

A nonparametric ensemble transform method for Bayesian inference

Article

Published Version

Reich, S. (2013) A nonparametric ensemble transform method for Bayesian inference. *SIAM Journal on Scientific Computing*, 35 (4). A2013-A2024. ISSN 1095-7197 doi: <https://doi.org/10.1137/130907367> Available at <https://centaur.reading.ac.uk/33551/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1137/130907367>

Publisher: Society for Industrial and Applied Mathematics

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

A NONPARAMETRIC ENSEMBLE TRANSFORM METHOD FOR BAYESIAN INFERENCE*

SEBASTIAN REICH†

Abstract. Many applications, such as intermittent data assimilation, lead to a recursive application of Bayesian inference within a Monte Carlo context. Popular data assimilation algorithms include sequential Monte Carlo methods and ensemble Kalman filters (EnKFs). These methods differ in the way Bayesian inference is implemented. Sequential Monte Carlo methods rely on importance sampling combined with a resampling step, while EnKFs utilize a linear transformation of Monte Carlo samples based on the classic Kalman filter. While EnKFs have proven to be quite robust even for small ensemble sizes, they are not consistent since their derivation relies on a linear regression ansatz. In this paper, we propose another transform method, which does not rely on any a priori assumptions on the underlying prior and posterior distributions. The new method is based on solving an optimal transportation problem for discrete random variables.

Key words. Bayesian inference, Monte Carlo method, sequential data assimilation, linear programming, resampling

AMS subject classifications. 65C05, 62M20, 93E11, 62F15, 86A22

DOI. 10.1137/130907367

1. Introduction. This paper is concerned with a particular implementation of Monte Carlo methods for Bayesian inference and its application to filtering and intermittent data assimilation [8]. More specifically, we consider the problem of estimating posterior expectation values under the assumption that a finite-size ensemble $\{x_i^f\}_{i=1}^M$ from the (generally unknown) prior distribution π_{X^f} is available. A standard approach for obtaining such estimators relies on the idea of importance sampling based on the likelihood $\pi_Y(y_0|x_i^f)$ of the samples x_i^f with regard to a given observation y_0 [6, 2, 3]. If applied recursively, it is necessary to combine importance sampling with a resampling step such as monomial or systematic resampling [2, 9]. More recently the ensemble Kalman filter (EnKF) has been introduced [7], which transforms the prior ensemble $\{x_i^f\}_{i=1}^M$ into a uniformly weighted posterior ensemble $\{x_i^a\}_{i=1}^M$ using the classic Kalman update step of linear filtering [8]. The EnKF leads, however, to a biased estimator even in the limit $M \rightarrow \infty$ [10]. In this paper, we propose a nonrandom ensemble transform (ET) method which is based on finite-dimensional optimal transportation in form of linear programming [17, 5]. We provide numerical and theoretical evidence that the new ET method leads to consistent posterior estimators. The new transform method can be applied to intermittent data assimilation, leading to a novel implementation of particle filters. We demonstrate this possibility for the chaotic Lorenz-63 model [12].

An outline of the paper is as follows. In section 2, Monte Carlo importance sampling is summarized in the context of Bayesian inference. Subsequently, importance sampling is put into the context of linear programming in section 3. This leads to

*Submitted to the journal's Methods and Algorithms for Scientific Computing section January 25, 2013; accepted for publication (in revised form) May 13, 2013; published electronically August 6, 2013.

<http://www.siam.org/journals/sisc/35-4/90736.html>

†Institut für Mathematik, Universität Potsdam, Am Neuen Palais 10, D-14469 Potsdam, Germany, and Department of Mathematics and Statistics, University of Reading, Whiteknights, PO Box 220, Reading RG6 6AX, UK (sreich@math.uni-potsdam.de).

a novel resampling method which maximizes the correlation between the prior and posterior ensemble members. We propose a further modification which turns the resampling step into a deterministic and linear transformation. Convergence of the proposed transformation step is demonstrated numerically by means of two examples. A theoretical convergence result is formulated based on results in [13]. Finally, the application to sequential Monte Carlo methods is discussed in section 4 and a novel ET filter is proposed. Numerical results are presented for the Lorenz-63 model.

2. Bayesian inference and importance sampling. We summarize the importance sampling approach to Bayesian inference. Given a prior (or in the context of dynamic models, forecast) random variable $X^f : \Omega \rightarrow \mathbb{R}^{N_x}$, we denote its probability density function (PDF) by $\pi_{X^f}(x)$, $x \in \mathbb{R}^{N_x}$, and consider the assimilation of an observed $y_0 \in \mathbb{R}^{N_y}$ with likelihood function $\pi_Y(y|x)$. According to Bayes' theorem the analyzed, posterior PDF is given by

$$(2.1) \quad \pi_{X^a}(x|y_0) = \frac{\pi_Y(y_0|x)\pi_{X^f}(x)}{\int_{\mathbb{R}^{N_x}} \pi_Y(y_0|x)\pi_{X^f}(x)dx}.$$

Typically, the forecast random variable X^f and its PDF are not available explicitly. Instead one assumes that an ensemble of forecasts $x_i^f \in \mathbb{R}^{N_x}$, $i = 1, \dots, M$, is given, which mathematically are considered as realizations $X_i^f(\omega)$, $\omega \in \Omega$, of M independent (or dependent) random variables $X_i^f : \Omega \rightarrow \mathbb{R}^{N_x}$ with law π_{X^f} . Then the expectation value $\bar{g}^f = \mathbb{E}_{X^f}[g]$ of a function $g : \mathbb{R}^{N_x} \rightarrow \mathbb{R}$ with respect to the prior PDF $\pi_{X^f}(x)$ can be estimated according to

$$\bar{G}_M^f = \frac{1}{M} \sum_{i=1}^M g(X_i^f)$$

with realization

$$\bar{g}_M^f = \bar{G}_M^f(\omega) = \frac{1}{M} \sum_{i=1}^M g(X_i^f(\omega)) = \frac{1}{M} \sum_{i=1}^M g(x_i^f)$$

for the ensemble $\{x_i^f = X_i^f(\omega)\}_{i=1}^M$. The estimator is unbiased for any $M > 0$ and its variance vanishes as $M \rightarrow \infty$ provided the variance of g is finite.

Following the idea of importance sampling [11], one obtains the following estimator with respect to the posterior PDF $\pi_{X^a}(x|y_0)$ using the forecast ensemble:

$$\bar{g}_M^a = \sum_{i=1}^M w_i g(x_i^f)$$

with weights

$$(2.2) \quad w_i = \frac{\pi_Y(y_0|x_i^f)}{\sum_{i=1}^M \pi_Y(y_0|x_i^f)}.$$

The estimator is no longer unbiased for finite M but remains consistent. Here an estimator is called consistent if the root mean square (RMS) error between the estimator \bar{g}_M^a and the exact expectation value $\bar{g}^a = \mathbb{E}_{X^a}[g]$ vanishes as $M \rightarrow \infty$.

3. An ET method based on linear programming. Alternatively to importance sampling, we may attempt to transform the samples $x_i^f = X_i^f(\omega)$ with $X_i^f \sim \pi_{X^f}$ into samples \hat{x}_i^a which follow the posterior distribution $\pi_{X^a}(x|y_0)$. Then we are back to an estimator

$$\bar{g}_M^a = \frac{1}{M} \sum_{i=1}^M g(\hat{x}_i^a)$$

with equal weights for posterior expectation values. For univariate random variables X^f and X^a with PDFs π_{X^f} and π_{X^a} , respectively, the transformation is characterized by

$$(3.1) \quad F_{X^a}(\hat{x}_i^a) = F_{X^f}(x_i^f),$$

where F_{X^f} and F_{X^a} denote the cumulative distribution functions of X^f and X^a , respectively, e.g.,

$$F_{X^f}(x) = \int_{-\infty}^x \pi_{X^f}(x') dx'.$$

Equation (3.1) requires knowledge of the associated PDFs, and its extension to multivariate random variables is nontrivial. In this section, we propose an alternative approach that does not require explicit knowledge of the underlying PDFs and that easily generalizes to multivariate random variables. To obtain the desired transformation we utilize the idea of optimal transportation [18] with respect to an appropriate distance $d(x, x')$ in \mathbb{R}^{N_x} . More precisely, we first seek a coupling between two discrete random variables $Z^f : \Omega' \rightarrow \mathcal{Z}$ and $Z^a : \Omega' \rightarrow \mathcal{Z}$ with realizations in $\mathcal{Z} = \{x_1^f, \dots, x_M^f\}$ and probability vector $p^f = (1/M, \dots, 1/M)^T$ for Z^f and $p^a = (w_1, \dots, w_M)^T$ for Z^a , respectively. A coupling between Z^f and Z^a is an $M \times M$ matrix \mathbf{T} with non-negative entries $t_{ij} = (\mathbf{T})_{ij} \geq 0$ such that

$$(3.2) \quad \sum_{i=1}^M t_{ij} = 1/M, \quad \sum_{j=1}^M t_{ij} = w_i.$$

We now aim for the coupling \mathbf{T}^* that minimizes the expected distance

$$(3.3) \quad \mathbb{E}_{Z^f Z^a}[d(z^f, z^a)] = \sum_{i,j=1}^M t_{ij} d(x_i^f, x_j^f).$$

The desired coupling \mathbf{T}^* is characterized by a linear programming problem [17]. Since (3.2) leads to $2M - 1$ independent constraints the matrix \mathbf{T}^* contains at most $2M - 1$ nonzero entries.

In this paper, we use the squared Euclidean distance, i.e.,

$$(3.4) \quad d(x_i^f, x_j^f) = \|x_i^f - x_j^f\|^2.$$

We recall that minimizing the expected distance with respect to the squared Euclidean distance is equivalent to maximizing $\mathbb{E}_{Z^f Z^a}[(z^f)^T z^a]$ since

$$\mathbb{E}_{Z^f Z^a}[\|z^f - z^a\|^2] = \mathbb{E}_{Z^f}[\|z^f\|^2] + \mathbb{E}_{Z^a}[\|z^a\|^2] - 2\mathbb{E}_{Z^f Z^a}[\langle z^f, z^a \rangle]$$

with $\langle z^f, z^a \rangle = (z^f)^T z^a$. Furthermore, the optimal coupling \mathbf{T}^* satisfies cyclical monotonicity [19], which is defined as follows. Let S denote the support of \mathbf{T}^* , i.e., the set of all $(x_i^f, x_j^a) \in \mathcal{Z} \times \mathcal{Z}$ such that $t_{ij}^* > 0$; then

$$(3.5) \quad \langle z_1^a, z_2^f - z_1^f \rangle + \langle z_2^a, z_3^f - z_2^f \rangle + \cdots + \langle z_K^a, z_1^f - z_K^f \rangle \leq 0$$

for any set of pairs $(z_i^f, z_i^a) \in S$, $k = 1, \dots, K$, and any integer $K > 1$. Any set $S \subset \mathbb{R}^{N_x} \times \mathbb{R}^{N_x}$ with this property is called cyclically monotone [19]. Furthermore, a standard result of convex analysis states (see Theorem 2.27 in [18]) that there is a convex function $\psi : \mathbb{R}^{N_x} \rightarrow \mathbb{R}$ such that S is contained in the subdifferential of ψ , i.e.,

$$S \subset \partial\psi.$$

We next introduce the Markov chain $\mathbf{P} \in \mathbb{R}^{M \times M}$ on \mathcal{Z} via

$$\mathbf{P} = M \mathbf{T}^*$$

with the property that

$$p^a = \mathbf{P} p^f.$$

Given realizations x_j^f , $j = 1, \dots, M$, from the prior PDF, a Monte Carlo resampling step proceeds now as follows: Solve (3.3) for an optimal coupling matrix \mathbf{T}^* and define discrete random variables

$$(3.6) \quad Z_j^a \sim \begin{pmatrix} p_{1j} \\ \vdots \\ p_{Mj} \end{pmatrix}$$

for $j = 1, \dots, M$. Here p_{ij} denotes the (i, j) th entry of \mathbf{P} . A new ensemble of size M is finally obtained by collecting a single realization from each random variable Z_j^a , i.e.,

$$x_j^a := Z_j^a(\omega)$$

for $j = 1, \dots, M$. This ensemble of equally weighted samples allows for the approximation of expectation values with respect to the posterior distribution $\pi_{X^a}(x|y_0)$.

The outlined procedure leads to a particular instance of resampling with replacement [2, 9]. The main difference from techniques such as monomial or systematic resampling is that the resampling is chosen such that the expected distance (3.3) between the prior and posterior samples is minimized.

We now propose a further modification which replaces the random resampling step by a linear transformation in the prior ensemble. This transformation exactly preserves the posterior ensemble mean $\bar{x}_M^a = \sum_{i=1}^M w_i x_i^f$ and is based on the observation that

$$(3.7) \quad \bar{x}_j^a = \mathbb{E}_{Z_j^a}[z] = \sum_{i=1}^M p_{ij} x_i^f$$

should provide a sufficient representation of the random variable Z_j^a . Hence we propose the deterministic transformation

$$(3.8) \quad x_j^a := \bar{x}_j^a = \sum_{i=1}^M p_{ij} x_i^f,$$

$j = 1, \dots, M$. The hope is that

$$(3.9) \quad \bar{g}_M^a = \frac{1}{M} \sum_{j=1}^M g(\bar{x}_j^a)$$

still provides a consistent estimator for $\mathbb{E}_{X^a}[g]$ as $M \rightarrow \infty$. For the special case $g(x) = x$ it is easy to verify that indeed

$$\bar{x}_M^a = \frac{1}{M} \sum_{j=1}^M x_j^a = \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^M p_{ij} x_i^f = \sum_{i,j} t_{ij}^* x_i^f = \sum_{i=1}^M w_i x_i^f.$$

Before investigating the theoretical properties of the proposed transformation (3.8) we consider two numerical examples which indicate that (3.8) indeed leads to a consistent approximation to (3.1) in the limit $M \rightarrow \infty$.

Example. We take the univariate Gaussian with mean $\bar{x} = 1$ and variance $\sigma^2 = 2$ as prior random variable X^f . Realizations of X^f are generated using

$$x_i^f = \sqrt{2} \operatorname{erf}^{-1}(2u_i - 1), \quad u_i = \frac{1}{2M} + \frac{i-1}{M}$$

for $i = 1, \dots, M$. The likelihood function is

$$\pi_Y(y|x) = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{(y-x)^2}{4}\right)$$

with assumed observed value $y_0 = 0.1$. Bayes' formula yields a posterior distribution which is Gaussian with mean $\bar{x} = 0.55$ and variance $\sigma^2 = 1$. The prior and posterior realizations from the transform method are shown for $M = 10$ in Figure 3.1. We also display the analytic transform, which is a straight line in case of Gaussian distributions, and the approximate transform using linear programming in Figure 3.2. The

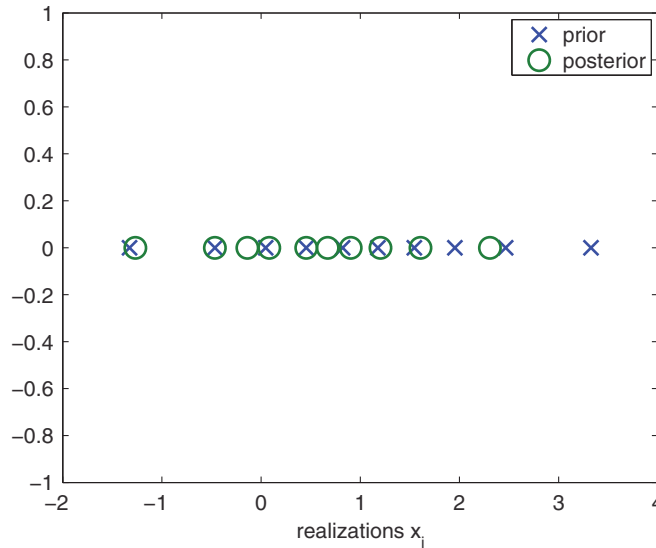


FIG. 3.1. Prior x_i^f and posterior x_i^a realizations from the transform method for $M = 10$.

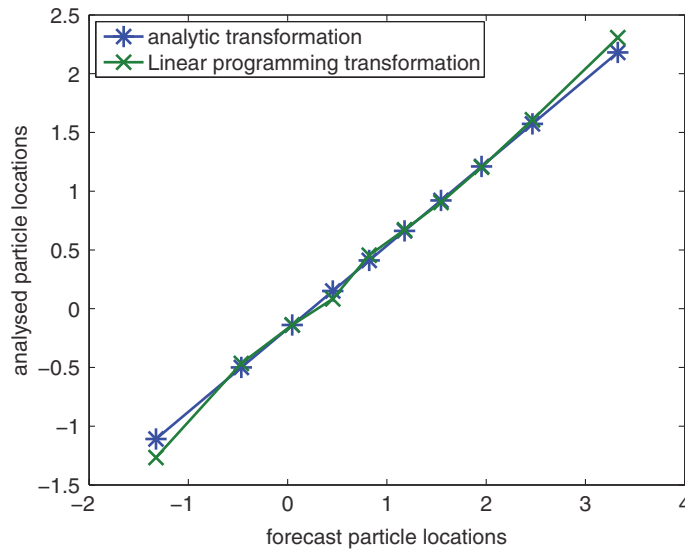


FIG. 3.2. Exact and numerical ET map for $M = 10$. The Gaussian case leads to the exact transformation being linear. The numerical approximation deviates from linearity mostly in both its tails.

structure of nonzero entries of the Markov chain matrix \mathbf{P} for $M = 40$ is displayed in Figure 3.3, which shows a banded structure of local interactions. The staircase-like arrangement is due to cyclical monotonicity of the support of \mathbf{T}^* . More generally, one obtains the posterior estimates for the first four moments displayed in Table 3.1, which indicate convergences as $M \rightarrow \infty$.

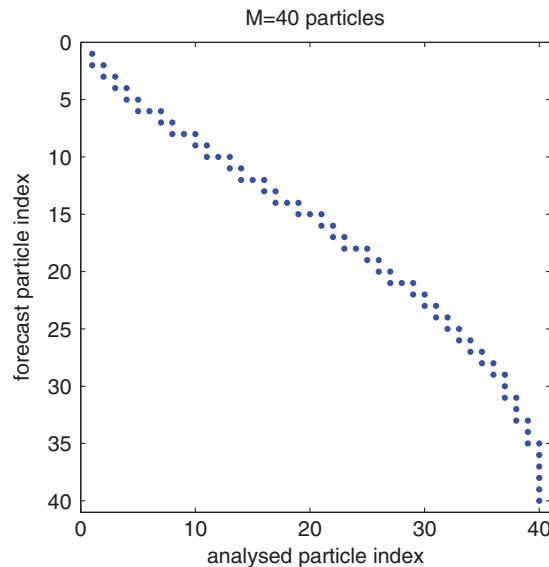


FIG. 3.3. Nonzero entries in the matrix \mathbf{P} for $M = 40$, i.e., the support of the coupling. There are a total of $2M - 1 = 79$ nonzero entries. The banded structure reveals the spatial locality and the cyclical monotonicity [18, 19] of the resampling step.

TABLE 3.1

Estimated posterior first- to fourth-order moments from the ET method applied to a Gaussian scalar Bayesian inference problem.

	\bar{x}	σ^2	$\mathbb{E}[(X - \bar{x})^3]$	$\mathbb{E}[(X - \bar{x})^4]$
$M = 10$	0.5361	1.0898	-0.0137	2.3205
$M = 40$	0.5473	1.0241	0.0058	2.7954
$M = 100$	0.5493	1.0098	-0.0037	2.9167

Example. As a further (non-Gaussian) example we consider a uniform prior on the interval $[0, 1]$ and use samples $x_i^f = u_i$ with the u_i 's as defined in the previous example. Given the observed value $y_0 = 0.1$, the posterior PDF is

$$\pi_{X^a}(x|0.1) = \begin{cases} \frac{1}{0.9427\dots} e^{-(x-0.1)^2/4}, & x \in [0, 1], \\ 0, & \text{else.} \end{cases}$$

The resulting posterior mean is $\bar{x} \approx 0.4836$ and its variance is $\sigma^2 \approx 0.0818$. The third and fourth moments are 0.0016 and 0.0122, respectively. The transform method yields the posterior estimates for the first four moments displayed in Table 3.2, which again indicate convergences as $M \rightarrow \infty$.

We now proceed with a theoretical investigation of the transformation (3.8). Our convergence result is based on the following lemma and general results from [13].

LEMMA 3.1. *The set \hat{S} consisting of all pairs (x_j^f, \bar{x}_j^a) , $j = 1, \dots, M$, with \bar{x}_j^a defined by (3.7), is cyclically monotone.*

Proof. Let $I(j)$ denote the set of indices i for which $p_{ij} > 0$. Since \mathbf{T}^* is cyclically monotone, (3.5) holds for sequences containing a term of type $\langle x_i^f, x_{j'}^f - x_j^f \rangle$ with $i \in I(j)$. By linearity of $\langle x_i^f, x_{j'}^f - x_j^f \rangle$ in each of its two arguments, (3.5) then also applies to linear combinations giving rise to

$$\begin{aligned} & \sum_{i=1}^M p_{ij} \left\{ \langle z_1^a, z_2^f - z_1^f \rangle + \langle z_2^a, z_3^f - z_2^f \rangle + \dots + \langle x_i^f, x_{j'}^f - x_j^f \rangle + \dots + \langle z_K^a, z_1^f - z_K^f \rangle \right\} \\ & = \langle z_1^a, z_2^f - z_1^f \rangle + \langle z_2^a, z_3^f - z_2^f \rangle + \dots + \langle \bar{x}_j^a, x_{j'}^f - x_j^f \rangle + \dots + \langle z_K^a, z_1^f - z_K^f \rangle \leq 0 \end{aligned}$$

since $\sum_{i=1}^M p_{ij} = 1$. The same procedure can be applied to all indices $k \in \{1, \dots, K\}$ resulting in

$$\langle \bar{z}_1^a, z_2^f - z_1^f \rangle + \langle \bar{z}_2^a, z_3^f - z_2^f \rangle + \dots + \langle \bar{z}_k^a, z_{k+1}^f - z_k^f \rangle + \dots + \langle \bar{z}_K^a, z_1^f - z_K^f \rangle \leq 0,$$

where $z_{k+1}^f := x_{j'}^f$ and $\bar{z}_k^a := \bar{x}_j^a$ if $z_k^f = x_j^f$. Hence the set \hat{S} is cyclically monotone. \square

THEOREM 3.2. *Assume that the ensemble $\mathcal{X}_M^f = \{x_i^f\}_{i=1}^M$ consists of realization from M independent and identically distributed random variables $X_i^f : \Omega \rightarrow \mathbb{R}^{N_x}$ with*

TABLE 3.2

Estimated posterior first- to fourth-order moments from the ET method applied to a non-Gaussian scalar Bayesian inference problem.

	\bar{x}	σ^2	$\mathbb{E}[(X - \bar{x})^3]$	$\mathbb{E}[(X - \bar{x})^4]$
$M = 10$	0.4838	0.0886	0.0014	0.0114
$M = 40$	0.4836	0.0838	0.0016	0.0121
$M = 100$	0.4836	0.0825	0.0016	0.0122

PDF π_{X^f} . Define the set $\mathcal{X}_M^a = \{\bar{x}_j^a\}_{j=1}^M$ with the \bar{x}_j^a 's given by (3.7). Then the associated maps $\Psi_M : \mathcal{X}_M^f \rightarrow \mathcal{X}_M^a$, defined for fixed M by

$$\bar{x}_j^a = \Psi_M(x_j^f), \quad j = 1, \dots, M,$$

converge weakly to a map $\Psi : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_x}$ for $M \rightarrow \infty$. Furthermore, the random variable defined by $X^a = \Psi(X^f)$ has distribution (2.1) and the expected distance between X^a and X^f is minimized among all such mappings.

Proof. The maps Ψ_M define a sequence of couplings between discrete random variables on $\mathcal{Z}^f = \mathcal{X}_M^f$ and $\mathcal{Z}^a = \mathcal{X}_M^a$, which satisfy cyclical monotonicity according to Lemma 3.1. We may now follow the proof of Theorem 6 in [13] to conclude that these couplings converge weakly to a continuous coupling, i.e., a probability measure $\mu_{\tilde{X}^f \tilde{X}^a}$ on $\mathbb{R}^{N_x} \times \mathbb{R}^{N_x}$ with marginals $\pi_{\tilde{X}^f}$ and $\pi_{\tilde{X}^a}$, respectively. By construction it is clear that $\pi_{\tilde{X}^f} = \pi_{X^f}$. We still need to show that $\pi_{\tilde{X}^a}(x) = \pi_{X^a}(x|y_0)$ and that the support of $\mu_{\tilde{X}^f \tilde{X}^a}$ is the graph of a map $\tilde{\Psi}$. The later property follows from the fact that $\mu_{\tilde{X}^f \tilde{X}^a}$ is cyclically monotone and that the probability measure for X^f is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^{N_x} . Hence the main theorem of [13] can be applied to guarantee the existence and uniqueness of the map $\tilde{\Psi}$. Alternatively, the coupling defined by the matrix \mathbf{T}^* also gives rise to a potential ψ_M such that

$$S \subset \partial\psi_M$$

due to cyclical monotonicity, i.e., the support S of \mathbf{T}^* is included in the subdifferential of ψ_M for all M . Since $\partial\psi_M(x)$ is a nonempty convex compact set, we also have

$$\hat{S} \subset \partial\psi_M$$

for the cyclically monotone set \hat{S} defined in Lemma 3.1. Furthermore, since \mathbf{T}^* couples π_{X^f} and $\pi_{X^a}(x|y_0)$ as $M \rightarrow \infty$ by construction, this then also implies that $\pi_{\tilde{X}^a}(x) = \pi_{X^a}(x|y_0)$. The coupling $\mu_{\tilde{X}^f \tilde{X}^a}$ solves the Monge–Kantorovitch problem with cost $c(x, y) = \|x - y\|^2$ [18, 19]. \square

Theorem 3.2 implies that (3.9) indeed provides a consistent estimator for expectation values $\mathbb{E}_{X^a}[g]$ of functions g with bounded variation. We cannot, however, expect (3.9) to be unbiased for finite ensemble sizes M .

One may replace the uniform probabilities in p^f by an appropriate random vector $p^f = (w_1^f, \dots, w_M^f)^T$, i.e., $w_i^f \geq 0$ and $\sum_{i=1}^M w_i^f = 1$. To clarify the notation we write $p^a = (w_1^a, \dots, w_M^a)^T$ for the posterior weights according to Bayes' formula. The linear programming problem (3.3) is adjusted accordingly and one obtains an optimal coupling \mathbf{T}^* and an induced Markov chain \mathbf{P} with entries

$$p_{ij} = \frac{t_{ij}}{w_j^f}.$$

Hence the transform method (3.3) is now replaced by

$$(3.10) \quad \bar{x}_j^a = \sum_{i=1}^M p_{ij} x_i^f$$

and the posterior ensemble mean satisfies

$$\bar{x}_M^a = \sum_{j=1}^M w_j^f \bar{x}_j^a = \sum_{j=1}^M \sum_{i=1}^M w_j^f \frac{t_{ij}}{w_j^f} x_i^f = \sum_{i=1}^M w_i^a x_i^f$$

as desired. More generally, posterior expectation values are given by

$$\bar{g}_M^a = \sum_{i=1}^M w_i^f g(x_i^a).$$

4. Application to sequential data assimilation. We now apply the proposed ET method (3.3) to sequential state estimation for ordinary differential equation models

$$(4.1) \quad \dot{x} = f(x)$$

with given PDF π_0 for the initial conditions $x(0) \in \mathbb{R}^{N_x}$ at time $t = 0$. Hence we treat solutions $x(t)$ as realizations of the random variables $X_t, t \geq 0$, determined by the flow of (4.1) and the initial PDF π_0 .

We assume the availability of observations $y(t_k) \in \mathbb{R}^{N_y}$ at discrete times $t_k = k\Delta t_{\text{obs}}, k > 0$, in intervals of $\Delta t_{\text{obs}} > 0$. The observations satisfy the stochastic forward model

$$Y(t_k) = h(x_{\text{ref}}(t_k)) + \Xi_k,$$

where $\Xi_k : \Omega \rightarrow \mathbb{R}^{N_y}$ represent independent and identically distributed centered Gaussian random variables with covariance matrix $R \in \mathbb{R}^{N_y \times N_y}, h : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_y}$ is the forward map, and $x_{\text{ref}}(t) \in \mathbb{R}^{N_x}$ denotes the desired reference solution. The forward model gives rise to the likelihood

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

A particle filter starts from an ensemble $\{x_i(0)\}_{i=1}^M$ of M realizations from the initial PDF π_0 . We evolve this ensemble of realizations under the model dynamics (4.1) till the first observation $y_{\text{obs}}(\Delta t_{\text{obs}})$ becomes available, at which point we apply the proposed ET method to the forecast ensemble members $x_i^f = x_i(\Delta t_{\text{obs}})$. If one furthermore collects these prior realizations into an $N_x \times M$ matrix

$$\mathbf{X}^f = [x_1^f \cdots x_M^f],$$

then, for given observation $y_0 = y(\Delta t_{\text{obs}})$, the ET method (3.10) leads to the posterior realizations simply given by

$$(4.2) \quad \mathbf{X}^a = \mathbf{X}^f \mathbf{P}, \quad [x_1^a \cdots x_M^a] = \mathbf{X}^a,$$

where \mathbf{P} is the Markov chain induced by the associated linear programming problem. The analyzed ensemble members $x_i^a, i = 1, \dots, M$, are now being used as new initial conditions for the model (4.1), and the process of alternating between propagation under model dynamics and assimilation of data is repeated for all $k > 1$.

It should be noted that a transformation similar to (4.2) arises from the ensemble square root filter (ESRF) [7]. However, the transform matrix $\mathbf{P} \in \mathbb{R}^{M \times M}$ used here is obtained in a completely different manner and does not rely on the assumption of the PDFs being Gaussian. We mention the work of [10] for an alternative approach to modify EnKFs in order to make them consistent with non-Gaussian distributions.

Ensemble inflation has become a popular technique for increasing the robustness of EnKFs [7]. Ensemble inflation simply replaces the forecast ensemble deviations $\delta x_i^f := x_i^f - \bar{x}_M^f$ by $\alpha \delta x_i^f$ with $\alpha > 1$. A related technique is particle rejuvenation

for sequential particle filters, where independent realizations ξ_i of a Gaussian with mean zero and covariance matrix $B \in \mathbb{R}^{N_x \times N_x}$ are added to the resampled ensemble members x_i^a . Particle rejuvenation also increases the ensemble spread but in addition avoids the creation of identical ensemble members under resampling. Particle rejuvenation can also be implemented for the proposed ET particle filter with the covariance matrix B determined by

$$B = \frac{\tau}{M-1} \sum_{i=1}^M (x_i^f - \bar{x}_M^f)(x_i^f - \bar{x}_M^f)^T.$$

Here $\tau > 0$ is the bandwidth parameter, which is formally related to the ensemble inflation factor α by $\alpha = \sqrt{1 + \tau}$.

We now provide a numerical example and compare an ESRF implementation with a particle filter using the new ET method.

Example. We consider the Lorenz-63 model [12]

$$\begin{aligned}\dot{x} &= \sigma(y - x), \\ \dot{y} &= x(\rho - z) - y, \\ \dot{z} &= xy - \beta z\end{aligned}$$

in the parameter and data assimilation setting of [1]. In particular, the state vector is $x = (x, y, z)^T \in \mathbb{R}^3$ and we observe all three variables every $\Delta t_{\text{obs}} = 0.12$ time units with a measurement error variance $R = 8$ in each observed solution component. The equations are integrated in time by the implicit midpoint rule with step-size $\Delta t = 0.01$. We implement an ESRF [7], a sequential importance-resampling (SIR) particle filter, and the new ET filter for ensemble sizes $M = 10, 20, 50, 75, 100, 200$, and we process a total of 2000 observations. The ESRF is furthermore implemented with optimized ensemble inflation factors α . Particle rejuvenation is used for both particle filters with $\tau = 0.5$ for the smallest ensemble size and $\tau = 0.05$ for $M = 200$. Both the SIR and the ET particle filter lead to filter divergence for $M = 10$, while the ESRF is stable for all given choices of M . The time averaged RMS errors over 2000 assimilation steps can be found in Figure 4.1. It is evident that the new ET filter

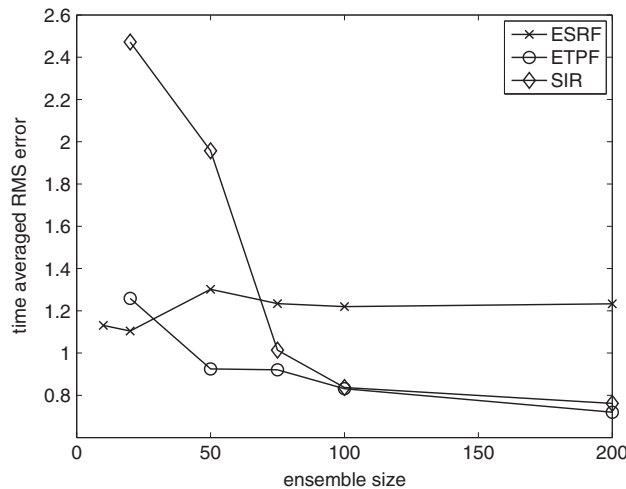


FIG. 4.1. Time averaged RMS errors for the Lorenz-63 model in the setting of [1] for an ESRF, the SIR particle filter, and the new ET filter for increasing ensemble sizes M .

leads to much lower RMS errors for $M \geq 50$ compared to the ESRF. The behavior of the SIR and the ET particle filters is similar for $M \geq 100$. However, for $M = 20$ and $M = 50$, the ET particle filter clearly outperforms the SIR filter. We may conclude that the ET particle filter might provide a viable alternative to both the ESRF and the SIR filter in a range of intermediate ensemble sizes. Our results also compare favorably to the ones displayed in [1] for the rank histogram filter [1] and the EnKF with perturbed observations [7].

5. Conclusions. We have explored the application of linear programming and optimal transportation to Bayesian inference and particle filters. We have demonstrated theoretically as well as numerically that the proposed ET method allows us to reproduce posterior expectation values in the limit $M \rightarrow \infty$ and a convergence to the associated continuum optimal transport problem [18, 19]. The application of continuous optimal transportation to Bayesian inference has been discussed by [14], [15, 16], [5]. However, a direct application of continuous optimal transportation to Bayesian inference in high-dimensional state spaces \mathbb{R}^{N_x} seems currently out of reach and efficient numerical techniques need to be developed.

Our numerical experiments indicate that the ET particle filter might be able to fill the performance gap between EnKFs and SIR particle filters as ensemble sizes increase from relatively small to relatively large. It remains to investigate what modifications are required (such as localization [7]) in order to implement the proposed ET method even if the ensemble sizes M are much smaller than the dimension of state space N_x (or the dimension of the attractor of (4.1) in the case of intermittent data assimilation).

A standard MATLAB implementation of the simplex algorithm was used for solving the linear programming problems in this paper. More efficient algorithms such as the auction algorithm [4] should be considered in future implementations of the ET method (3.3).

We finally mention that the ET particle filter can be extended from the classic importance sampling considered in this paper to Bayesian inference for weighted Gaussian mixtures as prior distributions.

Acknowledgments. I would like to thank Yann Brenier and Jacques Vanneste for inspiring discussions on the subject of this paper during an Oberwolfach workshop in 2010. The paper benefited furthermore from additional discussions with Wilhelm Stannat and Dan Crisan at another Oberwolfach workshop in 2012.

REFERENCES

- [1] J. ANDERSON, *A non-Gaussian ensemble filter update for data assimilation*, Monthly Weather Rev., 138 (2010), pp. 4186–4198.
- [2] M. ARULAMPALAM, S. MASKELL, N. GORDON, AND T. CLAPP, *A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking*, IEEE Trans. Signal Process., 50 (2002), pp. 174–188.
- [3] A. BAIN AND D. CRISAN, *Fundamentals of Stochastic Filtering*, Stoch. Model. Appl. Probab., Springer-Verlag, New York, 2008.
- [4] D. BERTSEKAS AND D. CASTANON, *The auction algorithm for the transportation problem*, Ann. Oper. Res., 20 (1989), pp. 67–96.
- [5] C. COTTER AND S. REICH, *Ensemble filter techniques for intermittent data assimilation*, Large Scale Inverse Problems. Computational Methods and Applications in the Earth Sciences, Radon Ser. Comput. Appl. Math., Walter de Gruyter, Berlin, 13 (2013), pp. 91–134.
- [6] A. DOUCET, N. DE FREITAS, AND N. GORDON, EDs., *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001.
- [7] G. EVENSEN, *Data Assimilation. The Ensemble Kalman Filter*, Springer-Verlag, New York, 2006.

- [8] A. JAZWINSKI, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
- [9] H. KÜNSCH, *Recursive Monte Carlo Filter: Algorithms and Theoretical Analysis*, Ann. Statist., 33 (2005), pp. 1983–2021.
- [10] J. LEI AND P. BICKEL, *A moment matching ensemble filter for nonlinear and non-Gaussian data assimilation*, Monthly Weather Rev., 139 (2011), pp. 3964–3973.
- [11] J. LIU, *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag, New York, 2001.
- [12] E. LORENZ, *Deterministic non-periodic flows*, J. Atmospheric Sci., 20 (1963), pp. 130–141.
- [13] R. MCCANN, *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J., 80 (1995), pp. 309–323.
- [14] T. E. MOSELHY AND Y. MARZOUK, *Bayesian inference with optimal maps*, J. Comput. Phys., 231 (2012), pp. 7815–7850.
- [15] S. REICH, *A dynamical systems framework for intermittent data assimilation*, BIT, 51 (2011), pp. 235–249.
- [16] S. REICH, *A Gaussian mixture ensemble transform filter*, Quart. J. Roy. Meteorol. Soc., 138 (2012), pp. 222–233.
- [17] G. STRANG, *Introduction to Applied Mathematics*, 2nd ed., Wellesley Cambridge Press, Wellesley, MA, 1986.
- [18] C. VILLANI, *Topics in Optimal Transportation*, AMS, Providence, RI, 2003.
- [19] C. VILLANI, *Optimal Transportation: Old and New*, Springer-Verlag, Heidelberg, 2009.