A Gaussian mixture ensemble transform filter for vector observations

Santosh Nannuru[a], Mark Coates[a] and Arnaud Doucet[b]

[a]McGill University, Montreal, Canada
[b]Oxford University, Oxford, UK

ABSTRACT

The ensemble Kalman filter relies on a Gaussian approximation being a reasonably accurate representation of the filtering distribution. Reich recently introduced a Gaussian mixture ensemble transform filter which can address scenarios where the prior can be modeled using a Gaussian mixture. Reichs derivation is suitable for a scalar measurement or a vector of uncorrelated measurements. We extend the derivation to the case of vector observations with arbitrary correlations. We illustrate through numerical simulation that implementation is challenging, because the filter is prone to instability.

Keywords: Ensemble transform filter, ensemble Kalman filter, differential equations

1. INTRODUCTION

Monte Carlo methods have been extensively used for solving the problem of state estimation in dynamical systems with noisy observations. This problem is referred to as the data assimilation or filtering problem. One of the most commonly studied Monte Carlo methods is the sequential Monte Carlo (SMC) filter, also called the particle filter [1]. In the SMC filter the system state density is approximated using a weighted particle measure. The sampling-importance-resampling principle is used to propagate the particles and weights over time as new observations are made available [1].

Another class of Monte Carlo methods, studied and applied in the geophysical and meteorological literature, are the ensemble transform methods [2–5] . They differ from the SMC methods in the way state density is approximated. A collection or ensemble of particles with equal weights is used approximate the density. Propagation of density is achieved by transporting or moving the particles appropriately to reflect the posterior state density. Many ensemble methods have been proposed in the literature with different underlying approximations and particle transportation techniques.

The ensemble transform methods have several advantages over SMC methods in certain application areas. Since the weight update stage in SMC filter is replaced by particle transport in ensemble method, there is no need for the particle resampling step and weight degeneracy is avoided. In problems with very high state dimensions such as weather prediction, SMC methods are not efficient because they require very large number of particles. Ensemble methods can work with fewer particles even if the state dimension is large.

The ensemble Kalman filter (EnKF) proposed by Evensen in [2] propagates the particle ensemble based on the Kalman filter equations. The empirical mean and covariance matrix computed from the ensemble converges to the mean and covariance matrix given by Kalman filter in the limit when the number of particles is increased to infinity. The ensemble represents samples from the true posterior distribution if the system dynamics are linear and the prior and posterior densities are Gaussian. Formulations of the EnKF which do not use perturbed observations have also been proposed and are called ensemble square root filters [3]. The continuous time formulation of the EnKF was introduced by Bergemann et al. in [4, 6] and is called the ensemble Kalman-Bucy filter. A good survey of the different ensemble filters is provided in the recent paper by Cotter et al. in [5].

Though the EnKF has been successfully used in many applications, its performance degrades when the system dynamics are highly non-linear or the densities involved are non-Gaussian. To handle problems with non-Gaussian densities extensions of EnKF have been proposed which propagate the particle ensemble under the assumptions of Gaussian mixture densities [7–9]. The ensemble Gaussian mixture filter (EGMF) [9] was developed by Reich based on the continuous time formulation of the analysis step in Bayes filter. This formulation models the

transition from the prior density to the posterior density using a continuity equation in artificial time [10]. The ensemble update is performed by numerically solving a differential equation over the artificial time parameter.

Reichs derivation of the EGMF is suitable for the case of scalar measurements or uncorrelated vector measurements, i.e., when the measurement noise covariance matrix is diagonal. In this paper we extend the derivation to include vector measurements with arbitrary correlations. This requires solving a general second order elliptical partial differential equation (PDE). Using an appropriate transformation of variables, this elliptical PDE is simplified to a system of second order PDE's which can be easily solved.

We derive the filter equations under the assumptions of linear system and observation model and Gaussian noise. Under these assumptions and using Gaussian mixture prior, the posterior density is also a Gaussian mixture. In this setting, a mixture of Kalman filters could be used for optimal data assimilation by propagating the mean and covariance matrices of each component and its weight. The ensemble Gaussian filter has some potential advantages compared to this approach; since it uses particles rather than an explicit mixture of Gaussians, it can more readily approximate non-linear state dynamics. A mixture of Kalman filters can also struggle when the state dimension is high (> 15); the particle representation has the potential to mitigate this limitation, although as we discuss later, successful implementation in higher dimensions remains challenging.

Our current work has similarities to the particle flow filters introduced by Daum and Huang [11, 12]. These filters are also based on the concept of transporting the particles from a prior density to the posterior density without any associated particle weights. In Daum-Huang filters a log-homotopy function is introduced between the prior and the posterior densities. This leads to an ordinary differential equation (ODE) defining the flow of particles. Numerical solution of the ODE migrates the particles from the prior to the posterior. Many different ODE formulations have been proposed by considering different underlying assumptions. Some of the recent particle flow algorithms are Coulomb's law particle flow [13] and small curvature particle flow [14]. In [15], Duam and Huang have provided a comparison of their particle flow methodology and the optimal transport approach used in this paper. As they highlight, the particle flow methods are more mature for dynamic state estimation and can already successfully address high-dimensional systems. Nevertheless, there are differences between the methods and there is value in trying to extend the applicability of the optimal transport approaches and to improve their implementation.

The paper is organized as follows. Section 2 provides an introduction to the filtering problem with intermittent observations and its continuous time formulation. The ensemble Gaussian mixture filter is discussed in Section 3. A brief overview of the derivation by Reich, valid when observations are scalars or uncorrelated vectors, is presented. Its extension to vector observations with arbitrary correlations is discussed next. Numerical simulations to demonstrate the proposed extension is provided in Section 4. We finally conclude with our observations and suggest directions for future research in Section 5.

2. PROBLEM FORMULATION

In this section we provide a mathematical background for the filtering problem with discrete time observations. The approach by Reich [10] is then applied to derive a continuous time formulation of the data assimilation step in artificial time $s \in [0, 1]$.

Consider a dynamical system with the discrete time state evolution given by the equation

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \mathbf{V}_n \tag{1}$$

where $\mathbf{X}_n \in \mathbb{R}^N$ represents the system state at time n, and $\mathbf{V}_n \sim \mathcal{N}(0, \mathbf{R}_X)$, $\mathbf{R}_X \in \mathbb{R}^{N\times N}$ are i.i.d. Gaussian random vectors independent of the state $\mathbf{X}_n$. Let the state evolution model be denoted $p_{n+1|n}(\mathbf{X}_{n+1}|\mathbf{X}_n)$. The prior state distribution at time $n = 0$ is given and is assumed to be an L component Gaussian mixture of the form

$$\pi_0(\mathbf{x}) = \sum_{l=1}^{L} \alpha_l^0 \cdot \pi_{l,0}(\mathbf{x}) \tag{2}$$

where $\pi_{l,0}(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_l^0, \mathbf{P}_l^0)$ with $\boldsymbol{\mu}_l^0 \in \mathbb{R}^N$ and $\mathbf{P}_l^0 \in \mathbb{R}^{N\times N}$ and $\sum_{l=1}^{L} \alpha_l^0 = 1$. Hence the state distribution at any time n is also an L component Gaussian mixture.

Though we know the state evolution model, the true system state is hidden and we have measurements of it available at each discrete time step n. The measurements are related to the state by

$$\mathbf{Y}_n = h(\mathbf{X}_n) + \mathbf{W}_n = \mathbf{H}\mathbf{X}_n + \mathbf{W}_n \tag{3}$$

where $\mathbf{Y}_n \in \mathbb{R}^M$ is the observation at time n, $\mathbf{W}_n \sim \mathcal{N}(0, \mathbf{R}_Y)$, $\mathbf{R}_Y \in \mathbb{R}^{M\times M}$ is the observation noise. The observation map $h()$ is assumed to be linear, i.e., $h(\mathbf{X}_n) = \mathbf{H}\mathbf{X}_n$ and $\mathbf{H} \in \mathbb{R}^{M\times N}$. Let the observation vector at time n be denoted by $\mathbf{y}_{n,obs}$ and the collection of observations up to time n be denoted $\mathbf{Y}_{obs}^{[n]} = [\mathbf{y}_{1,obs}, \mathbf{y}_{2,obs}, \ldots, \mathbf{y}_{n,obs}]$. The filtering problem at time n is to obtain the posterior state density given all observations up to time n, i.e., $\pi_n(\mathbf{x}|\mathbf{Y}_{obs}^{[n]})$. The posterior density at time $n-1$ and the system evolution model is used to obtain the predicted density at time n. This is called the prediction step or the forecast. The posterior density at time n is then obtained using the predicted density, the observation model and the observation at time n. This is called the update step or the analysis step. This recursive propagation over time is the Bayes recursive filter. The update step is implemented using the Bayes rule. Since the prior at time 0 is a Gaussian mixture, and the system and observation models are linear, the predicted and the posterior distributions at every time step n will also be Gaussian mixture distributions. Let the predicted density at time n be denoted as $\pi_{n|n-1}(\mathbf{x})$ and the posterior density at time n be given as

$$\pi_n(\mathbf{x}) = \sum_{l=1}^{L} \alpha_l^n \pi_{l,n}(\mathbf{x}) \tag{4}$$

where $\pi_{l,n}(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_l^n, \mathbf{P}_l^n)$ and $\sum_{l=1}^{L} \alpha_l^n = 1$. This posterior is approximated using a collection of particles with equal weights in the ensemble method.

2.1 Continuous time formulation

Bayes' rule is used to obtain the posterior density from the predicted density and is given by

$$\pi_n(\mathbf{x}) = \frac{f(\mathbf{y}_{n,obs}|\mathbf{x}) \cdot \pi_{n|n-1}(\mathbf{x})}{f(\mathbf{y}_{n,obs})} \tag{5}$$

where $f(\mathbf{y}_{n,obs}|\cdot)$ is the likelihood function. In traditional discrete-time filters, the transformation of density, from predicted to posterior, by application of Bayes' rule happens in a single step. The continuous time formulation of Bayes' rule by Reich [10] transforms the predicted density to the posterior density in a continuous fashion. This is achieved by use of the following result from statistical mechanics. If the system state obeys the time flow equation

$$\frac{d\mathbf{x}}{ds} = g(\mathbf{x}, s) \tag{6}$$

in the time interval $s \in [0, 1]$, then the evolution of the density $\pi(\mathbf{x}, s)$ of $\mathbf{x}$ over time s is given by the continuity equation or Liouville's equation:

$$\frac{\partial \pi}{\partial s} = -\nabla_{\mathbf{x}} \cdot (\pi g) \tag{7}$$

where the prior density at time $s = 0$, $\pi(\mathbf{x}, 0)$ is assumed known. Using the formulation by Reich [10] we can set $\pi(\mathbf{x}, 0) = \pi_{n|n-1}(\mathbf{x})$ and obtain $\pi(\mathbf{x}, 1) = \pi_n(\mathbf{x})$ by using the relation,

$$\frac{\partial \pi}{\partial s} = -\pi(S - \mathbb{E}_\pi[S]) \tag{8}$$

where S is the negative log-likelihood function, $S = -\log(f(\mathbf{y}_{n,obs}|\mathbf{x}))$. For the case of Gaussian observation noise we have $S = \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y}_{n,obs})^T \mathbf{R}_Y^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}_{n,obs})$. From equations (7) and (8)

$$\nabla_{\mathbf{x}} \cdot (\pi g) = \pi(S - \mathbb{E}_\pi[S]) \tag{9}$$

Solving for $g(\mathbf{x}, s)$ from the partial differential equation in (9) provides a way to numerically solve for $\mathbf{x}(s)$ using the ordinary differential equation in (6). The solution to equation (9) is not unique and we must make appropriate choices/assumptions to solve for $g()$. For example, we can impose the additional constraint that g is also the minimizer of the kinetic energy defined as

$$\mathcal{T}(v) = \frac{1}{2}\int v^T \mathbf{M}\, v\, d\pi \tag{10}$$

where $\mathbf{M} \in \mathbb{R}^{N\times N}$ is a positive definite matrix. Under such constraint it can be shown [16] that the flow function has the form $g = \mathbf{M}^{-1}\nabla_{\mathbf{x}}\psi$ where the potential $\psi(\mathbf{x}, s)$ is the solution of the elliptic PDE,

$$\nabla_{\mathbf{x}}\cdot(\pi\mathbf{M}^{-1}\nabla_{\mathbf{x}}\psi) = \pi(S - \mathbb{E}_{\pi}[S]) \tag{11}$$

In the simple case of a single component Gaussian mixture ($L = 1$) where the prior is distributed as $\mathcal{N}(\boldsymbol{\mu}(0), \mathbf{P}(0))$ and $\mathbf{M}^{-1} = \mathbf{P}(s)$ it can be shown [4, 6] that

$$\frac{d\mathbf{x}}{ds} = -\frac{1}{2}\mathbf{P}(s)\mathbf{H}^T\mathbf{R}_Y(\mathbf{H}\mathbf{x} + \mathbf{H}\boldsymbol{\mu}(s) - 2\mathbf{y}_{obs}) \tag{12}$$

3. ENSEMBLE GAUSSIAN MIXTURE FILTER

The continuous formulation of Bayes' rule with an artificial time parameter s allows us to develop the ensemble filter for the case of Gaussian mixture densities. This filter was derived by Reich in [9] for the case when observations are either scalars or uncorrelated vectors. We will now briefly discuss the derivation by Reich and then extend the filter to account for observation vectors with arbitrary correlations.

In case of a general L component Gaussian mixture density, consider the following decomposition of the flow vector field g:

$$\frac{d\mathbf{x}}{ds} = g(\mathbf{x}, s) = u_A(\mathbf{x}, s) + u_B(\mathbf{x}, s) \tag{13}$$

and define

$$u_A(\mathbf{x}) = \sum_{l=1}^{L}\frac{\alpha_l\pi_l(\mathbf{x})}{\pi(\mathbf{x})}\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{A,l}(\mathbf{x}) \tag{14}$$

$$u_B(\mathbf{x}) = \sum_{l=1}^{L}\frac{\alpha_l\pi_l(\mathbf{x})}{\pi(\mathbf{x})}\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x}) \tag{15}$$

where we have dropped the artificial time parameter s for brevity. Substituting these expressions into equation (9) we have:

$$\begin{aligned}\nabla_{\mathbf{x}}\cdot\left(\sum_{l=1}^{L}\alpha_l\pi_l(\mathbf{x})\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{A,l}(\mathbf{x}) + \sum_{l=1}^{L}\alpha_l\pi_l(\mathbf{x})\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x})\right) &= \sum_{l=1}^{L}\alpha_l\pi_l(\mathbf{x})(S - \mathbb{E}_{\pi}[S]) \\ &= \sum_{l=1}^{L}\alpha_l\pi_l(\mathbf{x})(S - \mathbb{E}_{\pi_l}[S]) + \sum_{l=1}^{L}\alpha_l\pi_l(\mathbf{x})(\mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S])\end{aligned} \tag{16}$$

Equating the individual components in the equation above we have the following set of equations

$$\nabla_{\mathbf{x}}\cdot\{\pi_l(\mathbf{x})\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{A,l}(\mathbf{x})\} = \pi_l(\mathbf{x})(S - \mathbb{E}_{\pi_l}[S]), \quad l = 1, 2, \ldots, L \tag{17}$$

$$\nabla_{\mathbf{x}}\cdot\{\pi_l(\mathbf{x})\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x})\} = \pi_l(\mathbf{x})(\mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S]), \quad l = 1, 2, \ldots, L \tag{18}$$

The equations in (17) are similar to the case of a single component Gaussian and hence from (12) and (14) we have the solution:

$$u_A(\mathbf{x}, s) = -\frac{1}{2}\sum_{l=1}^{L}\frac{\alpha_l(s)\pi_l(\mathbf{x}, s)}{\pi(\mathbf{x}, s)}\mathbf{P}_l(s)\mathbf{H}^T\mathbf{R}_Y^{-1}[\mathbf{H}\mathbf{x} + \mathbf{H}\boldsymbol{\mu}_l(s) - 2\mathbf{y}_{n,obs}] \tag{19}$$

The equation (18) needs to be solved to have the complete solution. We first summarize the solution to equation (18) for the case of scalar observations. It can be easily adapted to uncorrelated vector observations as well.

3.1 Scalar observations

When the observations are scalars, i.e., $y_{n,obs} \in \mathbb{R}$, the PDE (18) can be simplified and explicitly solved [9]. To derive this we assume the potential $\psi_{B,l}$ to be of the following form

$$\psi_{B,l}(\mathbf{x}) = \widehat{\psi}_{B,l}(\mathbf{H}\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_l) = \widehat{\psi}_{B,l}(y - y_l) \tag{20}$$

where $y = \mathbf{H}\mathbf{x}$ and $y_l = \mathbf{H}\boldsymbol{\mu}_l$. Hence we have $\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x}) = \mathbf{H}^T \frac{d\widehat{\psi}_{B,l}}{dy}(y - y_l)$. Thus the PDE in equation (18) simplifies to

$$-(y - y_l)\frac{d\widehat{\psi}_{B,l}}{dy}(y - y_l) + \mathbf{H}\mathbf{P}_l\mathbf{H}^T \frac{d^2\widehat{\psi}_{B,l}}{dy^2}(y - y_l) = \mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S] \tag{21}$$

Under the initial condition $\frac{d\widehat{\psi}_{B,l}}{dy}(y - y_l)|_{y=y_l} = 0$, we can solve the above differential equation to obtain

$$\frac{d\widehat{\psi}_{B,l}}{dy}(y - y_l) = \frac{1}{2}\frac{\mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S]}{\mathbf{H}\mathbf{P}_l\mathbf{H}^T}\frac{\mathrm{Erf}((y - y_l)/\sqrt{2\sigma_l^2})}{\pi_l(y)} \tag{22}$$

where the Gaussian PDF $\pi_l(y) = \mathcal{N}(y_l, \sigma_l^2)$ with $\sigma_l^2 = \mathbf{H}\mathbf{P}_l\mathbf{H}^T$ and $\mathrm{Erf}(\cdot)$ is the standard error function. Thus we have the following expression for $u_B(\mathbf{x}, s)$:

$$u_B(\mathbf{x}, s) = \frac{1}{2}\sum_{l=1}^{L}\frac{\alpha_l(s)\pi_l(\mathbf{x}, s)}{\pi(\mathbf{x}, s)}\mathbf{P}_l(s)\mathbf{H}^T\frac{\mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S]}{\mathbf{H}\mathbf{P}_l\mathbf{H}^T}\frac{\mathrm{Erf}((y - y_l)/\sqrt{2\sigma_l^2})}{\pi_l(y)} \tag{23}$$

For the scalar observation case with Gaussian noise variance of R_y we have $S = \frac{1}{2R_y}(\mathbf{H}\mathbf{x} - y_{n,obs})^2$. Thus the expectation of S can be shown to be

$$\mathbb{E}_{\pi_l}[S] = \frac{1}{2R_y}((y_{n,obs} - y_l)^2 + \sigma_l^2) \tag{24}$$

3.2 Extension to vector observations

For the general case of correlated vector observations the above technique cannot be applied to reduce the PDE. To simplify the problem we perform a transformation of the state $\mathbf{x}$. We can expand equation (18) as

$$\{\nabla_{\mathbf{x}}\pi_l(\mathbf{x})\}\cdot\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x}) + \pi_l(\mathbf{x})\nabla_{\mathbf{x}}\cdot\{\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x})\} = \pi_l(\mathbf{x})(\mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S]) \tag{25}$$

$$-\pi_l(\mathbf{x})(\mathbf{x} - \boldsymbol{\mu}_l)^T\mathbf{P}_l^{-1}\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x}) + \pi_l(\mathbf{x})\nabla_{\mathbf{x}}\cdot\{\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x})\} = \pi_l(\mathbf{x})(\mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S]) \tag{26}$$

$$-(\mathbf{x} - \boldsymbol{\mu}_l)^T\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x}) + \nabla_{\mathbf{x}}\cdot\{\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x})\} = \mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S] \tag{27}$$

This is a second order PDE with all the cross terms present in general. To simplify the problem we employ the following transformation of variables. Let $\mathbf{x} = \mathbf{Q}_l\mathbf{z}$ where $\mathbf{P}_l = \mathbf{Q}_l\mathbf{Q}_l^T$ is the Cholesky decomposition. This always exists since $\mathbf{P}_l$ is a positive definite matrix. The matrix $\mathbf{Q}_l$ is a lower triangular matrix with positive diagonal entries and is invertible. Hence we also have the inverse relation $\mathbf{z} = \mathbf{Q}_l^{-1}\mathbf{x}$ and we define $\boldsymbol{\gamma}_l = \mathbf{Q}_l^{-1}\boldsymbol{\mu}_l$. Let the transformation induce the function $\phi_l(\mathbf{z}) = \psi_{B,l}(\mathbf{x})|_{\mathbf{x}=\mathbf{Q}_l\mathbf{z}}$ in the variable $\mathbf{z}$ and we have the following relations:

$$\mathbf{Q}_l^T\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x}) = \nabla_{\mathbf{z}}\phi_l(\mathbf{z}) \tag{28}$$

$$\nabla_{\mathbf{x}}\cdot\{\mathbf{P}_l\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x})\} = \nabla_{\mathbf{z}}^2\phi_l(\mathbf{z}) \tag{29}$$

Thus equation (27) becomes

$$-(\mathbf{z} - \boldsymbol{\gamma}_l)^T\nabla_{\mathbf{z}}\phi_l(\mathbf{z}) + \nabla_{\mathbf{z}}^2\phi_l(\mathbf{z}) = \mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S] \tag{30}$$

If the vector $\mathbf{z} = (z_1, z_2, \ldots, z_N)$ and $\boldsymbol{\gamma}_l = (\gamma_{l,1}, \gamma_{l,2}, \ldots, \gamma_{l,N})$, then the above equation can be written as

$$-\sum_{i=1}^{N}(z_i - \gamma_{l,i})\frac{\partial \phi_l}{\partial z_i} + \sum_{i=1}^{N}\frac{\partial^2 \phi_l}{\partial z_i^2} = \mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S] \tag{31}$$

$$-\sum_{i=1}^{N}\{(z_i - \gamma_{l,i})\frac{\partial \phi_l}{\partial z_i} + \frac{\partial^2 \phi_l}{\partial z_i^2}\} = \mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S] = \sum_{i=1}^{N} C_{l,i} \tag{32}$$

$$-(z_i - \gamma_{l,i})\frac{\partial \phi_l}{\partial z_i} + \frac{\partial^2 \phi_l}{\partial z_i^2} = C_{l,i} \quad i = 1, 2...N \tag{33}$$

where $C_{l,i}$ are constants satisfying $\sum_{i=1}^{N} C_{l,i} = \mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S]$. If a solution $\phi_l(\mathbf{z})$ satisfies the system of PDE's given in (33), it also satisfies the PDE in equation (32). It can be seen that the individual equations in (33) are similar to equation (21) obtained in the scalar observation case. If we assume that $\frac{\partial \phi_l}{\partial z_i} = f_i(z_i)$, i.e., the individual partial derivatives of ϕ with respect to z_i are functions of only the variable z_i, then we can solve the individual equations in the above system independently. With the initial condition set to $f_i(z_i = \gamma_{l,i}) = 0$, we obtain the solution:

$$f_i(z_i) = \frac{\partial \phi_l}{\partial z_i} = \frac{1}{2}\frac{C_{l,i}}{\pi_s(z_i - \gamma_{l,i})}\left(\text{Erf}\left[\frac{(z_i - \gamma_{l,i})}{\sqrt{2}}\right]\right) \quad i = 1, 2...N \tag{34}$$

where π_s is the standard normal density function. To simplify the above expression we define the vectors $\mathbf{C}_l$ and $\boldsymbol{\kappa}_l$ as $\mathbf{C}_l = [C_{l,1}, C_{l,2}, \ldots, C_{l,N}]^T$ and $\boldsymbol{\kappa}_l = [\kappa_{l,1}, \kappa_{l,2}, \ldots, \kappa_{l,N}]^T$ where

$$\kappa_{l,i}(z_i - \gamma_{l,i}) = \frac{1}{\pi_s(z_i - \gamma_{l,i})}\text{Erf}\left[\frac{(z_i - \gamma_{l,i})}{\sqrt{2}}\right] \quad i = 1, 2...N \tag{35}$$

$$\text{and we have,} \quad \nabla_{\mathbf{z}}\phi_l(\mathbf{z}) = \frac{1}{2}\mathbf{C}_l * \boldsymbol{\kappa}_l(\mathbf{z} - \boldsymbol{\gamma}_l) \tag{36}$$

where $\mathbf{a}*\mathbf{b}$ denotes an element-wise product of vectors $\mathbf{a}$ and $\mathbf{b}$. The final solution is completed from equation (15) by observing that $\nabla_{\mathbf{x}}\psi_{B,l}(\mathbf{x}) = (\mathbf{Q}_l^T)^{-1}\nabla_{\mathbf{z}}\phi_l(\mathbf{z})|_{\mathbf{z}=\mathbf{Q}_l^{-1}\mathbf{x}}$. Hence we can write

$$\frac{d\mathbf{x}}{ds} = g(\mathbf{x}, s) = u_A(\mathbf{x}, s) + u_B(\mathbf{x}, s) \tag{37}$$

where,

$$u_A(\mathbf{x}, s) = -\frac{1}{2}\sum_{l=1}^{L}\frac{\alpha_l(s)\pi_l(\mathbf{x}, s)}{\pi(\mathbf{x}, s)}\mathbf{P}_l(s)\mathbf{H}^T\mathbf{R}_Y^{-1}[\mathbf{H}\mathbf{x} + \mathbf{H}\boldsymbol{\mu}_l(s) - 2\mathbf{y}_{n,obs}] \tag{38}$$

$$u_B(\mathbf{x}, s) = \frac{1}{2}\sum_{l=1}^{L}\frac{\alpha_l(s)\pi_l(\mathbf{x}, s)}{\pi(\mathbf{x}, s)}\mathbf{Q}_l(s)\mathbf{C}_l * \boldsymbol{\kappa}_l\Big(\mathbf{Q}_l(s)^{-1}[\mathbf{x} - \boldsymbol{\mu}_l(s)]\Big) \tag{39}$$

$$\text{and} \quad \sum_{i=1}^{N} C_{l,i} = \mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S] \tag{40}$$

For the case of vector observations with Gaussian noise covariance matrix $\mathbf{R}_Y$ we have

$$S = \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y}_{n,obs})^T\mathbf{R}_Y^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}_{n,obs}), \quad \text{and hence} \tag{41}$$

$$\mathbb{E}_{\pi_l}[S] = \frac{1}{2}\Big(tr\{\mathbf{H}^T\mathbf{R}_Y^{-1}\mathbf{H}(\mathbf{P}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T)\} - 2\mathbf{y}_{n,obs}^T\mathbf{R}_Y^{-1}\mathbf{H}\boldsymbol{\mu}_l + \mathbf{y}_{n,obs}^T\mathbf{R}_Y^{-1}\mathbf{y}_{n,obs}\Big) \tag{42}$$

where the trace of a matrix $tr\{A\}$ is defined as the sum of all the diagonal elements of matrix A. To compute $\mathbb{E}_{\pi}[S]$, we use the relation

$$\mathbb{E}_{\pi}[S] = \sum_{l=1}^{L}\alpha_l(s)\mathbb{E}_{\pi_l}[S]. \tag{43}$$

3.3 Implementation

In this section we present an algorithm implementation of the EGMF for the general case. Algorithm 1 provides pseudocode for the implementation. The filter is initialized by drawing samples from the known prior density π_0 which is an L component Gaussian mixture. The particles are propagated following the state evolution model $p_{n|n-1}(\cdot|\cdot)$ in the prediction step. In the update step particles are transported from the predicted density to the posterior density by numerically solving the differential equation (37).

The forward Euler method is used to numerically approximate the differential equation over the artificial time $s \in [0, 1]$. This time interval is divided into N_s equal intervals each of length $\Delta s = \frac{1}{N_s}$. The ensemble particles are propagated between the artificial time steps $s = i\Delta s, i = 0, 1, \ldots, N_s$ using the approximation

$$\mathbf{x}_{n,s}^{(j)} = \mathbf{x}_{n,s-\Delta s}^{(j)} + \Delta s\bigg(u_A(\mathbf{x}_{n,s-\Delta s}^{(j)}, s-\Delta s) + u_B(\mathbf{x}_{n,s-\Delta s}^{(j)}, s-\Delta s)\bigg), \quad s = i\Delta s, \;\; i = 1, \ldots, N_s \tag{44}$$

The larger the value of N_s, the closer the discrete numerical approximation is to the true solution of the continuous differential equation. The expectation maximization (EM) algorithm is used to fit a Gaussian mixture model with L components to the particle set at each artificial time step $i\Delta s$. The weights of each of the mixture component and its mean and covariance matrix parameters are calculated using the EM algorithm. These parameters are used in the expressions (38) and (39) to calculate the components u_A and u_B. The constants $C_{l,i}$ need to be specified to evaluate u_B. In our implementation we have used

$$C_{l,i} = \frac{1}{N}(\mathbb{E}_{\pi_l}[S] - \mathbb{E}_{\pi}[S]), \quad i = 1, \ldots, N \tag{45}$$

The expectations $\mathbb{E}_{\pi_l}[S]$ and $\mathbb{E}_{\pi}[S]$ are computed using equations (42) and (43).

Initialize an ensemble of N_p particles $\{\mathbf{x}_0^{(j)}\}_{j=1}^{N_p} \sim \pi_0(\mathbf{x})$
for $n = 1$ to T **do**
    Prediction
    for $j = 1$ to N_p **do**
        $\mathbf{x}_n^{(j)} \sim p_{n|n-1}(\mathbf{x}_n|\mathbf{x}_{n-1}^{(j)})$
    end for
    Update
    $\{\alpha_l, \boldsymbol{\mu}_l, \mathbf{P}_l\} = \text{EM}(\{\mathbf{x}_n^{(j)}\}, L)$
    $\{\alpha_{l,0}, \boldsymbol{\mu}_{l,0}, \mathbf{P}_{l,0}, \{\mathbf{x}_{k,0}^{(j)}\}\} = \{\alpha_l, \boldsymbol{\mu}_l, \mathbf{P}_l, \{\mathbf{x}_k^{(j)}\}\}$
    $\Delta s = \frac{1}{N_s}$
    for $s = \Delta s : \Delta s : 1$ **do**
        for $j = 1$ to N_p **do**
            $\mathbf{x}_{n,s}^{(j)} = \mathbf{x}_{n,s-\Delta s}^{(j)} + \Delta s\bigg(u_A(\mathbf{x}_{n,s-\Delta s}^{(j)}, s-\Delta s) + u_B(\mathbf{x}_{n,s-\Delta s}^{(j)}, s-\Delta s)\bigg)$
        end for
        $\{\alpha_{l,s}, \boldsymbol{\mu}_{l,s}, \mathbf{P}_{l,s}\} = \text{EM}(\{\mathbf{x}_{n,s}^{(j)}\}, L)$
    end for
    $\{\alpha_l, \boldsymbol{\mu}_l, \mathbf{P}_l, \{\mathbf{x}_n^{(j)}\}\} = \{\alpha_{l,1}, \boldsymbol{\mu}_{l,1}, \mathbf{P}_{l,1}, \{\mathbf{x}_{n,1}^{(j)}\}\}$
end for

Figure 1: Ensemble Gaussian mixture filter implementation.

4. NUMERICAL SIMULATIONS

In this section we discuss numerical simulations of the the proposed filter equations. The implementations are performed using MATLAB. We consider the simple problem of state estimation using noisy observations. The state dimension N is set to 2. The state evolution is modeled by the equation (1) with $\mathbf{R}_X = \sigma_{\mathbf{x}}^2 I_{2\times 2}$. In simulations we set $\sigma_{\mathbf{x}} = 0.5$. The prior distribution of the state at time $n = 0$ is a 2 component Gaussian

mixture. The parameters of the mixture are $\boldsymbol{\mu}_1^0 = [0,0]$, $\boldsymbol{\mu}_2^0 = [5,5]$, $\mathbf{P}_1^0 = \mathbf{P}_2^0 = I_{2\times 2}$ and $\alpha_1^0 = \alpha_2^0 = 0.5$. The observation matrix $\mathbf{H}$ is an identity matrix $I_{2\times 2}$. Though the filter equations derived above hold for the general case of correlated vectors, for the purpose of demonstration, we choose the observation noise covariance matrix to be diagonal, $\mathbf{R}_Y = \sigma_{\mathbf{y}}^2 I_{2\times 2}$. We use a value of $\sigma_{\mathbf{y}} = 0.5$ in our simulations.

In the filter implementation, we divide the artificial time $s \in [0,1]$ into $N_s = 20$ intervals with corresponding $\Delta s = 0.05$. It was observed that the components u_A and u_B can take very high numerical values. Hence to limit the amount by which particles are transported in a single iteration step we constrain its infinity norm to a maximum of $u_{A,cut}$ and $u_{B,cut}$ respectively. The EM algorithm is used to fit the Gaussian mixture model with $L = 2$ components to the particle ensemble. The EM algorithm is repeated 50 times with different random initializations and the obtained parameters corresponding to highest likelihood are used. In each repetition a maximum of 100 iterations are performed. We use an ensemble of $N_p = 200$ particles.

Shown in Figure 2a is a typical particle distribution and its evolution over the artificial time at the instants $s = i\Delta s, i = 1,2,3,15,16,17$ with $u_{A,cut} = u_{B,cut} = \frac{2}{\Delta s}$. The (red) arrows indicate diretion of particle motion as calculated by approximating the ODE. Figure 2b presents a magnified version of the same figure, zoomed in near the observation (black cross). We observe that though the particles move roughly in the direction of the current observation, the magnitude of motion is not significantly reduced as s increses. Thus there is an oscillatory movement of particles in the vicinity of the observations. Many particles which are far from the observation experience very little change in their location.

In Figure 3 we compare an example true state trajectory with the estimated state trajectories obtained using the general EGMF filter. The different estimated state trajectories are obtained by varying the cut off parameters as $u_{A,cut} = u_{B,cut} = \frac{2}{\Delta s}, \frac{5}{\Delta s}$ and $\frac{10}{\Delta s}$. The estimated trajectory obtained by using the Kalman filter with two component Gaussian mixture model is also shown. From the figures we observe that as the cut off parameter is increased, the estimation error of the EGMF filter is increasing.

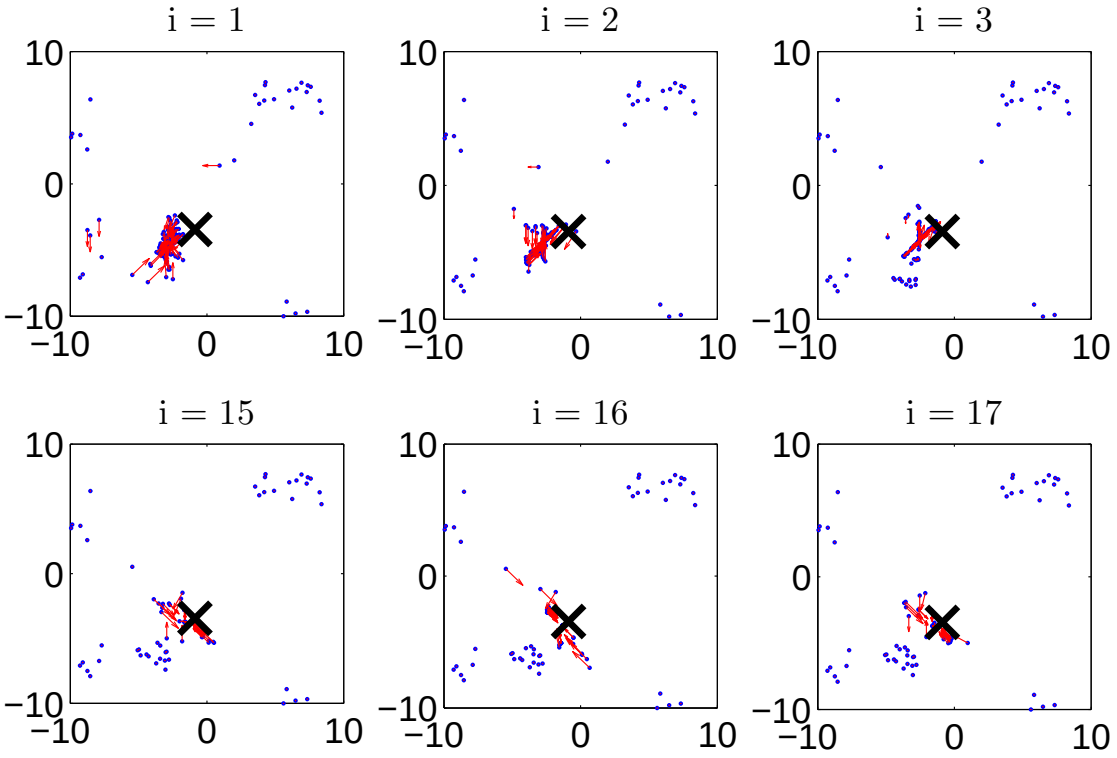
The above results indicate that we have stability issues in the implementation of the proposed EGMF filter and in its current form it has poor performance when compared to the Kalman filter for Gaussian mixture priors. One approach to address this issue would be employ more sophisticated numerical approximations to solve the partial differential equation. In our current implementation we have chosen the constants $C_{l,i}$ in the equation (32) to be equal. A choice of $C_{l,i}$ which takes into account the correlation between the components of the observations $\mathbf{y}_{obs}$ and its relation to the transformed variable $\mathbf{z}$ is expected to improve the filter performance.

5. CONCLUSION

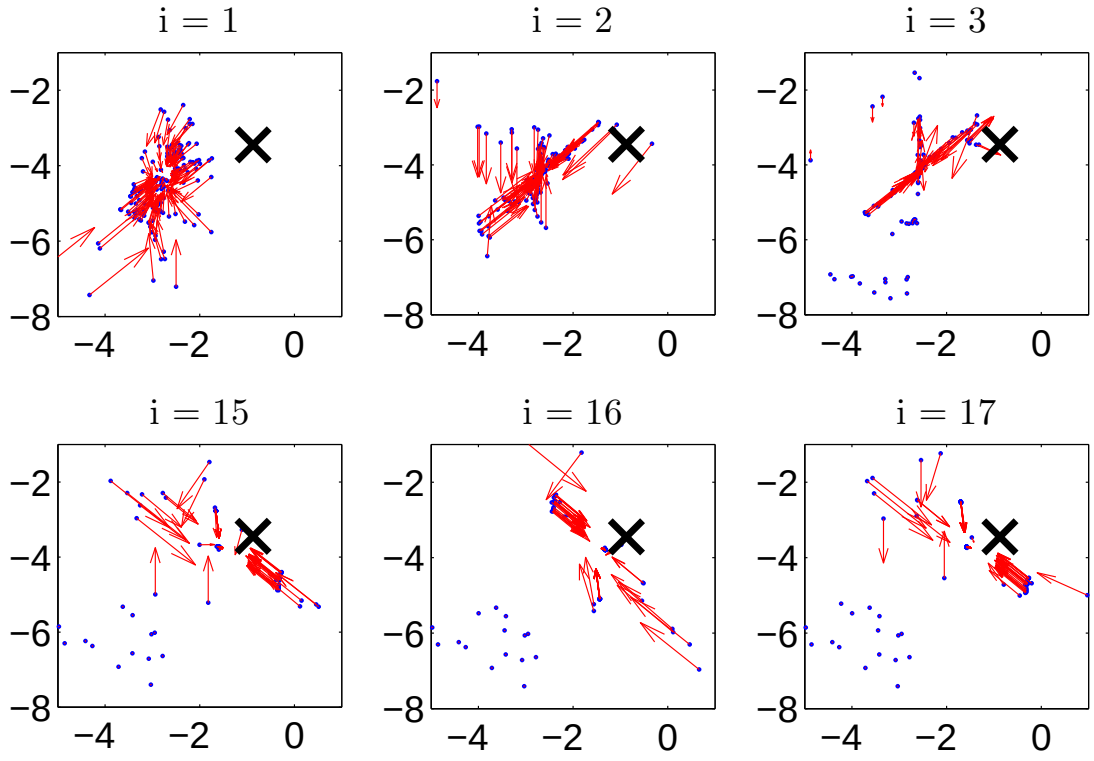
The ensemble Gaussian mixture filter is a generalization of the EnKF which accounts for non-Gaussian densities by propagating the Gaussian mixture model over time. The continuous time formulation of Bayes' rule allows smooth propagation of ensemble particles from the predicted to the posterior density. In this paper we extended the EGMF to include vector observations with arbitrary correlations. Transformation of the state variable simplifies a general second order PDE into a set of scalar PDEs which can be easily solved. We have presented an example numerical simulation of the extended filter and suggested some methods to address the stability issues in its implementation.

REFERENCES

1. A. Doucet, N. De Freitas, and N. Gordon, *An introduction to sequential Monte Carlo methods*, Springer, 2001.
2. G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *J. of Geophysical Research: Oceans* **99**(C5), pp. 10143–10162, 1994.
3. M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, "Ensemble square root filters*," *Monthly Weather Rev.* **131**(7), pp. 1485–1490, 2003.
4. K. Bergemann and S. Reich, "A localization technique for ensemble Kalman filters," *Quarterly J. of the Royal Meteorological Society* **136**(648), pp. 701–707, 2010.
5. C. J. Cotter and S. Reich, "Ensemble filter techniques for intermittent data assimilation-a survey." arXiv preprint arXiv:1208.6572, 2012.



(a)



(b)

Figure 2: Evolution of particle distribution (blue dots) over the artificial time steps $s = i\Delta s, i = 1, 2, 3, 15, 16, 17$. The red arrows indicate direction of motion of particles. The black marker represents the current observation. Figure (b) is magnified version of Figure (a).

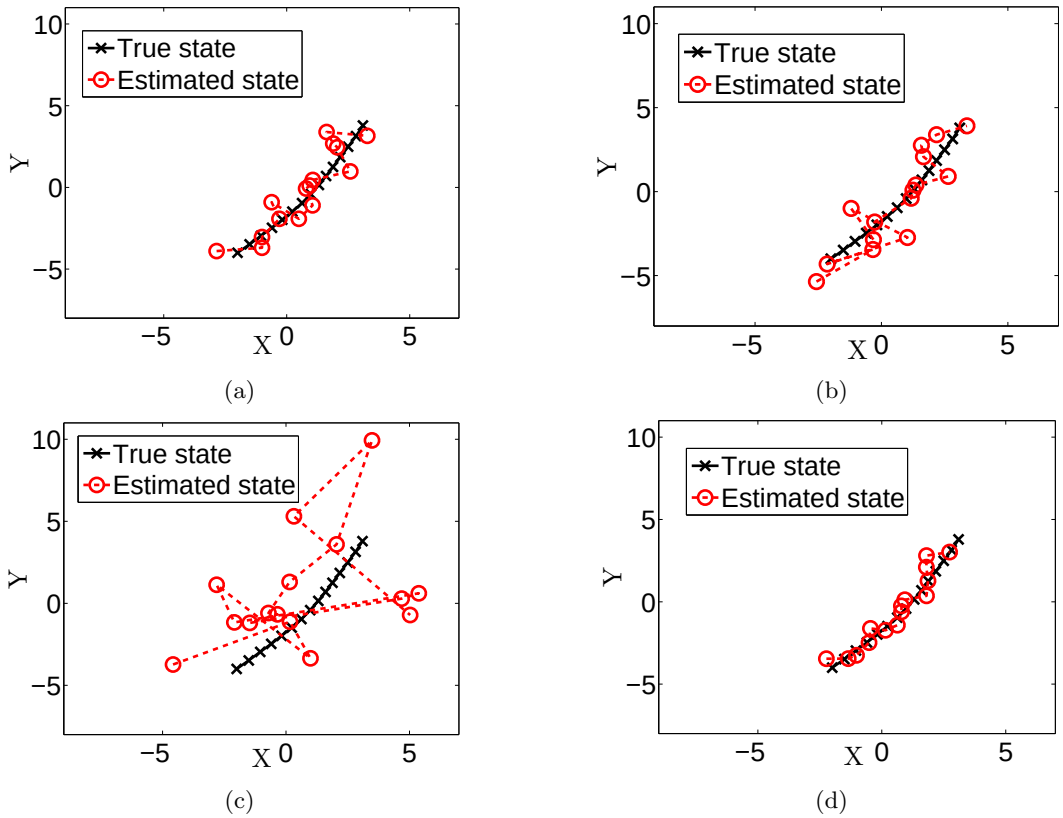


Figure 3: True state trajectory and estimated trajectories. The estimated trajectories obtained using EGMF filter are shown in (a),(b) and (c) with $u_{A,cut} = u_{B,cut} = \frac{2}{\Delta s}, \frac{5}{\Delta s}$ and $\frac{10}{\Delta s}$ respectively. Figure (d) is the estimated trajectory obtained using the Kalman filter.

6. K. Bergemann and S. Reich, "A mollified ensemble Kalman filter," *Quarterly J. of the Royal Meteorological Society* **136**(651), pp. 1636–1643, 2010.
7. A. S. Stordal, H. A. Karlsen, G. Nævdal, H. J. Skaug, and B. Vallès, "Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter," *Comp. Geosciences* **15**(2), pp. 293–305, 2011.
8. M. Frei and H. R. Künsch, "Mixture ensemble Kalman filters," *Comp. Stat. & Data Analysis* **58**, pp. 127–138, Feb. 2013.
9. S. Reich, "A Gaussian-mixture ensemble transform filter," *Quarterly J. of the Royal Meteorological Society* **138**(662), pp. 222–233, 2011.
10. S. Reich, "A dynamical systems framework for intermittent data assimilation," *BIT Numerical Mathematics* **51**(1), pp. 235–249, 2011.
11. F. Daum and J. Huang, "Nonlinear filters with log-homotopy," in *Proc. SPIE Conf. Signal and Data Processing of Small Targets*, (San Diego, CA, USA), Sept. 2007.
12. F. Daum and J. Huang, "Nonlinear filters with particle flow induced by log-homotopy," in *Proc. SPIE Conf. Signal Proc., Sensor Fusion, Target Recog.*, (Orlando, FL, USA), Apr. 2009.
13. F. Daum, J. Huang, and A. Noushin, "Coulomb's law particle flow for nonlinear filters," in *Proc. SPIE Conf. Signal and Data Processing of Small Targets*, (San Diego, CA, USA), Aug. 2011.
14. F. Daum and J. Huang, "Small curvature particle flow for nonlinear filters," in *Proc. SPIE Conf. Signal and Data Processing of Small Targets*, (Baltimore, MD, USA), Apr. 2012.
15. F. Daum and J. Huang, "Particle flow and Monge-Kantorovich transport," in *Proc. Int. Conf. Information Fusion*, (Singapore), Jul. 2012.
16. C. Villani, *Topics in optimal transportation*, vol. 58, American Mathematical Soc., 2003.