Adaptive Gaussian Sum Filter for Nonlinear Bayesian Estimation

Gabriel Terejanu, *Member, IEEE*, Puneet Singla, *Member, IEEE*, Tarunraj Singh, *Member, IEEE*, and Peter D. Scott, *Member, IEEE*

***Abstract*—A nonlinear filter is developed by representing the state probability density function by a finite sum of Gaussian density kernels whose mean and covariance are propagated from one time-step to the next using linear system theory methods such as extended Kalman filter or unscented Kalman filter. The novelty in the proposed method is that the weights of the Gaussian kernels are updated at every time-step, by solving a convex optimization problem posed by requiring the Gaussian sum approximation to satisfy the Fokker–Planck–Kolmogorov equation for continuous-time dynamical systems and the Chapman–Kolmogorov equation for discrete-time dynamical systems. The numerical simulation results show that updating the weights of different mixture components during propagation mode of the filter not only provides us with better state estimates but also with a more accurate state probability density function.**

***Index Terms*—Gaussian sum filter (GSF), Kalman filter, probability density function (pdf).**

I. Introduction

The state estimation problem has been extensively studied since the celebrated Kalman filter [1] appeared in 1960. The Kalman filter is the optimal estimator for linear dynamical systems with linear measurement models driven by Gaussian disturbances, and it finds the most probable state as the unbiased linear minimum variance estimate of the system. The optimal filtering problem for nonlinear systems poses a challenge since it requires maintaining the complete description of the conditional probability density function (pdf) which in general requires an infinite number of parameters [2]. Several finite dimensional suboptimal filters [3]–[8] have been developed over the last five decades while sampling based methods also known as Particle Filters (PF) [9] have become more popular for general nonlinear filtering problem. However, various factors like *volume of state space in which conditional pdf is non-vanishing, rate of decay of the conditional pdf in state space, stationarity of the problem, analytical structure of the problem (e.g., linear dynamics, bilinear dynamics, unimodal pdf, etc.), effective dimensionality of the problem etc.* strongly affect the computational complexity and performance of the PF [10] and hence, an efficient approach for general nonlinear filtering has to take into account the evolution of the state probability density function (pdf) using the Kolmogorov equation.

A Gaussian mixture approximation for the state pdf has been gaining increasing attention in the context of Bayesian estimation using continuous representation of state pdf [6], [11], [12]. For a nonlinear dynamical system with additive Gaussian white noise, the first two moments of the Gaussian components are propagated using the linearized model, and the weights of the new Gaussian components are set equal to the prior weights. In the case where observations are available both the moments and the weights are accordingly updated using Bayes rule to obtain an approximation of the a-posteriori pdf, yielding the so called Gaussian Sum Filter (GSF) [6], [12]. The literature on nonlinear filtering using Gaussian mixtures is rich in theoretical work on the GSF [12], GSF with a more advanced measurement update [6], mixture of Kalman filters [13] and Gaussian sum particle filtering [14]. Applications and improvements brought to the GSF are continuing endeavors in fields like target tracking [15]–[17], computer vision [18], [19] and geoscience [20]. However, in all of these methods *the weights of the Gaussian mixands are kept constant between two measurements which constrains the accuracy of the GSF algorithm in approximating the forecast pdf when measurements data is sparse* [12].

Initial results of the weight adaptation of the Gaussian mixtures in the nonlinear filtering problem have been studied in [21] where regularization was proposed as a means to circumvent the stationary pdf solution. The main focus of this technical note is to incorporate the complete evolution of state pdf between two measurements by solving the FPKE for continuous-time dynamical systems and the CKE for discrete-time dynamical systems in the GSF context.

The organization of the technical note is as follows: first, the conventional GSF is introduced in Section II, followed by the adaptive Gaussian sum filter method in Section III. Numerical results are presented in Section V followed by conclusions in Section VI.

II. Conventional Gaussian Sum Filter

Let us consider a general n-dimensional continuous-time dynamic system with uncertain initial conditions and discrete-time measurement model, given as:

$$\dot{\mathbf{x}}(t) = \mathbf{f}\left(t, \mathbf{x}(t)\right) + \mathbf{\Gamma}(t) \tag{1}$$

$$\mathbf{z}_k = \mathbf{h}(t_k, \mathbf{x}_k) + \mathbf{v}_k \tag{2}$$

where $\mathbf{x}_k = \mathbf{x}(t_k)$ represents the n-dimensional state vector, and $\mathbf{\Gamma}(t)$ represents a zero mean Gaussian white noise process with the correlation function $\mathbf{Q}\delta(t-\tau)$. The nonlinear function $\mathbf{h}(.)$ captures the measurement model and the random vector $\mathbf{v}_k$ denotes the measurement noise, which is temporally uncorrelated, zero-mean random sequence with known covariance, $\mathbf{R}_k$, and uncorrelated with $\mathbf{\Gamma}(t)$.

The total uncertainty associated with the state vector $\mathbf{x}(t)$ is characterized by the pdf $p(t, \mathbf{x}(t))$ and a nonlinear filtering problem corresponds to finding a-posteriori distribution for $\mathbf{x}_k$ given the measurement data $\mathbf{Z}_k = \{\mathbf{z}_i | i = 1, 2, \ldots, k\}$, i.e., $p(t, \mathbf{x}(t)|\mathbf{Z}_k)$ and a prior pdf $p(t_0, \mathbf{x}_0)$. In the GSF context, the conditional pdf is approximated by a finite sum of Gaussian kernels:

$$\hat{p}\left(t, \mathbf{x}(t)|\mathbf{Z}_k\right) = \sum_{i=1}^{N} w_{t|k}^i \underbrace{\mathcal{N}\left(\mathbf{x}(t); \boldsymbol{\mu}_{t|k}^i, \mathbf{P}_{t|k}^i\right)}_{p_i} \tag{3}$$

where $w_{t|k}^i$, $\boldsymbol{\mu}_{t|k}^i$, and $\mathbf{P}_{t|k}^i$ represent the conditional weight, mean and covariance of the i^{th} Gaussian kernel with respect to the first k measurements. The positivity and normalization constraint on the mixture pdf, $\hat{p}(t, \mathbf{x}|\mathbf{Z}_k)$, leads to the following constraints on the weights:

$$\sum_{i=1}^{N} w_{t|k}^i = 1, \quad w_{t|k}^i \geq 0, \quad \forall t \tag{4}$$

Since all the components of the mixture pdf (3) are Gaussian, only estimates of their mean and covariance need to be estimated between t_k

Manuscript received September 28, 2009; revised October 19, 2010 and November 07, 2010; accepted March 21, 2011. Date of publication April 07, 2011; date of current version September 08, 2011. This work was partially supported in part by ONR Contract HM1582-08-1-0012 and by the National Science Foundation under Award CMMI-0908403. Recommended by Associate Editor H. Zhang.

G. Terejanu is with the Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, Austin, TX 78712 USA (e-mail: terejanu@ices.utexas.edu).

P. Singla, T. Singh, and P. D. Scott are with the Department of Mechanical and Aerospace Engineering, University at Buffalo, Buffalo, NY 14260 USA (e-mail: psingla@buffalo.edu; tsingh@buffalo.edu; peter@buffalo.edu).

Color versions of one or more of the figures in this technical note are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2011.2141550

and t_{k+1} using, for example, the conventional Extended Kalman Filter (EKF) time update equations [12]:

$$\dot{\boldsymbol{\mu}}^i_{t|k} = \mathbf{f}\left(t, \boldsymbol{\mu}^i_{t|k}\right), \quad \mathbf{A}^i_{t|k} = \left.\frac{\partial \mathbf{f}\left(t, \mathbf{x}(t)\right)}{\partial \mathbf{x}(t)}\right|_{\boldsymbol{\mu}^i_{t|k}} \tag{5}$$

$$\dot{\mathbf{P}}^i_{t|k} = \mathbf{A}^i_{t|k}\mathbf{P}^i_{t|k} + \mathbf{P}^i_{t|k}\mathbf{A}^{i^T}_{t|k} + \mathbf{Q} \tag{6}$$

$$w^i_{t|k} = w^i_{k|k} \quad \text{for} \quad t_k \le t \le t_{k+1}. \tag{7}$$

The measurement update is done by making use of the Bayes's rule as explained in [12]

$$\boldsymbol{\mu}^i_{k+1|k+1} = \boldsymbol{\mu}^i_{k+1|k} + \mathbf{K}^i_k\left(\mathbf{z}_k - \mathbf{h}\left(t_k, \boldsymbol{\mu}^i_{k+1|k}\right)\right) \tag{8}$$

$$\mathbf{K}^i_k = \mathbf{P}^i_{k+1|k}\mathbf{H}^{i^T}_k\left(\mathbf{H}^i_k\mathbf{P}^i_{k+1|k}\mathbf{H}^{i^T}_k + \mathbf{R}_k\right)^{-1} \tag{9}$$

$$\mathbf{P}^i_{k+1|k+1} = \left(\mathbf{I} - \mathbf{K}^i_k\mathbf{H}^i_k\right)\mathbf{P}^i_{k+1|k},$$

$$\mathbf{H}^i_k = \left.\frac{\partial \mathbf{h}(t_k, \mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\boldsymbol{\mu}^i_{k+1|k}} \tag{10}$$

$$w^i_{k+1|k+1} = \frac{w^i_{k+1|k}\beta^i_k}{\sum_{i=1}^N w^i_{k+1|k}\beta^i_k}$$

$$\beta^i_k = \mathcal{N}\left(\mathbf{z}_k - \mathbf{h}\left(t_k, \boldsymbol{\mu}^i_{k+1|k}\right), \mathbf{H}^i_k\mathbf{P}^i_{k+1|k}\mathbf{H}^{i^T}_k + \mathbf{R}_k\right). \tag{11}$$

A point estimate for state is taken to be the mean of the posterior pdf and the corresponding covariance matrix is computed as

$$\boldsymbol{\mu}_{t|k} = \sum_{i=1}^N w^i_{t|k}\boldsymbol{\mu}^i_{t|k} \tag{12}$$

$$\mathbf{P}_{t|k} = \sum_{i=1}^N w^i_{t|k}\left[\mathbf{P}^i_{t|k} + \left(\boldsymbol{\mu}^i_{t|k} - \boldsymbol{\mu}_{t|k}\right)\left(\boldsymbol{\mu}^i_{t|k} - \boldsymbol{\mu}_{t|k}\right)^T\right]. \tag{13}$$

The Gaussian sum approximation for the conditional state pdf, $\hat{p}(t, \mathbf{x}|\mathbf{Z}_k)$, obtained by integrating (5)–(11), approaches the true conditional pdf under the assumption that there is a sufficient number of Gaussian kernels and the covariance of all Gaussian kernels is small enough such that the linearizations around the means are representative for the dynamics in the vicinity of the respective means [8]. Furthermore, the assumption that Gaussian component weights $w^i_{t|k}$ remain constant between measurement updates is appropriate if the underlying system dynamics is linear or the system is at worst marginally nonlinear between two measurement updates. In practice, this assumption may be easily violated resulting in a poor approximation of the conditional pdf. Practically, the dynamic system may exhibit strong nonlinearities between measurement updates and the state pdf may show multi-mode behavior in the absence of measurement data.

III. Adaptive Gaussian Sum Filter

Although different approaches [6], [14], [17] have been proposed to update Gaussian kernel weights using measurement data, the existing literature provides no means for adaptation of the Gaussian kernel weights during the propagation of state pdf between two measurements. The lack of adaptive algorithms for weights of Gaussian mixture is felt to be a serious disadvantage of existing GSF algorithms when measurement data is sparse or system dynamics is highly nonlinear. In this section, two different approaches are presented to update the Gaussian kernel weights between two measurements. The first approach makes use of the FPKE for state pdf evolution and hence addresses continuous-time dynamical systems. The second approach makes use of the CKE for state pdf evolution between two measurements and hence addresses discrete-time dynamical systems.

A. Continuous-Time Dynamical Systems

In essence, the problem of adaptation of weights of Gaussian kernels corresponds to finding the nature of time-evolution of the system-state pdf. For stochastic continuous dynamic system the exact evolution of state pdf is given by the FPKE [22]:

$$\frac{\partial}{\partial t}p(t, \mathbf{x}|\mathbf{Z}_k) = \mathcal{L}p(t, \mathbf{x}|\mathbf{Z}_k) = \frac{\partial \mathbf{f}^T(t, \mathbf{x})p(t, \mathbf{x}|\mathbf{Z}_k)}{\partial \mathbf{x}} + \frac{1}{2}\mathrm{Tr}\left[\mathbf{Q}(t)\frac{\partial^2 p(t, \mathbf{x}|\mathbf{Z}_k)}{\partial \mathbf{x}\partial \mathbf{x}^T}\right] \tag{14}$$

The FPKE is a formidable equation to solve, because of the following issues: 1) *Positivity* of the pdf, 2) *Normalization* constraint of the pdf: $\int_{\mathbb{R}^n} p(t, \mathbf{x}(t)|\mathbf{Z}_k)\mathrm{d}\mathbf{x} = 1$, and 3) *No fixed Solution Domain*: how to impose boundary conditions in a finite region with non-zero probability mass.

The key idea of the proposed approach is to approximate the state pdf by a finite sum of Gaussian density functions whose mean and covariance are propagated from one time-step to the next using (5). Unlike the GSF algorithm, the weights $w^i_{t|k}$ of the Gaussian kernels are not known at time t and must be computed as part of the solution process. The weights of the Gaussian kernels are updated at every time-step, by substituting the Gaussian mixture approximation, $\hat{p}(t, \mathbf{x}|\mathbf{Z}_k)$ of (3) in the FPKE (14) and minimizing the L_2-norm of the FPKE error while taking into account the constraints of (4) and (5). This leads to the following optimization problem:

$$\min_{w^i_{t|k}} J = \frac{1}{2}\int e^2(t, \mathbf{x})\mathrm{d}\mathbf{x}, \quad i = 1, \cdots, N$$

$$\text{s.t.} \sum_{i=1}^N w^i_{t|k} = 1, \quad w^i_{t|k} \ge 0. \tag{15}$$

Here $e(t, \mathbf{x}(t))$ represents the FPKE error:

$$e(t, \mathbf{x}) = \frac{\partial \hat{p}(t, \mathbf{x}|\mathbf{Z}_k)}{\partial t} - \underbrace{\mathcal{L}\hat{p}(t, \mathbf{x}|\mathbf{Z}_k)}_{\mathbf{l}^T_{\mathbf{FP}}\mathbf{w}_{t|k}} \tag{16}$$

where $\mathbf{w}_{t|k} \in \mathbb{R}^N$ is the vector of Gaussian kernel weights and the elements of $\mathbf{l}_{\mathbf{FP}} \in \mathbb{R}^N$ are given as

$$\mathcal{L}p_i = -\frac{\partial p_i^T}{\partial \mathbf{x}}\mathbf{f}(t, \mathbf{x}) - p_i\mathrm{Tr}\left[\frac{\partial \mathbf{f}(t, \mathbf{x})}{\partial \mathbf{x}}\right] + \frac{1}{2}\mathrm{Tr}\left[\mathbf{Q}\frac{\partial^2 p_i}{\partial \mathbf{x}\partial \mathbf{x}^T}\right]. \tag{17}$$

Furthermore, the first term in (16) is computed as

$$\frac{\partial \hat{p}(t, \mathbf{x}|\mathbf{Z}_k)}{\partial t} = \sum_{i=1}^N\left(\dot{w}^i_{t|k}p_i + w^i_{t|k}\frac{\partial p_i^T}{\partial \boldsymbol{\mu}^i_{t|k}}\dot{\boldsymbol{\mu}}^i_{t|k} + w^i_{t|k}\mathrm{Tr}\left[\frac{\partial p_i}{\partial \mathbf{P}^i_{t|k}}\dot{\mathbf{P}}^i_{t|k}\right]\right). \tag{18}$$

Here, $\dot{\boldsymbol{\mu}}^i_{t|k}$ and $\dot{\mathbf{P}}^i_{t|k}$ are given by (5). One can use Galerkin projection method to find the system of deterministic differential equations to solve for $w^i_{t|k}$. This is equivalent to minimization of L_2-norm of $e(t, \mathbf{x})$ without the constraints of (4). Since the satisfaction of these constraints is essential to guarantee a unique solution for the FPKE, this can create many spurious modes in the numerical solution for FPKE [23]. To find unknown weights, $\dot{w}^i_{t|k}$, is approximated as

$$\dot{w}^i_{t|k} = \frac{1}{\Delta t}\left(w^i_{t'|k} - w^i_{t|k}\right) \quad \text{where} \quad t' = t + \Delta t. \tag{19}$$

Notice that the time interval Δt represents our choice of time scale for changing the mixand weights between two measurements. Finally, the

substitution of (19) into (18) leads to

$$\begin{aligned}\frac{\partial \hat{p}(t,\mathbf{x}|\mathbf{Z}_k)}{\partial t} &= \sum_{i=1}^{N}\frac{1}{\Delta t}p_i w^i_{t'|k} \\ &\quad + \sum_{i=1}^{N}\underbrace{\left(\frac{\partial p_i^T}{\partial \boldsymbol{\mu}^i_{t|k}}\dot{\boldsymbol{\mu}}^i_{t|k} + \mathrm{Tr}\left[\frac{\partial p_i}{\partial \mathbf{P}^i_{t|k}}\dot{\mathbf{P}}^i_{t|k}\right] - \frac{1}{\Delta t}p_i\right)}_{m^i_{DT}} w^i_{t|k} \\ &= \frac{1}{\Delta t}\mathbf{p}_\mathbf{g}^T\mathbf{w}_{t'|k} + \mathbf{m}_\mathbf{DT}^T\mathbf{w}_{t|k} \end{aligned} \tag{20}$$

where $\mathbf{w}_{t'|k} \in \mathbb{R}^{N\times 1}$ is the vector of new weights to be found, $\mathbf{p_g} \in \mathbb{R}^{N\times 1}$ is the vector of Gaussian components. The derivatives in the aforementioned equations can be analytically computed [24]. Finally, the FPKE error of (16) can be compactly written as

$$e(t,\mathbf{x}) = \frac{1}{\Delta t}\mathbf{p}_\mathbf{g}^T\mathbf{w}_{t'|k} + (\mathbf{m_{DT}} - \mathbf{l_{FP}})^T\mathbf{w}_{t|k}. \tag{21}$$

Since the FPKE error (21) is linear in the new Gaussian weights, $w^i_{t'|k}$, the optimization problem of (15) can be written as the following Quadratic Programming (QP) problem:

$$\begin{aligned}\mathbf{w}^*_{t'|k} = \arg\min_{\mathbf{w}_{t'|k}}\ &\frac{1}{2}\mathbf{w}^T_{t'|k}\mathbf{M_c}\mathbf{w}_{t'|k} + \mathbf{w}^T_{t'|k}\mathbf{N_c}\mathbf{w}_{t|k} \\ &\text{s.t } \mathbf{1}^T_{N\times 1}\mathbf{w}_{t'|k} = 1, \mathbf{w}_{t'|k} \geq \mathbf{0}_{N\times 1}\end{aligned} \tag{22}$$

where $\mathbf{1}_{N\times 1} \in \mathbb{R}^N$ is a vector of ones, $\mathbf{0}_N \in \mathbb{R}^N$ is a vector of zeros and the matrices $\mathbf{M_c} \in \mathbb{R}^{N\times N}$ and $\mathbf{N_c} \in \mathbb{R}^{N\times N}$ are given as

$$\begin{aligned}M_{c_{ij}} &= \frac{1}{\Delta t^2}\mathcal{N}\left(\boldsymbol{\mu}^j_{t|k}; \boldsymbol{\mu}^i_{t|k}, \mathbf{P}^i_{t|k} + \mathbf{P}^j_{t|k}\right) \\ N_{c_{ij}} &= \frac{1}{\Delta t}p_i\int_V\left(\frac{\partial p_j^T}{\partial \boldsymbol{\mu}^j_{t|k}}\dot{\boldsymbol{\mu}}^j_{t|k} + \mathrm{Tr}\left[\frac{\partial p_j}{\partial \mathbf{P}^j_{t|k}}\dot{\mathbf{P}}^j_{t|k}\right] - \frac{1}{\Delta t}p_j\right. \\ &\qquad + \frac{\partial p_j^T}{\partial \mathbf{x}}\mathbf{f}(t,\mathbf{x}) + p_j\mathrm{Tr}\left[\frac{\partial \mathbf{f}(t,\mathbf{x})}{\partial \mathbf{x}}\right] \\ &\qquad \left.-\frac{1}{2}\mathrm{Tr}\left[\mathbf{Q}\frac{\partial^2 p_j}{\partial \mathbf{x}\partial \mathbf{x}^T}\right]\right)\mathrm{d}\mathbf{x}.\end{aligned}$$

It is easy to see that the matrix $\mathbf{M}_c$ is positive semi-definite and the cost function J is lower bounded due to linear equality and inequality constraints. As a consequence of this, the aforementioned optimization problem can be posed as a convex QP problem [24]. The solution of this minimization problem (22) will substitute the weight update of (7), in the conventional GSF. Ideally, one can seek to optimize for the mean and covariance of each Gaussian kernel along with the weights. This leads to a nonlinear optimization problem with $(((n+2)(n+1))/2)N$ variables rather than N variables here. To make matter worse there are no guarantees for a global optimal for the solution of the resulting nonlinear optimization problem as opposed to the guaranteed solution here for the resulting QP problem.

The expectation integrals involved in the matrix $\mathbf{N}_c$ may be computed using Gaussian Quadrature (GQ), Monte Carlo integration or via the Unscented Transformation (UT). While in lower dimensions the UT is essentially equivalent to GQ for evaluating integrals, in higher dimensions the UT is computationally more appealing in evaluating integrals because the required number of evaluation points grows only linearly with the number of dimensions.

Although, knowledge of the Jacobian matrix, $\mathbf{A}^i_{t|k}$, is required in (5) to propagate the covariance of each of the Gaussian component, one can easily use other methods such as the Unscented Kalman Filter (UKF) [7]. In [25], the efficacy of using the UKF to propagate the mean and covariance of the mixture components is shown for a finite dimensional dynamical system.

B. Discrete-Time Nonlinear Dynamical System

In this section, the time-evolution of the system-state pdf between two measurements is considered for discrete-time dynamic systems given by the following stochastic difference equation:

$$\mathbf{x}_{k+1} = \mathbf{f}(t_k, \mathbf{x}_k) + \boldsymbol{\eta}_k \tag{23}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ represents the state vector and $\boldsymbol{\eta}_k$ represents the zero mean white noise with covariance function $\mathbf{Q}_k$. The time evolution of the conditional pdf $p(t_{k+1}, \mathbf{x}_{k+1}|\mathbf{Z}_k)$ from time-step k to time-step $k+1$ is given by the CKE [5]

$$p(t_{k+1}, \mathbf{x}_{k+1}|\mathbf{Z}_k) = \int p(t_{k+1}, \mathbf{x}_{k+1}|t_k, \mathbf{x}_k)p(t_k, \mathbf{x}_k|\mathbf{Z}_k)\mathrm{d}\mathbf{x}_k \tag{24}$$

where, $p(t_{k+1}, \mathbf{x}_{k+1}|t_k, \mathbf{x}_k)$ is the conditional state transition pdf which corresponds to the pdf for $\boldsymbol{\eta}_k$. The properties of the CKE are similar to that of the FPKE and like the FPKE the ordinary problem of finding the solution to the CKE is not so tractable in a systematic manner. The Gaussian mixture model approximation of the forecast pdf can be written as

$$\hat{p}(t_{k+1}, \mathbf{x}_{k+1}|\mathbf{Z}_k) = \sum_{i=1}^{N} w^i_{k+1}\mathcal{N}\left(\mathbf{x}_{k+1}; \boldsymbol{\mu}^i_{k+1|k}, \mathbf{P}^i_{k+1|k}\right). \tag{25}$$

In this equation, $\boldsymbol{\mu}^i_{k+1|k}$ and $\mathbf{P}^i_{k+1|k}$ represent the mean and covariance of the i^{th} Gaussian kernel and are computed by making use of the EKF equations

$$\boldsymbol{\mu}^i_{k+1|k} = \mathbf{f}\left(k, \boldsymbol{\mu}^i_{k|k}\right) \tag{26}$$

$$\mathbf{P}^i_{k+1|k} = \mathbf{A}^i_k\mathbf{P}^i_{k|k}\mathbf{A}^{i^T}_k + \mathbf{Q}_k,\ \ \mathbf{A}^i_k = \left.\frac{\partial \mathbf{f}(k,\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\boldsymbol{\mu}^i_{k|k}}. \tag{27}$$

To determine the unknown weights, w^i_{k+1}, the following integral square difference between the true pdf, $p(t_{k+1}, \mathbf{x}_{k+1}|\mathbf{Z}_k)$, and its approximation, $\hat{p}(t_{k+1}, \mathbf{x}_{k+1}|\mathbf{Z}_k)$, is minimized:

$$\begin{aligned}\min_{w^i_{k+1}}\ &\frac{1}{2}\int\left|p(t_{k+1}, \mathbf{x}_{k+1}|\mathbf{Z}_k) - \hat{p}(t_{k+1}, \mathbf{x}_{k+1}|\mathbf{Z}_k)\right|^2\mathrm{d}\mathbf{x}_{k+1} \\ &\text{s.t. } \sum_{i=1}^{N} w^i_{k+1} = 1, \quad w^i_{k+1} \geq 0.\end{aligned} \tag{28}$$

Again making use of the CKE for $p(t_{k+1}, \mathbf{x}_{k+1}|\mathbf{Z}_k)$, and Gaussian mixture approximation of (25) for $\hat{p}(t_{k+1}, \mathbf{x}_{k+1}|\mathbf{Z}_k)$, the cost function can be rewritten as [24]

$$J = \frac{1}{2}\mathbf{w}^T_{k+1}\mathbf{M}\mathbf{w}_{k+1} - \mathbf{w}^T_{k+1}\mathbf{y} \tag{29}$$

where $\mathbf{w}_{k+1} = [w^1_{k+1}w^2_{k+1}\ldots w^N_{k+1}]^T$, and the components of matrix $\mathbf{M}$ are given by [24]

$$M_{ij} = \mathcal{N}\left(\boldsymbol{\mu}^i_{k+1} - \boldsymbol{\mu}^j_{k+1}, \mathbf{P}^i_{k+1} + \mathbf{P}^j_{k+1}\right). \tag{30}$$

The components of the vector $\mathbf{y} \in \mathbb{R}^{N\times 1}$ are given by

$$y_i = \sum_{j=1}^{N} w^j_k N_{ij} \tag{31}$$

$$\begin{aligned}N_{ij} = \int &\mathcal{N}\left(\mathbf{f}(t_k, \mathbf{x}_k); \boldsymbol{\mu}^i_{k+1}, \mathbf{P}^i_{k+1} + \mathbf{Q}_k\right) \\ &\times \mathcal{N}\left(\mathbf{x}_k; \boldsymbol{\mu}^j_k, \mathbf{P}^j_k\right)\mathrm{d}\mathbf{x}_k.\end{aligned} \tag{32}$$

That is, y_i is a sum of expectations of the component functions. Substitution of (31) into (29) leads to the following quadratic programming problem which can be readily solved:

$$\min_{\mathbf{w}_{k+1}} \; J = \frac{1}{2}\mathbf{w}_{k+1}^T \mathbf{M}\mathbf{w}_{k+1} - \mathbf{w}_{k+1}^T \mathbf{N}\mathbf{w}_k$$
$$\text{s.t} \quad \mathbf{1}^T \mathbf{w}_{k+1} = 1, \quad \mathbf{w}_{k+1} \geq \mathbf{0}. \tag{33}$$

Similar to the continuous-time case, this minimization problem along with measurement update equations of (8)–(11) constitute the nonlinear filter. It should be noticed that the matrix $\mathbf{M}$ is positive semi-definite and the cost function J is lower bounded and hence, the aforementioned optimization problem is indeed a convex optimization problem.

IV. Computational Complexity

In this section, the asymptotic numerical efficiency of the Adaptive Gaussian Sum Filter (AGSF) algorithm proposed in the previous section is considered. A key question is to determine the effect on the running time of a particular iteration of the filter with respect to the state dimension—n, the dimension of the observations—m, and the number of Gaussian kernels N.

As reference the discrete-time version of the filter is considered. It is assumed that the evaluations of the functions $\mathbf{f}(\cdot)$ and $\mathbf{h}(\cdot)$, Jacobians and likelihood function is performed in $O(1)$. The overall naive computational complexity of the EKF due to its time update step and measurement update step is $O(n^3 + m^3 + n^2 m + n m^2)$. The complexity to obtain the matrix $\mathbf{M}$ is $O(N^2 n^3)$ and the computational complexity to obtain the matrix $\mathbf{N}$ using Unscented Transformation is $O(N n^3 + N^2 L)$ where L is the number of sigma points. $O(N n^3)$ is due to sigma point selection scheme where the square root of the covariance is needed and $O(N^2 L)$ is due to function evaluations. Furthermore, the convex optimization problem to find mixand weights can be solved in polynomial time with $O(SN^5)$ elementary operations [26] where S is the input length of the QP problem encoding in binary. Thus the total running time of the AGSF per time step is given by

$$O(n^3 + m^3 + n^2 m + n m^2 + N^2 n^3 + N^2 L + S N^5). \tag{34}$$

The above calculation reveals that the complexity is very sensitive to the number of Gaussian kernels, but still polynomial in nature. This emphasizes the need to limit N even for multi-modal distributions. Hence, the selection of Gaussian components has to be done judiciously in order to obtain an accurate and relevant approximation to the state pdf while scaling the computational burden to the temporal response requirements of the task at hand.

Finally, the proposed AGSF algorithms adapt the mixand weights under the assumption that the number of components remains constant during propagation. A more robust scheme can be obtained using a component refining and coarsening scheme. In [27] an iterative split-merge procedure is presented which selects the mixand with the largest contribution to the uncertainty propagation error and splits it along the direction with most nonlinear dynamics.

V. Numerical Results

The performance of the AGSF, both in continuous and discrete-time dynamical systems, have been compared with the conventional GSF where the weights are kept constant between two measurements. Two performance measures are considered to compare the proposed algorithms with other nonlinear filtering algorithms. First, Root Mean Squared Error (RMSE) averaged over R Monte Carlo runs is considered to assess the accuracy for the point state estimates

$$\mathrm{RMSE}_{Truth}(t) = \sqrt{\frac{1}{R}\sum_{j=1}^{R}\left\|\mathbf{x}_t^j - \boldsymbol{\mu}_{t|k}^j\right\|_2^2}. \tag{35}$$

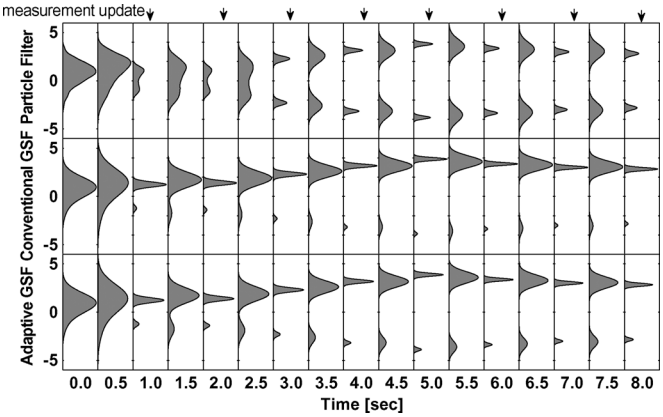

Fig. 1. Example 1: evolution of the conditional pdf.

Where x_t^j is the true value of the state at time t for the j^{th} run and R is the total number of runs. Since our objective is not only to obtain more accurate state estimates but also to obtain a better approximation to the true conditional pdf, the cross entropy with respect to the bootstrap PF is considered

$$\mathrm{H}(t) = -\sum_{j=1}^{M} w_j^{pf} \log \sum_{i=1}^{N} w_i \mathcal{N}\left(\mathbf{x}_t^j; \boldsymbol{\mu}_{t|k}^i, \mathbf{P}_{t|k}^i\right). \tag{36}$$

Here, $\mathbf{x}_t^j$ is the jth particle, w_j^{pf} is its corresponding weight and, M is the total number of particles used in the bootstrap PF. Since we are interested to see at each time step which estimate is closer to the truth, the normalized cross entropy is computed

$$\mathrm{NH}_{GSF}(t) = \frac{\mathrm{H}_{GSF}(t)}{\sqrt{\mathrm{H}_{GSF}(t)^2 + \mathrm{H}_{AGSF}(t)^2}}. \tag{37}$$

A similar expression is computed for $\mathrm{NH}_{AGSF}(t)$. Both performance measures, the RMSE and the normalized cross entropy have been averaged over 100 Monte Carlo runs for simulation results considered here.

Example 1: Let us consider the following continuous-time dynamical system with uncertain initial condition and discrete measurement model given as

$$\begin{aligned} \dot{x} &= \sin x + \Gamma(t) \\ z_k &= x_k^2 + v_k \\ p(x_0) &\sim 0.1\mathcal{N}(-1,1) + 0.9\mathcal{N}(1,1), \quad Q_k = R_k = 1 \end{aligned}.$$

The total simulation time is set to 8 s, with measurement updates every 1 s and the weights are updated every $\Delta t = 0.1$ s. Due to the square form of the measurement model, and bimodal nature of the forecast pdf, the measurements do not offer sufficient information to choose one mode of the conditional pdf, thus the state estimate maintain its bimodal nature. In such situations an accurate propagation makes the difference in providing better estimates and a more accurate conditional pdf. Fig. 1 presents the evolution of the conditional pdf for one particular run for the bootstrap PF, conventional GSF and the proposed AGSF algorithm. Fig. 1 shows the increase of the weight of the first component which is in agreement with the conditional pdf given by the bootstrap PF.

Fig. 2(a) shows the plots for the RMS error for the conventional GSF, bootstrap PF and the AGSF based upon both FPKE and CKE error feedback. Furthermore, Fig. 2(b) shows the plots for normalized cross entropy for the conventional GSF and AGSF with respect to the bootstrap PF solution. From these plots, it is clear that the AGSF solution is closer to the bootstrap PF solution as compared to the conventional GSF. It should be mentioned that 1000 particles have been used for

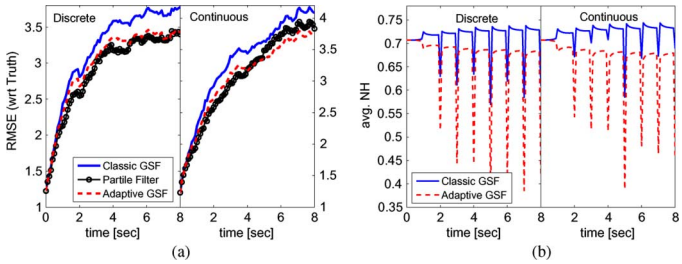

Fig. 2. Example 1: simulation results. (a) RMSE; (b) average normalized cross entropy.

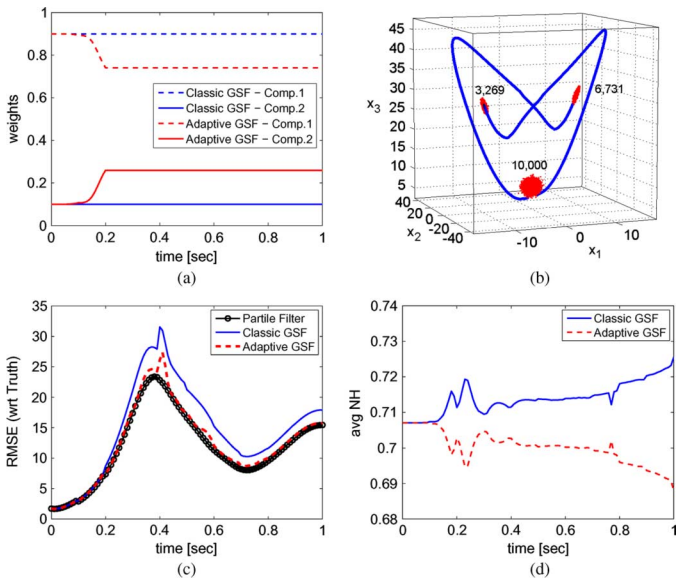

Fig. 3. Example 2: numerical simulation results. (a) Weights comparison; (b) dynamics; (c) RMSE; (d) average normalized cross entropy.

the bootstrap PF with resampling after each measurements, and the expectations needed for the quadratic programming problem have been computed using GQ with 200 quadrature points.

Example 2: To show the efficacy of the proposed approach, let us consider the Lorenz system with uncertain initial condition and discrete measurement model given as

$$\begin{aligned}\dot{x}_1 &= \alpha(-x_1 + x_2)\\ \dot{x}_2 &= \beta x_1 - x_2 - x_1 x_3\\ \dot{x}_3 &= -\gamma x_3 + x_1 x_2 + \Gamma(t)\\ z_k &= \sqrt{x_1(t_k)^2 + x_2(t_k)^2 + x_3(t_k)^2} + v_k\\ \alpha &= 10, \beta = 28, \gamma = \frac{8}{3}, R_k = Q_k = 1. \end{aligned} \tag{38}$$

The total simulation time is 1 s with weights being updated every $\Delta t = 0.01$ s. The initial state pdf is set to $p(\mathbf{x}(t_0)) \sim 0.9\mathcal{N}([-0.2,-0.2,8]^T, \sqrt{0.35}\mathbf{I}_{3\times 3}) + 0.1\mathcal{N}([0.2,0.2,8]^T, \sqrt{0.35}\mathbf{I}_{3\times 3})$. Fig. 3(a) shows the evolution of the weights of different Gaussian components for conventional GSF and AGSF while Fig. 3(b) shows the evolution of the particles of the bootstrap PF as well as the evolution of the mean of the Gaussian components (blue lines) for one particular run. It is clear that the weights found by the AGSF are in agreement with the distribution of the particles. The values for Gaussian mixand weights in the conventional GSF have not changed due to the symmetry in the two Gaussian components and the quadratic measurement model. Fig. 3(c) and 3(d) show the plots for RMSE and normalized cross entropy. As expected, updating the weights of the Gaussian mixture leads to lower distance between the estimated conditional pdf and the conditional pdf given by the bootstrap PF. Hence, the adaptation of the weights during propagation leads to a more accurate conditional pdf approximation than without weight update. For this example 10,000 particles have been used for the bootstrap PF with resampling after each measurement, and the expectations needed for the quadratic programming problem have been computed using UT with $6n + 1$ sigma points.

VI. Conclusion

Two related approaches have been described to adapt the weights of Gaussian mixands between two measurements in the context of Gaussian sum filter. This is accomplished by minimizing the L_2-norm of the Fokker–Planck–Kolmogorov equation or the Chapman–Kolmogorov equation error resulting in a convex optimization problem. Although any other L_p-norm can be used to define the resulting optimization problem to find the updated weight, the L_2-norm has been considered in this work since it helps in capturing the state pdf well in high probability areas. Since the means and variances of each Gaussian kernel evolve independently, the adaptive Gaussian sum filter can easily be parallelized, and the end-to-end compute time can be managed efficiently.

The update methods presented are particularly useful in the case of pure forecast, systems characterized by temporally sparse observations, large measurement noise or unobservable systems. The numerical results presented in this technical note serve to illustrate the fact that a better approximation to the conditional pdf can be achieved by updating the forecast weights between two measurements. Future efforts are focused on studying the methods for automatic selection of Gaussian kernels and use of L_1-norm to define the optimization problem.

References

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME-J. Basic Eng.*, ser. D, vol. 82, pp. 35–45, 1960.
[2] H. J. Kushner, "Dynamical equations for optimal nonlinear filter," *J. Differential Equations*, vol. 3, no. 2, pp. 179–190, 1967.
[3] H. J. Kushner, "Approximations to optimal nonlinear filters," *IEEE Trans. Autom. Control*, vol. AC-12, no. 5, pp. 546–556, Oct. 1967.
[4] H. W. Sorenson and A. R. Stubberud, "Non-linear filtering by approximation of the a posteriori density," *Int. J. Control*, vol. 8, no. 1, pp. 33–51, 1968.
[5] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, R. Bellman, Ed. New York: Academic Press, 1970.
[6] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Autom. Control*, vol. 45, no. 5, pp. 910–927, May 2000.
[7] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.
[8] B. D. Anderson and J. B. Moore, *Optimal Filtering*, T. Kailath, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1979.
[9] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2001.
[10] F. Daum and J. Huang, "Curse of dimensionality and particle filters," in *Proc. IEEE Aerosp. Conf.*, Mar. 8–15, 2003, vol. 4, pp. 1979–1993.
[11] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465–479, 1971.
[12] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Autom. Control*, vol. AC-17, no. 4, pp. 439–448, Aug. 1972.
[13] R. Chen and J. Liu, "Mixture Kalman filters," *J. Royal Stat. Soc. B* vol. 62, pp. 493–508, 2000 [Online]. Available: citeseer.ist.psu.edu/chen00mixture.html
[14] J. Kotecha and P. Djuric, "Gaussian sum particle filtering," *IEEE Trans. Signal Processing*, vol. 51, no. 10, pp. 2602–2612, Oct. 2003.
[15] H. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Autom. Control*, vol. AC-33, no. 8, pp. 780–783, Aug. 1988.

[16] W. I. Tam, K. N. Plataniotis, and D. Hatzinakos, "An adaptive Gaussian sum algorithm for radar tracking," *Signal Processing*, vol. 77, pp. 85–104, 1999.
[17] D. Musicki, B. L. Scala, and R. Evans, "The integrated track splitting filter—Efficient multi-scan single target tracking in clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 4, pp. 1409–1425, Jul. 2007.
[18] P. Brasnett, L. Mihaylova, N. Canagarajah, and D. Bull, "Particle filtering with multiple cues for object tracking in video sequences," in *Proc. SPIE's Annu. Symp. EI ST, 5685*, 2005 [Online]. Available: citeseer.ist.psu.edu/brasnett05particle.html
[19] B. Han, Y. Zhu, D. Comaniciu, and L. Davis, "Kernel-based Bayesian filtering for object tracking," in *Proc. IEEE Conf. Comp. Vision Pattern Recog.*, 2005, pp. 227–234.
[20] I. Santamaria, C. Pantaleon, J. Ibanez, and A. Artes, "Deconvolution of seismic data using adaptive Gaussian mixtures," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 855–859, Mar. 1999.
[21] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "A novel Gaussian sum filter method for accurate solution to nonlinear filtering problem," in *Proc. 11th Int. Conf. Inform. Fusion*, Cologne, Germany, Jun. 2008, pp. 1–8.
[22] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*. New York: Springer, 1989.
[23] M. Kumar, S. Chakravorty, P. Singla, and J. L. Junkins, "The partition of unity finite element approach with hp-refinement for the stationary Fokker-Planck equation," *J. Sound Vibr.*, vol. 327, no. 1–2, pp. 144–162, 2009.
[24] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "Uncertainty propagation for nonlinear dynamical systems using Gaussian mixture models," *AIAA J. Guid., Control, Dynam.*, vol. 31, pp. 1623–1633, 2008.
[25] D. Giza, P. Singla, and M. Jah, "An approach for nonlinear uncertainty propagation: Application to orbital mechanics," in *Proc. AIAA Guid. Navig. Control Conf.*, Chicago, IL, Aug. 10–13, 2009, pp. 1–19.
[26] M. Kozlov, S. Tarasov, and L. Khachiyan, "Polynomial solvability of convex quadratic programming," *Doklady Akademiia Nauk SSSR*, vol. 248, pp. 1108–1111, 1979.
[27] G. Terejanu, "An adaptive split-merge scheme for uncertainty propagation using Gaussian mixture models," in *Proc. 49th AIAA Aerosp. Sci. Meeting*, Orlando, FL, 2011, pp. 1–10.

Optimal Smoothing for Finite State Hidden Reciprocal Processes

Langford B White, *Senior Member, IEEE*, and Francesco Carravetta

***Abstract*—This technical note addresses modelling and estimation of a class of finite state random processes called hidden reciprocal chains (HRC). A hidden reciprocal chain consists of a finite state reciprocal process, together with an observation process conditioned on the reciprocal process much as in the case of a hidden Markov model (HMM). The key difference between Markov models and reciprocal models is that reciprocal models are non-causal. The technical note presents a characterization of a HRC by a finite set of hidden Markov bridges, which are HMMs with the final state fixed. The technical note then uses this characterization to derive the optimal fixed interval smoother for a HRC. Performance of linear and optimal smoothers derived for both HMM and HRC are compared (using simulations) for a class of HRC derived from underlying Markov transitions. These experiments suggest that, not surprisingly, the performance of the optimal HMM and HRC smoothers are signifcantly better than their linear counterparts, and that some performance improvement is obtained using the HRC smoothers compared to the HMM smoothers. The technical note concludes by mentioning some ongoing and future work which exploits this new Markov bridge characterization of a HRC.**

***Index Terms*—Finite state systems, hidden Markov models (HMMs), Markov processes, optimal smoothing, reciprocal processes (RP).**

Manuscript received October 07, 2010; accepted April 04, 2011. Date of publication April 11, 2011; date of current version September 08, 2011. Recommended by Associate Editor Z. Wang.

L. B. White is with the School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide 5067, Australia (e-mail: lang.white@adelaide.edu.au).

F. Carravetta is with the Insituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti," Consiglio Nazionale della Ricerche, Roms 00185, Italy (e-mail: francesco.carravetta@iasi.cnr.it; carravetta@iasi.cnr.it).

Digital Object Identifier 10.1109/TAC.2011.2141510

I. Introduction

Reciprocal processes (RP) are one-dimensional Markov random fields (MRF), although they are not Markov processes in the usual sense. However, any Markov process is reciprocal. Reciprocal processes share the fundamental property of MRFs in that their statistical properties may be specified by a *nearest neighbour* condition. Reciprocal processes were originally studied in detail in the 1930s, in particular by Bernstein [1] who formulated the concept of a RP, and also by Schrödinger [2], who studied the quantum mechanical behavior of an electron on a finite interval of the real line. The study of reciprocal processes became prominent in the 1960s and 1970s, when several researchers such as Slepian [3] and Jamison [4] studied Gaussian RPs. The general theory of RPs was presented by Jamison in [5] where important relationships between RPs and Markov processes were established.

In the last 2 decades of the twentieth century, a considerable amount of related work was published by Krener [6], Krener, *et al.* [7] and Adams, *et al.* [8] in particular. These papers dealt specifically with continuous time reciprocal processes with the underlying models being constructed from stochastic differential equations. Optimal linear estimation was addressed in [8]. An interesting example which demonstrates one significant application of RPs is described in [9] which showed how tracking with predictive information may be dealt with in the RP framework. That paper also dealt with the continuous time, Gaussian case. We also note related work in physics also by Levy and Krener [10], [11] where relationships between reciprocal processes and quantum mechanics are explored.

In [12], a complete description of discrete time Gaussian reciprocal processes from a modelling and optimal estimation perspective is given by Levy, *et al.*. In particular, it is shown that any Gaussian reciprocal process can be represented by a second order difference equation driven by a Gaussian moving average (MA) noise process of order one. A global matrix-vector equation describing the state of the process is derived and a forward-backward procedure for generating realizations of the process in a causal manner is obtained from the global equation via LU factorization. Using a similar approach, the optimal fixed-interval smoother is obtained, and is realized by a similar forward-backward method, not unlike the familiar fixed-interval smoother for Gauss–Markov processes.

More recently, Vats and Moura [13] also considered the Gaussian reciprocal case. In their approach, they also used an LU decomposition of a global model to derive a forward-backward representation which, unlike [12], incorporated the boundary conditions explicitly into these recursions. Optimal smoothing was also addressed in [13].

This technical note is the first to be concerned with *finite state* discrete time reciprocal processes, although non-recursive models for finite state MRFs in two dimensions were considered in [14]. In the sequel, we shall sometimes refer to a finite state reciprocal process as a *reciprocal chain*. Analogously to a hidden Markov chain (HMC), a HRC consists of a reciprocal chain, and an observation process statistically dependent on the RC. The main contributions of this technical