shown in particular that the numerical implementation of the feedback particle filter can be achieved naturally via the approximation of an associated Schrödinger problem using the Sinkhorn algorithm.

In the appendices we provide additional background material on mesh-free approximations of the Fokker–Planck and backward Kolmogorov equations (Appendix A.1), on the regularized Störmer–Verlet time-stepping methods for the hybrid Monte Carlo method, applicable to Bayesian inference

problems over path spaces (Appendix A.2), on the ensemble Kalman filter (Appendix A.3), and on the numerical approximation of forward–backward stochastic differential equations (SDEs) (Appendix A.4).

1.2. Summary of essential notations

We typically denote the probability density function (PDF) of a random variable Z by π. Realizations of Z will be denoted by $z = Z(\omega)$.

Realizations of a random variable can also be continuous functions/paths, in which case the associated probability measure on path space is denoted by $\mathbb{Q}$. We will primarily consider continuous functions over the unit time interval and denote the associated random variable by $Z_{[0,1]}$ and its realizations $Z_{[0,1]}(\omega)$ by $z_{[0,t]}$. The restriction of $Z_{[0,1]}$ to a particular instance $t \in [0,1]$ is denoted by Z_t with marginal distribution π_t and realizations $z_t = Z_t(\omega)$.

For a random variable Z having only finitely many outcomes z^i, $i = 1, \ldots, M$, with probabilities p_i, that is,

$$\mathbb{P}[Z(\omega) = z^i] = p_i,$$

we will work with either the probability vector $p = (p_1, \ldots, p_M)^{\mathrm{T}}$ or the empirical measure

$$\pi(z) = \sum_{i=1}^{M} p_i \, \delta(z - z^i),$$

where $\delta(\cdot)$ denotes the standard Dirac delta function.

We use the shorthand

$$\pi[f] = \int f(z) \, \pi(z) \, \mathrm{d}z$$

for the expectation of a function f under a PDF π. Similarly, integration with respect to a probability measure $\mathbb{Q}$, not necessarily absolutely continuous with respect to Lebesgue, will be denoted by

$$\mathbb{Q}[f] = \int f(z) \, \mathbb{Q}(\mathrm{d}z).$$

The notation $\mathbb{E}[f]$ is used if we do not wish to specify the measure explicitly.

The PDF of a Gaussian random variable Z with mean $\bar{z}$ and covariance matrix B will be abbreviated by $\mathrm{n}(z; \bar{z}, B)$. We also use $Z \sim \mathrm{N}(\bar{z}, B)$.

Let $u \in \mathbb{R}^N$, then $D(u) \in \mathbb{R}^{N \times N}$ denotes the diagonal matrix with entries $(D(u))_{ii} = u_i$, $i = 1, \ldots, N$. We also denote the $N \times 1$ vector of ones by $\mathbb{1}_N = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^N$.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

A matrix $P \in \mathbb{R}^{L\times M}$ is called bi-stochastic if all its entries are non-negative, which we will abbreviate by $P \geq 0$, and

$$\sum_{l=1}^{L} q_{li} = p_0, \quad \sum_{i=1}^{M} q_{li} = p_1,$$

where both $p_1 \in \mathbb{R}^L$ and $p_0 \in \mathbb{R}^M$ are probability vectors. A matrix $Q \in \mathbb{R}^{M\times M}$ defines a discrete Markov chain if all its entries are non-negative and

$$\sum_{l=1}^{L} q_{li} = 1.$$

The Kullback–Leibler divergence between two bi-stochastic matrices $P \in \mathbb{R}^{L\times M}$ and $Q \in \mathbb{R}^{L\times M}$ is defined by

$$\mathrm{KL}\,(P||Q) := \sum_{l,j} p_{lj} \log \frac{p_{lj}}{q_{lj}}.$$

Here we have assumed for simplicity that $q_{lj} > 0$ for all entries of Q. This definition extends to the Kullback–Leibler divergence between two discrete Markov chains.

The transition probability going from state z_0 at time $t = 0$ to state z_1 at time $t = 1$ is denoted by $q_+(z_1|z_0)$. Hence, given an initial PDF $\pi_0(z_0)$ at $t = 0$, the resulting (prediction or forecast) PDF at time $t = 1$ is provided by

$$\pi_1(z_1) := \int q_+(z_1|z_0)\, \pi_0(z_0)\, \mathrm{d}z_0. \tag{1.1}$$

Given a twisting function $\psi(z) > 0$, the twisted transition kernel $q_+^\psi(z_1|z_0)$ is defined by

$$q_+^\psi(z_1|z_0) := \psi(z_1)\, q_+(z_1|z_0)\, \widehat{\psi}(z_0)^{-1} \tag{1.2}$$

provided

$$\widehat{\psi}(z_0) := \int q_+(z_1|z_0)\, \psi(z_1)\, \mathrm{d}z_1 \tag{1.3}$$

is non-zero for all z_0. See Definition 2.8 for more details.

If transitions are characterized by a discrete Markov chain $Q_+ \in \mathbb{R}^M$, then a twisted Markov chain is provided by

$$Q_+^u = D(u)\, Q_+\, D(v)^{-1}$$

for given twisting vector $u \in \mathbb{R}^M$ with positive entries u_i, *i.e.* $u > 0$, and the vector $v \in \mathbb{R}^M$ determined by

$$v = (D(u)\, Q_+)^{\mathrm{T}}\, \mathbb{1}_M.$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

The conditional probability of observing y given z is denoted by $\pi(y|z)$ and the likelihood of z given an observed y is abbreviated by $l(z) = \pi(y|z)$. We will also use the abbreviations

$$\widehat{\pi}_1(z_1) = \pi_1(z_1|y_1)$$

and

$$\widehat{\pi}_0(z_0) = \pi_0(z_0|y_1)$$

to denote the conditional PDFs of a process at time $t = 1$ given data at time $t = 1$ (filtering) and the conditional PDF at time $t = 0$ given data at time $t = 1$ (smoothing), respectively. Finally, we also introduce the evidence

$$\beta := \pi_1[l] = \int p(y_1|z_1)\pi_1(z_1)\,\mathrm{d}z_1$$

of observing y_1 under the given model as represented by the forecast PDF (1.1). A more precise definition of these expressions will be given in the following section.

2. Mathematical foundation of discrete-time DA

Let us assume that we are given partial and noisy observations y_k, $k = 1, \ldots, K$, of a stochastic process in regular time intervals of length $T = 1$. Given a likelihood function $\pi(y|z)$, a Markov transition kernel $q_+(z'|z)$ and an initial distribution π_0, the associated prior and posterior PDFs are given by

$$\pi(z_{0:K}) := \pi_0(z_0) \prod_{k=1}^{K} q_+(z_k|z_{k-1}) \tag{2.1}$$

and

$$\pi(z_{0:K}|y_{1:K}) := \frac{\pi_0(z_0) \prod_{k=1}^{K} \pi(y_k|z_k)\, q_+(z_k|z_{k-1})}{\pi(y_{1:K})}, \tag{2.2}$$

respectively (Jazwinski 1970, Särkkä 2013). While it is of broad interest to approximate the posterior or smoothing PDF (2.2), we will focus on the recursive approximation of the filtering PDFs $\pi(z_k|y_{1:k})$ using sequential particle filters in this paper. More specifically, we wish to address the following computational task.

Problem 2.1. We have M equally weighted Monte Carlo samples z_{k-1}^i, $i = 1, \ldots, M$, from the filtering PDF $\pi(z_{k-1}|y_{1:k-1})$ at time $t = k - 1$ available and we wish to produce M equally weighted samples from the filtering PDF $\pi(z_k|y_{1:k})$ at time $t = k$ having access to the transition kernel $q_+(z_k|z_{k-1})$ and the likelihood $\pi(y_k|z_k)$ only. Since the computational task is exactly the same for all indices $k \geq 1$, we simply set $k = 1$ throughout this paper.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

We introduce some notations before we discuss several possibilities of addressing Problem 2.1. Since we do not have direct access to the filtering distribution at time $k = 0$, the PDF at t_0 becomes

$$\pi_0(z_0) := \frac{1}{M} \sum_{i=1}^{M} \delta(z_0 - z_0^i), \tag{2.3}$$

where $\delta(z)$ denotes the Dirac delta function and z_0^i, $i = 1, \ldots, M$, are M given Monte Carlo samples representing the actual filtering distribution. Recall that we abbreviate the resulting filtering PDF $\pi(z_1|y_1)$ at $t = 1$ by $\widehat{\pi}_1(z_1)$ and the likelihood $\pi(y_1|z_1)$ by $l(z_1)$. Because of (1.1), the forecast PDF is given by

$$\pi_1(z_1) = \frac{1}{M} \sum_{i=1}^{M} q_+(z_1|z_0^i) \tag{2.4}$$

and the filtering PDF at time $t = 1$ by

$$\widehat{\pi}_1(z_1) := \frac{l(z_1)\, \pi_1(z_1)}{\pi_1[l]} = \frac{1}{\pi_1[l]} \frac{1}{M} \sum_{i=1}^{M} l(z_1)\, q_+(z_1|z_0^i) \tag{2.5}$$

according to Bayes' theorem.

Remark 2.2. The normalization constant $\pi(y_{1:K})$ in (2.2), also called the evidence, can be determined recursively using

$$\begin{aligned} \pi(y_{1:k}) &= \pi(y_{1:k-1}) \int \pi(y_k, z_{k-1})\, \pi(z_{k-1}|y_{1:k-1})\, \mathrm{d}z_{k-1} \\ &= \pi(y_{1:k-1}) \int \int \pi(y_k|z_k)\, q_+(z_k|z_{k-1})\, \pi(z_{k-1}|y_{1:k-1})\, \mathrm{d}z_{k-1}\, \mathrm{d}z_k \\ &= \pi(y_{1:k-1}) \int \pi(y_k|z_k)\, \pi(z_k|y_{1:k-1})\, \mathrm{d}z_k \end{aligned} \tag{2.6}$$

(Särkkä 2013, Reich and Cotter 2015). Since, as for the state estimation problem, the computational task is the same for each index $k \geq 1$, we simply set $k = 1$ and formally use $\pi(y_{1:0}) \equiv 1$. We are then left with

$$\beta := \pi_1[l] = \frac{1}{M} \sum_{i=1}^{M} \int l(z_1)\, q_+(z_1|z_0^i)\, \mathrm{d}z_1 \tag{2.7}$$

within the setting of Problem 2.1, and β becomes a shorthand for $\pi(y_1)$. If the model depends on parameters, λ, or different models are to be compared, then it is important to evaluate the evidence (2.7) for each parameter value λ or model, respectively. More specifically, if $q_+(z_1|z_0; \lambda)$, then $\beta = \beta(\lambda)$ in (2.7) and larger values of $\beta(\lambda)$ indicate a better fit of the transition kernel to the data for that parameter value. One can then perform Bayesian parameter inference based upon appropriate approximations to the likelihood

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

$\pi(y_1|\lambda) = \beta(\lambda)$ and a given prior PDF $\pi(\lambda)$. The extension to the complete data set $y_{1:K}$, $K > 1$, is straightforward using (2.6) and an appropriate data assimilation algorithm, *i.e.* algorithms that can tackle Problem 2.1 sequentially.

Alternatively, one can treat a combined state–parameter estimation problem as a particular case of Problem 2.1 by introducing the extended state variable (z, λ) and augmented transition probabilities $Z_1 \sim q_+(\cdot|z_0, \lambda_0)$ and $\mathbb{P}[\Lambda_1 = \lambda_0] = 1$. The state augmentation technique allows one to extend all approaches discussed in this paper for Problem 2.1 to combined state–parameter estimation.

See Kantas *et al.* (2015) for a detailed survey of the topic of combined state and parameter estimation.

The filtering distribution $\widehat{\pi}_1$ at time $t = 1$ implies a smoothing distribution at time $t = 0$, which is given by

$$\widehat{\pi}_0(z_0) := \frac{1}{\beta} \int l(z_1)\, q_+(z_1|z_0)\, \pi_0(z_0)\, \mathrm{d}z_1 = \frac{1}{M} \sum_{i=1}^{M} \gamma^i \delta(z_0 - z_0^i) \tag{2.8}$$

with weights

$$\gamma^i := \frac{1}{\beta} \int l(z_1)\, q_+(z_1|z_0^i)\, \mathrm{d}z_1. \tag{2.9}$$

It is important to note that the filtering PDF $\widehat{\pi}_1$ can be obtained from $\widehat{\pi}_0$ using the transition kernels

$$\widehat{q}_+(z_1|z_0^i) := \frac{l(z_1)\, q_+(z_1|z_0^i)}{\beta\, \gamma^i}, \tag{2.10}$$

that is,

$$\widehat{\pi}_1(z_1) = \frac{1}{M} \sum_{i=1}^{M} \widehat{q}_+(z_1|z_0^i)\, \gamma^i.$$

See Figure 2.1 for a schematic illustration of these distributions and their mutual relationships.

Remark 2.3. The modified transition kernel (2.10) can be seen as a particular instance of a twisted transition kernel (1.2) with $\psi(z) = l(z)/\beta$ and $\widehat{\psi}(z_0^i) = \gamma^i$. Such twisting kernels will play a prominent role in this survey, not only in the context of optimal proposals (Doucet *et al.* 2001, Arulampalam, Maskell, Gordon and Clapp 2002) but also in the context of the Schrödinger approach to data assimilation, *i.e.* scenario (C) below.

The following scenarios of how to tackle Problem 2.1, that is, how to produce the desired samples $\widehat{z}_1^i$, $i = 1, \ldots, M$, from the filtering PDF (2.5), will be considered in this paper.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

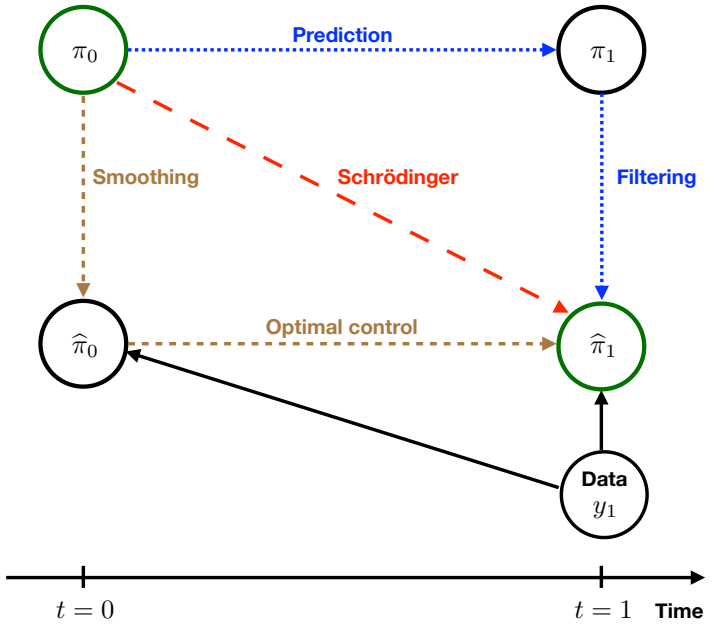


Figure 2.1. Schematic illustration of a single data assimilation cycle. The distribution π_0 characterizes the distribution of states conditioned on all observations up to and including t_0, which we set here to $t = 0$ for simplicity. The predictive distribution at time $t_1 = 1$, as generated by the model dynamics, is denoted by π_1. Upon assimilation of the data y_1 and application of Bayes' formula, one obtains the filtering distribution $\widehat{\pi}_1$. The conditional distribution of states at time t_0 conditioned on all the available data including y_1 is denoted by $\widehat{\pi}_0$. Control theory provides the adjusted model dynamics for transforming $\widehat{\pi}_0$ into $\widehat{\pi}_1$. Finally, the Schrödinger problem links π_0 and $\widehat{\pi}_1$ in the form of a penalized boundary value problem in the space of joint probability measures. Data assimilation scenario (A) corresponds to the dotted lines, scenario (B) to the short-dashed lines, and scenario (C) to the long-dashed line.

Definition 2.4. We define the following three scenarios of how to tackle Problem 2.1.

(A) We first produce samples, z_1^i, from the forecast PDF π_1 and then transform those samples into samples, $\widehat{z}_1^i$, from $\widehat{\pi}_1$. This can be viewed as introducing a Markov transition kernel $q_1(\widehat{z}_1|z_1)$ with the property that

$$\widehat{\pi}_1(\widehat{z}_1) = \int q_1(\widehat{z}_1|z_1)\, \pi_1(z_1)\, \mathrm{d}z_1. \tag{2.11}$$

Techniques from optimal transportation can be used to find appropriate transition kernels (Villani 2003, Villani 2009, Reich and Cotter 2015).

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

(B) We first produce M samples from the smoothing PDF (2.8) via resampling with replacement and then sample from $\widehat{\pi}_1$ using the smoothing transition kernels (2.10). The resampling can be represented in terms of a Markov transition matrix $Q_0 \in \mathbb{R}^{M\times M}$ such that

$$\gamma = Q_0\, p.$$

Here we have introduced the associated probability vectors

$$\gamma = \left(\frac{\gamma^1}{M},\ldots,\frac{\gamma^M}{M}\right)^{\mathrm{T}} \in \mathbb{R}^M, \quad p = \left(\frac{1}{M},\ldots,\frac{1}{M}\right)^{\mathrm{T}} \in \mathbb{R}^M. \tag{2.12}$$

Techniques from optimal transport will be explored to find such Markov transition matrices in Section 3.

(C) We directly seek Markov transition kernels $q_+^*(z_1|z_0^i)$, $i = 1,\ldots,M$, with the property that

$$\widehat{\pi}_1(z_1) = \frac{1}{M}\sum_{i=1}^M q_+^*(z_1|z_0^i) \tag{2.13}$$

and then draw a single sample, $\widehat{z}_1^i$, from each kernel $q_+^*(z_1|z_0^i)$. Such kernels can be found by solving a Schrödinger problem (Leonard 2014, Chen *et al.* 2014) as demonstrated in Section 2.3.

Scenario (A) forms the basis of the classical bootstrap particle filter (Doucet *et al.* 2001, Liu 2001, Bain and Crisan 2008, Arulampalam *et al.* 2002) and also provides the starting point for many currently used ensemble-based data assimilation algorithms (Evensen 2006, Reich and Cotter 2015, Law *et al.* 2015). Scenario (B) is also well known in the context of particle filters under the notion of optimal proposal densities (Doucet *et al.* 2001, Arulampalam *et al.* 2002, Fearnhead and Künsch 2018). Recently there has been renewed interest in scenario (B) from the perspective of optimal control and twisting approaches (Guarniero, Johansen and Lee 2017, Heng, Bishop, Deligiannidis and Doucet 2018, Kappen and Ruiz 2016, Ruiz and Kappen 2017). Finally, scenario (C) has not yet been explored in the context of particle filters and data assimilation, primarily because the required kernels q_+^* are typically not available in closed form or cannot be easily sampled from. However, as we will argue in this paper, progress on the numerical solution of Schrödinger's problem (Cuturi 2013, Peyre and Cuturi 2018) turns scenario (C) into a viable option in addition to providing a unifying mathematical framework for data assimilation.

We emphasize that not all existing particle methods fit into these three scenarios. For example, the methods put forward by van Leeuwen (2015) are based on proposal densities which attempt to overcome limitations of scenario (B) and which lead to less variable particle weights, thus attempting

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

to obtain particle filter implementations closer to what we denote here as scenario (C). More broadly speaking, the exploration of alternative proposal densities in the context of data assimilation has started only recently. See, for example, Vanden-Eijnden and Weare (2012), Morzfeld, Tu, Atkins and Chorin (2012), van Leeuwen (2015), Pons Llopis, Kantas, Beskos and Jasra (2018) and van Leeuwen *et al.* (2018).

The accuracy of an ensemble-based data assimilation method can be characterized in terms of its effective sample size $M_{\rm eff}$ (Liu 2001). The relevant effective sample size for scenario (B) is, for example, given by

$$M_{\rm eff} = \frac{M^2}{\sum_{i=1}^M (\gamma^i)^2} = \frac{1}{\|\gamma\|^2}.$$

We find that $M \geq M_{\rm eff} \geq 1$ and the accuracy of a data assimilation step decreases with decreasing $M_{\rm eff}$, that is, the convergence rate $1/\sqrt{M}$ of a standard Monte Carlo method is replaced by $1/\sqrt{M_{\rm eff}}$ (Agapiou, Papaspiliopoulos, Sanz-Alonso and Stuart 2017). Scenario (C) offers a route around this problem by bridging π_0 with $\widehat{\pi}_1$ directly, that is, solving the Schrödinger problem delivers the best possible proposal densities leading to equally weighted particles without the need for resampling.[1]

Example 2.5. We illustrate the three scenarios with a simple example. The prior samples are given by $M = 11$ equally spaced particles $z_0^i \in \mathbb{R}$ from the interval $[-1, 1]$. The forecast PDF π_1 is provided by

$$\pi_1(z) = \frac{1}{M} \sum_{i=1}^M \frac{1}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{1}{2\sigma^2}(z - z_0^i)^2\right)$$

with variance $\sigma^2 = 0.1$. The likelihood function is given by

$$\pi(y_1|z) = \frac{1}{(2\pi R)^{1/2}} \exp\left(-\frac{1}{2R}(y_1 - z)^2\right)$$

with $R = 0.1$ and $y_1 = -0.5$. The implied filtering and smoothing distributions can be found in Figure 2.2. Since $\widehat{\pi}_1$ is in the form of a weighted Gaussian mixture distribution, the Markov chain leading from $\widehat{\pi}_0$ to $\widehat{\pi}_1$ can be stated explicitly, that is, (2.10) is provided by

$$\widehat{q}_+(z_1|z_0^i) = \frac{1}{(2\pi)^{1/2}\widehat{\sigma}} \exp\left(-\frac{1}{2\widehat{\sigma}^2}(\bar{z}_1^i - z_1)^2\right) \tag{2.14}$$

with

$$\widehat{\sigma}^2 = \sigma^2 - \frac{\sigma^4}{\sigma^2 + R}, \quad \bar{z}_1^i = z_0^i - \frac{\sigma^2}{\sigma^2 + R}(z_0^i - y_1).$$

[1] The kernel (2.10) is called the optimal proposal in the particle filter community. However, the kernel (2.10) is suboptimal in the broader framework considered in this paper.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

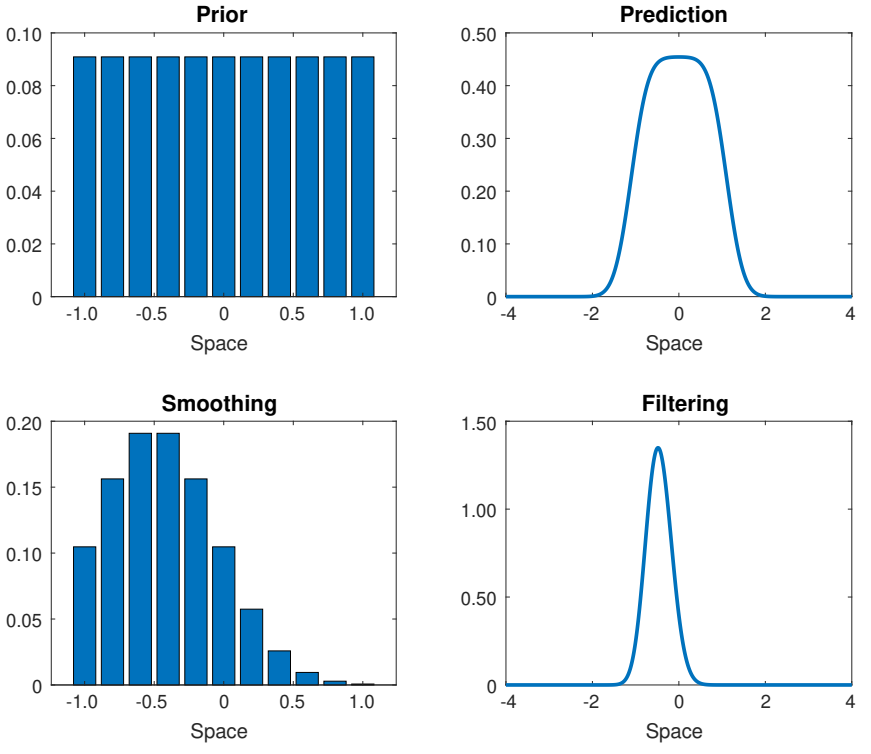


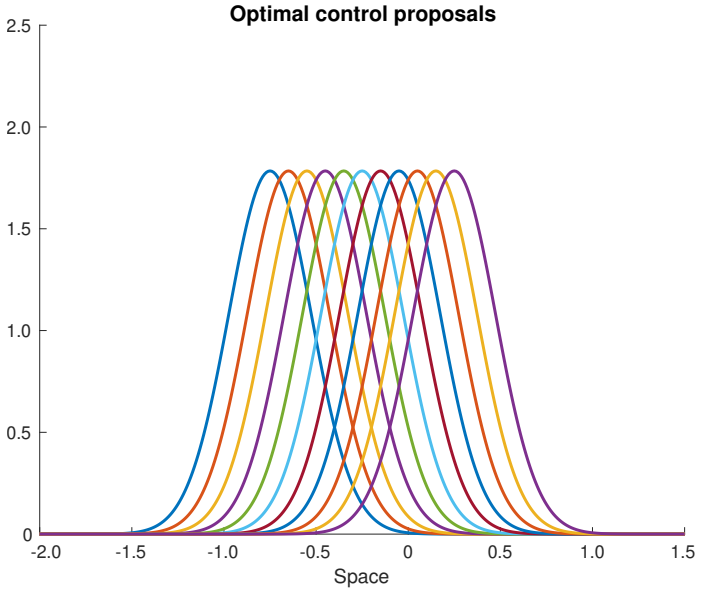
Figure 2.2. The initial PDF π_0, the forecast PDF π_1, the filtering PDF $\widehat{\pi}_1$, and the smoothing PDF $\widehat{\pi}_0$ for a simple Gaussian transition kernel.

The resulting transition kernels are displayed in Figure 2.3 together with the corresponding transition kernels for the Schrödinger approach, which connects π_0 directly with $\widehat{\pi}_1$.

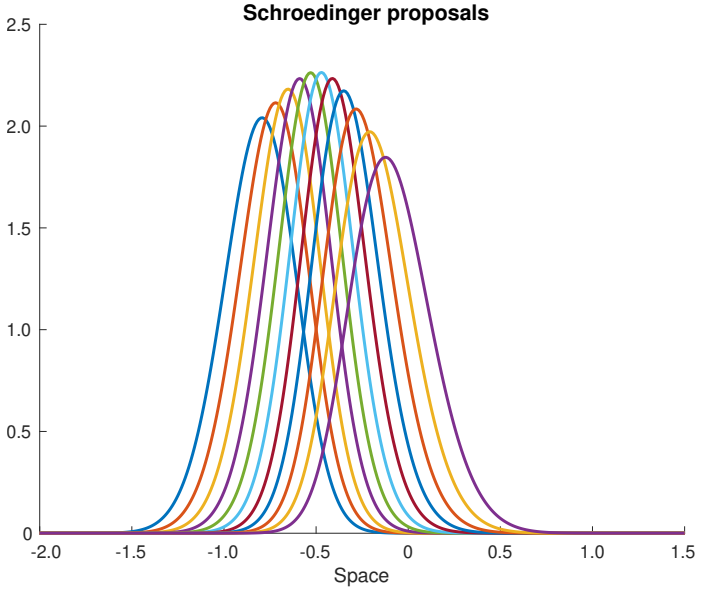
Remark 2.6. It is often assumed in optimal control or rare event simulations arising from statistical mechanics that π_0 in (2.1) is a point measure, that is, the starting point of the simulation is known exactly. See, for example, Hartmann, Richter, Schütte and Zhang (2017). This corresponds to (2.3) with $M = 1$. It turns out that the associated smoothing problem becomes equivalent to Schrödinger's problem under this particular setting since the distribution at $t = 0$ is fixed.

The remainder of this section is structured as follows. We first recapitulate the pure prediction problem for discrete-time Markov processes and continuous-time diffusion processes, after which we discuss the filtering and smoothing problem for a single data assimilation step as relevant for scenarios (A) and (B). The final subsection is devoted to the Schrödinger

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press



(a)



(b)

Figure 2.3. (a) The transition kernels (2.14) for the $M = 11$ different particles z_0^i. These correspond to the optimal control path in Figure 2.1. (b) The corresponding transition kernels, which lead directly from π_0 to $\widehat{\pi}_1$. These correspond to the Schrödinger path in Figure 2.1. Details of how to compute these Schrödinger transition kernels, $q_+^*(z_1|z_0^i)$, can be found in Section 3.4.1.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

problem (Leonard 2014, Chen *et al.* 2014) of bridging the filtering distribution, π_0, at $t = 0$ directly with the filtering distribution, $\widehat{\pi}_1$, at $t = 1$, thus leading to scenario (C).

2.1. Prediction

We assume under the chosen computational setting that we have access to M samples $z_0^i \in \mathbb{R}^{N_z}$, $i = 1, \ldots, M$, from the filtering distribution at $t = 0$. We also assume that we know (explicitly or implicitly) the forward transition probabilities $q_+(z_1|z_0^i)$ of the underlying Markovian stochastic process. This leads to the forecast PDF π_1 as given by (2.4).

Before we consider two specific examples, we introduce two concepts related to the forward transition kernel which we will need later in order to address scenarios (B) & (C) from Definition 2.4.

We first introduce the backward transition kernel $q_-(z_0|z_1)$, which is defined via the equation

$$q_-(z_0|z_1)\,\pi_1(z_1) = q_+(z_1|z_0)\,\pi_0(z_0).$$

Note that $q_-(z_0|z_1)$ as well as π_0 are not absolutely continuous with respect to the underlying Lebesgue measure, that is,

$$q_-(z_0|z_1) = \frac{1}{M}\sum_{i=1}^{M} \frac{q_+(z_1|z_0^i)}{\pi_1(z_1)}\,\delta(z_0 - z_0^i). \tag{2.15}$$

The backward transition kernel $q_-(z_1|z_0)$ reverses the prediction process in the sense that

$$\pi_0(z_0) = \int q_-(z_0|z_1)\,\pi_1(z_1)\,\mathrm{d}z_1.$$

Remark 2.7. Let us assume that the detailed balance

$$q_+(z_1|z_0)\,\pi(z_0) = q_+(z_0|z_1)\,\pi(z_1)$$

holds for some PDF π and forward transition kernel $q_+(z_1|z_0)$. Then $\pi_1 = \pi$ for $\pi_0 = \pi$ (invariance of π) and $q_-(z_0|z_1) = q_+(z_1|z_0)$.

We next introduce a class of forward transition kernels using the concept of twisting (Guarniero *et al.* 2017, Heng *et al.* 2018), which is an application of Doob's H-transform technique (Doob 1984).

Definition 2.8. Given a non-negative twisting function $\psi(z_1)$ such that the modified transition kernel (1.2) is well-defined, one can define the twisted forecast PDF

$$\pi_1^{\psi}(z_1) := \frac{1}{M}\sum_{i=1}^{M} q_+^{\psi}(z_1|z_0^i) = \frac{1}{M}\sum_{i=1}^{M} \frac{\psi(z_1)}{\widehat{\psi}(z_0^i)}\,q_+(z_1|z_0^i). \tag{2.16}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

The PDFs π_1 and π_1^ψ are related by

$$\frac{\pi_1(z_1)}{\pi_1^\psi(z_1)} = \frac{\sum_{i=1}^M q_+(z_1|z_0^i)}{\sum_{i=1}^M \frac{\psi(z_1)}{\widehat{\psi}(z_0^i)} q_+(z_1|z_0^i)}. \tag{2.17}$$

Equation (2.17) gives rise to importance weights

$$w^i \propto \frac{\pi_1(z_1^i)}{\pi_1^\psi(z_1^i)} \tag{2.18}$$

for samples $z_1^i = Z_1^i(\omega)$ drawn from the twisted forecast PDF, that is,

$$Z_1^i \sim q_+^\psi(\cdot\,|z_0^i)$$

and

$$\pi_1(z) \approx \frac{1}{M}\sum_{i=1}^M w^i\, \delta(z - z_1^i)$$

in a weak sense. Here we have assumed that the normalization constant in (2.18) is chosen such that

$$\sum_{i=1}^M w^i = M. \tag{2.19}$$

Such twisted transition kernels will become important when looking at the filtering and smoothing as well as the Schrödinger problem later in this section.

Let us now discuss a couple of specific models which give rise to transition kernels $q_+(z_1|z_0)$. These models will be used throughout this paper to illustrate mathematical and algorithmic concepts.

2.1.1. Gaussian model error

Let us consider the discrete-time stochastic process

$$Z_1 = \Psi(Z_0) + \gamma^{1/2}\Xi_0 \tag{2.20}$$

for given map $\Psi : \mathbb{R}^{N_z} \to \mathbb{R}^{N_z}$, scaling factor $\gamma > 0$, and Gaussian distributed random variable Ξ_0 with mean zero and covariance matrix $B \in \mathbb{R}^{N_z \times N_z}$. The associated forward transition kernel is given by

$$q_+(z_1|z_0) = \mathrm{n}(z_1; \Psi(z_0), \gamma B). \tag{2.21}$$

Recall that we have introduced the shorthand $\mathrm{n}(z; \bar{z}, P)$ for the PDF of a Gaussian random variable with mean $\bar{z}$ and covariance matrix P.

Let us consider a twisting potential ψ of the form

$$\psi(z_1) \propto \exp\left(-\frac{1}{2}(Hz_1 - d)^{\mathrm{T}} R^{-1}(Hz_1 - d)\right)$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

for given $H \in \mathbb{R}^{N_z \times N_d}$, $d \in \mathbb{R}^{N_d}$, and covariance matrix $R \in \mathbb{R}^{N_d \times N_d}$. We define

$$K := BH^{\mathrm{T}}(HBH^{\mathrm{T}} + \gamma^{-1}R)^{-1} \tag{2.22}$$

and

$$\bar{B} := B - KHB, \quad \bar{z}_1^i := \Psi(z_0^i) - K(H\Psi(z_0^i) - d). \tag{2.23}$$

The twisted transition kernels are given by

$$q_+^{\psi}(z_1|z_0^i) = \mathrm{n}(z_1; \bar{z}_1^i, \gamma\bar{B})$$

and

$$\widehat{\psi}(z_0^i) \propto \exp\left(-\frac{1}{2}(H\Psi(z_0^i) - d)^{\mathrm{T}}(R + \gamma HBH^{\mathrm{T}})^{-1}(H\Psi(z_0^i) - d)\right)$$

for $i = 1, \ldots, M$.

2.1.2. SDE models

Consider the (forward) SDE (Pavliotis 2014)

$$\mathrm{d}Z_t^+ = f_t(Z_t^+)\,\mathrm{d}t + \gamma^{1/2}\,\mathrm{d}W_t^+ \tag{2.24}$$

with initial condition $Z_0^+ = z_0$ and $\gamma > 0$. Here W_t^+ stands for standard Brownian motion in the sense that the distribution of $W_{t+\Delta t}^+$, $\Delta t > 0$, conditioned on $w_t^+ = W_t^+(\omega)$ is Gaussian with mean w_t^+ and covariance matrix $\Delta t\, I$ (Pavliotis 2014) and the process Z_t^+ is adapted to W_t^+.

The resulting time-t transition kernels $q_t^+(z|z_0)$ from time zero to time t, $t \in (0, 1]$, satisfy the Fokker–Planck equation (Pavliotis 2014)

$$\partial_t q_t^+(\cdot\,|z_0) = -\nabla_z \cdot (q_t^+(\cdot\,|z_0) f_t) + \frac{\gamma}{2}\Delta_z q_t^+(\cdot\,|z_0)$$

with initial condition $q_0^+(z|z_0) = \delta(z - z_0)$, and the time-one forward transition kernel $q_+(z_1|z_0)$ is given by

$$q_+(z_1|z_0) = q_1^+(z_1|z_0).$$

We introduce the operator $\mathcal{L}_t$ by

$$\mathcal{L}_t g := \nabla_z g \cdot f_t + \frac{\gamma}{2}\Delta_z g$$

and its adjoint $\mathcal{L}_t^\dagger$ by

$$\mathcal{L}_t^\dagger \pi := -\nabla_z \cdot (\pi\, f_t) + \frac{\gamma}{2}\Delta_z \pi \tag{2.25}$$

(Pavliotis 2014). We call $\mathcal{L}_t^\dagger$ the Fokker–Planck operator and $\mathcal{L}_t$ the generator of the Markov process associated to the SDE (2.24).

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

Solutions (realizations) $z_{[0,1]} = Z^+_{[0,1]}(\omega)$ of the SDE (2.24) with initial conditions drawn from π_0 are continuous functions of time, that is, $z_{[0,1]} \in \mathcal{C} := C([0,1], \mathbb{R}^{N_z})$, and define a probability measure $\mathbb{Q}$ on $\mathcal{C}$, that is,

$$Z^+_{[0,1]} \sim \mathbb{Q}.$$

We note that the marginal distributions π_t of $\mathbb{Q}$, given by

$$\pi_t(z_t) = \int q^+_t(z_t|z_0)\, \pi_0(z_0)\, \mathrm{d}z_0,$$

also satisfy the Fokker–Planck equation, that is,

$$\partial_t \pi_t = \mathcal{L}^\dagger_t\, \pi_t = -\nabla_z \cdot (\pi_t\, f_t) + \frac{\gamma}{2} \Delta_z \pi_t \tag{2.26}$$

for given PDF π_0 at time $t = 0$.

Furthermore, we can rewrite the Fokker–Planck equation (2.26) in the form

$$\partial \pi_t = -\nabla_z \cdot (\pi_t (f_t - \gamma \nabla_z \log \pi_t)) - \frac{\gamma}{2} \Delta_z \pi_t, \tag{2.27}$$

which allows us to read off from (2.27) the backward SDE

$$\begin{aligned} \mathrm{d}Z^-_t &= f_t(Z^-_t)\, \mathrm{d}t - \gamma \nabla_z \log \pi_t\, \mathrm{d}t + \gamma^{1/2}\, \mathrm{d}W^-_t, \\ &= b_t(Z^-_t)\, \mathrm{d}t + \gamma^{1/2}\, \mathrm{d}W^-_t \end{aligned} \tag{2.28}$$

with final condition $Z^-_1 \sim \pi_1$, W^-_t backward Brownian motion, and density-dependent drift term

$$b_t(z) := f_t(z) - \gamma \nabla_z \log \pi_t$$

(Nelson 1984, Chen *et al.* 2014). Here backward Brownian motion is to be understood in the sense that the distribution of $W^-_{t-\Delta\tau}$, $\Delta\tau > 0$, conditioned on $w^-_t = W^-_t(\omega)$ is Gaussian with mean w^-_t and covariance matrix $\Delta\tau\, I$ and all other properties of Brownian motion appropriately adjusted. The process Z^-_t is adapted to W^-_t.

Lemma 2.9. The backward SDE (2.28) induces a corresponding backward transition kernel from time one to time $t = 1 - \tau$ with $\tau \in [0,1]$, denoted by $q^-_\tau(z|z_1)$, which satisfies the time-reversed Fokker–Planck equation

$$\partial_\tau q^-_\tau(\cdot\,|z_1) = \nabla_z \cdot (q^-_\tau(\cdot\,|z_1)\, b_{1-\tau}) + \frac{\gamma}{2} \Delta_z q^-_\tau(\cdot\,|z_1)$$

with boundary condition $q^-_0(z|z_1) = \delta(z - z_1)$ at $\tau = 0$ (or, equivalently, at $t = 1$). The induced backward transition kernel $q_-(z_0|z_1)$ is then given by

$$q_-(z_0|z_1) = q^-_1(z_0|z_1)$$

and satisfies (2.15).

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

Proof. The lemma follows from the fact that the backward SDE (2.28) implies the Fokker–Planck equation (2.27) and that we have reversed time by introducing $\tau = 1 - t$. □

Remark 2.10. The notion of a backward SDE also arises in a different context where the driving Brownian motion is still adapted to the past, *i.e.* W_t^+ in our notation, and a final condition is prescribed as for (2.28). See (2.54) below as well as Appendix A.4 and Carmona (2016) for more details.

We note that the mean-field equation,

$$\frac{\mathrm{d}}{\mathrm{d}t} z_t = f_t(z_t) - \frac{\gamma}{2} \nabla_z \log \pi_t(z_t) = \frac{1}{2}(f_t(z_t) + b_t(z_t)), \tag{2.29}$$

resulting from (2.26), leads to the same marginal distributions π_t as the forward and backward SDEs, respectively. It should be kept in mind, however, that the path measure generated by (2.29) is different from the path measure $\mathbb{Q}$ generated by (2.24).

Please also note that the backward SDE and the mean-field equation (2.29) become singular as $t \to 0$ for the given initial PDF (2.3). A meaningful solution can be defined via regularization of the Dirac delta function, that is,

$$\pi_0(z) \approx \frac{1}{M} \sum_{i=1}^{M} \mathrm{n}(z; z_0^i, \epsilon I),$$

and taking the limit $\epsilon \to 0$.

We will find later that it can be advantageous to modify the given SDE (2.24) by a time-dependent drift term $u_t(z)$, that is,

$$\mathrm{d}Z_t^+ = f_t(Z_t^+)\,\mathrm{d}t + u_t(Z_t^+)\,\mathrm{d}t + \gamma^{1/2}\,\mathrm{d}W_t^+. \tag{2.30}$$

In particular, such a modification leads to the time-continuous analogue of the twisted transition kernel (1.2) introduced in Section 2.1.

Lemma 2.11. Let $\psi_t(z)$ denote the solutions of the backward Kolmogorov equation

$$\partial_t \psi_t = -\mathcal{L}_t \psi_t = -\nabla_z \psi_t \cdot f_t - \frac{\gamma}{2} \Delta_z \psi_t \tag{2.31}$$

for given final $\psi_1(z) > 0$ and $t \in [0, 1]$. The controlled SDE (2.30) with

$$u_t(z) := \gamma \nabla_z \log \psi_t(z) \tag{2.32}$$

leads to a time-one forward transition kernel $q_+^\psi(z_1|z_0)$ which satisfies

$$q_+^\psi(z_1|z_0) = \psi_1(z_1)\, q_+(z_1|z_0)\, \psi_0(z_0)^{-1},$$

where $q_+(z_1|z_0)$ denotes the time-one forward transition kernel of the uncontrolled forward SDE (2.24).

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

Proof. A proof of this lemma has, for example, been given by Dai Pra (1991, Theorem 2.1). □

More generally, the modified forward SDE (2.30) with $Z_0^+ \sim \pi_0$ generates a path measure which we denote by $\mathbb{Q}^u$ for given functions $u_t(z)$, $t \in [0,1]$. Realizations of this path measure are denoted by $z^u_{[0,1]}$. According to Girsanov's theorem (Pavliotis 2014), the two path measures $\mathbb{Q}$ and $\mathbb{Q}^u$ are absolutely continuous with respect to each other, with Radon–Nikodym derivative

$$\frac{\mathrm{d}\mathbb{Q}^u}{\mathrm{d}\mathbb{Q}}_{|z^u_{[0,1]}} = \exp\left(\frac{1}{2\gamma}\int_0^1 (\|u_t\|^2\,\mathrm{d}t + 2\gamma^{1/2}u_t \cdot \mathrm{d}W_t^+)\right), \tag{2.33}$$

provided that the Kullback–Leibler divergence $\mathrm{KL}(\mathbb{Q}^u||\mathbb{Q})$ between $\mathbb{Q}^u$ and $\mathbb{Q}$, given by

$$\mathrm{KL}(\mathbb{Q}^u||\mathbb{Q}) := \int \left[\frac{1}{2\gamma}\int_0^1 \|u_t\|^2\,\mathrm{d}t\right]\mathbb{Q}^u(\mathrm{d}z^u_{[0,1]}), \tag{2.34}$$

is finite. Recall that the Kullback–Leibler divergence between two path measures $\mathbb{P} \ll \mathbb{Q}$ on $\mathcal{C}$ is defined by

$$\mathrm{KL}(\mathbb{P}||\mathbb{Q}) = \int \log\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\,\mathbb{P}(\mathrm{d}z_{[0,1]}).$$

If the modified SDE (2.30) is used to make predictions, then its solutions $z^u_{[0,1]}$ need to be weighted according to the inverse Radon–Nikodym derivative

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{Q}^u}_{|z^u_{[0,1]}} = \exp\left(-\frac{1}{2\gamma}\int_0^1 (\|u_t\|^2\,\mathrm{d}t + 2\gamma^{1/2}u_t \cdot\, \mathrm{d}W_t^+)\right) \tag{2.35}$$

in order to reproduce the desired forecast PDF π_1 of the original SDE (2.24).

Remark 2.12. A heuristic derivation of equation (2.33) can be found in Section 3.1.2 below, where we discuss the numerical approximation of SDEs by the Euler–Maruyama method. Equation (2.34) follows immediately from (2.33) by noting that the expectation of Brownian motion under the path measure $\mathbb{Q}^u$ is zero.

2.2. Filtering and smoothing

We now incorporate the likelihood

$$l(z_1) = \pi(y_1|z_1)$$

of the data y_1 at time $t_1 = 1$. Bayes' theorem tells us that, given the forecast PDF π_1 at time t_1, the posterior PDF $\widehat{\pi}_1$ is given by (2.5). The distribution $\widehat{\pi}_1$ solves the filtering problem at time t_1 given the data y_1. We also recall

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

the definition of the evidence (2.7). The quantity $\mathcal{F} = -\log \beta$ is called the free energy in statistical physics (Hartmann *et al.* 2017).

An appropriate transition kernel $q_1(\widehat{z}_1|z_1)$, satisfying (2.11), is required in order to complete the transition from π_0 to $\widehat{\pi}_1$ following scenario (A) from Definition 2.4. A suitable framework for finding such transition kernels is via the theory of optimal transportation (Villani 2003). More specifically, let $\Pi_{\rm c}$ denote the set of all joint probability measures $\pi(z_1, \widehat{z}_1)$ with marginals

$$\int \pi(z_1, \widehat{z}_1)\, {\rm d}\widehat{z}_1 = \pi_1(z_1), \quad \int \pi(z_1, \widehat{z}_1)\, {\rm d}z_1 = \widehat{\pi}_1(\widehat{z}_1).$$

We seek the joint measure $\pi^*(z_1, \widehat{z}_1) \in \Pi_{\rm c}$ which minimizes the expected Euclidean distance between the two associated random variables Z_1 and $\widehat{Z}_1$, that is,

$$\pi^* = \arg \inf_{\pi \in \Pi_{\rm c}} \int\int \|z_1 - \widehat{z}_1\|^2\, \pi(z_1, \widehat{z}_1)\, {\rm d}z_1\, {\rm d}\widehat{z}_1. \tag{2.36}$$

The minimizing joint measure is of the form

$$\pi^*(z_1, \widehat{z}_1) = \delta(\widehat{z}_1 - \nabla_z \Phi(z_1))\, \pi_1(z_1) \tag{2.37}$$

with suitable convex potential Φ under appropriate conditions on the PDFs π_1 and $\widehat{\pi}_1$ (Villani 2003). These conditions are satisfied for dynamical systems with Gaussian model errors and typical SDE models. Once the potential Φ (or an approximation) is available, samples z_1^i, $i = 1, \ldots, M$, from the forecast PDF π_1 can be converted into samples $\widehat{z}_1^i$, $i = 1, \ldots, M$, from the filtering distribution $\widehat{\pi}_1$ via the deterministic transformation

$$\widehat{z}_1^i = \nabla_z \Phi(z_1^i). \tag{2.38}$$

We will discuss in Section 3 how to approximate the transformation (2.38) numerically. We will find that many popular data assimilation schemes, such as the ensemble Kalman filter, can be viewed as approximations to (2.38) (Reich and Cotter 2015).

We recall at this point that classical particle filters start from the importance weights

$$w^i \propto \frac{\widehat{\pi}_1(z_1^i)}{\pi_1(z_1^i)} = \frac{l(z_1^i)}{\beta},$$

and obtain the desired samples $\widehat{z}^i$ by an appropriate resampling with replacement scheme (Doucet *et al.* 2001, Arulampalam *et al.* 2002, Douc and Cappe 2005) instead of applying a deterministic transformation of the form (2.38).

Remark 2.13. If one replaces the forward transition kernel $q_+(z_1|z_0)$ with a twisted kernel (1.2), then, using (2.17), the filtering distribution (2.5)

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

satisfies

$$\frac{\widehat{\pi}_1(z_1)}{\pi_1^{\psi}(z_1)} = \frac{l(z_1)\sum_{j=1}^M q_+(z_1|z_0^j)}{\beta\sum_{j=1}^M \frac{\psi(z_1)}{\widehat{\psi}(z_0^j)}\, q_+(z_1|z_0^j)}. \tag{2.39}$$

Hence drawing samples z_1^i, $i = 1, \ldots, M$, from π_1^{ψ} instead of π_1 leads to modified importance weights

$$w^i \propto \frac{l(z_1^i)\sum_{j=1}^M q_+(z_1^i|z_0^j)}{\beta\sum_{j=1}^M \frac{\psi(z_1^i)}{\widehat{\psi}(z_0^j)}\, q_+(z_1^i|z_0^j)}. \tag{2.40}$$

We will demonstrate in Section 2.3 that finding a twisting potential ψ such that $\widehat{\pi}_1 = \pi_1^{\psi}$, leading to importance weights $w^i = 1$ in (2.40), is equivalent to solving the Schrödinger problem (2.58)–(2.61).

The associated smoothing distribution at time $t = 0$ can be defined as follows. First introduce

$$\psi(z_1) := \frac{\widehat{\pi}_1(z_1)}{\pi_1(z_1)} = \frac{l(z_1)}{\beta}. \tag{2.41}$$

Next we set

$$\widehat{\psi}(z_0) := \int q_+(z_1|z_0)\,\psi(z_1)\,\mathrm{d}z_1 = \beta^{-1}\int q_+(z_1|z_0)\,l(z_1)\,\mathrm{d}z_1, \tag{2.42}$$

and introduce $\widehat{\pi}_0 := \pi_0\,\widehat{\psi}$, that is,

$$\begin{aligned}\widehat{\pi}_0(z_0) &= \frac{1}{M}\sum_{i=1}^M \widehat{\psi}(z_0^i)\,\delta(z_0 - z_0^i) \\ &= \frac{1}{M}\sum_{i=1}^M \gamma^i\,\delta(z_0 - z_0^i)\end{aligned} \tag{2.43}$$

since $\widehat{\psi}(z_0^i) = \gamma^i$ with γ^i defined by (2.9).

Lemma 2.14. The smoothing PDFs $\widehat{\pi}_0$ and $\widehat{\pi}_1$ satisfy

$$\widehat{\pi}_0(z_0) = \int q_-(z_0|z_1)\,\widehat{\pi}_1(z_1)\,\mathrm{d}z_1 \tag{2.44}$$

with the backward transition kernel defined by (2.15). Furthermore,

$$\widehat{\pi}_1(z_1) = \int \widehat{q}_+(z_1|z_0)\,\widehat{\pi}_0(z_0)\,\mathrm{d}z_0$$

with twisted forward transition kernels

$$\widehat{q}_+(z_1|z_0^i) := \psi(z_1)\,q_+(z_1|z_0^i)\,\widehat{\psi}(z_0^i)^{-1} = \frac{l(z_1)}{\beta\,\gamma^i}q_+(z_1|z_0^i) \tag{2.45}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

and γ^i, $i = 1, \ldots, M$, defined by (2.9).

Proof. We note that

$$q_-(z_0|z_1)\,\widehat{\pi}_1(z_1) = \frac{\pi_0(z_0)}{\pi_1(z_1)} q_+(z_1|z_0)\widehat{\pi}_1(z_1) = \frac{l(z_1)}{\beta} q_+(z_1|z_0)\,\pi_0(z_0),$$

which implies the first equation. The second equation follows from $\widehat{\pi}_0 = \widehat{\psi}\,\pi_0$ and

$$\int \widehat{q}_+(z_1|z_0)\,\widehat{\pi}_0(z_0)\,\mathrm{d}z_0 = \frac{1}{M}\sum_{i=1}^{M} \frac{l(z_1)}{\beta}\, q_+(z_1|z_0^i).$$

In other words, we have defined a twisted forward transition kernel of the form (1.2). □

Seen from a more abstract perspective, we have provided an alternative formulation of the joint smoothing distribution

$$\widehat{\pi}(z_0, z_1) := \frac{l(z_1)\, q_+(z_1|z_0)\,\pi_0(z_0)}{\beta} \tag{2.46}$$

in the form of

$$\begin{aligned}\widehat{\pi}(z_0, z_1) &= \frac{l(z_1)}{\beta}\frac{\psi(z_1)}{\psi(z_1)} q_+(z_1|z_0)\frac{\widehat{\psi}(z_0)}{\widehat{\psi}(z_0)}\pi_0(z_0) \\ &= \widehat{q}_+(z_1|z_0)\,\widehat{\pi}_0(z_0)\end{aligned} \tag{2.47}$$

because of (2.41). Note that the marginal distributions of $\widehat{\pi}$ are provided by $\widehat{\pi}_0$ and $\widehat{\pi}_1$, respectively.

One can exploit these formulations computationally as follows. If one has generated M equally weighted particles $\widehat{z}_0^j$ from the smoothing distribution (2.43) at time $t = 0$ via resampling with replacement, then one can obtain equally weighted samples $\widehat{z}_1^j$ from the filtering distribution $\widehat{\pi}_1$ using the modified transition kernels (2.45). This is the idea behind the optimal proposal particle filter (Doucet *et al.* 2001, Arulampalam *et al.* 2002, Fearnhead and Künsch 2018) and provides an implementation of scenario (B) as introduced in Definition 2.4.

Remark 2.15. We remark that backward simulation methods use (2.44) in order to address the smoothing problem (2.2) in a sequential forward–backward manner. Since we are not interested in the general smoothing problem in this paper, we refer the reader to the survey by Lindsten and Schön (2013) for more details.

Lemma 2.16. Let $\psi(z) > 0$ be a twisting potential such that

$$l^\psi(z_1) := \frac{l(z_1)}{\beta\,\psi(z_1)}\pi_0[\widehat{\psi}]$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

is well-defined with $\widehat{\psi}$ given by (1.3). Then the smoothing PDF (2.46) can be represented as

$$\widehat{\pi}(z_0, z_1) = l^{\psi}(z_1)\, q_+^{\psi}(z_1|z_0)\, \pi_0^{\psi}(z_0), \tag{2.48}$$

where the modified forward transition kernel $q_+^{\psi}(z_1|z_0)$ is defined by (1.2) and the modified initial PDF by

$$\pi_0^{\psi}(z_0) := \frac{\widehat{\psi}(z_0)\, \pi_0(z_0)}{\pi_0[\widehat{\psi}]}.$$

Proof. This follows from the definition of the smoothing PDF $\widehat{\pi}(z_0, z_1)$ and the twisted transition kernel $q_+^{\psi}(z_1|z_0)$. □

Remark 2.17. As mentioned before, the choice (2.41) implies $l^{\psi} = \text{const.}$, and leads to the well-known optimal proposal density for particle filters. The more general formulation (2.48) has recently been explored and expanded by Guarniero *et al.* (2017) and Heng *et al.* (2018) in order to derive efficient proposal densities for the general smoothing problem (2.2). Within the simplified formulation (2.48), such approaches reduce to a change of measure from π_0 to π_0^{ψ} at t_0 followed by a forward transition according to q_+^{ψ} and subsequent reweighting by a modified likelihood l^{ψ} at t_1 and hence lead to particle filters that combine scenarios (A) and (B) as introduced in Definition 2.4.

2.2.1. Gaussian model errors (cont.)
We return to the discrete-time process (2.20) and assume a Gaussian measurement error leading to a Gaussian likelihood

$$l(z_1) \propto \exp\left(-\frac{1}{2}(Hz_1 - y_1)^T R^{-1}(Hz_1 - y_1)\right).$$

We set $\psi_1 = l/\beta$ in order to derive the optimal forward kernel for the associated smoothing/filtering problem. Following the discussion from Section 2.1.1, this leads to the modified transition kernels

$$\widehat{q}_+(z_1|z_0^i) := \mathrm{n}(z_1; \bar{z}_1^i, \gamma \bar{B})$$

with $\bar{B}$ and K defined by (2.23) and (2.22), respectively, and

$$\bar{z}_1^i := \Psi(z_0^i) - K(H\Psi(z_0^i) - y_1).$$

The smoothing distribution $\widehat{\pi}_0$ is given by

$$\widehat{\pi}_0(z_0) = \frac{1}{M}\sum_{i=1}^{M} \gamma^i\, \delta(z - z_0^i)$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

with coefficients

$$\gamma^i \propto \exp\left(-\frac{1}{2}(H\Psi(z_0^i) - y_1)^{\mathrm{T}}(R + \gamma HBH^{\mathrm{T}})^{-1}(H\Psi(z_0^i) - y_1)\right).$$

It is easily checked that, indeed,

$$\widehat{\pi}_1(z_1) = \int q_+^{\psi}(z_1|z_0)\,\widehat{\pi}_0(z_0)\,\mathrm{d}z_0.$$

The results from this subsection have been used in simplified form in Example 2.5 in order to compute (2.14). We also note that a non-optimal, *i.e.* $\psi(z_1) \neq l(z_1)/\beta$, but Gaussian choice for ψ leads to a Gaussian l^{ψ} and the transition kernels $q_+^{\psi}(z_1|z_0^i)$ in (2.48) remain Gaussian as well. This is in contrast to the Schrödinger problem, which we discuss in the following Section 2.3 and which leads to forward transition kernels of the form (2.66) below.

2.2.2. SDE models (cont.)

The likelihood $l(z_1)$ introduces a change of measure over path space $z_{[0,1]} \in \mathcal{C}$ from the forecast measure $\mathbb{Q}$ with marginals π_t to the smoothing measure $\widehat{\mathbb{P}}$ via the Radon–Nikodym derivative

$$\frac{\mathrm{d}\widehat{\mathbb{P}}}{\mathrm{d}\mathbb{Q}}_{|z_{[0,1]}} = \frac{l(z_1)}{\beta}. \tag{2.49}$$

We let $\widehat{\pi}_t$ denote the marginal distributions of the smoothing measure $\widehat{\mathbb{P}}$ at time t.

Lemma 2.18. Let ψ_t denote the solution of the backward Kolmogorov equation (2.31) with final condition $\psi_1(z) = l(z)/\beta$ at $t = 1$. Then the controlled forward SDE

$$\mathrm{d}Z_t^+ = (f_t(Z_t^+) + \gamma\nabla_z \log \psi_t(Z_t^+))\,\mathrm{d}t + \gamma^{1/2}\,\mathrm{d}W_t^+, \tag{2.50}$$

with $Z_0^+ \sim \widehat{\pi}_0$ at time $t = 0$, implies $Z_1^+ \sim \widehat{\pi}_1$ at final time $t = 1$.

Proof. The lemma is a consequence of Lemma 2.11 and definition (2.45) of the smoothing kernel $\widehat{q}_+(z_1|z_0)$ with $\psi(z_1) = \psi_1(z_1)$ and $\widehat{\psi}(z_0) = \psi_0(z_0)$. □

The SDE (2.50) is obviously a special case of (2.30) with control law u_t given by (2.32). Note, however, that the initial distributions for (2.30) and (2.50) are different. We will reconcile this fact in the following subsection by considering the associated Schrödinger problem (Föllmer and Gantert 1997, Leonard 2014, Chen *et al.* 2014).

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

Lemma 2.19. The solution ψ_t of the backward Kolmogorov equation (2.31) with final condition $\psi_1(z) = l(z)/\beta$ at $t = 1$ satisfies

$$\psi_t(z) = \frac{\widehat{\pi}_t(z)}{\pi_t(z)} \tag{2.51}$$

and the PDFs $\widehat{\pi}_t$ coincide with the marginal PDFs of the backward SDE (2.28) with final condition $Z_1^- \sim \widehat{\pi}_1$.

Proof. We first note that (2.51) holds at final time $t = 1$. Furthermore, equation (2.51) implies

$$\partial_t \widehat{\pi}_t = \pi_t \, \partial_t \psi_t - \psi_t \, \partial_t \pi_t.$$

Since π_t satisfies the Fokker–Planck equation (2.26) and ψ_t the backward Kolmogorov equation (2.31), it follows that

$$\partial_t \widehat{\pi}_t = -\nabla_z \cdot (\widehat{\pi}_t (f - \gamma \nabla_z \log \pi_t)) - \frac{\gamma}{2} \Delta_z \widehat{\pi}_t, \tag{2.52}$$

which corresponds to the Fokker–Planck equation (2.27) of the backward SDE (2.28) with final condition $Z_1^- \sim \widehat{\pi}_1$ and marginal PDFs denoted by $\widehat{\pi}_t$ instead of π_t. □

Note that (2.52) is equivalent to

$$\partial \widehat{\pi}_t = -\nabla_z \cdot (\widehat{\pi}_t (f - \gamma \nabla_z \log \pi_t + \gamma \nabla_z \log \widehat{\pi}_t)) + \frac{\gamma}{2} \Delta_z \widehat{\pi}_t,$$

which in turn is equivalent to the Fokker–Planck equation of the forward smoothing SDE (2.50) since $\psi_t = \widehat{\pi}_t / \pi_t$.

We conclude from the previous lemma that one can either solve the backward Kolmogorov equation (2.31) with $\psi_1(z) = l(z)/\beta$ or the backward SDE (2.28) with $Z_1^- \sim \widehat{\pi}_1 = l \, \pi_1 / \beta$ in order to derive the desired control law $u_t(z) = \gamma \nabla_z \log \psi_t$ in (2.50).

Remark 2.20. The notion of a backward SDE used throughout this paper is different from the notion of a backward SDE in the sense of Carmona (2016), for example. More specifically, Itô's formula

$$\mathrm{d}\psi_t = \partial_t \psi_t \, \mathrm{d}t + \nabla_z \psi_t \cdot \mathrm{d}Z_t^+ + \frac{\gamma}{2} \Delta_z \psi_t \, \mathrm{d}t$$

and the fact that ψ_t satisfies the backward Kolmogorov equation (2.31) imply that

$$\mathrm{d}\psi_t = \gamma^{1/2} \nabla_z \psi_t \cdot \mathrm{d}W_t^+ \tag{2.53}$$

along solutions of the forward SDE (2.24). In other words, the quantities ψ_t are materially advected along solutions Z_t^+ of the forward SDEs in expectation or, in the language of stochastic analysis, ψ_t is a martingale. Hence, by the martingale representation theorem, we can reformulate the problem of

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

determining ψ_t as follows. Find the solution (Y_t, V_t) of the backward SDE

$$\mathrm{d}Y_t = V_t \cdot \mathrm{d}W_t^+ \tag{2.54}$$

subject to the final condition $Y_1 = l(Z_1^+)/\beta$ at $t = 1$. Here (2.54) has to be understood as a backward SDE in the sense of Carmona (2016), for example, where the solution (Y_t, V_t) is adapted to the forward SDE (2.24), *i.e.* to the past $s \le t$, whereas the solution Z_t^- of the backward SDE (2.28) is adapted to the future $s \ge t$. The solution of (2.54) is given by $Y_t = \psi_t(Z_t^+)$ and $V_t = \gamma^{1/2}\,\nabla_z \psi_t(Z_t^+)$ in agreement with (2.53) (compare Carmona 2016, page 42). See Appendix A.4 for the numerical treatment of (2.54).

A variational characterization of the smoothing path measure $\widehat{\mathbb{P}}$ is given by the Donsker–Varadhan principle

$$\widehat{\mathbb{P}} = \arg \inf_{\mathbb{P} \ll \mathbb{Q}} \{-\mathbb{P}[\log(l)] + \mathrm{KL}(\mathbb{P}||\mathbb{Q})\}, \tag{2.55}$$

that is, the distribution $\widehat{\mathbb{P}}$ is chosen such that the expected loss, $-\mathbb{P}[\log(l)]$, is minimized subject to the penalty introduced by the Kullback–Leibler divergence with respect to the original path measure $\mathbb{Q}$. Note that

$$\inf_{\mathbb{P} \ll \mathbb{Q}} \{\mathbb{P}[-\log(l)] + \mathrm{KL}(\mathbb{P}||\mathbb{Q})\} = -\log \beta \tag{2.56}$$

with $\beta = \mathbb{Q}[l]$. The connection between smoothing for SDEs and the Donsker–Varadhan principle has, for example, been discussed by Mitter and Newton (2003). See also Hartmann *et al.* (2017) for an in-depth discussion of variational formulations and their numerical implementation in the context of rare event simulations for which it is generally assumed that $\pi_0(z) = \delta(z - z_0)$ in (2.3), that is, the ensemble size is $M = 1$ when viewed within the context of this paper.

Remark 2.21. One can choose ψ_t differently from the choice made in (2.51) by changing the final condition for the backward Kolmogorov equation (2.31) to any suitable ψ_1. As already discussed for twisted discrete-time smoothing, such modifications give rise to alternative representations of the smoothing distribution $\widehat{\mathbb{P}}$ in terms of modified forward SDEs, likelihood functions and initial distributions. See Kappen and Ruiz (2016) and Ruiz and Kappen (2017) for an application of these ideas to importance sampling in the context of partially observed diffusion processes. More specifically, let u_t denote the associated control law (2.32) for the forward SDE (2.30) with given initial distribution $Z_0^+ \sim q_0$. Then

$$\frac{\mathrm{d}\widehat{\mathbb{P}}}{\mathrm{d}\mathbb{Q}}_{|z^u_{[0,1]}} = \frac{\mathrm{d}\widehat{\mathbb{P}}}{\mathrm{d}\mathbb{Q}^u}_{|z^u_{[0,1]}} \frac{\mathrm{d}\mathbb{Q}^u}{\mathrm{d}\mathbb{Q}}_{|z^u_{[0,1]}},$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

which, using (2.33) and (2.49), implies the modified Radon–Nikodym derivative

$$\frac{\mathrm{d}\widehat{\mathbb{P}}}{\mathrm{d}\mathbb{Q}^u}_{|z^u_{[0,1]}} = \frac{l(z^u_1)}{\beta}\frac{\pi_0(z^u_0)}{q_0(z^u_0)}\exp\left(-\frac{1}{2\gamma}\int_0^1 (\|u_t\|^2\,\mathrm{d}t + 2\gamma^{1/2}u_t\cdot \mathrm{d}W_t^+)\right). \quad (2.57)$$

Recall from Lemma 2.18 that the control law (2.32) with ψ_t defined by (2.51) together with $q_0 = \widehat{\pi}_0$ leads to $\widehat{\mathbb{P}} = \mathbb{Q}^u$.

2.3. Schrödinger problem

In this subsection we show that scenario (C) from Definition 2.4 leads to a certain boundary value problem first considered by Schrödinger (1931). More specifically, we state the so-called Schrödinger problem and show how it is linked to data assimilation scenario (C).

In order to introduce the Schrödinger problem, we return to the twisting potential approach as utilized in Section 2.2, with two important modifications. These modifications are, first, that the twisting potential ψ is determined implicitly and, second, that the modified transition kernel q_+^ψ is applied to π_0 instead of the tilted initial density π_0^ψ as in (2.48). More specifically, we have the following.

Definition 2.22. We seek the pair of functions $\widehat{\psi}(z_0)$ and $\psi(z_1)$ which solve the boundary value problem

$$\pi_0(z_0) = \pi_0^\psi(z_0)\,\widehat{\psi}(z_0), \quad (2.58)$$

$$\widehat{\pi}_1(z_1) = \pi_1^\psi(z_1)\,\psi(z_1), \quad (2.59)$$

$$\pi_1^\psi(z_1) = \int q_+(z_1|z_0)\,\pi_0^\psi(z_0)\,\mathrm{d}z_0, \quad (2.60)$$

$$\widehat{\psi}(z_0) = \int q_+(z_1|z_0)\,\psi(z_1)\,\mathrm{d}z_1, \quad (2.61)$$

for given marginal (filtering) distributions π_0 and $\widehat{\pi}_1$ at $t = 0$ and $t = 1$, respectively. The required modified PDFs π_0^ψ and π_1^ψ are defined by (2.58) and (2.59), respectively. The solution $(\widehat{\psi}, \psi)$ of the Schrödinger system (2.58)–(2.61) leads to the modified transition kernel

$$q_+^*(z_1|z_0) := \psi(z_1)\,q_+(z_1|z_0)\,\widehat{\psi}(z_0)^{-1}, \quad (2.62)$$

which satisfies

$$\widehat{\pi}_1(z_1) = \int q_+^*(z_1|z_0)\,\pi_0(z_0)\,\mathrm{d}z_0$$

by construction.

The modified transition kernel $q_+^*(z_1|z_0)$ couples the two marginal distributions π_0 and $\widehat{\pi}_1$ with the twisting potential ψ implicitly defined. In other

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

words, q_+^* provides the transition kernel for going from the initial distribution (2.3) at time t_0 to the filtering distribution at time t_1 without the need for any reweighting, *i.e.* the desired transition kernel for scenario (C). See Leonard (2014) and Chen *et al.* (2014) for more mathematical details on the Schrödinger problem.

Remark 2.23. Let us compare the Schrödinger system to the twisting potential approach (2.48) for the smoothing problem from Section 2.2 in some more detail. First, note that the twisting potential approach to smoothing replaces (2.58) with

$$\pi_0^\psi(z_0) = \pi_0(z_0)\,\widehat{\psi}(z_0)$$

and (2.59) with

$$\pi_1^\psi(z_1) = \pi_1(z_1)\,\psi(z_1),$$

where ψ is a given twisting potential normalized such that $\pi_1[\psi] = 1$. The associated $\widehat{\psi}$ is determined by (2.61) as in the twisting approach. In both cases, the modified transition kernel is given by (2.62). Finally, (2.60) is replaced by the prediction step (2.4).

In order to solve the Schrödinger system for our given initial distribution (2.3) and the associated filter distribution $\widehat{\pi}_1$, we make the *ansatz*

$$\pi_0^\psi(z_0) = \frac{1}{M}\sum_{i=1}^M \alpha^i\,\delta(z_0 - z_0^i), \quad \sum_{i=1}^M \alpha^i = M.$$

This *ansatz* together with (2.58)–(2.61) immediately implies

$$\widehat{\psi}(z_0^i) = \frac{1}{\alpha^i}, \quad \pi_1^\psi(z_1) = \frac{1}{M}\sum_{i=1}^M \alpha^i\,q_+(z_1|z_0^i),$$

as well as

$$\psi(z_1) = \frac{\widehat{\pi}_1(z_1)}{\pi_1^\psi(z_1)} = \frac{l(z_1)}{\beta}\frac{\frac{1}{M}\sum_{j=1}^M q_+(z_1|z_0^j)}{\frac{1}{M}\sum_{j=1}^M \alpha^j\,q_+(z_1|z_0^j)}. \tag{2.63}$$

Hence we arrive at the equations

$$\begin{aligned}\widehat{\psi}(z_0^i) &= \frac{1}{\alpha^i} &(2.64)\\ &= \int \psi(z_1)\,q_+(z_1|z_0^i)\,\mathrm{d}z_1\\ &= \int \frac{l(z_1)}{\beta}\frac{\sum_{j=1}^M q_+(z_1|z_0^j)}{\sum_{j=1}^M \alpha^j\,q_+(z_1|z_0^j)}\,q_+(z_1|z_0^i)\,\mathrm{d}z_1 &(2.65)\end{aligned}$$

for $i = 1, \ldots, M$. These M equations have to be solved for the M unknown

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

coefficients $\{\alpha^i\}$. In other words, the Schrödinger problem becomes finite-dimensional in the context of this paper. More specifically, we have the following result.

Lemma 2.24. The forward Schrödinger transition kernel (2.62) is given by

$$q_+^*(z_1|z_0^i) = \frac{\widehat{\pi}_1(z_1)}{\pi_1^\psi(z_1)} q_+(z_1|z_0^i)\,\alpha^i$$
$$= \frac{\alpha^i \sum_{j=1}^M q_+(z_1|z_0^j)}{\sum_{j=1}^M \alpha^j\, q_+(z_1|z_0^j)}\, \frac{l(z_1)}{\beta}\, q_+(z_1|z_0^i), \tag{2.66}$$

for each particle z_0^i with the coefficients α^j, $j = 1, \ldots, M$, defined by (2.64)–(2.65).

Proof. Because of (2.64)–(2.65), the forward transition kernels (2.66) satisfy

$$\int q_+^*(z_1|z_0^i)\,\mathrm{d}z_1 = 1 \tag{2.67}$$

for all $i = 1, \ldots, M$ and

$$\int q_+^*(z_1|z_0)\,\pi_0(z_0)\,\mathrm{d}z = \frac{1}{M}\sum_{i=1}^M q_+^*(z_1|z_0^i)$$
$$= \frac{1}{M}\sum_{i=1}^M \alpha^i q_+(z_1|z_0^i)\, \frac{\widehat{\pi}_1(z_1)}{\frac{1}{M}\sum_{j=1}^M \alpha^j\, q_+(z_1|z_0^j)}$$
$$= \widehat{\pi}_1(z_1), \tag{2.68}$$

as desired. □

Numerical implementations will be discussed in Section 3. Note that knowledge of the normalizing constant β is not required *a priori* for solving (2.64)–(2.65) since it appears as a common scaling factor.

We note that the coefficients $\{\alpha^j\}$ together with the associated potential ψ from the Schrödinger system provide the optimally twisted prediction kernel (1.2) with respect to the filtering distribution $\widehat{\pi}_1$, that is, we set $\widehat{\psi}(z_0^i) = 1/\alpha^i$ in (2.16) and define the potential ψ by (2.63).

Lemma 2.25. The Schrödinger transition kernel (2.62) satisfies the following constrained variational principle. Consider the joint PDFs given by $\pi(z_0, z_1) := q_+(z_1|z_0)\pi_0(z_0)$ and $\pi^*(z_0, z_1) := q_+^*(z_1|z_0)\pi_0(z_0)$. Then

$$\pi^* = \arg \inf_{\widetilde{\pi} \in \Pi_S} \mathrm{KL}(\widetilde{\pi}||\pi). \tag{2.69}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

Here a joint PDF $\widetilde{\pi}(z_0, z_1)$ is an element of Π_{S} if

$$\int \widetilde{\pi}(z_0, z_1)\,\mathrm{d}z_1 = \pi_0(z_0), \quad \int \widetilde{\pi}(z_0, z_1)\,\mathrm{d}z_0 = \widehat{\pi}_1(z_1).$$

Proof. See Föllmer and Gantert (1997) for a proof and Remark 2.31 for a heuristic derivation in the case of discrete measures. □

The constrained variational formulation (2.69) of Schrödinger's problem should be compared to the unconstrained Donsker–Varadhan variational principle

$$\widehat{\pi} = \arg\inf\{-\widetilde{\pi}[\log(l)] + \mathrm{KL}(\widetilde{\pi}||\pi)\} \tag{2.70}$$

for the associated smoothing problem. See Remark 2.27 below.

Remark 2.26. The Schrödinger problem is closely linked to optimal transportation (Cuturi 2013, Leonard 2014, Chen *et al.* 2014). For example, consider the Gaussian transition kernel (2.21) with $\Psi(z) = z$ and $B = I$. Then the solution (2.66) to the associated Schrödinger problem of coupling π_0 and $\widehat{\pi}_1$ reduces to the solution π^* of the associated optimal transport problem

$$\pi^* = \arg\inf_{\widetilde{\pi}\in\Pi_{\mathrm{S}}} \int\int \|z_0 - z_1\|^2\,\widetilde{\pi}(z_0, z_1)\,\mathrm{d}z_0\,\mathrm{d}z_1$$

in the limit $\gamma \to 0$.

2.3.1. SDE models (cont.)

At the SDE level, Schrödinger's problem amounts to continuously bridging the given initial PDF π_0 with the PDF $\widehat{\pi}_1$ at final time using an appropriate modification of the stochastic process $Z^+_{[0,1]} \sim \mathbb{Q}$ defined by the forward SDE (2.24) with initial distribution π_0 at $t = 0$. The desired modified stochastic process $\mathbb{P}^*$ is defined as the minimizer of

$$\mathcal{L}(\widetilde{\mathbb{P}}) := \mathrm{KL}(\widetilde{\mathbb{P}}||\mathbb{Q})$$

subject to the constraint that the marginal distributions $\widetilde{\pi}_t$ of $\widetilde{\mathbb{P}}$ at time $t = 0$ and $t = 1$ satisfy π_0 and $\widehat{\pi}_1$, respectively (Föllmer and Gantert 1997, Leonard 2014, Chen *et al.* 2014).

Remark 2.27. We note that the Donsker–Varadhan variational principle (2.55), characterizing the smoothing path measure $\widehat{\mathbb{P}}$, can be replaced by

$$\mathbb{P}^* = \arg\inf_{\widetilde{\mathbb{P}}\in\Pi} \{-\widetilde{\pi}_1[\log(l)] + \mathrm{KL}(\widetilde{\mathbb{P}}||\mathbb{Q})\}$$

with

$$\Pi = \{\widetilde{\mathbb{P}} \ll \mathbb{Q} :\ \widetilde{\pi}_1 = \widehat{\pi}_1,\ \widetilde{\pi}_0 = \pi_0\}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

in the context of Schrödinger's problem. The associated

$$-\log \beta^* := \inf_{\widetilde{\mathbb{P}} \in \Pi} \{-\widetilde{\pi}_1[\log(l)] + \mathrm{KL}(\widetilde{\mathbb{P}}||\mathbb{Q})\} = -\widehat{\pi}[\log(l)] + \mathrm{KL}(\mathbb{P}^*||\mathbb{Q})$$

can be viewed as a generalization of (2.56) and gives rise to a generalized evidence β^*, which could be used for model comparison and parameter estimation.

The Schrödinger process $\mathbb{P}^*$ corresponds to a Markovian process across the whole time domain $[0,1]$ (Leonard 2014, Chen *et al.* 2014). More specifically, consider the controlled forward SDE (2.30) with initial conditions

$$Z_0^+ \sim \pi_0$$

and a given control law u_t for $t \in [0,1]$. Let $\mathbb{P}^u$ denote the path measure associated to this process. Then, as discussed in detail by Dai Pra (1991), one can find time-dependent potentials ψ_t with associated control laws (2.32) such that the marginal of the associated path measure $\mathbb{P}^u$ at times $t = 1$ satisfies

$$\pi_1^u = \widehat{\pi}_1$$

and, more generally,

$$\mathbb{P}^* = \mathbb{P}^u.$$

We summarize this result in the following lemma.

Lemma 2.28. The Schrödinger path measure $\mathbb{P}^*$ can be generated by a controlled SDE (2.30) with control law (2.32), where the desired potential ψ_t can be obtained as follows. Let $(\widehat{\psi}, \psi)$ denote the solution of the associated Schrödinger system (2.58)–(2.61), where $q_+(z_1|z_0)$ denotes the time-one forward transition kernel of (2.24). Then ψ_t in (2.32) is the solution of the backward Kolmogorov equation (2.31) with prescribed $\psi_1 = \psi$ at final time $t = 1$.

Remark 2.29. As already pointed out in the context of smoothing, the desired potential ψ_t can also be obtained by solving an appropriate backward SDE. More specifically, given the solution $(\widehat{\psi}, \psi)$ and the implied PDF $\widetilde{\pi}_0^+ := \pi_0^\psi = \pi_0/\widehat{\psi}$ of the Schrödinger system (2.58)–(2.61), let $\widetilde{\pi}_t^+$, $t \geq 0$, denote the marginals of the forward SDE (2.24) with $Z_0^+ \sim \widetilde{\pi}_0^+$. Furthermore, consider the backward SDE (2.28) with drift term

$$b_t(z) = f_t(z) - \gamma \nabla_z \log \widetilde{\pi}_t^+(z), \tag{2.71}$$

and final time condition $Z_1^- \sim \widehat{\pi}_1$. Then the choice of $\widetilde{\pi}_0^+$ ensures that $Z_0^- \sim \pi_0$. Furthermore the desired control in (2.30) is provided by

$$u_t = \gamma \nabla_z \log \frac{\widetilde{\pi}_t^-}{\widetilde{\pi}_t^+},$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

where $\widetilde{\pi}_t^-$ denotes the marginal distributions of the backward SDE (2.28) with drift term (2.71) and $\widetilde{\pi}_1^- = \widehat{\pi}_1$. We will return to this reformulation of the Schrödinger problem in Section 3 when considering it as the limit of a sequence of smoothing problems.

Remark 2.30. The solution to the Schrödinger problem for linear SDEs and Gaussian marginal distributions has been discussed in detail by Chen, Georgiou and Pavon (2016*b*).

2.3.2. Discrete measures

We finally discuss the Schrödinger problem in the context of finite-state Markov chains in more detail. These results will be needed in the following sections on the numerical implementation of the Schrödinger approach to sequential data assimilation.

Let us therefore consider an example which will be closely related to the discussion in Section 3. We are given a bi-stochastic matrix $Q \in \mathbb{R}^{L\times M}$ with all entries satisfying $q_{lj} > 0$ and two discrete probability measures represented by vectors $p_1 \in \mathbb{R}^L$ and $p_0 \in \mathbb{R}^M$, respectively. Again we assume for simplicity that all entries in p_1 and p_0 are strictly positive. We introduce the set of all bi-stochastic $L \times M$ matrices with those discrete probability measures as marginals, that is,

$$\Pi_{\mathrm{s}} := \{P \in \mathbb{R}^{L\times M} : P \geq 0,\, P^{\mathrm{T}}\mathbb{1}_L = p_0,\, P\mathbb{1}_M = p_1\}. \tag{2.72}$$

Solving Schrödinger's system (2.58)–(2.61) corresponds to finding two non-negative vectors $u \in \mathbb{R}^L$ and $v \in \mathbb{R}^M$ such that

$$P^* := D(u)\, Q\, D(v)^{-1} \in \Pi_{\mathrm{s}}.$$

In turns out that P^* is uniquely determined and minimizes the Kullback–Leibler divergence between all $P \in \Pi_{\mathrm{s}}$ and the reference matrix Q, that is,

$$P^* = \arg\min_{P\in\Pi_{\mathrm{s}}} \mathrm{KL}\,(P||Q). \tag{2.73}$$

See Peyre and Cuturi (2018) and the following remark for more details.

Remark 2.31. If one makes the *ansatz*

$$p_{lj} = \frac{u_l q_{lj}}{v_j},$$

then the minimization problem (2.73) becomes equivalent to

$$P^* = \arg\min_{(u,v)>0} \sum_{l,j} p_{lj}(\log u_l - \log v_j)$$

subject to the additional constraints

$$P\,\mathbb{1}_M = D(u)QD(v)^{-1}\,\mathbb{1}_M = p_1,\; P^{\mathrm{T}}\,\mathbb{1}_L = D(v)^{-1}Q^{\mathrm{T}}D(u)\,\mathbb{1}_L = p_0.$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

Note that these constraint determine $u > 0$ and $v > 0$ up to a common scaling factor. Hence (2.73) can be reduced to finding $(u, v) > 0$ such that

$$u^{\mathrm{T}} \mathbb{1}_L = 1, \quad P \mathbb{1}_M = p_1, \quad P^{\mathrm{T}} \mathbb{1}_L = p_0.$$

Hence we have shown that solving the Schrödinger system is equivalent to solving the minimization problem (2.73) for discrete measures. Thus

$$\min_{P \in \Pi_{\mathrm{s}}} \mathrm{KL}\,(P||Q) = p_1^{\mathrm{T}} \log u - p_0^{\mathrm{T}} \log v.$$

Lemma 2.32. The Sinkhorn iteration (Sinkhorn 1967)

$$u^{k+1} := D(P^{2k} \mathbb{1}_M)^{-1}\, p_1, \tag{2.74}$$

$$P^{2k+1} := D(u^{k+1})\, P^{2k}, \tag{2.75}$$

$$v^{k+1} := D(p_0)^{-1}\, (P^{2k+1})^{\mathrm{T}}\, \mathbb{1}_L, \tag{2.76}$$

$$P^{2k+2} := P^{2k+1}\, D(v^{k+1})^{-1}, \tag{2.77}$$

$k = 0, 1, \ldots$, with initial $P^0 = Q \in \mathbb{R}^{L \times M}$ provides an algorithm for computing P^*, that is,

$$\lim_{k \to \infty} P^k = P^*. \tag{2.78}$$

Proof. See, for example, Peyre and Cuturi (2018) for a proof of (2.78), which is based on the contraction property of the iteration (2.74)–(2.77) with respect to the Hilbert metric on the projective cone of positive vectors. □

It follows from (2.78) that

$$\lim_{k \to \infty} u^k = \mathbb{1}_L, \quad \lim_{k \to \infty} v^k = \mathbb{1}_M.$$

The essential idea of the Sinkhorn iteration is to enforce

$$P^{2k+1}\, \mathbb{1}_M = p_1, \quad (P^{2k})^{\mathrm{T}}\, \mathbb{1}_L = p_0$$

at each iteration step and that the matrix P^* satisfies both constraints simultaneously in the limit $k \to \infty$. See Cuturi (2013) for a computationally efficient and robust implementation of the Sinkhorn iteration.

Remark 2.33. One can introduce a similar iteration for the Schrödinger system (2.58)–(2.61). For example, pick $\widehat{\psi}(z_0) = 1$ initially. Then (2.58) implies $\pi_0^{\psi} = \pi_0$ and (2.60) $\pi_1^{\psi} = \pi_1$. Hence $\psi = l/\beta$ in the first iteration. The second iteration starts with $\widehat{\psi}$ determined by (2.61) with $\psi = l/\beta$. We again cycle through (2.58), (2.60) and (2.59) in order to find the next approximation to ψ. The third iteration takes now this ψ and computes the associated $\widehat{\psi}$ from (2.61) *etc.* A numerical implementation of this procedure requires the approximation of two integrals which essentially leads back to a Sinkhorn type algorithm in the weights of an appropriate quadrature rule.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

3. Numerical methods

Having summarized the relevant mathematical foundation for prediction, filtering (data assimilation scenario (A)) and smoothing (scenario (B)) and the Schrödinger problem (scenario (C)), we now discuss numerical approximations suitable for ensemble-based data assimilation. It is clearly impossible to cover all available methods, and we will focus on a selection of approaches which are built around the idea of optimal transport, ensemble transform methods and Schrödinger systems. We will also focus on methods that can be applied or extended to problems with high-dimensional state spaces even though we will not explicitly cover this topic in this survey. See Reich and Cotter (2015), van Leeuwen (2015) and Asch *et al.* (2017) instead.

3.1. Prediction

Generating samples from the forecast distributions $q_+(\cdot\,|z_0^i)$ is in most cases straightforward. The computational expenses can, however, vary dramatically, and this impacts on the choice of algorithms for sequential data assimilation. We demonstrate in this subsection how samples from the prediction PDF π_1 can be used to construct an associated finite-state Markov chain that transforms π_0 into an empirical approximation of π_1.

Definition 3.1. Let us assume that we have $L \geq M$ independent samples z_1^l from the M forecast distributions $q_+(\cdot\,|z_0^j)$, $j = 1, \ldots, M$. We introduce the $L \times M$ matrix Q with entries

$$q_{lj} := q_+(z_1^l|z_0^j). \tag{3.1}$$

We now consider the associated bi-stochastic matrix $P^* \in \mathbb{R}^{L\times M}$, as defined by (2.73), with the two probability vectors in (2.72) given by $p_1 = \mathbb{1}_L/L \in \mathbb{R}^L$ and $p_0 = \mathbb{1}_M/M \in \mathbb{R}^M$, respectively. The finite-state Markov chain

$$Q_+ := MP^* \tag{3.2}$$

provides a sample-based approximation to the forward transition kernel $q_+(z_1|z_0)$.

More precisely, the ith column of Q_+ provides an empirical approximation to $q_+(\cdot|z_0^i)$ and

$$Q_+\, p_0 = p_1 = \frac{1}{L}\mathbb{1}_L,$$

which is in agreement with the fact that the z_1^l are equally weighted samples from the forecast PDF π_1.

Remark 3.2. Because of the simple relation between a bi-stochastic matrix $P \in \Pi_s$ with p_0 in (2.72) given by $p_0 = \mathbb{1}_M/M$ and its associated finite-state Markov chain $Q_+ = MP$, one can reformulate the minimization

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

problem (2.73) in those cases directly in terms of Markov chains $Q_+ \in \Pi_{\rm M}$ with the definition of $\Pi_{\rm s}$ adjusted to

$$\Pi_{\rm M} := \left\{ Q \in \mathbb{R}^{L\times M} : \, Q \geq 0, \, Q^{\rm T} \mathbb{1}_L = \mathbb{1}_M, \, \frac{1}{M} Q \mathbb{1}_M = p_1 \right\}. \tag{3.3}$$

Remark 3.3. The associated backward transition kernel $Q_- \in \mathbb{R}^{M\times L}$ satisfies

$$Q_- \, D(p_1) = (Q_+ \, D(p_0))^{\rm T}$$

and is hence given by

$$Q_- = (Q_+ \, D(p_0))^{\rm T} \, D(p_1)^{-1} = \frac{L}{M} Q_+^{\rm T}.$$

Thus

$$Q_- \, p_1 = D(p_0) \, Q_+^{\rm T} \, \mathbb{1}_L = D(p_0) \, \mathbb{1}_M = p_0,$$

as desired.

Definition 3.4. We can extend the concept of twisting to discrete Markov chains such as (3.2). A twisting potential ψ gives rise to a vector $u \in \mathbb{R}^L$ with normalized entries

$$u_l = \frac{\psi(z_1^l)}{\sum_{k=1}^L \psi(z_1^k)},$$

$l = 1, \ldots, L$. The twisted finite-state Markov kernel is now defined by

$$Q_+^\psi := D(u) \, Q_+ \, D(v)^{-1}, \quad v := (D(u) \, Q_+)^{\rm T} \, \mathbb{1}_L \in \mathbb{R}^M, \tag{3.4}$$

and thus $\mathbb{1}_L^{\rm T} \, Q_+^\psi = \mathbb{1}_M^{\rm T}$, as required for a Markov kernel. The twisted forecast probability is given by

$$p_1^\psi := Q_+^\psi \, p_0$$

with $p_0 = \mathbb{1}_M / M$. Furthermore, if we set $p_0 = v$ then $p_1^\psi = u$.

3.1.1. Gaussian model errors (cont.)

The proposal density is given by (2.21) and it is easy to produce $K > 1$ samples from each of the M proposals $q_+(\cdot \, | z_0^j)$. Hence we can make the total sample size $L = K \, M$ as large as desired. In order to produce M samples, $\widetilde{z}_1^j$, from a twisted finite-state Markov chain (3.4), we draw a single realization from each of the M associated discrete random variables $\widetilde{Z}_1^j$, $j = 1, \ldots, M$, with probabilities

$$\mathbb{P}[\widetilde{Z}_1^j(\omega) = z_1^l] = (Q_+^\psi)_{lj}.$$

We will provide more details when discussing the Schrödinger problem in the context of Gaussian model errors in Section 3.4.1.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

3.1.2. SDE models (cont.)
The Euler–Maruyama method (Kloeden and Platen 1992)

$$Z_{n+1}^{+} = Z_n^{+} + f_{t_n}(Z_n^{+})\,\Delta t + (\gamma \Delta t)^{1/2}\,\Xi_n, \quad \Xi_n \sim \mathrm{N}(0, I), \tag{3.5}$$

$n = 0, \ldots, N-1$, will be used for the numerical approximation of (2.24) with step-size $\Delta t := 1/N$, $t_n = n\Delta t$. In other words, we replace $Z_{t_n}^{+}$ with its numerical approximation Z_n^{+}. A numerical approximation (realization) of the whole solution path $z_{[0,1]}$ will be denoted by $z_{0:N} = Z_{0:N}^{+}(\omega)$ and can be computed recursively due to the Markov property of the Euler–Maruyama scheme. The marginal PDFs of Z_n^{+} are denoted by π_n.

For any finite number of time-steps N, we can define a joint PDF $\pi_{0:N}$ on $\mathcal{U}_N = \mathbb{R}^{N_z \times (N+1)}$ via

$$\pi_{0:N}(z_{0:N}) \propto \exp\left(-\frac{1}{2\Delta t}\sum_{n=0}^{N-1} \|\eta_n\|^2\right)\pi_0(z_0) \tag{3.6}$$

with

$$\eta_n := \gamma^{-1/2}(z_{n+1} - z_n - f_{t_n}(z_n)\Delta t) \tag{3.7}$$

and $\eta_n = \Delta t^{1/2}\Xi_n(\omega)$. Note that the joint PDF $\pi_{0:N}(z_{0:N})$ can also be expressed in terms of z_0 and $\eta_{0:N-1}$.

The numerical approximation of SDEs provides an example for which the increase in computational cost for producing $L > M$ samples from the PDF $\pi_{0:N}$ versus $L = M$ is non-trivial, in general.

We now extend Definition 3.1 to the case of temporally discretized SDEs in the form of (3.5).

Definition 3.5. Let us assume that we have $L = M$ independent numerical solutions $z_{0:N}^{i}$ of (3.5). We introduce an $M \times M$ matrix Q_n for each $n = 1, \ldots, N$ with entries

$$q_{lj} = q_{+}(z_n^{l} | z_{n-1}^{j}) := \mathrm{n}(z_n^{l}; z_{n-1}^{j} + \Delta t f(z_{n-1}^{j}), \gamma \Delta t\, I).$$

With each Q_n we associate a finite-state Markov chain Q_n^{+} as defined by (3.2) for general transition densities q_{+} in Definition 3.1. An approximation of the Markov transition from time $t_0 = 0$ to $t_1 = 1$ is now provided by

$$Q_{+} := \prod_{n=1}^{N} Q_n^{+}. \tag{3.8}$$

Remark 3.6. The approximation (3.2) can be related to the diffusion map approximation of the infinitesimal generator of Brownian dynamics

$$\mathrm{d}Z_t^{+} = -\nabla_z U(Z_t^{+})\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t^{+} \tag{3.9}$$

with potential $U(z) = -\log \pi^{*}(z)$ in the following sense. First note that

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

π^* is invariant under the associated Fokker–Planck equation (2.26) with (time-independent) operator $\mathcal{L}^\dagger$ written in the form

$$\mathcal{L}^\dagger \pi = \nabla_z \cdot \left(\pi^* \nabla_z \frac{\pi}{\pi^*} \right).$$

Let z^i, $i = 1, \ldots, M$, denote M samples from the invariant PDF π^* and define the symmetric matrix $Q \in \mathbb{R}^{M \times M}$ with entries

$$q_{lj} = \mathrm{n}(z^l; z^j, 2\Delta t\, I).$$

Then the associated (symmetric) matrix (3.2), as introduced in Definition 3.1, provides a discrete approximation to the evolution of a probability vector $p_0 \propto \pi_0/\pi^*$ over a time-interval Δt and, hence, to the semigroup operator $\mathrm{e}^{\Delta t \mathcal{L}}$ with the infinitesimal generator $\mathcal{L}$ given by

$$\mathcal{L} g = \frac{1}{\pi^*} \nabla_z \cdot (\pi^* \nabla_z g). \tag{3.10}$$

We formally obtain

$$\mathcal{L} \approx \frac{Q_+ - I}{\Delta t} \tag{3.11}$$

for Δt sufficiently small. The symmetry of Q_+ reflects the fact that $\mathcal{L}$ is self-adjoint with respect to the weighted inner product

$$\langle f, g \rangle_{\pi^*} = \int f(z)\, g(z)\, \pi^*(z)\, \mathrm{d}z.$$

See Harlim (2018) for a discussion of alternative diffusion map approximations to the infinitesimal generator $\mathcal{L}$ and Appendix A.1 for an application to the feedback particle filter formulation of continuous-time data assimilation.

We also consider the discretization

$$Z^+_{n+1} = Z^+_n + (f_{t_n}(Z^+_n) + u_{t_n}(Z^+_n))\Delta t + (\gamma \Delta t)^{1/2}\, \Xi_n, \tag{3.12}$$

$n = 0, \ldots, N-1$, of a controlled SDE (2.30) with associated PDF $\pi^u_{0:N}$ defined by

$$\pi^u_{0:N}(z^u_{0:N}) \propto \exp\left(-\frac{1}{2\Delta t} \sum_{n=0}^{N-1} \|\eta^u_n\|^2 \right) \pi_0(z_0), \tag{3.13}$$

where

$$\begin{aligned} \eta^u_n &:= \gamma^{-1/2} \{ z^u_{n+1} - z^u_n - (f_{t_n}(z^u_n) + u_{t_n}(z^u_n))\Delta t \} \\ &= \eta_n - \frac{\Delta t}{\gamma^{1/2}} u_{t_n}(z^u_n). \end{aligned}$$

Here $z^u_{0:N}$ denotes a realization of the discretization (3.12) with control laws

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

u_{t_n}. We find that

$$\frac{1}{2\Delta t}\|\eta_n^u\|^2 = \frac{1}{2\Delta t}\|\eta_n\|^2 - \frac{1}{\gamma^{1/2}}u_{t_n}(z_n^u)^{\mathrm{T}}\eta_n + \frac{\Delta t}{2\gamma}\|u_{t_n}(z_n^u)\|^2$$
$$= \frac{1}{2\Delta t}\|\eta_n\|^2 - \frac{1}{\gamma^{1/2}}u_{t_n}(z_n^u)^{\mathrm{T}}\eta_n^u - \frac{\Delta t}{2\gamma}\|u_{t_n}(z_n^u)\|^2,$$

and hence

$$\frac{\pi^u_{0:N}(z^u_{0:N})}{\pi_{0:N}(z^u_{0:N})} = \exp\left(\frac{1}{2\gamma}\sum_{n=0}^{N-1}(\|u_{t_n}(z_n^u)\|^2\Delta t + 2\gamma^{1/2}u_{t_n}(z_n^u)^{\mathrm{T}}\eta_n^u)\right), \quad (3.14)$$

which provides a discrete version of (2.33) since $\eta_n^u = \Delta t^{1/2}\,\Xi_n(\omega)$ are increments of Brownian motion over time intervals of length Δt.

Remark 3.7. Instead of discretizing the forward SDE (2.24) in order to produce samples from the forecast PDF π_1, one can also start from the mean-field formulation (2.29) and its time discretization, for example,

$$z_{n+1}^i = z_n^i + (f_{t_n}(z_n^i) + u_{t_n}(z_n^i))\Delta t \quad (3.15)$$

for $i = 1, \ldots, M$ and

$$u_{t_n}(z) = -\frac{\gamma}{2}\nabla_z \log \widetilde{\pi}_n(z).$$

Here $\widetilde{\pi}_n$ stands for an approximation to the marginal PDF π_{t_n} based on the available samples z_n^i, $i = 1, \ldots, M$. A simple approximation is obtained by the Gaussian PDF

$$\widetilde{\pi}_n(z) = \mathrm{n}(z; \bar{z}_n, P_n^{zz})$$

with empirical mean

$$\bar{z}_n = \frac{1}{M}\sum_{i=1}^M z_n^i$$

and empirical covariance matrix

$$P_n^{zz} = \frac{1}{M-1}\sum_{i=1}^M z_n^i(z_n^i - \bar{z}_n)^{\mathrm{T}}.$$

The system (3.15) becomes

$$z_{n+1}^i = z_n^i + (f_{t_n}(z_n^i) + \gamma(P_n^{zz})^{-1}(z_n^i - \bar{z}_n))\Delta t,$$

$i = 1, \ldots, M$, and provides an example of an interacting particle approximation. Similar mean-field formulations can be found for the backward SDE (2.28).

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

3.2. Filtering

Let us assume that we are given M samples, z_1^i, from the forecast PDF using forward transition kernels $q_+(\cdot\,|z_0^i)$, $i = 1, \ldots, M$. The likelihood function $l(z)$ leads to importance weights

$$w^i \propto l(z_1^i). \tag{3.16}$$

We also normalize these importance weights such that (2.19) holds.

Remark 3.8. The model evidence β can be estimated from the samples, z_1^i, and the likelihood $l(z)$ as follows:

$$\widetilde{\beta} := \frac{1}{M}\sum_{i=1}^M l(z_1^i).$$

If the likelihood is of the form

$$l(z) \propto \exp\left(-\frac{1}{2}(y_1 - h(z))^{\mathrm{T}} R^{-1}(y_1 - h(z)\right)$$

and the prior distribution in $y = h(z)$ can be approximated as being Gaussian with covariance

$$P^{hh} := \frac{1}{M-1}\sum_{i=1}^M h(z_1^i)(h(z_1^i) - \bar{h})^{\mathrm{T}}, \quad \bar{h} := \frac{1}{M}\sum_{i=1}^M h(z_1^i),$$

then the evidence can be approximated by

$$\widetilde{\beta} \approx \frac{1}{(2\pi)^{N_y/2}|P^{yy}|^{1/2}} \exp\left(-\frac{1}{2}(y_1 - \bar{h})^{\mathrm{T}}(P^{yy})^{-1}(y_1 - \bar{h})\right)$$

with

$$P^{yy} := R + P^{hh}.$$

Such an approximation has been used, for example, in Carrassi, Bocquet, Hannart and Ghil (2017). See also Reich and Cotter (2015) for more details on how to compute and use model evidence in the context of sequential data assimilation.

Sequential data assimilation requires that we produce M equally weighted samples $\widehat{z}_1^j \sim \widehat{\pi}_1$ from the M weighted samples $z_1^i \sim \pi_1$ with weights w^i. This is a standard problem in Monte Carlo integration and there are many ways to tackle this problem, among which are multinomial, residual, systematic and stratified resampling (Douc and Cappe 2005). Here we focus on those resampling methods which are based on a discrete Markov chain $P \in \mathbb{R}^{M\times M}$ with the property that

$$w = \frac{1}{M} P\,\mathbb{1}_M, \quad w = \left(\frac{w^1}{M}, \ldots, \frac{w^M}{M}\right)^{\mathrm{T}}. \tag{3.17}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

The Markov property of P implies that $P^{\mathrm{T}}\mathbb{1}_M = \mathbb{1}_M$. We now consider the set of all Markov chains Π_{M}, as defined by (3.3), with $p_1 = w$. Any Markov chain $P \in \Pi_{\mathrm{M}}$ can now be used for resampling, but we seek the Markov chain $P^* \in \Pi_{\mathrm{M}}$ which minimizes the expected distance between the samples, that is,

$$P^* = \arg\min_{P\in\Pi_{\mathrm{M}}} \sum_{i,j=1}^{M} p_{ij}\|z_1^i - z_1^j\|^2. \tag{3.18}$$

Note that (3.18) is a special case of the optimal transport problem (2.36) with the involved probability measures being discrete measures. Resampling can now be performed according to

$$\mathbb{P}[\widehat{Z}_1^j(\omega) = z_1^i] = p_{ij}^* \tag{3.19}$$

for $j = 1, \ldots, M$.

Since it is known that (3.18) converges to (2.36) as $M \to \infty$ (McCann 1995) and since (2.36) leads to a transformation (2.38), the resampling step (3.19) has been replaced by

$$\widehat{z}_1^j = \sum_{i=1}^{M} z_1^i\, p_{ij}^* \tag{3.20}$$

in the so-called ensemble transform particle filter (ETPF) (Reich 2013, Reich and Cotter 2015). In other words, the ETPF replaces resampling with probabilities p_{ij}^* by its mean (3.20) for each $j = 1, \ldots, M$. The ETPF leads to a biased approximation to the resampling step which is consistent in the limit $M \to \infty$.

The general formulation (3.20) with the coefficients p_{ij}^* chosen appropriately[2] leads to a large class of so-called ensemble transform particle filters (Reich and Cotter 2015). Ensemble transform particle filters generally result in biased and inconsistent but robust estimates, which have found applications to high-dimensional state space models (Evensen 2006, Vetra-Carvalho *et al.* 2018) for which traditional particle filters fail due to the 'curse of dimensionality' (Bengtsson, Bickel and Li 2008). More specifically, the class of ensemble transform particle filters includes the popular ensemble Kalman filters (Evensen 2006, Reich and Cotter 2015, Vetra-Carvalho *et al.* 2018, Carrassi, Bocquet, Bertino and Evensen 2018) and so-called second-order accurate particle filters with coefficients p_{ij}^* in (3.20) chosen such that the weighted ensemble mean

$$\bar{z}_1 := \frac{1}{M}\sum_{i=1}^{M} w^i z_1^i$$

[2] The coefficients p_{ij}^* of an ensemble transform particle filter do not need to be non-negative and only satisfy $\sum_{i=1}^{M} p_{ij}^* = 1$ (Acevedo, de Wiljes and Reich 2017).

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

and the weighted ensemble covariance matrix

$$\widetilde{P}^{zz} := \frac{1}{M}\sum_{i=1}^M w^i (z_1^i - \bar{z}_1)(z_1^i - \bar{z}_1)^{\mathrm{T}}$$

are exactly reproduced by the transformed and equally weighted particles $\widehat{z}_1^j$, $j = 1, \ldots, M$, defined by (3.20), that is,

$$\frac{1}{M}\sum_{j=1}^M \widehat{z}_1^j = \bar{z}_1, \quad \frac{1}{M-1}\sum_{j=1}^M (\widehat{z}_1^i - \bar{z}_1)(\widehat{z}_1^i - \bar{z}_1)^{\mathrm{T}} = \widetilde{P}^{zz}.$$

See the survey paper by Vetra-Carvalho *et al.* (2018) and the paper by Acevedo *et al.* (2017) for more details. A summary of the ensemble Kalman filter can be found in Appendix A.3.

In addition, hybrid methods (Frei and Künsch 2013, Chustagulprom, Reich and Reinhardt 2016), which bridge between classical particle filters and the ensemble Kalman filter, have recently been successfully applied to atmospheric fluid dynamics (Robert, Leuenberger and Künsch 2018).

Remark 3.9. Another approach for transforming samples, z_1^i, from the forecast PDF π_1 into samples, $\widehat{z}_1^i$, from the filtering PDF $\widehat{\pi}_1$ is provided through the mean-field interpretation

$$\frac{\mathrm{d}}{\mathrm{d}s}\breve{Z}_s = -\nabla_z \log \frac{\breve{\pi}_s(\breve{Z}_s)}{\widehat{\pi}_1(\breve{Z}_s)} \tag{3.21}$$

of the Fokker–Planck equation (2.26) for a random variable $\breve{Z}_s$ with law $\breve{\pi}_s$, drift term $f_s(z) = \nabla_z \log \widehat{\pi}_1$ and $\gamma = 2$, that is,

$$\partial_s \breve{\pi}_s = \nabla_z \cdot \left(\breve{\pi}_s \nabla_z \log \frac{\breve{\pi}_s}{\widehat{\pi}_1} \right)$$

in artificial time $s \geq 0$. It holds under fairly general assumptions that

$$\lim_{s\to\infty} \breve{\pi}_s = \widehat{\pi}_1$$

(Pavliotis 2014) and one can set $\breve{\pi}_0 = \pi_1$. The more common approach would be to solve Brownian dynamics

$$\mathrm{d}\breve{Z}_s = \nabla_z \log \widehat{\pi}_1(Z_s)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_s^+$$

for each sample z_1^i from π_1, *i.e.* $\breve{Z}_0(\omega) = z_1^i$, $i = 1, \ldots, M$, at initial time and

$$\widehat{z}_1^i = \lim_{s\to\infty} \breve{Z}_s(\omega).$$

In other words, formulation (3.21) replaces stochastic Brownian dynamics with a deterministic interacting particle system. See Appendix A.1 and Remark 4.3 for further details.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

3.3. Smoothing

Recall that the joint smoothing distribution $\widehat{\pi}(z_0, z_1)$ can be represented in the form (2.47) with modified transition kernel (2.10) and smoothing distribution (2.8) at time t_0 with weights γ^i determined by (2.9).

Let us assume that it is possible to sample from $\widehat{q}_+(z_1|z_0^i)$ and that the weights γ^i are available. Then we can utilize (2.47) in sequential data assimilation as follows. We first resample the z_0^i at time t_0 using a discrete Markov chain $P \in \mathbb{R}^{M\times M}$ satisfying

$$\widehat{p}_0 = \frac{1}{M} P\, \mathbb{1}_M, \quad \widehat{p}_0 := \gamma, \tag{3.22}$$

with γ defined in (2.12). Again optimal transportation can be used to identify a suitable P. More explicitly, we now consider the set of all Markov chains Π_{M}, as defined by (3.3), with $p_1 = \gamma$. Then the Markov chain P^* arising from the associated optimal transport problem (3.18) can be used for resampling, that is,

$$\mathbb{P}[\widetilde{Z}_0^j(\omega) = z_0^i] = p_{ij}^*.$$

Once equally weighted samples $\widehat{z}_0^i$, $i = 1, \ldots, M$, from $\widehat{\pi}_0$ have been determined, the desired samples $\widehat{z}_1^i$ from $\widehat{\pi}_1$ are simply given by

$$\widehat{z}_1^i := \widehat{Z}_1^i(\omega), \quad \widehat{Z}_1^i \sim \widehat{q}_+(\cdot\,|\widehat{z}_0^i),$$

for $i = 1, \ldots, M$.

The required transition kernels (2.10) are explicitly available for state space models with Gaussian model errors and Gaussian likelihood functions. In many other cases, these kernels are not explicitly available or are difficult to draw from. In such cases, one can resort to sample-based transition kernels.

For example, consider the twisted discrete Markov kernel (3.4) with twisting potential $\psi(z) = l(z)$. The associated vector v from (3.4) then gives rise to a probability vector $\widehat{p}_0 = c\, v \in \mathbb{R}^M$ with $c > 0$ an appropriate scaling factor, and

$$\widehat{p}_1 := Q_+^{\psi} \widehat{p}_0 \tag{3.23}$$

approximates the filtering distribution at time t_1. The Markov transition matrix $Q_+^{\psi} \in \mathbb{R}^{L\times M}$ together with $\widehat{p}_0$ provides an approximation to the smoothing kernel $\widehat{q}_+(z_1|z_0)$ and $\widehat{\pi}_0$, respectively.

The approximations $Q_+^{\psi} \in \mathbb{R}^{L\times M}$ and $\widehat{p}_0 \in \mathbb{R}^M$ can be used to first generate equally weighted samples $\widehat{z}_0^i \in \{z_0^1, \ldots, z_0^M\}$ with distribution $\widehat{p}_0$ via, for example, resampling with replacement. If $\widehat{z}_0^i = z_0^k$ for an index $k = k(i) \in \{1, \ldots, M\}$, then

$$\mathbb{P}[\widehat{Z}_1^i(\omega) = z_1^l] = (Q_+^{\psi})_{lk}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

for each $i = 1, \ldots, M$. The $\widehat{z}_1^i$ are equally weighted samples from the discrete filtering distribution $\widehat{p}_1$, which is an approximation to the continuous filtering PDF $\widehat{\pi}_1$.

Remark 3.10. One has to take computational complexity and robustness into account when deciding whether to utilize methods from Section 3.2 or this subsection to advance M samples z_0^i from the prior distribution π_0 into M samples $\widehat{z}_1^i$ from the posterior distribution $\widehat{\pi}_1$. While the methods from Section 3.2 are easier to implement, the methods of this subsection benefit from the fact that

$$M > \frac{1}{\|\gamma\|^2} \geq \frac{1}{\|w\|^2} \geq 1,$$

in general, where the importance weights $\gamma \in \mathbb{R}^M$ and $w \in \mathbb{R}^M$ are defined in (2.12) and (3.17), respectively. In other words, the methods from this subsection lead to larger effective sample sizes (Liu 2001, Agapiou *et al.* 2017).

Remark 3.11. We mention that finding efficient methods for solving the more general smoothing problem (2.2) is an active area of research. See, for example, the recent contributions by Guarniero *et al.* (2017) and Heng *et al.* (2018) for discrete-time Markov processes, and Kappen and Ruiz (2016) as well as Ruiz and Kappen (2017) for smoothing in the context of SDEs. Ensemble transform methods of the form (3.20) can also be extended to the general smoothing problem. See, for example, Evensen (2006) and Carrassi *et al.* (2018) for extensions of the ensemble Kalman filter, and Kirchgessner, Tödter, Ahrens and Nerger (2017) for an extension of the nonlinear ensemble transform filter to the smoothing problem.

3.3.1. SDE models (cont.)
After discretization in time, smoothing leads to a change from the forecast PDF (3.6) to

$$\begin{aligned}\widehat{\pi}_{0:N}(z_{0:N}) &:= \frac{l(z_N)\pi_{0:N}(z_{0:N})}{\pi_{0:N}[l]} \\ &\propto \exp\left(-\frac{1}{2\Delta t}\sum_{n=0}^{N-1}\|\xi_n\|^2\right)\pi_0(z_0)\, l(z_N)\end{aligned}$$

with ξ_n given by (3.7), or, alternatively,

$$\frac{\widehat{\pi}_{0:N}}{\pi_{0:N}}(z_{0:N}) = \frac{l(z_N)}{\pi_{0:N}[l]}.$$

Remark 3.12. Efficient MCMC methods for sampling high-dimensional smoothing distributions can be found in Beskos *et al.* (2017) and Beskos, Pinski, Sanz-Serna and Stuart (2011). Improved sampling can also be

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

achieved by using regularized Störmer–Verlet time-stepping methods (Reich and Hundertmark 2011) in a hybrid Monte Carlo method (Liu 2001). See Appendix A.2 for more details.

3.4. Schrödinger problem

Recall that the Schrödinger system (2.58)–(2.61) reduces in our context to solving equations (2.64)–(2.65) for the unknown coefficients α^i, $i = 1, \ldots, M$. In order to make this problem tractable we need to replace the required expectation values with respect to $q_+(z_1|z_0^j)$ by Monte Carlo approximations. More specifically, let us assume that we have $L \geq M$ samples z_1^l from the forecast PDF π_1. The associated $L \times M$ matrix Q with entries (3.1) provides a discrete approximation to the underlying Markov process defined by $q_+(z_1|z_0)$ and initial PDF (2.3).

The importance weights in the associated approximation to the filtering distribution

$$\widehat{\pi}_1(z) = \frac{1}{L} \sum_{l=1}^{L} w^l \, \delta(z - z_1^l)$$

are given by (3.16) with the weights normalized such that

$$\sum_{l=1}^{L} w^l = L. \tag{3.24}$$

Finding the coefficients $\{\alpha^i\}$ in (2.64)–(2.65) can now be reformulated as finding two vectors $u \in \mathbb{R}^L$ and $v \in \mathbb{R}^M$ such that

$$P^* := D(u) Q D(v)^{-1} \tag{3.25}$$

satisfies $P^* \in \Pi_{\mathrm{M}}$ with $p_1 = w$ in (3.3), that is, more explicitly

$$\Pi_{\mathrm{M}} = \left\{ P \in \mathbb{R}^{L \times M} : \, p_{lj} \geq 0, \, \sum_{l=1}^{L} p_{lj} = 1, \, \frac{1}{M} \sum_{j=1}^{M} p_{lj} = \frac{w^l}{L} \right\}. \tag{3.26}$$

We note that (3.26) are discrete approximations to (2.67) and (2.68), respectively. The scaling factor $\widehat{\psi}$ in (2.58) is approximated by the vector v up to a normalization constant, while the vector u provides an approximation to ψ in (2.59). Finally, the desired approximations to the Schrödinger transition kernels $q_+^*(z_1|z_0^i)$, $i = 1, \ldots, M$, are provided by the columns of P^*, that is,

$$\mathbb{P}[\widehat{Z}_1^i(\omega) = z_1^l] = p_{li}^*$$

characterizes the desired equally weighted samples $\widehat{z}_1^i$, $i = 1, \ldots, M$, from the filtering distribution $\widehat{\pi}_1$. See the following subsection for more details.

The required vectors u and v can be computed using the iterative Sinkhorn algorithm (2.74)–(2.76) (Cuturi 2013, Peyre and Cuturi 2018).

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

Remark 3.13. Note that one can replace the forward transition kernel $q_+(z_1|z_0)$ in (3.1) with any suitable twisted prediction kernel (1.2). This results in a modified matrix Q in (3.25) and weights w^l in 3.26. The resulting matrix (3.25) still provides an approximation to the Schrödinger problem.

Remark 3.14. The approximation (3.25) can be extended to an approximation of the Schrödinger forward transition kernels (2.66) in the following sense. We use $\alpha^i = 1/v_i$ in (2.66) and note that the resulting approximation satisfies (2.68) while (2.67) now longer holds exactly. However, since the entries u_l of the vector u appearing in (3.25) satisfy

$$u_l = \frac{w^l}{L} \frac{1}{\sum_{j=1}^M q_+(z_1^l|z_0^j)/v_j},$$

it follows that

$$\begin{aligned}\int q_+^*(z_1|z_0^i)\,\mathrm{d}z_1 &\approx \frac{1}{L}\sum_{l=1}^L \frac{l(z_1^l)}{\beta} \frac{q_+(z_1^l|z_0^i)/v_i}{\sum_{j=1}^M q_+(z_1^l|z_0^j)/v_j} \\ &\approx \frac{1}{L}\sum_{l=1}^L w^l \frac{q_+(z_1^l|z_0^i)/v_i}{\sum_{j=1}^M q_+(z_1^l|z_0^j)/v_j} = \sum_{l=1}^L p_{li}^* = 1.\end{aligned}$$

Furthermore, one can use such continuous approximations in combination with Monte Carlo sampling methods which do not require normalized target PDFs.

3.4.1. Gaussian model errors (cont.)

One can easily generate L, $L \geq M$, i.i.d. samples z_1^l from the forecast PDF (2.21), that is,

$$Z_1^l \sim \frac{1}{M}\sum_{j=1}^M \mathrm{n}(\cdot\,; \Psi(z_0^j), \gamma B),$$

and with the filtering distribution $\widehat{\pi}_1$ characterized through the importance weights (3.16).

We define the distance matrix $D \in \mathbb{R}^{L\times M}$ with entries

$$d_{lj} := \frac{1}{2}\|z_1^l - \Psi(z_0^j)\|_B^2, \quad \|z\|_B^2 := z^{\mathrm{T}} B^{-1} z,$$

and the matrix $Q \in \mathbb{R}^{L\times M}$ with entries

$$q_{lj} := \mathrm{e}^{-d_{lj}/\gamma}.$$

The Markov chain $P^* \in \mathbb{R}^{L\times M}$ is now given by

$$P^* = \arg\min_{P\in\Pi_{\mathrm{M}}} \mathrm{KL}(P||Q)$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

with the set Π_{M} defined by (3.26).

Once P^* has been computed, the desired Schrödinger transitions from π_0 to $\widehat{\pi}_1$ can be represented as follows. The Schrödinger transition kernels $q_+^*(z_1|z_0^i)$ are approximated for each z_0^i by

$$\tilde{q}_+^*(z_1|z_0^i) := \sum_{l=1}^{L} p_{li}^* \, \delta(z_1 - z_1^l), \quad i = 1, \ldots, M. \tag{3.27}$$

The empirical measure in (3.27) converges weakly to the desired $q_+^*(z_1|z_0^i)$ as $L \to \infty$ and

$$\widehat{\pi}_1(z_1) \approx \frac{1}{M} \sum_{i=1}^{M} \delta(z - \widehat{z}_1^i),$$

with

$$\widehat{z}_1^i = \widehat{Z}_1^i(\omega), \quad \widehat{Z}_1^i \sim \tilde{q}_+^*(\cdot\,|z_0^i), \tag{3.28}$$

provides the desired approximation of $\widehat{\pi}_1$ by M equally weighted particles $\widehat{z}_1^i$, $i = 1, \ldots, M$.

We remark that (3.27) has been used to produce the Schrödinger transition kernels for Example 2.5 and Figure 2.3(b) in particular. More specifically, we have $M = 11$ and used $L = 11\,000$. Since the particles $z_1^l \in \mathbb{R}$, $l = 1, \ldots, L$, are distributed according to the forecast PDF π_1, a function representation of $\tilde{q}_+^*(z_1|z_0^i)$ over all of $\mathbb{R}$ is provided by interpolating p_{li}^* onto $\mathbb{R}$ and multiplication of this interpolated function by $\pi_1(z)$.

For $\gamma \ll 1$, the measure in (3.27) can also be approximated by a Gaussian measure with mean

$$\bar{z}_1^i := \sum_{l=1}^{L} z_1^l p_{li}^*$$

and covariance matrix γB, that is, we replace (3.28) with

$$\widehat{Z}_1^i \sim \mathrm{N}(\bar{z}_1^i, \gamma B)$$

for $i = 1, \ldots, M$.

3.4.2. SDE (cont.)

One can also apply (3.25) in order to approximate the Schrödinger problem associated with SDE models. We typically use $L = M$ in this case and utilize (3.8) in place of Q in (3.25). The set Π_{M} is still given by (3.26).

Example 3.15. We consider scalar-valued motion of a Brownian particle in a bimodal potential, that is,

$$\mathrm{d}Z_t^+ = Z_t^+ \, \mathrm{d}t - (Z_t^+)^3 \, \mathrm{d}t + \gamma^{1/2} \, \mathrm{d}W_t^+ \tag{3.29}$$

with $\gamma = 0.5$ and initial distribution $Z_0 \sim \mathrm{N}(-1, 0.3)$. At time $t = 2$ we

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

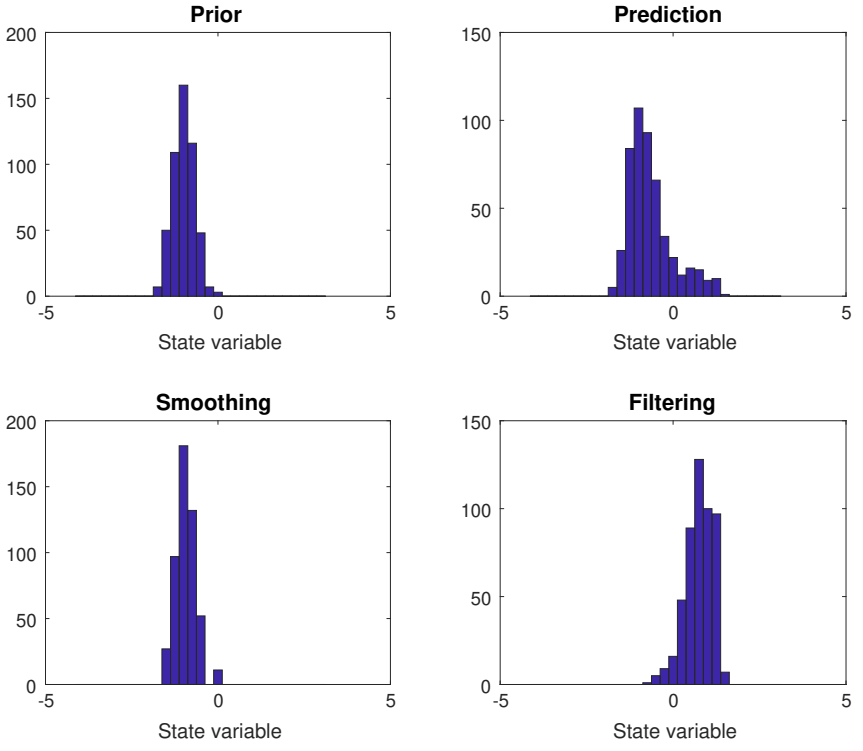


Figure 3.1. Histograms produced from $M = 200$ Monte Carlo samples of the initial PDF π_0, the forecast PDF π_2 at time $t = 2$, the filtering distribution $\widehat{\pi}_2$ at time $t = 2$, and the smoothing PDF $\widehat{\pi}_0$ at time $t = 0$ for a Brownian particle moving in a double well potential.

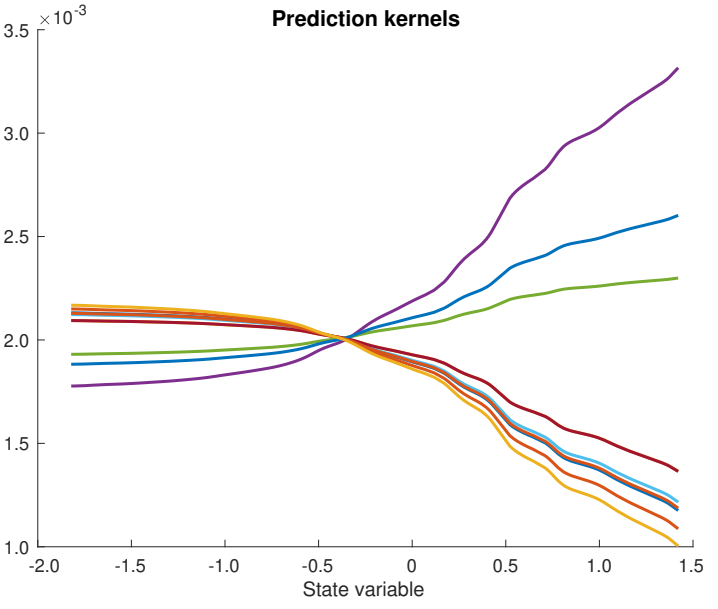
measure the location $y = 1$ with measurement error variance $R = 0.2$. We simulate the dynamics using $M = 200$ particles and a time-step of $\Delta t = 0.01$ in the Euler–Maruyama discretization (3.5). One can find histograms produced from the Monte Carlo samples in Figure 3.1. The samples from the filtering and smoothing distributions are obtained by resampling with replacement from the weighted distributions with weights given by (3.16). Next we compute (3.8) from the $M = 200$ Monte Carlo samples of (3.29). Eleven out of the 200 transition kernels from π_0 to π_2 (prediction problem) and π_0 to $\widehat{\pi}_2$ (Schrödinger problem) are displayed in Figure 3.2.

The Sinkhorn approach requires relatively large sample sizes M in order to lead to useful approximations. Alternatively we may assume that there is an approximative control term $u_t^{(0)}$ with associated forward SDE

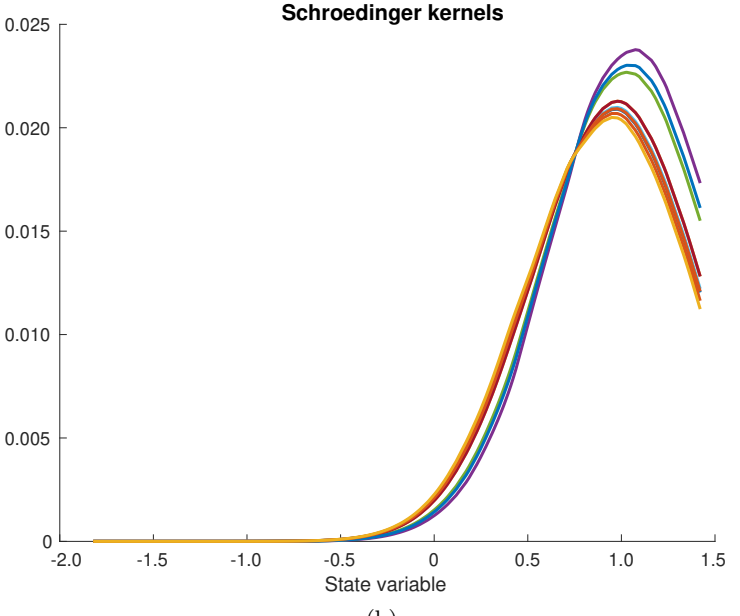
$$\mathrm{d}Z_t^+ = f_t(Z_t)\,\mathrm{d}t + u_t^{(0)}(Z_t^+)\,\mathrm{d}t + \gamma^{1/2}\,\mathrm{d}W_t^+, \quad t \in [0,1], \tag{3.30}$$

and $Z_0^+ \sim \pi_0$. We denote the associated path measure by $\mathbb{Q}^{(0)}$. Girsanov's

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press



(a)



(b)

Figure 3.2. (a) Approximations of typical transition kernels from time π_0 to π_2 under the Brownian dynamics model (3.29). (b) Approximations of typical Schrödinger transition kernels from π_0 to $\widehat{\pi}_2$. All approximations were computed using the Sinkhorn algorithm and by linear interpolation between the $M = 200$ data points.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

theorem implies that the Radon–Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{Q}^{(0)}$ is given by (compare (2.33))

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{Q}^{(0)}}_{|z^{(0)}_{[0,1]}} = \exp(-V^{(0)}),$$

where $V^{(0)}$ is defined via the stochastic integral

$$V^{(0)} := \frac{1}{2\gamma}\int_0^1 (\|u_t^{(0)}\|^2\,\mathrm{d}t + 2\gamma^{1/2}u_t^{(0)}\cdot\,\mathrm{d}W_t^+)$$

along solution paths $z^{(0)}_{[0,1]}$ of (3.30). Because of

$$\frac{\mathrm{d}\widehat{\mathbb{P}}}{\mathrm{d}\mathbb{Q}^{(0)}}_{|z^{(0)}_{[0,1]}} = \frac{\mathrm{d}\widehat{\mathbb{P}}}{\mathrm{d}\mathbb{Q}}_{|z^{(0)}_{[0,1]}}\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{Q}^{(0)}}_{|z^{(0)}_{[0,1]}} \propto l(z_1^{(0)})\,\exp(-V^{(0)}),$$

we can now use (3.30) to importance-sample from the filtering PDF $\widehat{\pi}_1$. The control $u_t^{(0)}$ should be chosen such that the variance in the modified likelihood function

$$l^{(0)}(z^{(0)}_{[0,1]}) := l(z_1^{(0)})\,\exp(-V^{(0)}) \tag{3.31}$$

is reduced compared to the uncontrolled case $u_t^{(0)} \equiv 0$. In particular, the filter distribution $\widehat{\pi}_1$ at time $t = 1$ satisfies

$$\widehat{\pi}_1(z_1^{(0)}) \propto l^{(0)}(z^{(0)}_{[0,1]})\,\pi_1^{(0)}(z_1^{(0)}),$$

where $\pi_t^{(0)}$, $t \in (0,1]$, denote the marginal PDFs generated by (3.30).

We now describe an iterative algorithm for the associated Schrödinger problem in the spirit of the Sinkhorn iteration from Section 2.3.2.

Lemma 3.16. The desired optimal control law can be computed iteratively,

$$u_t^{(k+1)} = u_t^{(k)} + \gamma\nabla_z \log\psi_t^{(k)}, \quad k = 0, 1, \ldots, \tag{3.32}$$

for given $u_t^{(k)}$ and the potential $\psi_t^{(k)}$ obtained as the solutions to the backward Kolmogorov equation

$$\partial_t\psi_t^{(k)} = -\mathcal{L}_t^{(k)}\psi_t^{(k)}, \quad \mathcal{L}_t^{(k)}g := \nabla_z g\cdot(f_t + u_t^{(k)}) + \frac{\gamma}{2}\Delta_z g, \tag{3.33}$$

with final time condition

$$\psi_1^{(k)}(z) := \frac{\widehat{\pi}_1(z)}{\pi_1^{(k)}(z)}. \tag{3.34}$$

Here $\pi_1^{(k)}$ denotes the time-one marginal of the path measure $\mathbb{Q}^{(k)}$ induced by (2.30) with control term $u_t = u_t^{(k)}$ and initial PDF π_0. The recursion

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

(3.32) is stopped whenever the final time condition (3.34) is sufficiently close to a constant function.

Proof. The extension of the Sinkhorn algorithm to continuous PDFs and its convergence has been discussed by Chen, Georgiou and Pavon (2016*a*). □

Remark 3.17. Note that $\psi_t^{(k)}$ needs to be determined up to a constant of proportionality only since the associated control law is determined from $\psi_t^{(k)}$ by (3.32). One can also replace (3.33) with any other method for solving the smoothing problem associated to the SDE (2.30) with $Z_0^+ \sim \pi_0$, control law $u_t = u_t^{(k)}$, and likelihood function $l(z) = \psi_1^{(k)}(z)$. See Appendix A.4 for a forward–backward SDE formulation in particular.

We need to restrict the class of possible control laws $u_t^{(k)}$ in order to obtain a computationally feasible implementations in practice. For example, a simple class of control laws is provided by linear controls of the form

$$u_t^{(k)}(z) = -B_t^{(k)}(z - m_t^{(k)})$$

with appropriately chosen symmetric positive definite matrices $B_t^{(k)}$ and vectors $m_t^{(k)}$. Such approximations can, for example, be obtained from the smoother extensions of ensemble transform methods mentioned earlier. See also the recent work by Kappen and Ruiz (2016) and Ruiz and Kappen (2017) on numerical methods for the SDE smoothing problem.

4. DA for continuous-time data

In this section we focus on the continuous-time filtering problem over the time interval $[0, 1]$, that is, on the assimilation of data that arrive continuously in time. If one is only interested in transforming samples from the prior distribution at $t = 0$ into samples of the filtering distribution at time $t = 1$, then all methods from the previous sections can be applied once the associated filtering distribution $\widehat{\pi}_1$ is available. However, it is more natural to consider the associated filtering distributions $\widehat{\pi}_t$ for all $t \in (0, 1]$ and to derive appropriate transformations in the form of mean-field equations in continuous time. We distinguish between smooth and non-smooth data y_t, $t \in [0, 1]$.

4.1. Smooth data

We start from a forward SDE model (2.24) with associated path measure $\mathbb{Q}$ over the space of continuous functions $\mathcal{C}$. However, contrary to the previous sections, the likelihood l is defined along a whole solution path $z_{[0,1]}$ as

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

follows:

$$\frac{\mathrm{d}\widehat{\mathbb{P}}}{\mathrm{d}\mathbb{Q}}_{|z_{[0,1]}} \propto l(z_{[0,1]}), \quad l(z_{[0,1]}) := \exp\left(-\int_0^1 V_t(z_t)\,\mathrm{d}t\right)$$

with the assumption that $\mathbb{Q}[l] < \infty$ and $V_t(z) \geq 0$. A specific example of a suitable V_t is provided by

$$V_t(z) = \frac{1}{2}\|h(z) - y_t\|^2, \tag{4.1}$$

where the data function $y_t \in \mathbb{R}$, $t \in [0,1]$, is a smooth function of time and $h(z)$ is a forward operator connecting the model states to the observations/data. The associated estimation problem has, for example, been addressed by Mortensen (1968) and Fleming (1997) from an optimal control perspective and has led to what is called the minimum energy estimator. Recall that the filtering PDF $\widehat{\pi}_1$ is the marginal PDF of $\widehat{\mathbb{P}}$ at time $t = 1$.

The associated time-continuous smoothing/filtering problems are based on the time-dependent path measures $\widehat{\mathbb{P}}_t$ defined by

$$\frac{\mathrm{d}\widehat{\mathbb{P}}_t}{\mathrm{d}\mathbb{Q}}_{|z_{[0,1]}} \propto l(z_{[0,t]}), \quad l(z_{[0,t]}) := \exp\left(-\int_0^t V_s(z_s)\,\mathrm{d}s\right)$$

for $t \in (0,1]$. We let $\widehat{\pi}_t$ denote the marginal PDF of $\widehat{\mathbb{P}}_t$ at time t. Note that $\widehat{\pi}_t$ is the filtering PDF, *i.e.* the marginal PDF at time t conditioned on all the data available until time t. Also note that $\widehat{\pi}_t$ is different from the marginal (smoothing) PDF of $\widehat{\mathbb{P}}$ at time t.

We now state a modified Fokker–Planck equation which describes the time evolution of the filtering PDFs $\widehat{\pi}_t$.

Lemma 4.1. The marginal distributions $\widehat{\pi}_t$ of $\widehat{\mathbb{P}}_t$ satisfy the modified Fokker–Planck equation

$$\partial_t \widehat{\pi}_t = \mathcal{L}_t^\dagger \widehat{\pi}_t - \widehat{\pi}_t(V_t - \widehat{\pi}_t[V_t]) \tag{4.2}$$

with $\mathcal{L}_t^\dagger$ defined by (2.25).

Proof. This can be seen by setting $\gamma = 0$ and $f_t \equiv 0$ in (2.24) for simplicity and by considering the incremental change of measure induced by the likelihood, that is,

$$\frac{\widehat{\pi}_{t+\delta t}}{\widehat{\pi}_t} \propto \mathrm{e}^{-V_t \delta t} \approx 1 - V_t\,\delta t,$$

and taking the limit $\delta t \to 0$ under the constraint that $\widehat{\pi}_t[1] = 1$ is preserved. □

We now derive a mean-field interpretation of (4.2) and rewrite (4.2) in the form

$$\partial_t \widehat{\pi}_t = \mathcal{L}_t^\dagger \widehat{\pi}_t + \nabla_z \cdot (\widehat{\pi}_t \nabla_z \phi_t), \tag{4.3}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

where the potential $\phi_t : \mathbb{R}^{N_z} \to \mathbb{R}$ satisfies the elliptic PDE

$$\nabla_z \cdot (\widehat{\pi}_t \nabla_z \phi_t) = -\widehat{\pi}_t (V_t - \widehat{\pi}_t[V_t]). \tag{4.4}$$

Remark 4.2. Necessary conditions for the elliptic PDE (4.4) to be solvable and to lead to bounded gradients $\nabla_z \phi$ for given π_t have been discussed by Laugesen, Mehta, Meyn and Raginsky (2015). It is an open problem to demonstrate that continuous-time data assimilation problems actually satisfy such conditions.

With (4.3) in place, we formally obtain the mean-field equation

$$\mathrm{d}Z_t^+ = \{f_t(Z_t^+) - \nabla_z \phi_t(Z_t^+)\}\,\mathrm{d}t + \gamma^{1/2}\,\mathrm{d}W_t^+, \tag{4.5}$$

and the marginal distributions π_t^u of this controlled SDE agree with the marginals $\widehat{\pi}_t$ of the path measures $\widehat{\mathbb{P}}_t$ at times $t \in (0, 1]$.

The control u_t is not uniquely determined. For example, one can replace (4.4) with

$$\nabla_z \cdot (\pi_t M_t \nabla_z \phi_t) = -\pi_t (V_t - \pi_t[V_t]), \tag{4.6}$$

where M_t is a symmetric positive definite matrix. More specifically, let us assume that π_t is Gaussian with mean $\bar{z}_t$ and covariance matrix P_t^{zz} and that $h(z)$ is linear, *i.e.* $h(z) = Hz$. Then (4.6) can be solved analytically for $M_t = P_t^{zz}$ with

$$\nabla_z \phi_t(z) = \frac{1}{2} H^{\mathrm{T}} (Hz + H\bar{z}_t - 2y_t).$$

The resulting mean-field equation becomes

$$\mathrm{d}Z_t^+ = \left\{ f_t(Z_t^+) - \frac{1}{2} P_t^{zz} H^{\mathrm{T}} (HZ_t^+ + H\bar{z}_t - 2y_t) \right\} \mathrm{d}t + \gamma^{1/2}\,\mathrm{d}W_t^+, \tag{4.7}$$

which gives rise to the ensemble Kalman–Bucy filter upon Monte Carlo discretization (Bergemann and Reich 2012). See Section 5 below and Appendix A.3 for further details.

Remark 4.3. The approach described in this subsection can also be applied to standard Bayesian inference without model dynamics. More specifically, let us assume that we have samples z_0^i, $i = 1, \ldots, M$, from a prior distribution π_0 which we would like to transform into samples from a posterior distribution

$$\pi^*(z) := \frac{l(z)\,\pi_0(z)}{\pi_0[l]}$$

with likelihood $l(z) = \pi(y|z)$. One can introduce a homotopy connecting π_0 with π^*, for example, via

$$\breve{\pi}_s(z) := \frac{l(z)^s\,\pi_0(z)}{\pi_0[l^s]} \tag{4.8}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

with $s \in [0,1]$. We find that

$$\frac{\partial \breve{\pi}_s}{\partial s} = \breve{\pi}_s(\log l - \breve{\pi}_s[\log l]). \tag{4.9}$$

We now seek a differential equation

$$\frac{\mathrm{d}}{\mathrm{d}s}\breve{Z}_s = u_s(\breve{Z}_s) \tag{4.10}$$

with $\breve{Z}_0 \sim \pi_0$ such that its marginal distributions $\breve{\pi}_s$ satisfy (4.9) and, in particular, $\breve{Z}_1 \sim \pi^*$. This condition together with Liouville's equation for the time evolution of marginal densities under a differential equation (4.10) leads to

$$-\nabla_z \cdot (\breve{\pi}_s\, u_s) = \breve{\pi}_s(\log l - \breve{\pi}_s[\log l]). \tag{4.11}$$

In order to define u_s in (4.10) uniquely, we make the *ansatz*

$$u_s(z) = -\nabla_z \phi_s(z) \tag{4.12}$$

which leads to the elliptic PDE

$$\nabla_z \cdot (\breve{\pi}_s\, \nabla_z \phi_s) = \breve{\pi}_s(\log l - \breve{\pi}_s[\log l]) \tag{4.13}$$

in the potential ϕ_s. The desired samples from π^* are now obtained as the time-one solutions of (4.10) with 'control law' (4.12) satisfying (4.13) and initial conditions z_0^i, $i = 1, \ldots, M$. There are many modifications of this basic procedure (Daum and Huang 2011, Reich 2011, El Moselhy and Marzouk 2012), some of them leading to explicit expressions for (4.10) such as Gaussian PDFs (Bergemann and Reich 2010) and Gaussian mixture PDFs (Reich 2012). We finally mention that the limit $s \to \infty$ in (4.8) leads, formally, to the PDF $\breve{\pi}_\infty = \delta(z - z_{\mathrm{ML}})$, where z_{ML} denotes the minimizer of $V(z) = -\log \pi(y|z)$, *i.e.* the maximum likelihood estimator, which we assume here to be unique, for example, V is convex. In other words these homotopy methods can be used to solve optimization problems via derivative-free mean-field equations and their interacting particle approximations. See, for example, Zhang, Taghvaei and Mehta (2019) and Schillings and Stuart (2017) as well as Appendices A.1 and A.3 for more details.

4.2. Random data

We now replace (4.1) with an observation model of the form

$$\mathrm{d}Y_t = h(Z_t^+)\,\mathrm{d}t + \mathrm{d}V_t^+,$$

where we set $Y_t \in \mathbb{R}$ for simplicity and V_t^+ denotes standard Brownian motion. The forward operator $h : \mathbb{R}^{N_z} \to \mathbb{R}$ is also assumed to be known. The marginal PDFs $\widehat{\pi}_t$ for Z_t conditioned on all observations y_s with $s \in$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

$[0,t]$ satisfy the Kushner–Stratonovitch equation (Jazwinski 1970)

$$\mathrm{d}\widehat{\pi}_t = \mathcal{L}_t^\dagger \widehat{\pi}_t \,\mathrm{d}t + (h - \widehat{\pi}_t[h])(\mathrm{d}Y_t - \widehat{\pi}_t[h]\,\mathrm{d}t) \tag{4.14}$$

with $\mathcal{L}^\dagger$ defined by (2.25). The following observation is important for the subsequent discussion.

Remark 4.4. Consider state-dependent diffusion

$$\mathrm{d}Z_t^+ = \gamma_t(Z_t^+) \circ \mathrm{d}U_t^+, \tag{4.15}$$

in its Stratonovitch interpretation (Pavliotis 2014), where U_t^+ is scalar-valued Brownian motion and $\gamma_t(z) \in \mathbb{R}^{N_z \times 1}$. Here the Stratonovitch interpretation is to be applied to the implicit time-dependence of $\gamma_t(z)$ through Z_t^+ only, that is, the explicit time-dependence of γ_t remains to be Itô-interpreted. The associated Fokker–Planck equation for the marginal PDFs π_t takes the form

$$\partial_t \pi_t = \frac{1}{2}\nabla_z \cdot (\gamma_t \nabla_z \cdot (\pi_t \gamma_t)), \tag{4.16}$$

and expectation values $\bar{g} = \pi_t[g]$ evolve in time according to

$$\pi_t[g] = \pi_0[g] + \int_0^t \pi_s[\mathcal{A}_t g]\,\mathrm{d}s \tag{4.17}$$

with operator $\mathcal{A}_t$ defined by

$$\mathcal{A}_t g = \frac{1}{2}\gamma_t^{\mathrm{T}} \nabla_z (\gamma_t^{\mathrm{T}} \nabla_z g).$$

Now consider the mean-field equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{Z}_t = -\frac{1}{2}\gamma_t(\widetilde{Z}_t)\, J_t, \quad J_t := \widetilde{\pi}_t^{-1} \nabla_z \cdot (\widetilde{\pi}_t \gamma_t), \tag{4.18}$$

with $\widetilde{\pi}_t$ the law of $\widetilde{Z}_t$. The associated Liouville equation is

$$\partial_t \widetilde{\pi}_t = \frac{1}{2}\nabla_z \cdot (\widetilde{\pi}_t \gamma_t J_t) = \frac{1}{2}\nabla_z \cdot (\gamma_t \nabla_z \cdot (\gamma_t \widetilde{\pi}_t)).$$

In other words, the marginal PDFs and the associated expectation values evolve identically under (4.15) and (4.18), respectively.

We now state a formulation of the continuous-time filtering problem in terms of appropriate mean-field equations. These equations follow the framework of the feedback particle filter (FPF) as first introduced by Yang *et al.* (2013) and theoretically justified by Laugesen *et al.* (2015). See Crisan and Xiong (2010) and Xiong (2011) for an alternative formulation.

Lemma 4.5. The mean-field SDE

$$\mathrm{d}Z_t^+ = f_t(Z_t^+)\,\mathrm{d}t + \gamma^{1/2}\,\mathrm{d}W_t^+ - K_t(Z_t^+) \circ \mathrm{d}I_t \tag{4.19}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

with

$$\mathrm{d}I_t := h(Z_t^+)\,\mathrm{d}t - \mathrm{d}Y_t + \mathrm{d}U_t^+,$$

U_t^+ standard Brownian motion, and $K_t := \nabla_z \phi_t$, where the potential ϕ_t satisfies the elliptic PDE

$$\nabla_z \cdot (\pi_t \nabla_z \phi_t) = -\pi_t(h - \pi_t[h]), \tag{4.20}$$

leads to the same evolution of its conditional marginal distributions π_t as (4.14).

Proof. We set $\gamma = 0$ and $f_t \equiv 0$ in (4.19) for simplicity. Then, following (4.16) with $\gamma_t = K_t$, the Fokker–Planck equation for the marginal distributions π_t of (4.19) conditioned on $\{Y_s\}_{s\in[0,t]}$ is given by

$$\begin{aligned}
\mathrm{d}\pi_t &= \nabla_z \cdot (\pi_t K_t (h(z)\,\mathrm{d}t - \mathrm{d}Y_t)) + \nabla_z \cdot (K_t \nabla_z \cdot (\pi_t K_t))\,\mathrm{d}t && (4.21)\\
&= (\pi_t[h]\,\mathrm{d}t - \mathrm{d}Y_t)\,\nabla_z \cdot (\pi_t K_t) + \nabla_z \cdot (\pi_t K_t (h(z) - \pi_t[h])\,\mathrm{d}t \\
&\qquad + \nabla_z \cdot (K_t \nabla_z \cdot (\pi_t K_t))\,\mathrm{d}t && (4.22)\\
&= \pi_t(h - \pi_t[h])(\mathrm{d}Y_t - \pi_t[h]\,\mathrm{d}t), && (4.23)
\end{aligned}$$

as desired, where we have used (4.20) twice to get from (4.22) to (4.23). Also note that both Y_t and U_t^+ contributed to the diffusion-induced final term in (4.21) and hence the factor $1/2$ in (4.16) is replaced by one. □

Remark 4.6. Using the reformulation (4.18) of (4.15) in Stratonovitch form with $\gamma_t = K_t$ together with (4.20), one can replace $K_t \circ \mathrm{d}U_t^+$ with $\frac{1}{2}K_t(\pi_t[h] - h)\,\mathrm{d}t$, which leads to the alternative

$$\mathrm{d}I_t = \frac{1}{2}(h + \pi_t[h])\,\mathrm{d}t - \mathrm{d}Y_t$$

for the innovation I_t, as originally proposed by Yang *et al.* (2013) in their FPF formulation. We also note that the feedback particle formulation (4.19) can be extended to systems for which the measurement and model errors are correlated. See Nüsken, Reich and Rozdeba (2019) for more details.

The ensemble Kalman–Bucy filter (Bergemann and Reich 2012) with the Kalman gain factor K_t being independent of the state variable z and of the form

$$K_t = P_t^{zh} \tag{4.24}$$

can be viewed as a special case of an FPF. Here P_t^{zh} denotes the covariance matrix between Z_t and $h(Z_t)$ at time t.

5. Numerical methods

In this section we discuss some numerical implementations of the mean-field approach to continuous-time data assimilation. An introduction to standard particle filter implementations can, for example, be found in Bain

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

and Crisan (2008). We start with the continuous-time formulation of the ensemble Kalman filter and state a numerical implementation of the FPF using a Schrödinger formulation in the second part of this section. See also Appendix A.1 for some more details on a particle-based solution of the elliptic PDEs (4.4), (4.13) and (4.20), respectively.

5.1. Ensemble Kalman–Bucy filter

Let us start with the ensemble Kalman–Bucy filter (EnKBF), which arises naturally from the mean-field equations (4.7) and (4.19), respectively, with Kalman gain (4.24) (Bergemann and Reich 2012). We state the EnKBF here in the form

$$\mathrm{d}Z_t^i = f_t(Z_t^i)\,\mathrm{d}t + \gamma^{1/2}\,\mathrm{d}W_t^+ - K_t^M\,\mathrm{d}I_t^i \tag{5.1}$$

for $i = 1, \ldots, M$ and

$$K_t^M := \frac{1}{M-1}\sum_{i=1}^M Z_t^i(h(Z_t^i) - \bar{h}_t^M)^{\mathrm{T}}, \quad \bar{h}_t^M := \frac{1}{M}\sum_{i=1}^M h(Z_t^i).$$

The innovations $\mathrm{d}I_t^i$ take different forms depending on whether the data are smooth in time, that is,

$$\mathrm{d}I_t^i = \frac{1}{2}(h(Z_t^i) + \bar{h}_t^M - 2y_t)\,\mathrm{d}t,$$

or contains stochastic contributions, that is,

$$\mathrm{d}I_t^i = \frac{1}{2}(h(Z_t^i) + \bar{h}_t^M)\,\mathrm{d}t - \mathrm{d}y_t, \tag{5.2}$$

or, alternatively,

$$\mathrm{d}I_t^i = h(Z_t^i)\,\mathrm{d}t + \mathrm{d}U_t^i - \mathrm{d}y_t,$$

where U_t^i denotes standard Brownian motion. The SDEs (5.1) can be discretized in time by any suitable time-stepping method such as the Euler–Maruyama scheme (Kloeden and Platen 1992). However, one has to be careful with the choice of the step-size Δt due to potentially stiff contributions from $K_t^M\,\mathrm{d}I_t^i$. See, for example, Amezcua, Kalnay, Ide and Reich (2014) and Blömker, Schillings and Wacker (2018).

Remark 5.1. It is of broad interest to study the stability and accuracy of interacting particle filter algorithms such as the discrete-time EnKF and the continuous-time EnKBF for fixed particle numbers M. On the negative side, it has been shown by Kelly, Majda and Tong (2015) that such algorithms can undergo finite-time instabilities, while it has also been demonstrated (González-Tokman and Hunt 2013, Kelly, Law and Stuart 2014, Tong, Majda and Kelly 2016, de Wiljes, Reich and Stannat 2018) that such algorithms can be stable and accurate under appropriate conditions on the dynamics

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

and measurement process. Asymptotic properties of the EnKF and EnKBF in the limit of $M \to \infty$ have also been studied, for example, by Le Gland, Monbet and Tran (2011), Kwiatowski and Mandel (2015) and de Wiljes *et al.* (2018).

5.2. Feedback particle filter

A Monte Carlo implementation of the FPF (4.19) faces two main obstacles. First, one needs to approximate the potential ϕ_t in (4.20) with the density π_t, which is only available in terms of an empirical measure

$$\pi_t(z) = \frac{1}{M}\sum_{i=1}^M \delta(z - z_t^i).$$

Several possible approximations have been discussed by Taghvaei and Mehta (2016) and Taghvaei, de Wiljes, Mehta and Reich (2017). Here we would like to mention in particular an approximation based on diffusion maps, which we summarize in Appendix A.1. Second, one needs to apply suitable time-stepping methods for the SDE (4.19) in Stratonovitch form. Here we suggest using the Euler–Heun method (Burrage, Burrage and Tian 2004),

$$\begin{aligned}
\tilde{z}_{n+1}^i &= z_n^i + \Delta t f_{t_n}(z_n^i) + (\gamma\Delta t)^{1/2}\xi_n^i - K_n(z_n^i)\Delta I_n^i,\\
z_{n+1}^i &= z_n^i + \Delta t f_{t_n}(z_n^i) + (\gamma\Delta t)^{1/2}\xi_n^i - \frac{1}{2}(K_n(z_n^i) + K_n(\tilde{z}_{n+1}^i))\Delta I_n^i,
\end{aligned}$$

$i = 1, \ldots, M$, with, for example,

$$\Delta I_n^i = \frac{1}{2}(h(z_n^i) + \bar{h}_n^M)\Delta t - \Delta y_n.$$

While the above implementation of the FPF requires one to solve the elliptic PDE (4.20) twice per time-step, we now suggest a time-stepping approach in terms of an associated Schrödinger problem. Let us assume that we have M equally weighted particles z_n^i representing the conditional filtering distribution at time t_n. We first propagate these particles forward under the drift term alone, that is,

$$\hat{z}_{n+1}^i := z_n^i + \Delta t\, f_{t_n}(z_n^i), \quad i = 1, \ldots, M.$$

In the next step, we draw $L = KM$ with $K \geq 1$ samples $\widetilde{z}_{n+1}^l$ from the forecast PDF

$$\widetilde{\pi}(z) := \frac{1}{M}\sum_{i=1}^M \mathrm{n}(z; \hat{z}_{n+1}^i, \gamma\Delta t I)$$

and assign importance weights

$$w_{n+1}^l \propto \exp\left(-\frac{\Delta t}{2}(h(\widetilde{z}_{n+1}^l))^2 + \Delta y_n h(\widetilde{z}_{n+1}^l)\right)$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

with normalization (3.24). Recall that we assumed that $y_t \in \mathbb{R}$ for simplicity and $\Delta y_n := y_{t_{n+1}} - y_{t_n}$. We then solve the Schrödinger problem

$$P^* = \arg \min_{P \in \Pi_{\mathrm{M}}} \mathrm{KL}(P||Q) \tag{5.3}$$

with the entries of $Q \in \mathbb{R}^{L \times M}$ given by

$$q_{li} = \exp\left(-\frac{1}{2\gamma \Delta t} \| \widetilde{z}^l_{n+1} - \hat{z}^i_{n+1} \|^2 \right)$$

and the set Π_{M} defined by (3.26). The desired particles $z^i_{n+1} = Z^i_{n+1}(\omega)$ are finally given as realizations of

$$Z^i_{n+1} = \sum_{l=1}^{L} \widetilde{z}^l_{n+1} p^*_{li} + (\gamma \Delta t)^{1/2} \Xi^i_n, \quad \Xi^i_n \sim \mathrm{N}(0, I), \tag{5.4}$$

for $i = 1, \ldots, M$.

The update (5.4), with P^* defined by (5.3), can be viewed as data-driven drift correction combined with a standard approximation to the Brownian diffusion part of the underlying SDE model. It remains to be investigated in what sense (5.4) can be viewed as an approximation to the FPF formulation (4.19) as $M \to \infty$ and $\Delta t \to 0$.

Remark 5.2. One can also use the matrix P^* from (5.3) to implement a resampling scheme

$$\mathbb{P}[Z^i_{n+1}(\omega) = \widetilde{z}^l_{n+1}] = p^*_{li} \tag{5.5}$$

for $i = 1, \ldots, M$. Note that, contrary to classical resampling schemes based on weighted particles $(\widetilde{z}^l_{n+1}, w^l_{n+1})$, $l = 1, \ldots, L$, the sampling probabilities p^*_{li} take into account the underlying geometry of the forecasts $\hat{z}^i_{n+1}$ in state space.

Example 5.3. We consider the SDE formulation

$$\mathrm{d}Z_t = f(Z_t)\,\mathrm{d}t + \gamma^{1/2}\,\mathrm{d}W_t$$

of a stochastically perturbed Lorenz-63 model (Lorenz 1963, Reich and Cotter 2015, Law *et al.* 2015) with diffusion constant $\gamma = 0.1$. The system is fully observed according to

$$\mathrm{d}Y_t = f(Z_t)\,\mathrm{d}t + R^{1/2}\,\mathrm{d}V_t$$

with measurement error variance $R = 0.1$, and the system is simulated over a time interval $t \in [0, 40\,000]$ with step-size $\Delta t = 0.01$. We implemented a standard particle filter with resampling performed after each time-step and compared the resulting RMS errors with those arising from using (5.4)

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

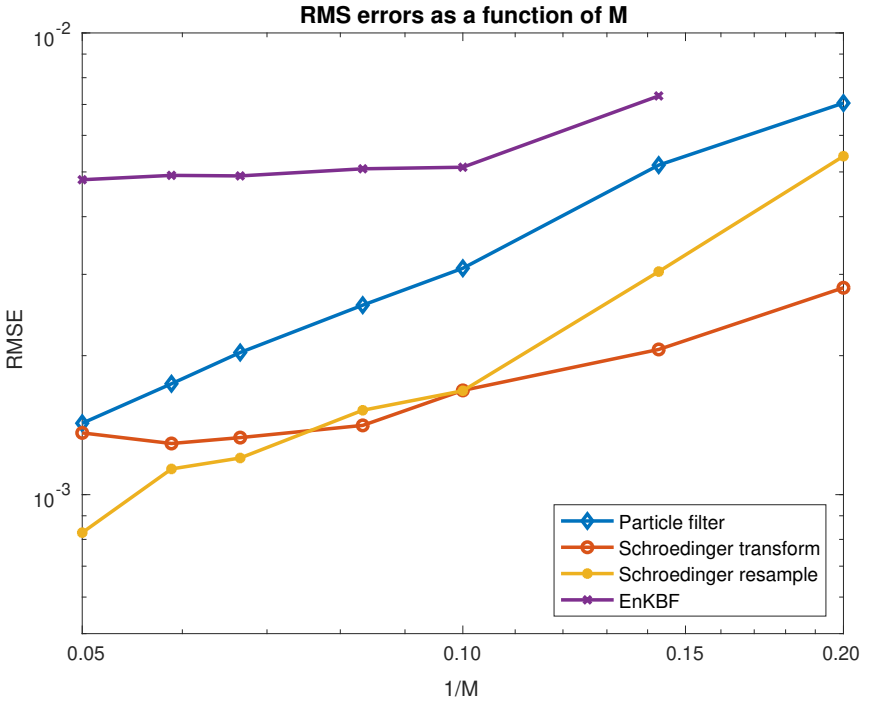


Figure 5.1. RMS errors as a function of sample size, M, for a standard particle filter, the EnKBF, and implementations of (5.4) (Schrödinger transform) and (5.5) (Schrödinger resample), respectively. Both Schrödinger-based methods outperform the standard particle filter for small ensemble sizes. The EnKBF diverged for the smallest ensemble size of $M = 5$ and performed worse than all other methods for this highly nonlinear problem.

(Schrödinger transform) and (5.5) (Schrödinger resample), respectively. See Figure 5.1. It can be seen that the Schrödinger-based methods outperform the standard particle filter in terms of RMS errors for small ensemble sizes. The Schrödinger transform method is particularly robust for very small ensemble sizes while Schrödinger resample performs better at larger sample sizes. We also implemented the EnKBF (5.1) and found that it diverged for the smallest ensemble size of $M = 5$ and performed worse than the other methods for larger ensemble sizes.

6. Conclusions

We have summarized sequential data assimilation techniques suitable for state estimation of discrete- and continuous-time stochastic processes. In addition to algorithmic approaches based on the standard filtering and smoothing framework of stochastic analysis, we have drawn a connection

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

to a boundary value problem over joint probability measures first formulated by Erwin Schrödinger. We have argued that sequential data assimilation essentially needs to approximate such a boundary value problem with the boundary conditions given by the filtering distributions at consecutive observation times.

Application of these techniques to high-dimensional problems arising, for example, from the spatial discretization of PDEs requires further approximations in the form of localization and inflation, which we have not discussed in this survey. See, for example, Evensen (2006), Reich and Cotter (2015) and Asch *et al.* (2017) for further details. In particular, the localization framework for particle filters as introduced by Chen and Reich (2015) and Reich and Cotter (2015) in the context of scenario (A) could be generalized to scenarios (B) and (C) from Definition 2.4.

Finally, the approaches and computational techniques discussed in this paper are also relevant to combined state and parameter estimation.

Acknowledgement

This research has been partially funded by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1294 'Data Assimilation'. Feedback on earlier versions of this paper by Nikolas Kantas, Prashant Mehta and Tim Sullivan has been invaluable.

Appendices

A.1. Mesh-free approximations to Fokker–Planck and backward Kolmogorov equations

In this appendix we discuss two closely related approximations, first to the Fokker–Planck equation (2.26) with the (time-independent) operator (2.25) taking the special form

$$\mathcal{L}^\dagger \pi = -\nabla_z \cdot (\pi \nabla_z \log \pi^*) + \Delta_z \pi = \nabla_z \cdot \left(\pi^* \nabla_z \frac{\pi}{\pi^*} \right)$$

and, second, to its adjoint operator $\mathcal{L}$ given by (3.10).

The approximation to the Fokker–Planck equation (2.26) with drift term

$$f_t(z) = \nabla_z \log \pi^*(z) \tag{A.1}$$

can be used to transform samples x_0^i, $i = 1, \ldots, M$ from a (prior) PDF π_0 into samples from a target (posterior) PDF π^* using an evolution equation of the form

$$\frac{\mathrm{d}}{\mathrm{d}s} \breve{Z}_s = F_s(\breve{Z}_s), \tag{A.2}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

with $\breve{Z}_0 \sim \breve{\pi}_0 := \pi_0$ such that

$$\lim_{s\to\infty} \breve{Z}_s \sim \pi^*.$$

The evolution of the marginal PDFs $\breve{\pi}_s$ is given by Liouville's equation

$$\partial_s \breve{\pi}_s = -\nabla_z \cdot (\breve{\pi}_s F_s). \tag{A.3}$$

We now choose F_s such that the Kullback–Leibler divergence $\mathrm{KL}\,(\breve{\pi}_s||\pi^*)$ is non-increasing in time, that is,

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathrm{KL}\,(\breve{\pi}_s||\pi^*) = \int \breve{\pi}_s \left\{ F_s \cdot \nabla_z \log \frac{\breve{\pi}_s}{\pi^*} \right\} \mathrm{d}z \le 0. \tag{A.4}$$

A natural choice is

$$F_s(z) := -\nabla_z \log \frac{\breve{\pi}_s}{\pi^*}(z),$$

which renders (A.3) formally equivalent to the Fokker–Planck equation (2.26) with drift term (A.1) (Reich and Cotter 2015, Peyre and Cuturi 2018).

Let us now approximate the evolution equation (A.2) over a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ with kernel $k(z - z')$ and inner product $\langle f, g\rangle_{\mathcal{H}}$, which satisfies the reproducing property

$$\langle k(\cdot - z'), f\rangle_{\mathcal{H}} = f(z'). \tag{A.5}$$

Following Russo (1990) and Degond and Mustieles (1990), we first introduce the approximation

$$\widetilde{\pi}_s(z) := \frac{1}{M}\sum_{i=1}^M k(z - z_s^i) \tag{A.6}$$

to the marginal densities $\breve{\pi}_s$. Note that (A.5) implies that

$$\langle f, \widetilde{\pi}_s\rangle_{\mathcal{H}} = \frac{1}{M}\sum_{i=1}^M f(z_s^i).$$

Given some evolution equations

$$\frac{\mathrm{d}}{\mathrm{d}s} z_s^i = u_s^i$$

for the particles z_s^i, $i = 1, \ldots, M$, we find that (A.6) satisfies Liouville's equation, that is,

$$\partial_s \widetilde{\pi}_s = -\nabla_z \cdot (\widetilde{\pi}_s \widetilde{F}_s)$$

with

$$\widetilde{F}_s(z) = \frac{\sum_{i=1}^M k(z - z_s^i)\, u_s^i}{\sum_{i=1}^M k(z - z_s^i)}.$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

We finally introduce the functional

$$\mathcal{V}(\{z_s^l\}) := \left\langle \widetilde{\pi}_s, \log \frac{\widetilde{\pi}_s}{\pi^*} \right\rangle_{\mathcal{H}} = \frac{1}{M} \sum_{i=1}^{M} \log \frac{\frac{1}{M}\sum_{j=1}^{M} k(z_s^i - z_s^j)}{\pi^*(z_s^i)}$$

as an approximation to the Kullback–Leibler divergence in the RKHS $\mathcal{H}$ and set

$$u_s^i := -M \nabla_{z_s^i} \mathcal{V}(\{z_s^l\}), \tag{A.7}$$

which constitutes the desired particle approximation to the Fokker–Planck equation (2.26) with drift term (A.1). Time-stepping methods for such gradient flow systems have been discussed by Pathiraja and Reich (2019).

We also remark that an alternative interacting particle system, approximating the same asymptotic PDF π^* in the limit $s \to \infty$, has been proposed recently by Liu and Wang (2016) under the notion of Stein variational descent. See Lu, Lu and Nolen (2019) for a theoretical analysis of Stein variational descent, which implies in particular that Stein variational descent can be viewed as a Lagrangian particle approximation to the modified evolution equation

$$\partial_s \breve{\pi}_s = \nabla_z \cdot \left(\breve{\pi}_s^2 \nabla_z \log \frac{\breve{\pi}_s}{\pi^*}(z) \right) = \nabla_z \cdot (\breve{\pi}_s (\nabla_z \breve{\pi}_s - \breve{\pi}_s \nabla_z \log \pi^*))$$

in the marginal PDFs $\breve{\pi}_s$, that is, one uses

$$F_s(z) := -\breve{\pi}_s \nabla_z \log \frac{\breve{\pi}_s}{\pi^*}(z)$$

in (A.2). The Kullback–Leibler divergence is still non-increasing since (A.4) becomes

$$\frac{\mathrm{d}}{\mathrm{d}s} \mathrm{KL}\,(\breve{\pi}_s || \pi^*) = -\int \|F_s\|^2 \,\mathrm{d}z \le 0.$$

A numerical discretization is obtained through the approximation

$$F_s(z') \approx \int F_s(z)\, k(z - z')\,\mathrm{d}z,$$

that is, one views the kernel $k(z - z')$ as a regularized Dirac delta function. This approximation leads to another vector field

$$\begin{aligned} \widehat{F}_s(z') &:= -\int \breve{\pi}_s(z) \{\nabla_z \log \breve{\pi}_s(z) - \nabla_z \log \pi^*(z)\} k(z - z')\,\mathrm{d}z \\ &= \int \breve{\pi}_s(z) \{\nabla_z k(z - z') + k(z - z')\, \nabla_z \log \pi^*(z)\}\,\mathrm{d}z. \end{aligned}$$

On extending the RKHS $\mathcal{H}$ and its reproducing property (A.5) componentwise to vector-valued functions, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}s} \mathrm{KL}\,(\breve{\pi}_s || \pi^*) = -\int F_s \cdot \widehat{F}_s \,\mathrm{d}z = -\langle \widehat{F}_s, \widehat{F}_s \rangle_{\mathcal{H}} \le 0$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

along transformations induced by the vector field $\widehat{F}_s$. See Liu and Wang (2016) for more details.

We now turn our attention to the dual operator $\mathcal{L}_t$, defined by (3.10), which also arises from (4.6) and (4.20), respectively. More specifically, let us rewrite (4.20) in the form

$$\mathcal{A}_t\phi_t = -(h - \pi_t[h]) \tag{A.8}$$

with the operator $\mathcal{A}_t$ defined by

$$\mathcal{A}_t g := \frac{1}{\pi_t}\nabla_z \cdot (\pi_t \nabla_z g).$$

Then we find that $\mathcal{A}_t$ is of the form of $\mathcal{L}_t$ with π_t taking the role of π^*.

We also recall that (3.11) provides an approximation to $\mathcal{L}_t$ and hence to $\mathcal{A}_t$. This observation allows one to introduce a sample-based method for approximating the potential ϕ defined by the elliptic partial differential equation (A.8) for a given function $h(z)$.

Here we instead follow the presentation of Taghvaei and Mehta (2016) and Taghvaei *et al.* (2017) and assume that we have M samples z^i from a PDF π. The method is based on

$$\frac{\phi - \mathrm{e}^{\epsilon\mathcal{A}}\phi}{\epsilon} \approx h - \pi[h] \tag{A.9}$$

for $\epsilon > 0$ sufficiently small and upon replacing $\mathrm{e}^{\epsilon\mathcal{A}}$ with a diffusion map approximation (Harlim 2018) of the form

$$\mathrm{e}^{\epsilon\mathcal{A}}\phi(z) \approx T_\epsilon\phi(z) := \sum_{i=1}^M k_\epsilon(z, z^i)\,\phi(z^i). \tag{A.10}$$

The required kernel functions $k_\epsilon(z, z^i)$ are defined as follows. Let

$$n_\epsilon(z) := \mathrm{n}(z; 0, 2\epsilon\, I)$$

and

$$p_\epsilon(z) := \frac{1}{M}\sum_{j=1}^M n_\epsilon(z - z^j) = \frac{1}{M}\sum_{j=1}^M \mathrm{n}(z; z^j, 2\epsilon\, I).$$

Then

$$k_\epsilon(z, z^i) := \frac{n_\epsilon(z - z^i)}{c_\epsilon(z)\, p_\epsilon(z^i)^{1/2}}$$

with normalization factor

$$c_\epsilon(z) := \sum_{l=1}^M \frac{n_\epsilon(z - z^l)}{p_\epsilon(z^l)^{1/2}}.$$

In other words, the operator T_ϵ reproduces constant functions.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

The approximations (A.9) and (A.10) lead to the fixed-point problem[3]

$$\phi_j = \sum_{i=1}^{M} k_\epsilon(z^j, z^i)\,\phi_i + \epsilon \Delta h_i\,, \quad j = 1, \ldots, M, \tag{A.11}$$

in the scalar coefficients ϕ_j, $j = 1, \ldots, M$, for given

$$\Delta h_i := h(z^i) - \bar{h}, \quad \bar{h} := \frac{1}{M}\sum_{l=1}^{M} h(z^l).$$

Since T_ϵ reproduces constant functions, (A.11) determines ϕ_i up to a constant contribution, which we fix by requiring

$$\sum_{i=1}^{M} \phi_i = 0.$$

The desired functional approximation $\widetilde{\phi}$ to the potential ϕ is now provided by

$$\widetilde{\phi}(z) = \sum_{i=1}^{M} k_\epsilon(z, z^i)\{\phi_i + \epsilon \Delta h_i\}. \tag{A.12}$$

Furthermore, since

$$\begin{aligned}\nabla_z k_\epsilon(z, z^i) &= \frac{-1}{2\epsilon} k_\epsilon(z, z^i)\left((z - z^i) - \sum_{l=1}^{M} k_\epsilon(z, z^l)(z - z^l)\right) \\ &= \frac{1}{2\epsilon} k_\epsilon(z, z^i)\left(z^i - \sum_{l=1}^{M} k_\epsilon(z, z^l) z^l\right),\end{aligned}$$

we obtain

$$\nabla_z \widetilde{\phi}(z^j) = \sum_{i=1}^{M} \nabla_z k_\epsilon(z^j, z^i)\, r_i = \sum_{i=1}^{M} z^i\, a_{ij},$$

with

$$r_i = \phi_i + \epsilon\, \Delta h_i$$

and

$$a_{ij} := \frac{1}{2\epsilon} k_\epsilon(z^j, z^i)\left(r_i - \sum_{l=1}^{M} k_\epsilon(z^j, z^l)\, r_l\right).$$

[3] It would also be possible to employ the approximation (3.11) in the fixed-point problem (A.11), that is, to replace $k_\epsilon(z^j, z^i)$ by $(Q_+)_{ji}$ in (3.11) with $\Delta t = \epsilon$ and $\pi^* = \pi_t$.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

We note that

$$\sum_{i=1}^{M} a_{ij} = 0$$

and

$$\lim_{\epsilon\to\infty} a_{ij} = \frac{1}{M}\Delta h_i$$

since

$$\lim_{\epsilon\to\infty} k_\epsilon(z^j, z^i) = \frac{1}{M}.$$

In other words,

$$\lim_{\epsilon\to\infty} \nabla_z \widetilde{\phi}(z^j) = \frac{1}{M}\sum_{i=1}^{M} z^i\,(h(z^i) - \bar{h}) = K^M$$

independent of z^j, which is equal to an empirical estimator for the covariance between z and $h(z)$ and which, in the context of the FPF, leads to the EnKBF formulations (5.1) of Section 5.1. See Taghvaei *et al.* (2017) for more details and Taghvaei, Mehta and Meyn (2019) for a convergence analysis.

A.2. Regularized Störmer–Verlet for HMC

One is often faced with the task of sampling from a high-dimensional PDF of the form

$$\pi(x) \propto \exp(-V(x)), \quad V(x) := \frac{1}{2}(x-\bar{x})^{\mathrm{T}} B^{-1}(x-\bar{x}) + U(x),$$

for known $\bar{x} \in \mathbb{R}^{N_x}$, $B \in \mathbb{R}^{N_x\times N_x}$, and $U : \mathbb{R}^{N_x} \to \mathbb{R}$. The hybrid Monte Carlo (HMC) method (Neal 1996, Liu 2001, Bou-Rabee and Sanz-Serna 2018) has emerged as a popular Markov chain Monte Carlo (MCMC) method for tackling this problem. HMC relies on a symplectic discretization of the Hamiltonian equations of motion

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\tau}x &= M^{-1}p,\\
\frac{\mathrm{d}}{\mathrm{d}\tau}p &= -\nabla_x V(x) = -B^{-1}(x-\bar{x}) - \nabla_x U(x)
\end{aligned}$$

in an artificial time τ (Leimkuhler and Reich 2005). The conserved energy (or Hamiltonian) is provided by

$$\mathcal{H}(x,p) = \frac{1}{2}p^{\mathrm{T}}M^{-1}p + V(x). \tag{A.13}$$

The symmetric positive definite mass matrix $M \in \mathbb{R}^{N_x\times N_x}$ can be chosen arbitrarily, and a natural choice in terms of sampling efficiency is $M = B^{-1}$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

(Beskos *et al.* 2011). However, when also taking into account computational efficiency, a Störmer–Verlet discretization

$$p_{n+1/2} = p_n - \frac{\Delta\tau}{2}\nabla_x V(x_n), \tag{A.14}$$

$$q_{n+1} = q_n + \Delta\tau \widetilde{M}^{-1} p_{n+1/2}, \tag{A.15}$$

$$p_{n+1} = p_{n+1/2} - \frac{\Delta\tau}{2}\nabla_x V(x_{n+1}), \tag{A.16}$$

with step-size $\Delta\tau > 0$, mass matrix $M = I$ in (A.13) and modified mass matrix

$$\widetilde{M} = I + \frac{\Delta\tau^2}{4} B^{-1} \tag{A.17}$$

in (A.15) emerges as an attractive alternative, since it implies

$$\mathcal{H}(x_n, p_n) = \mathcal{H}(x_{n+1}, p_{n+1})$$

for all $\Delta\tau > 0$ provided $U(x) \equiv 0$. The Störmer–Verlet formulation (A.14)–(A.16) is based on a regularized formulation of Hamiltonian equations of motion for highly oscillatory systems as discussed, for example, by Reich and Hundertmark (2011).

Energy-conserving time-stepping methods for linear Hamiltonian systems have become an essential building block for applications of HMC to infinite-dimensional inference problems, where B^{-1} corresponds to the discretization of a positive, self-adjoint and trace-class operator $\mathcal{B}$. See, for example, Beskos *et al.* (2017).

Note that the Störmer–Verlet discretization (A.14)–(A.16) together with (A.17) can be easily extended to inference problems with constraints $g(x) = 0$ (Leimkuhler and Reich 2005) and that (A.14)–(A.16) conserves equilibria,[4] that is, points x_* with $\nabla V(x_*) = 0$, regardless of the step-size $\Delta\tau$.

HMC methods, based on (A.14)–(A.16) and (A.17), can be used to sample from the smoothing distribution of an SDE as considered in Sections 2.2 and 3.3.

A.3. Ensemble Kalman filter

We summarize the formulation of an ensemble Kalman filter in the form (3.20). We start with the stochastic ensemble Kalman filter (Evensen 2006), which is given by

$$\widehat{Z}_1^j = z_1^j - K(h(z_1^j) + \Theta^j - y_1), \quad \Theta^j \sim \mathrm{N}(0, R), \tag{A.18}$$

[4] Note that equilibria of the Hamiltonian equations of motion correspond to MAP estimators of the underlying Bayesian inference problem.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

with Kalman gain matrix

$$K = P^{zh}(P^{hh} + R)^{-1} = \frac{1}{M-1}\sum_{i=1}^{M} z_1^i \, (h(z_1^i) - \bar{h})^{\mathrm{T}}(P^{hh} + R)^{-1}$$

and

$$P^{hh} := \frac{1}{M-1}\sum_{l=1}^{M} h(z_1^l) \, (h(z_1^l) - \bar{h})^{\mathrm{T}}, \quad \bar{h} := \frac{1}{M}\sum_{l=1}^{M} h(z_1^l).$$

Formulation (A.18) can be rewritten in the form (3.20) with

$$p_{ij}^* = \delta_{ij} - \frac{1}{M-1}(h(z_1^i) - \bar{h})^{\mathrm{T}}(P^{hh} + R)^{-1}(h(z_1^j) - y_1 + \Theta^j), \tag{A.19}$$

where δ_{ij} denotes the Kronecker delta, that is, $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$.

More generally, one can think about ensemble Kalman filters and their generalizations (Anderson 2010) as first defining appropriate updates $\widehat{y}_1^i$ to the predicted $y_1^i = h(z_1^i)$ using the observed y_1, which is then extrapolated to the state variable z via linear regression, that is,

$$\widehat{z}_1^j = z_1^j + \frac{1}{M-1}\sum_{i=1}^{M} z_1^i (h(z_1^i) - \bar{h})^{\mathrm{T}}(P^{hh})^{-1}(\widehat{y}_1^j - y_1^j), \tag{A.20}$$

which can be reformulated in the form (3.20) (Reich and Cotter 2015). Note that the consistency result

$$H\widehat{z}_1^i = \widehat{y}_1^i$$

follows from (A.20) for linear forward maps $h(z) = Hz$.

Within such a linear regression framework, one can easily derive ensemble transformations for the particles z_0^i at time $t = 0$. We simply take the coefficients p_{ij}^*, as defined for example by an ensemble Kalman filter (A.19), and apply them to z_0^i, that is,

$$\widehat{z}_0^j = \sum_{i=1}^{M} z_0^i \, p_{ij}^*.$$

These transformed particles can be used to approximate the smoothing distribution $\widehat{\pi}_0$. See, for example, Evensen (2006) and Kirchgessner *et al.* (2017) for more details.

Finally, one can also interpret the ensemble Kalman filter as a continuous update in artificial time $s \geq 0$ of the form

$$\mathrm{d}z_s^i = -P^{zh}R^{-1} \, \mathrm{d}I_s^i \tag{A.21}$$

with the innovations I_s^i given either by

$$\mathrm{d}I_s^i = \frac{1}{2}(h(z_s^i) + \bar{h}_s) \, \mathrm{d}s - y_1 \, \mathrm{d}s \tag{A.22}$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

or, alternatively, by

$$\mathrm{d}I_s^i = h(z_s^i)\,\mathrm{d}s + R^{1/2}\,\mathrm{d}V_s^i - y_1\,\mathrm{d}s,$$

where V_s^i stands for standard Brownian motion (Bergemann and Reich 2010, Reich 2011, Bergemann and Reich 2012). Equation (A.21) with innovation (A.22) can be given a gradient flow structure (Bergemann and Reich 2010, Reich and Cotter 2015) of the form

$$\frac{1}{\mathrm{d}s}\mathrm{d}z_s^i = -P^{zz}\nabla_{z^i}\mathcal{V}(\{z_s^j\}), \tag{A.23}$$

with potential

$$\mathcal{V}(\{z^j\}) := \frac{1-\alpha}{4}\sum_{j=1}^{M}(h(z^j)-y_1)^{\mathrm{T}}R^{-1}(h(z^j)-y_1) + \frac{(1+\alpha)M}{4}(\bar{h}-y_1)^{\mathrm{T}}R^{-1}(\bar{h}-y_1)$$

and $\alpha = 0$ for the standard ensemble Kalman filter, while $\alpha \in (0,1)$ can be seen as a form of variance inflation (Reich and Cotter 2015).

A theoretical study of such dynamic formulations in the limit of $s \to \infty$ and $\alpha = -1$ has been initiated by Schillings and Stuart (2017). There is an interesting link to stochastic gradient methods (Bottou, Curtis and Nocedal 2018), which find application in situations where the dimension of the data y_1 is very high and the computation of the complete gradient $\nabla_z h(z)$ becomes prohibitive. More specifically, the basic concepts of stochastic gradient methods can be extended to (A.23) if R is diagonal, in which case one would pick at random paired components of h and y_1 at the kth time-step of a discretization of (A.23), with the step-size Δs_k chosen appropriately. Finally, we also point to a link between natural gradient methods and Kalman filtering (Ollivier 2018) which can be explored further in the context of the continuous-time ensemble Kalman filter formulation (A.23).

A.4. Numerical treatment of forward–backward SDEs

We discuss a numerical approximation of the forward–backward SDE problem defined by the forward SDE (2.24) and the backward SDE (2.54) with initial condition $Z_0^+ \sim \pi_0$ at time $t = 0$ and final condition $Y_1(Z_1^+) = l(Z_1^+)/\beta$ at time $t = 1$. Discretization of the forward SDE (2.24) by the Euler–Maruyama method (3.5) leads to M numerical solution paths $z_{0:N}^i$, $i = 1, \ldots, M$, which, according to Definition 3.5, lead to N discrete Markov transition matrices $Q_n^+ \in \mathbb{R}^{M\times M}$, $n = 1, \ldots, N$.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

The Euler–Maruyama method is now also applied to the backward SDE (2.54) and yields

$$Y_n = Y_{n+1} - \Delta t^{1/2} \Xi_n^{\mathrm{T}} V_n.$$

Upon taking conditional expectation we obtain

$$Y_n(Z_n^+) = \mathbb{E}[Y_{n+1}|Z_n^+] \tag{A.24}$$

and

$$\Delta t^{1/2} \mathbb{E}[\Xi_n \Xi_n^{\mathrm{T}}] V_n(Z_n^+) = \mathbb{E}[(Y_{n+1} - Y_n)\Xi_n | Z_n^+],$$

respectively. The last equation leads to

$$V_n(Z_n^+) = \Delta t^{-1/2} \mathbb{E}[(Y_{n+1} - Y_n)\Xi_n | Z_n^+]. \tag{A.25}$$

We also have $Y_N(Z_N^+) = l(Z_N^+)/\beta$ at final time $t = 1 = N\Delta t$. See page 45 in Carmona (2016) for more details.

We finally need to approximate the conditional expectation values in (A.24) and (A.25), for which we employ the discrete Markov transition matrix Q_{n+1}^+ and the discrete increments

$$\zeta_{ij} := \frac{1}{(\gamma \Delta t)^{1/2}} (z_{n+1}^i - z_n^j - \Delta t f_{t_n}(z_n^j)) \in \mathbb{R}^M.$$

Given $y_{n+1}^j \approx Y(z_{n+1}^j)$ at time level t_{n+1}, we then approximate (A.24) by

$$y_n^j := \sum_{i=1}^M y_{n+1}^i (Q_{n+1}^+)_{ij} \tag{A.26}$$

for $n = N-1, \ldots, 0$. The backward iteration is initiated by setting $y_N^i = l(z_N^i)/\beta$, $i = 1, \ldots, M$. Furthermore, a Monte Carlo approximation to (A.25) at z_n^j is provided by

$$\Delta t^{1/2} \sum_{i=1}^M \{\xi_{ij} (\xi_{ij})^{\mathrm{T}} (Q_{n+1}^+)_{ij}\} v_n^j = \sum_{i=1}^M (y_{n+1}^i - y_n^j) \xi_{ij} (Q_{n+1}^+)_{ij}$$

and, upon assuming invertibility, we obtain the explicit expression

$$v_n^j := \Delta t^{-1/2} \left(\sum_{i=1}^M \xi_{ij} (\xi_{ij})^{\mathrm{T}} (Q_{n+1}^+)_{ij} \right)^{-1} \sum_{i=1}^M (y_{n+1}^i - y_n^j) \xi_{ij} (Q_{n+1}^+)_{ij} \tag{A.27}$$

for $n = N-1, \ldots, 0$.

Recall from Remark 2.20 that $y_n^j \in \mathbb{R}$ provides an approximation to $\psi_{t_n}(z_n^j)$ and $v_n^j \in \mathbb{R}^{N_z}$ an approximation to $\gamma^{1/2} \nabla_z \psi_{t_n}(z_n^j)$, respectively, where ψ_t denotes the solution of the backward Kolmogorov equation (2.31) with final condition $\psi_1(z) = l(z)/\beta$. Hence, the forward solution paths $z_{0:N}^i$,

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

$i = 1, \ldots, M$, together with the backward approximations (A.26) and (A.27) provide a mesh-free approximation to the backward Kolmogorov equation (2.31). Furthermore, the associated control law (2.32) can be approximated by

$$u_{t_n}(z_n^i) \approx \frac{\gamma^{1/2}}{y_n^i} v_n^i.$$

The division by y_n^i can be avoided by means of the following alternative formulation. We introduce the potential

$$\phi_t := \log \psi_t, \tag{A.28}$$

which satisfies the modified backward Kolmogorov equation

$$0 = \partial_t \phi_t + \mathcal{L}_t \phi_t + \frac{\gamma}{2} \|\nabla_z \phi_t\|^2$$

with final condition $\phi_1(z) = \log l(z)$, where we have ignored the constant $\log \beta$. Hence Itô's formula applied to $\phi_t(Z_t^+)$ leads to

$$\mathrm{d}\phi_t = -\frac{\gamma}{2} \|\nabla_z \phi_t\|^2 \, \mathrm{d}t + \gamma^{1/2} \nabla_z \phi_t \cdot \mathrm{d}W_t^+ \tag{A.29}$$

along solutions Z_t^+ of the forward SDE (2.24). It follows from

$$\frac{\mathrm{d}\widehat{\mathbb{P}}}{\mathrm{d}\mathbb{Q}^u}_{|z_{[0,1]}} = \frac{l(z_1)}{\beta} \frac{\pi_0(z_0)}{q_0(z_0)} \exp\left(\frac{1}{2\gamma} \int_0^1 (\|u_t\|^2 \, \mathrm{d}t - 2\gamma^{1/2} u_t \cdot \mathrm{d}W_t^+) \right),$$

with $u_t = \gamma \nabla_z \phi_t$, $q_0 = \widehat{\pi}_0$, and

$$\log \frac{l(z_1)}{\beta} - \log \frac{\widehat{\pi}_0(z_0)}{\pi_0(z_0)} = \int_0^1 \mathrm{d}\phi_t$$

that $\widehat{\mathbb{P}} = \mathbb{Q}^u$, as desired.

The backward SDE associated with (A.29) becomes

$$\mathrm{d}Y_t = -\frac{1}{2} \|V_t\|^2 \, \mathrm{d}t + V_t \cdot \mathrm{d}W_t^+ \tag{A.30}$$

and its Euler–Maruyama discretization is

$$Y_n = Y_{n+1} + \frac{\Delta t}{2} \|V_n\|^2 - \Delta t^{1/2} \Xi_n^{\mathrm{T}} V_n.$$

Numerical values (y_n^i, v_n^i) can be obtained as before with (A.26) replaced by

$$y_n^j := \sum_{i=1}^M \left(y_{n+1}^i + \frac{\Delta t}{2} \|v_n^j\|^2 \right) (Q_{n+1}^+)_{ij}$$

and the control law (2.32) is now approximated by

$$u_{t_n}(z_n^i) \approx \gamma^{1/2} v_n^i.$$

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

We re-emphasize that the backward SDE (A.30) arises naturally from an optimal control perspective onto the smoothing problem. See Carmona (2016) for more details on the connection between optimal control and backward SDEs. In particular, this connection leads to the following alternative approximation

$$u_{t_n}(z_n^j) \approx \sum_{i=1}^M z_{n+1}^i \{(\widehat{Q}_{n+1}^+)_{ij} - (Q_{n+1}^+)_{ij}\}$$

of the control law (2.32). Here $\widehat{Q}_{n+1}^+$ denotes the twisted Markov transition matrix defined by

$$\widehat{Q}_{n+1}^+ = D(y_{n+1})\, Q_{n+1}^+\, D(y_n)^{-1}, \quad y_n = (y_n^1, \ldots, y_n^M)^{\mathrm{T}}.$$

Remark. The backward SDE (A.30) can also be utilized to reformulate the Schrödinger system (2.58)–(2.61). More specifically, one seeks an initial π_0^ψ which evolves under the forward SDE (2.24) with $Z_0^+ \sim \pi_0^\psi$ such that the solution Y_t of the associated backward SDE (A.30) with final condition

$$Y_1(z) = \log \widehat{\pi}_1(z) - \log \pi_1^\psi(z)$$

implies $\pi_0 \propto \pi_0^\psi \exp(Y_0)$. The desired control law in (2.30) is provided by

$$u_t(z) = \gamma \nabla_z Y_t(z) = \gamma^{1/2} V_t(z).$$

REFERENCES[5]

W. Acevedo, J. de Wiljes and S. Reich (2017), 'Second-order accurate ensemble transform particle filters', *SIAM J. Sci. Comput.* **39**, A1834–A1850.

S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso and A. Stuart (2017), 'Importance sampling: Computational complexity and intrinsic dimension', *Statist. Sci.* **32**, 405–431.

J. Amezcua, E. Kalnay, K. Ide and S. Reich (2014), 'Ensemble transform Kalman–Bucy filters', *Q. J. Roy. Meteorol. Soc.* **140**, 995–1004.

J. L. Anderson (2010), 'A non-Gaussian ensemble filter update for data assimilation', *Monthly Weather Rev.* **138**, 4186–4198.

M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp (2002), 'A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking', *IEEE Trans. Signal Process.* **50**, 174–188.

M. Asch, M. Bocquet and M. Nodet (2017), *Data Assimilation: Methods, Algorithms and Applications*, SIAM.

[5] The URLs cited in this work were correct at the time of going to press, but the publisher and the authors make no undertaking that the citations remain live or are accurate or appropriate.

https://doi.org/10.1017/S0962492919000011 Published online by Cambridge University Press

A. Bain and D. Crisan (2008), *Fundamentals of Stochastic Filtering*, Vol. 60 of Stochastic Modelling and Applied Probability, Springer.

T. Bengtsson, P. Bickel and B. Li (2008), Curse of dimensionality revisited: Collapse of the particle filter in very large scale systems. In *Probability and Statistics: Essays in Honor of David F. Freedman*, Vol. 2 of IMA Collections, Institute of Mathematical Sciences, pp. 316–334.

K. Bergemann and S. Reich (2010), 'A mollified ensemble Kalman filter', *Q. J. Roy. Meteorol. Soc.* **136**, 1636–1643.

K. Bergemann and S. Reich (2012), 'An ensemble Kalman–Bucy filter for continuous data assimilation', *Meteorol. Z.* **21**, 213–219.

A. Beskos, M. Girolami, S. Lan, P. Farrell and A. Stuart (2017), 'Geometric MCMC for infinite-dimensional inverse problems', *J. Comput. Phys.* **335**, 327–351.

A. Beskos, F. J. Pinski, J. M. Sanz-Serna and A. M. Stuart (2011), 'Hybrid Monte Carlo on Hilbert spaces', *Stochastic Process. Appl.* **121**, 2201–2230.

D. Blömker, C. Schillings and P. Wacker (2018), 'A strongly convergent numerical scheme for ensemble Kalman inversion', *SIAM J. Numer. Anal.* **56**, 2537–2562.

L. Bottou, F. E. Curtis and J. Nocedal (2018), 'Optimization methods for large-scale machine learning', *SIAM Rev.* **60**, 223–311.

N. Bou-Rabee and J. M. Sanz-Serna (2018), Geometric integrators and the Hamiltonian Monte Carlo method. In *Acta Numerica*, Vol. 27, Cambridge University Press, pp. 113–206.

K. Burrage, P. M. Burrage and T. Tian (2004), 'Numerical methods for strong solutions of stochastic differential equations: An overview', *Proc. Roy. Soc. Lond. A* **460**, 373–402.

R. Carmona (2016), *Lectures on BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications*, SIAM.

A. Carrassi, M. Bocquet, L. Bertino and G. Evensen (2018), 'Data assimilation in the geosciences: An overview of methods, issues, and perspectives', *WIREs Clim Change.* **9**, e535.

A. Carrassi, M. Bocquet, A. Hannart and M. Ghil (2017), 'Estimation model evidence using data assimilation', *Q. J. Roy. Meteorol. Soc.* **143**, 866–880.

Y. Chen and S. Reich (2015), Assimilating data into scientific models: An optimal coupling perspective. In *Frontiers in Applied Dynamical Systems: Reviews and Tutorials* (P. J. van Leeuwen *et al.* eds), Vol. 2, Springer, pp. 75–118.

Y. Chen, T. T. Georgiou and M. Pavon (2014), 'On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint', *J. Optim. Theory Appl.* **169**, 671–691.

Y. Chen, T. T. Georgiou and M. Pavon (2016*a*), 'Entropic and displacement interpolation: A computational approach using the Hilbert metric', *SIAM J. Appl. Math.* **76**, 2375–2396.

Y. Chen, T. T. Georgiou and M. Pavon (2016*b*), 'Optimal steering of a linear stochastic system to a final probability distribution, Part I', *Trans. Automat. Control* **61**, 1158–1169.

N. Chustagulprom, S. Reich and M. Reinhardt (2016), 'A hybrid ensemble transform filter for nonlinear and spatially extended dynamical systems', *SIAM/ASA J. Uncertain. Quantif.* **4**, 592–608.

D. Crisan and J. Xiong (2010), 'Approximate McKean–Vlasov representation for a class of SPDEs', *Stochastics* **82**, 53–68.
M. Cuturi (2013), Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)* (C. J. C. Burges *et al.*, eds), pp. 2292–2300.
P. Dai Pra (1991), 'A stochastic control approach to reciprocal diffusion processes', *Appl. Math. Optim.* **23**, 313–329.
F. Daum and J. Huang (2011), Particle filter for nonlinear filters. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5920–5923.
J. de Wiljes, S. Reich and W. Stannat (2018), 'Long-time stability and accuracy of the ensemble Kalman–Bucy filter for fully observed processes and small measurement noise', *SIAM J. Appl. Dyn. Syst.* **17**, 1152–1181.
P. Degond and F.-J. Mustieles (1990), 'A deterministic approximation of diffusion equations using particles', *SIAM J. Sci. Comput.* **11**, 293–310.
P. del Moral (2004), *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Springer.
J. L. Doob (1984), *Classical Potential Theory and its Probabilistic Counterpart*, Springer.
R. Douc and O. Cappe (2005), Comparison of resampling schemes for particle filtering. In *4th International Symposium on Image and Signal Processing and Analysis (ISPA 2005)*, pp. 64–69.
A. Doucet, N. de Freitas and N. Gordon, eds (2001), *Sequential Monte Carlo Methods in Practice*, Springer.
T. A. El Moselhy and Y. M. Marzouk (2012), 'Bayesian inference with optimal maps', *J. Comput. Phys.* **231**, 7815–7850.
G. Evensen (2006), *Data Assimilation: The Ensemble Kalman Filter*, Springer.
P. Fearnhead and H. R. Künsch (2018), 'Particle filters and data assimilation', *Annu. Rev. Statist. Appl.* **5**, 421–449.
W. H. Fleming (1997), 'Deterministic nonlinear filtering', *Ann. Scuola Norm. Super. Pisa* **25**, 435–454.
H. Föllmer and N. Gantert (1997), 'Entropy minimization and Schrödinger processes in infinite dimensions', *Ann. Probab.* **25**, 901–926.
M. Frei and H. R. Künsch (2013), 'Bridging the ensemble Kalman and particle filters', *Biometrika* **100**, 781–800.
C. González-Tokman and B. R. Hunt (2013), 'Ensemble data assimilation for hyperbolic systems', *Phys. D* **243**, 128–142.
P. Guarniero, A. M. Johansen and A. Lee (2017), 'The iterated auxiliary particle filter', *J. Amer. Statist. Assoc.* **112**, 1636–1647.
J. Harlim (2018), *Data-Driven Computational Methods*, Cambridge University Press.
C. Hartmann, L. Richter, C. Schütte and W. Zhang (2017), 'Variational characterization of free energy: Theory and algorithms', *Entropy* **19**, 629.
J. Heng, A. N. Bishop, G. Deligiannidis and A. Doucet (2018), Controlled sequential Monte Carlo. Technical report, Harvard University. arXiv:1708.08396v2
A. H. Jazwinski (1970), *Stochastic Processes and Filtering Theory*, Academic Press.

N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski and N. Chopin (2015), 'On particle methods for parameter estimation in state-space models', *Statist. Sci.* **30**, 328–351.
H. J. Kappen and H. C. Ruiz (2016), 'Adaptive importance sampling for control and inference', *J. Statist. Phys.* **162**, 1244–1266.
H. J. Kappen, V. Gomez and M. Opper (2012), 'Optimal control as a graphical model inference problem', *Machine Learning* **87**, 159–182.
D. Kelly, A. J. Majda and X. T. Tong (2015), 'Concrete ensemble Kalman filters with rigorous catastrophic filter divergence', *Proc. Natl Acad. Sci. USA* **112**, 10589–10594.
D. T. Kelly, K. J. H. Law and A. Stuart (2014), 'Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time', *Nonlinearity* **27**, 2579–2604.
P. Kirchgessner, J. Tödter, B. Ahrens and L. Nerger (2017), 'The smoother extension of the nonlinear ensemble transform filter', *Tellus A* **69**, 1327766.
P. E. Kloeden and E. Platen (1992), *Numerical Solution of Stochastic Differential Equations*, Springer.
E. Kwiatowski and J. Mandel (2015), 'Convergence of the square root ensemble Kalman filter in the large ensemble limit', *SIAM/ASA J. Uncertain. Quantif.* **3**, 1–17.
R. S. Laugesen, P. G. Mehta, S. P. Meyn and M. Raginsky (2015), 'Poisson's equation in nonlinear filtering', *SIAM J. Control Optim.* **53**, 501–525.
K. Law, A. Stuart and K. Zygalakis (2015), *Data Assimilation: A Mathematical Introduction*, Springer.
F. Le Gland, V. Monbet and V.-D. Tran (2011), Large sample asymptotics for the ensemble Kalman filter. In *The Oxford Handbook of Nonlinear Filtering* D. Crisan and B. Rozovskii, eds), Oxford University Press, pp. 598–631.
B. Leimkuhler and S. Reich (2005), *Simulating Hamiltonian Dynamics*, Cambridge University Press.
C. Leonard (2014), 'A survey of the Schrödinger problem and some of its connections with optimal transportation', *Discrete Contin. Dyn. Syst. A* **34**, 1533–1574.
F. Lindsten and T. B. Schön (2013), 'Backward simulation methods for Monte Carlo statistical inference', *Found. Trends Machine Learning* **6**, 1–143.
J. S. Liu (2001), *Monte Carlo Strategies in Scientific Computing*, Springer.
Q. Liu and D. Wang (2016), Stein variational gradient descent: A general purpose Bayesian inference algorithm. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)* (D. D. Lee *et al.*, eds), pp. 2378–2386.
E. N. Lorenz (1963), 'Deterministic non-periodic flows', *J. Atmos. Sci.* **20**, 130–141.
J. Lu, Y. Lu and J. Nolen (2019), 'Scaling limit of the Stein variational gradient descent: The mean field regime', *SIAM J. Math. Anal.* **51**, 648–671.
R. J. McCann (1995), 'Existence and uniqueness of monotone measure-preserving maps', *Duke Math. J.* **80**, 309–323.
S. K. Mitter and N. J. Newton (2003), 'A variational approach to nonlinear estimation', *SIAM J. Control Optim.* **42**, 1813–1833.
R. E. Mortensen (1968), 'Maximum-likelihood recursive nonlinear filtering', *J. Optim. Theory Appl.* **2**, 386–394.

M. Morzfeld, X. Tu, E. Atkins and A. J. Chorin (2012), 'A random map implementation of implicit filters', *J. Comput. Phys.* **231**, 2049–2066.
R. M. Neal (1996), *Bayesian Learning for Neural Networks*, Springer.
E. Nelson (1984), *Quantum Fluctuations*, Princeton University Press.
N. Nüsken, S. Reich and P. Rozdeba (2019), State and parameter estimation from observed signal increments. Technical report, University of Potsdam. arXiv:1903.10717
Y. Ollivier (2018), Online natural gradient as a Kalman filter. *Electron. J. Statist.* **12**, 2930–2961.
S. Pathiraja and S. Reich (2019), Discrete gradients for computational Bayesian inference. Technical report, University of Potsdam. arXiv:1903.00186
G. A. Pavliotis (2014), *Stochastic Processes and Applications*, Springer.
G. Peyre and M. Cuturi (2018), Computational optimal transport. Technical report, CNRS, ENS, CREST, ENSAE. arXiv:1803.00567
F. Pons Llopis, N. Kantas, A. Beskos and A. Jasra (2018), 'Particle filtering for stochastic Navier–Stokes signal observed with linear additive noise', *SIAM J. Sci. Comput.* **40**, A1544–A1565.
S. Reich (2011), 'A dynamical systems framework for intermittent data assimilation', *BIT Numer. Math.* **51**, 235–249.
S. Reich (2012), 'A Gaussian mixture ensemble transform filter', *Q. J. Roy. Meterol. Soc.* **138**, 222–233.
S. Reich (2013), 'A nonparametric ensemble transform method for Bayesian inference', *SIAM J. Sci. Comput.* **35**, A2013–A2024.
S. Reich and C. J. Cotter (2015), *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press.
S. Reich and T. Hundertmark (2011), 'On the use of constraints in molecular and geophysical fluid dynamics', *Eur. Phys. J. Spec. Top.* **200**, 259–270.
S. Robert, D. Leuenberger and H. R. Künsch (2018), 'A local ensemble transform Kalman particle filter for convective-scale data assimilation', *Q. J. Roy. Meteorol. Soc.* **144**, 1279–1296.
H. C. Ruiz and H. J. Kappen (2017), 'Particle smoothing for hidden diffusion processes: Adaptive path integral smoother', *IEEE Trans. Signal Process.* **62**, 3191–3203.
G. Russo (1990), 'Deterministic diffusion of particles', *Commun. Pure Appl. Math.* **43**, 697–733.
S. Särkkä (2013), *Bayesian Filtering and Smoothing*, Cambridge University Press.
C. Schillings and A. M. Stuart (2017), 'Analysis of the ensemble Kalman filter for inverse problems', *SIAM J. Numer. Anal.* **55**, 1264–1290.
E. Schrödinger (1931), 'Über die Umkehrung der Naturgesetze', *Sitzungsberichte der Preußischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse* **IX**, 144–153.
R. Sinkhorn (1967), 'Diagonal equivalence to matrices with prescribed row and column sums', *Amer. Math. Monthly* **74**, 402–405.
A. Taghvaei and P. G. Mehta (2016), Gain function approximation in the feedback particle filter. In *IEEE 55th Conference on Decision and Control (CDC)*, IEEE, pp. 5446–5452.

A. Taghvaei, J. de Wiljes, P. G. Mehta and S. Reich (2017), 'Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem', *ASME. J. Dyn. Sys. Meas. Control.* **140**, 030904.

A. Taghvaei, P. Mehta and S. Meyn (2019), Gain function approximation in the feedback particle fitler. Technical report, University of Illinois at Urbana-Champaign. arXiv:1902.07263

S. Thijssen and H. J. Kappen (2015), 'Path integral control and state-dependent feedback', *Phys. Rev. E* **91**, 032104.

X. T. Tong, A. J. Majda and D. Kelly (2016), 'Nonlinear stability and ergodicity of ensemble based Kalman filters', *Nonlinearity* **29**, 657.

P. J. van Leeuwen (2015), Nonlinear data assimilation for high-dimensional systems. In *Frontiers in Applied Dynamical Systems: Reviews and Tutorials* (P. J. van Leeuwen *et al.* eds), Vol. 2, Springer, pp. 1–73.

P. J. van Leeuwen, H. R. Künsch, L. Nerger, R. Potthast and S. Reich (2018), Particle filter and applications in geosciences. Technical report, University of Reading. arXiv:1807.10434

E. Vanden-Eijnden and J. Weare (2012), 'Data assimilation in the low noise regime with application to the Kuroshio', *Monthly Weather Rev.* **141**, 1822–1841.

S. Vetra-Carvalho, P. J. van Leeuwen, L. Nerger, A. Barth, M. U. Altaf, P. Brasseur, P. Kirchgessner and J.-M. Beckers (2018), 'State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems', *Tellus A* **70**, 1445364.

C. Villani (2003), *Topics in Optimal Transportation*, American Mathematical Society.

C. Villani (2009), *Optimal Transportation: Old and New*, Springer.

J. Xiong (2011), Particle approximations to the filtering problem in continuous time. In *The Oxford Handbook of Nonlinear Filtering* (D. Crisan and B. Rozovskii, eds), Oxford University Press, pp. 635–655.

T. Yang, P. G. Mehta and S. P. Meyn (2013), 'Feedback particle filter', *IEEE Trans. Automat. Control* **58**, 2465–2480.

C. Zhang, A. Taghvaei and P. G. Mehta (2019), 'A mean-field optimal control formulation for global optimization', *IEEE Trans. Automat. Control* **64**, 282–289.