TAMPEERE TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Matematiikan laitos. Tutkimusraportti 89
Institute of Mathematics. Research Report 89

Simo Ali-Löytty

On the Convergence of the Gaussian Mixture Filter

Tampere 2008

Tampereen teknillinen yliopisto. Matematiikan laitos.
Tutkimusraportti 89
Tampere University of Technology. Institute of Mathematics.
Research Report 89

Simo Ali-Löytty

On the Convergence of the Gaussian Mixture Filter

Tampereen teknillinen yliopisto. Matematiikan laitos
Tampere 2008

ISBN 978-952-15-2097-6 (printed)
ISBN 978-952-15-2568-1 (PDF)
ISSN 1459-3750

On the Convergence of the Gaussian Mixture Filter

Simo Ali-Löytty

***Abstract*—This paper presents convergence results for the Box Gaussian Mixture Filter (BGMF). BGMF is a Gaussian Mixture Filter (GMF) that is based on a bank of Extended Kalman Filters. The critical part of GMF is the approximation of probability density function (pdf) as pdf of Gaussian mixture such that its components have small enough covariance matrices. Because GMF approximates prior and posterior as Gaussian mixture it is enough if we have a method to approximate arbitrary Gaussian (mixture) as a Gaussian mixture such that the components have small enough covariance matrices. In this paper, we present the Box Gaussian Mixture Approximation (BGMA) that partitions the state space into specific boxes and matches weights, means and covariances of the original Gaussian in each box to a GM approximation. If the original distribution is Gaussian mixture, BGMA does this approximation separately for each component of the Gaussian mixture. We show that BGMA converges weakly to the original Gaussian (mixture). When we apply BGMA in a Gaussian mixture filtering framework we get BGMF. We show that GMF, and also BGMF, converges weakly to the correct/exact posterior distribution.**

***Index Terms*—Extended Kalman Filter, Filter banks, Filtering techniques, Filtering theory, Gaussian distribution**

I. Introduction

The problem of estimating the state of a stochastic system from noisy measurement data is considered. We consider the discrete-time nonlinear non-Gaussian system

$$x_k = F_{k-1}x_{k-1} + w_{k-1}, \tag{1a}$$

$$y_k = h_k(x_k) + v_k, \tag{1b}$$

where the vectors $x_k \in \mathbb{R}^{n_x}$ and $y_k \in \mathbb{R}^{n_{y_k}}$ represent the state of the system and the measurement at time t_k, $k \in \mathbb{N}\backslash\{0\}$, respectively. The state transition matrix F_{k-1} is assumed to be non-singular. We assume that errors w_k and v_k are white, mutually independent and independent of the initial state x_0. The errors as well as the initial state are assumed to have Gaussian mixture distributions. We assume that initial state x_0 and measurement errors v_k have density functions p_{x_0} and p_{v_k}, respectively. We do not assume that state model errors w_k have density functions. These assumptions guarantee that the prior (the conditional probability density function given all past measurements $y_{1:k-1} \triangleq \{y_1, \ldots, y_{k-1}\}$) and the posterior (the conditional probability density function given all current and past measurements $y_{1:k} \triangleq \{y_1, \ldots, y_k\}$) have density functions $p(x_k|y_{1:k-1})$ and $p(x_k|y_{1:k})$, respectively. We use the notation $x^-_{k,\text{exact}}$ for a random variable whose density function is $p(x_k|y_{1:k-1})$ (prior) and $x^+_{k,\text{exact}}$ for a random variable whose density function is $p(x_k|y_{1:k})$ (posterior). The posterior can be determined recursively according to the following relations [1], [2].

Prediction (prior):

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})\mathrm{d}x_{k-1}; \tag{2}$$

Update (posterior):

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{\int p(y_k|x_k)p(x_k|y_{1:k-1})\mathrm{d}x_k}, \tag{3}$$

where the transitional density

$$p(x_k|x_{k-1}) = p_{w_{k-1}}(x_k - \mathrm{F}_{k-1}x_{k-1})$$

and the likelihood

$$p(y_k|x_k) = p_{v_k}(y_k - h_k(x_k)).$$

The initial condition for the recursion is given by the pdf of the initial state $p_{x_0}(x_0) \triangleq p(x_0|y_{1:0})$. Knowledge of the posterior distribution (3) enables one to compute an optimal state estimate with respect to any criterion. For example, the minimum mean-square error (MMSE) estimate is the conditional mean of x_k [2], [3]. Unfortunately, in general and in our case, the conditional probability density function cannot be determined analytically.

There are many different methods (filters) to compute the approximation of the posterior. One popular approximation is the so-called Extended Kalman Filter [2]–[10], that linearizes the measurement function around the prior mean. EKF works quite well in many applications, where the system model is almost linear and the errors Gaussian but there are plenty of examples where EKF does not work satisfactorily. For example, in satellite positioning systems, EKF works quite well, but in a positioning system based on the range measurements of nearby base stations EKF may diverge [11].

There are also other Kalman Filter extensions to the nonlinear problem, which try to compute the mean and covariance of the posterior, for example Second Order Extended Kalman Filter (EKF2) [3], [4], [11], Iterated Extended Kalman Filter (IEKF) [3], [10] and Unscented Kalman Filters (UKF) [12], [13]. These extensions usually (not always) give better performance than the conventional EKF. However, if the true posterior has multiple peaks, one-component filters that compute only the mean and covariance do not achieve good performance, and because of that we have to use more sophisticated nonlinear filters. Here sophisticated nonlinear filter mean filter that has some convergence results. Possible filters are e.g.

S. Ali-Löytty is with the Department of Mathematics, Tampere University of Technology, Finland e-mail: simo.ali-loytty@tut.fi.

a grid based method (e.g. Point Mass Filter) [2], [14]–[17], Particle Filter [1], [2], [18], [19] and Gaussian Mixture Filter (GMF) [6], [20], [21]. Some comparison of different filters may be found for example in [22], [23].

In this paper we consider Gaussian Mixture Filter, also called Gaussian Sum Filter, which is a filter whose approximate prior and posterior densities are Gaussian Mixtures (GMs), a convex combination of Gaussian densities. One motivation to use GMF is that any continuous density function p_x may be approximated as a density function of GM p_{gm} as closely as we wish in the Lissack-Fu distance sense, which is also norm in $L^1(\mathbb{R}^n)$-space [21] [24, Chapter 18]:

$$\int |p_x(x) - p_{\mathrm{gm}}(x)| \mathrm{d}x. \tag{4}$$

Because the set of all continuous functions, with compact support is dense in $L^1(\mathbb{R}^n)$ [25, Theorem 3.14], we can approximate any density function p_x as a density function of GM [26]. The outline of the conventional GMF algorithm for the system (1) is given as Algorithm 1. In Algorithm 1

Algorithm 1 Gaussian mixture filter

Approximate initial state x_0 as GM x_0^+.
for $k = 1$ to n_{meas} **do**
1) *Prediction*: Compute prior approximation x_k^-.
2) Approximate x_k^- as a new GM $\bar{x}_k^-$ if necessary.
3) *Update*: Compute GM posterior approximation $\bar{x}_k^+$.
4) Reduce the number of components of $\bar{x}_k^+$ and get x_k^+.
end for

all random variables x_0^+, x_k^-, $\bar{x}_k^-$, $\bar{x}_k^+$, and x_k^+ are GMs and approximations of the exact random variables $x_0 \stackrel{\triangle}{=} x_{0,\mathrm{exact}}^+$, $x_{k,\mathrm{exact}}^-$, $x_{k,\mathrm{exact}}^-$, $x_{k,\mathrm{exact}}^+$, and $x_{k,\mathrm{exact}}^+$, respectively. This algorithm stops at time $t_{n_{\mathrm{meas}}}$.

The major contribution of this paper is a new method to approximate a Gaussian mixture as a Gaussian mixture, such that the components have arbitrary small covariance matrices. We call this method the Box Gaussian Mixture Approximation (BGMA) (Section V). We show that BGMA converges weakly to the original GM. One big advantage of BGMA compared to other GM approximations [6], [20], [21] is that BGMA does not require that the norm of the covariance matrices approach zero when the number of mixture components increases. It is sufficient that only parts of the covariance matrices approaches zero when the number of mixture components increases. Thus, BGMA subdivides only those dimensions where we get nonlinear measurements. For example, in positioning applications, nonlinear measurements often depend only on the position. So, using BGMA, it is possible to split only position dimension into boxes instead of the whole state space, which contains usually at least the position vector and the velocity vector. This means that significantly fewer mixture components are needed than in the previous GM approximations.

Another major contribution of this paper is the proof that the general version of the Gaussian Mixture Filter converges weakly to the exact posterior distribution. Especially, the Box Gaussian Mixture Filter (BGMF), which is GMF filter (Algorithm 1) that uses BGMA in Step 2, converges weakly to the exact posterior distribution. In this work BGMF is a generalization of the filter having the same name (BGMF) in our earlier work [27].

An outline of the paper is as follows. In Section II, we study the basics of the GM. In Section III, we give the general algorithm of GMF, which is also the algorithm of BGMF. In Section IV, we present the convergence results of GMF. In Section V, we present the BGMA, show some of its properties and that it converges weakly to the original Gaussian (mixture). In Section VI, we combine the previous sections and present BGMF. Finally in Section VII, we present a small one-step simulation where we compare BGMF and a particle filter [18].

II. Gaussian Mixture

In this section, we define the Gaussian Mixture (GM) distribution and present some of its properties, such as the mean, covariance, linear transformation and sum. Because GM is a convex combination of Gaussians, we first define the Gaussian distribution.

Definition 1 (Gaussian): An n-dimensional random variable x_j is Gaussian if its characteristic function has the form

$$\varphi_{x_j}(t) = \exp\left(it^T\mu_j - \frac{1}{2}t^T\Sigma_j t \right), \tag{5}$$

where $\mu_j \in \mathbb{R}^n$ and $\Sigma_j \in \mathbb{R}^{n\times n}$ is symmetric positive semidefinite ($\Sigma_j \geq 0$)[1]. We use the abbreviation

$$x_j \sim \mathrm{N}_n(\mu_j, \Sigma_j) \quad \text{or} \quad x_j \sim \mathrm{N}(\mu_j, \Sigma_j).$$

Gaussian random variable is well defined, that is the function (5) is a proper characteristic function [28, p.297].

Theorem 2 (Mean and Covariance of Gaussian): Assume that $x_j \sim \mathrm{N}(\mu_j, \Sigma_j)$. Then $\mathrm{E}(x_j) = \mu_j$ and $\mathrm{V}(x_j) = \Sigma_j$

Proof: We use the properties of the characteristic function [29, p.34] to get

$$\begin{aligned}\mathrm{E}(x_j) &= \frac{1}{i}\left(\varphi'_{x_j}(t)|_{t=0}\right)^T \\ &= \frac{1}{i}\exp\left(it^T\mu_j - \frac{1}{2}t^T\Sigma_j t \right)(i\mu_j - \Sigma_j t)\Big|_{t=0} \\ &= \mu_j\end{aligned}$$

and

$$\begin{aligned}\mathrm{V}(x_j) &= \mathrm{E}(x_j x_j^T) - \mathrm{E}(x_j)\,\mathrm{E}(x_j)^T \\ &= -\varphi''_{x_j}(t)|_{t=0} - \mu_j\mu_j^T \\ &= -\Bigg[\left((i\mu_j - \Sigma_j t)(i\mu_j - \Sigma_j t)^T - \Sigma_j\right)\cdot\ldots \\ &\qquad \exp\left(it^T\mu_j - \frac{1}{2}t^T\Sigma_j t \right)\Bigg]\Bigg|_{t=0} - \mu_j\mu_j^T \\ &= \mu_j\mu_j^T + \Sigma_j - \mu_j\mu_j^T \\ &= \Sigma_j.\end{aligned}$$

□

[1] If $\mathrm{A} \geq \mathrm{B}$ then both matrices A and B are symmetric and $x^T(\mathrm{A}-\mathrm{B})x \geq 0$ for all x.

Theorem 3 (Density function of non-singular Gaussian): Assume that $x_j \sim \mathrm{N}(\mu_j, \Sigma_j)$, where $\Sigma_j > 0$ (positive definite matrix)[2]. Then the density function of the random variable x is

$$p_{x_j}(\xi) \stackrel{\triangle}{=} \mathrm{N}^{\mu_j}_{\Sigma_j}(\xi) = \frac{\exp\left(-\frac{1}{2}\|\xi-\mu_j\|^2_{\Sigma_j^{-1}}\right)}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\Sigma_j)}},$$

where $\|\xi-\mu_j\|^2_{\Sigma_j^{-1}} = (\xi-\mu_j)^T\,\Sigma_j^{-1}\,(\xi-\mu_j)$.

Proof: We know that the characteristic function $\varphi_{x_j}(t)$ is absolutely integrable. Thus using the properties of the characteristic function [29, p.33] we get

$$\begin{aligned}
p_{x_j}(\xi) &= \frac{1}{(2\pi)^n}\int \exp\left(-it^T\xi\right)\varphi_{x_j}(t)\mathrm{d}t \\
&= \frac{1}{(2\pi)^n}\int \exp\left(it^T(\mu_j-\xi)-\frac{1}{2}t^T\Sigma_j t\right)\mathrm{d}t \\
&= \frac{\frac{\sqrt{\det(\Sigma_j)}}{(2\pi)^{\frac{n}{2}}}\int \exp\left(it^T(\mu_j-\xi)-\frac{1}{2}t^T\Sigma_j t\right)\mathrm{d}t}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\Sigma_j)}} \\
&\stackrel{\star}{=} \frac{\exp\left(-\frac{1}{2}\left(\xi-\mu_j\right)^T\Sigma_j^{-1}\left(\xi-\mu_j\right)\right)}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\Sigma_j)}}
\end{aligned}$$

$\star$ see [28, p.297]. □

Definition 4 (Gaussian Mixture): An n-dimensional random variable x is an N-component Gaussian Mixture if its characteristic function has the form

$$\varphi_x(t) = \sum_{j=1}^{N}\alpha_j \exp\left(it^T\mu_j - \frac{1}{2}t^T\Sigma_j t\right), \tag{6}$$

where $\mu_j \in \mathbb{R}^n$, $\Sigma_j \in \mathbb{R}^{n\times n}$ is symmetric positive semidefinite, $\alpha_j \geq 0$, and $\sum_{j=1}^N \alpha_j = 1$. We use the abbreviation

$$x \sim \mathrm{M}(\alpha_j, \mu_j, \Sigma_j)_{(j,N)}.$$

We show that GM is well defined, which means that function (6) is in fact a characteristic function. First, assume that all matrices Σ_j are positive definite. We know that function

$$p(\xi) = \sum_{j=1}^{N}\alpha_j\,\mathrm{N}^{\mu_j}_{\Sigma_j}(\xi), \tag{7}$$

is a density function, that is $\int p(\xi)\mathrm{d}\xi = 1$ and $p(\xi)\geq 0$ for all ξ. Because

$$\begin{aligned}
\int \exp(it^T\xi)p(\xi)\mathrm{d}\xi &= \int \exp(it^T\xi)\left(\sum_{j=1}^{N}\alpha_j\,\mathrm{N}^{\mu_j}_{\Sigma_j}(\xi)\right)\mathrm{d}\xi \\
&= \sum_{j=1}^{N}\alpha_j\int \exp(it^T\xi)\,\mathrm{N}^{\mu_j}_{\Sigma_j}(\xi)\mathrm{d}\xi \\
&\stackrel{(5)}{=} \sum_{j=1}^{N}\alpha_j \exp\left(it^T\mu_j - \frac{1}{2}t^T\Sigma_j t\right),
\end{aligned}$$

[2]If $\mathrm{A} > \mathrm{B}$ then both matrices A and B are symmetric and $x^T(\mathrm{A}-\mathrm{B})x > 0$ for all $x \neq 0$.

function (6) is the characteristic function of a continuous n-dimensional Gaussian Mixture. The density function of this distribution is given in equation (7).

Now, let at least one of the covariance matrices Σ_j be singular. Take $\epsilon > 0$ and consider the positive definite symmetric matrices $\Sigma_j^\epsilon = \Sigma_j + \epsilon\mathrm{I}$. Then by what has been proved,

$$\varphi_{x_\epsilon}(t) = \sum_{j=1}^{N}\alpha_j \exp\left(it^T\mu_j - \frac{1}{2}t^T\Sigma_j^\epsilon t\right)$$

is a characteristic function. Because function (6) is the limit of characteristic functions

$$\lim_{\epsilon\to 0}\varphi_{x_\epsilon}(t) = \sum_{j=1}^{N}\alpha_j \exp\left(it^T\mu_j - \frac{1}{2}t^T\Sigma_j t\right),$$

and it is continuous at $t = 0$, then this function (6) is a characteristic function [28, p.298].

Theorem 5 (Mean and Covariance of mixture): Assume that

$$\varphi_x(t) = \sum_{j=1}^{N}\alpha_j\varphi_{x_j}(t)$$

where $\mathrm{E}(x_j) = \mu_j \in \mathbb{R}^n$, $\mathrm{V}(x_j) = \Sigma_j \in \mathbb{R}^{n\times n}$, $\alpha_j \geq 0$, and $\sum_{j=1}^N \alpha_j = 1$. Then

$$\mathrm{E}(x) = \sum_{j=1}^{N}\alpha_j\mu_j \stackrel{\triangle}{=} \mu \quad \text{and}$$

$$\mathrm{V}(x) = \sum_{j=1}^{N}\alpha_j\left(\Sigma_j + (\mu_j-\mu)(\mu_j-\mu)^T\right).$$

Proof: We use the properties of the characteristic function [29, p.34] to get

$$\begin{aligned}
\mathrm{E}(x) &= \frac{1}{i}\left(\varphi'_x(t)|_{t=0}\right)^T \\
&= \sum_{j=1}^{N}\alpha_j\frac{1}{i}\left(\varphi'_{x_j}(t)|_{t=0}\right)^T \\
&= \sum_{j=1}^{N}\alpha_j\mu_j \stackrel{\triangle}{=} \mu
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{V}(x) &= -\mathrm{E}(x)\,\mathrm{E}(x)^T - \varphi''_x(t)|_{t=0} \\
&= -\mu\mu^T + \sum_{j=1}^{N}\alpha_j\left(-\varphi''_{x_j}(t)|_{t=0}\right) \\
&= -\mu\mu^T + \sum_{j=1}^{N}\alpha_j\left(\Sigma_j + \mu_j\mu_j^T\right) \\
&= \sum_{j=1}^{N}\alpha_j\left(\Sigma_j + \mu_j\mu_j^T - \mu\mu^T\right) \\
&= \sum_{j=1}^{N}\alpha_j\left(\Sigma_j + (\mu_j-\mu)(\mu_j-\mu)^T\right).
\end{aligned}$$

□

Note that Theorem 5 does not assume that the distribution is a Gaussian mixture, these results are valid for all mixtures.

Theorem 6 (Linear transformation and sum of GM): Assume that an n-dimensional random variable

$$x \sim \mathrm{M}(\alpha_j, \mu_j, \Sigma_j)_{(j,N)}$$

and an m-dimensional random variable

$$v \sim \mathrm{M}(\beta_k, r_k, \mathrm{R}_k)_{(k,M)}$$

are independent. Define a random variable $y = \mathrm{H}x + v$, where matrix $\mathrm{H} \in \mathbb{R}^{m\times n}$. Then

$$y \sim \mathrm{M}(\alpha_{j(l)}\beta_{k(l)}, \mathrm{H}\mu_{j(l)} + r_{k(l)}, \mathrm{H}\Sigma_{j(l)}\mathrm{H}^T + \mathrm{R}_{k(l)})_{(l,NM)},$$

where $j(l) = [(l-1) \mod N]+1$ and $k(l) = \lceil \frac{l}{N} \rceil$.[3] We also use the abbreviation

$$y \sim \mathrm{M}(\alpha_j\beta_k, \mathrm{H}\mu_j + r_k, \mathrm{H}\Sigma_j\mathrm{H}^T + \mathrm{R}_k)_{(j*k,NM)}.$$

Proof: Since x and v are independent, also $\mathrm{H}x$ and v are independent.

$$\begin{aligned}
\varphi_{\mathrm{H}x+v}(t) &\overset{\text{ind.}}{=} \varphi_{\mathrm{H}x}(t)\varphi_v(t) \\
&= \mathrm{E}\left(\exp\left(it^T(\mathrm{H}x)\right)\right)\varphi_v(t) \\
&= \mathrm{E}\left(\exp\left(i\left(\mathrm{H}^Tt\right)^T x\right)\right)\varphi_v(t) \\
&= \varphi_x(\mathrm{H}^Tt)\varphi_v(t) \\
&= \sum_{j=1}^{N}\alpha_j \exp\left(it^T\mathrm{H}\mu_j - \frac{1}{2}t^T\mathrm{H}\Sigma_j\mathrm{H}^Tt\right)\cdots \\
&\quad \sum_{k=1}^{M}\beta_k \exp\left(it^Tr_k - \frac{1}{2}t^T\mathrm{R}_kt\right) \\
&= \sum_{l=1}^{NM}\alpha_{j(l)}\beta_{k(l)} \exp\left(it^T\left(\mathrm{H}\mu_{j(l)} + r_{k(l)}\right)\cdots\right. \\
&\qquad \left. - \frac{1}{2}t^T\left(\mathrm{H}\Sigma_{j(l)}\mathrm{H}^T + \mathrm{R}_{k(l)}\right)t\right).
\end{aligned}$$

□

Corollary 7: Assume that an n-dimensional random variable

$$x \sim \mathrm{M}(\alpha_j, \mu_j, \Sigma_j)_{(j,N)}$$

and

$$y = \mathrm{A}x + b,$$

where $\mathrm{A} \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Then

$$y \sim \mathrm{M}(\alpha_j, \mathrm{A}\mu_j + b, \mathrm{A}\Sigma_j\mathrm{A}^T)_{(j,N)}.$$

Proof: Now $b \sim \mathrm{M}(1, b, 0)_{(k,1)}$. Constant random variable b and x are independent, so using Theorem 6 we get

$$y \sim \mathrm{M}(\alpha_j, \mathrm{A}\mu_j + b, \mathrm{A}\Sigma_j\mathrm{A}^T)_{(j,N)}.$$

□

Note that if $x \sim \mathrm{N}(\mu_1, \Sigma_1)$ then $x \sim \mathrm{M}(1, \mu_j, \Sigma_j)_{(j,1)}$. So Theorem 6 and Corollary 7 hold also for Gaussian distributions.

[3] Ceiling function $\lceil x \rceil = \min\{n \in \mathbb{Z} | n \geq x\}$ and modulo function ($a \mod n) = a + n\lceil -\frac{a}{n} \rceil$.

III. Algorithm of Gaussian Mixture Filter

In this section, we give the algorithm of Gaussian Mixture Filter for the system (1) (Algorithm 2). The subsections III-A–III-D present the details of this algoritm. Algorithm 2 uses the following assumptions:

1) Initial state
$$x_0 \sim \mathrm{M}(\alpha^+_{i,0}, \mu^+_{i,0}, \Sigma^+_{i,0})_{(i,n_0)}$$
is a continuous Gaussian Mixture, that is, $\Sigma^+_{i,0} > 0$ for all i.
2) Errors are GMs
$$\begin{aligned} w_k &\sim \mathrm{M}(\gamma_{j,k}, \bar{w}_{j,k}, \mathrm{Q}_{j,k})_{(j,n_{w_k})} \text{ and} \\ v_k &\sim \mathrm{M}(\beta_{j,k}, \bar{v}_{j,k}, \mathrm{R}_{j,k})_{(j,n_{v_k})}, \end{aligned}$$
where all $\mathrm{R}_{j,k} > 0$.
3) Measurement functions are of the form
$$h_k(x) = \bar{h}_k(x_{1:d}) + \bar{\mathrm{H}}_kx. \tag{8}$$
This means that the nonlinear part $\bar{h}_k(x_{1:d})$ only depends on the first d dimensions ($d \leq n_x$). We assume that functions $\bar{h}_k(x_{1:d})$ are twice continuously differentiable in $\mathbb{R}^d \setminus \{s_1, \ldots, s_{n_s}\}$.[4]

Algorithm 2 Gaussian mixture filter

Initial state at time t_0: $x^+_0 \sim \mathrm{M}(\alpha^+_{i,0}, \mu^+_{i,0}, \Sigma^+_{i,0})_{(i,n_0)}$
for $k = 1$ to n_{meas} **do**

1) Prediction (see Sec. III-A):
$$x^-_k \sim \mathrm{M}(\alpha^-_{i*j,k}, \mu^-_{i*j,k}, \Sigma^-_{i*j,k})_{(i*j,n^-_k)}$$
2) Approximate x^-_k as a new GM $\bar{x}^-_k$ if necessary (see Sec. III-B):
$$\bar{x}^-_k \sim \mathrm{M}(\bar{\alpha}^-_{i,k}, \bar{\mu}^-_{i,k}, \bar{\Sigma}^-_{i,k})_{(i,\bar{n}^-_k)}$$
3) Update (see Sec. III-C):
$$\bar{x}^+_k \sim \mathrm{M}(\bar{\alpha}^+_{i*j,k}, \bar{\mu}^+_{i*j,k}, \bar{\Sigma}^+_{i*j,k})_{(i*j,\bar{n}^+_k)}$$
4) Reduce the number of components (see Sec. III-D):
$$x^+_k \sim \mathrm{M}(\alpha^+_{i,k}, \mu^+_{i,k}, \Sigma^+_{i,k})_{(i,n_k)}$$

end for

A. Prediction, Step (1)

Prediction is based on Eq. (1a) and Thm. 6 (see also Eq. (2)).

$$x^-_k \sim \mathrm{M}(\alpha^-_{i*j,k}, \mu^-_{i*j,k}, \Sigma^-_{i*j,k})_{(i*j,n^-_k)},$$

where

$$\begin{aligned}
n^-_k &= n_{k-1}n_{w_{k-1}}, \\
\alpha^-_{i*j,k} &= \alpha^+_{i,k-1}\gamma_{j,k-1}, \\
\mu^-_{i*j,k} &= \mathrm{F}_{k-1}\mu^+_{i,k-1} + \bar{w}_{j,k-1} \text{ and} \\
\Sigma^-_{i*j,k} &= \mathrm{F}_{k-1}\Sigma^+_{i,k-1}\mathrm{F}_{k-1}{}^T + \mathrm{Q}_{j,k-1}.
\end{aligned}$$

[4] For example, in positioning applications that are based on range measurements and a constant velocity model $n_x = 6$ (position+velocity), $d = 3$ (position) and s_i is position vector of the ith base station [11], [27]

B. Approximate GM as a new GM, Step (2)

There are different methods to compute Step (2). Here we present one conventional method. Another method, namely, the Box Gaussian Mixture Approximation, is given in Section V. The density function of a new GM approximation $p_{\bar{x}_k^-}$ is [20]

$$p_{\bar{x}_k^-}(\xi) \propto \sum_{i=1}^{\bar{n}_{k,\mathrm{g}}^-} p_{x_k^-}(\xi_{\mathrm{g}}^{(i)})\, \mathrm{N}_{c_{\mathrm{g}}\mathrm{I}}^{\xi_{\mathrm{g}}^{(i)}}(\xi), \tag{9}$$

where the mean values $\xi_{\mathrm{g}}^{(i)}$ are used to establish a grid in the region of the state space that contains the significant part of the probability mass, $\bar{n}_{k,\mathrm{g}}^-$ is the number of grid points and $c_{\mathrm{g}} > 0$ is determined such that the error in the approximation, e.g. the Lissack-Fu distance (4), is minimized. So

$$\bar{x}_k^- \sim \mathrm{M}(\bar{\alpha}_{i,k}^-, \bar{\mu}_{i,k}^-, \bar{\Sigma}_{i,k}^-)_{(i,\bar{n}_k^-)},$$

where

$$\begin{aligned} \bar{n}_k^- &= \bar{n}_{k,\mathrm{g}}^-, \\ \bar{\alpha}_{i,k}^- &= \frac{p_{x_k^-}(\xi_{\mathrm{g}}^{(i)})}{\sum_{i=1}^{\bar{n}_{k,\mathrm{g}}^-} p_{x_k^-}(\xi_{\mathrm{g}}^{(i)})}, \\ \bar{\mu}_{i,k}^- &= \xi_{\mathrm{g}}^{(i)} \text{ and} \\ \bar{\Sigma}_{i,k}^- &= c_{\mathrm{g}}\mathrm{I}. \end{aligned}$$

It can be shown that $p_{\bar{x}_k^-}(x)$ converges almost everywhere uniformly to the density function of x_k^- as the number of components $\bar{n}_k^-$ increases and c_{g} approaches zero [20], [21]. Moreover, the Lissack-Fu distance (4) of the approximation converges to zero.

Step (2) is executed only when necessary. If it is not necessary then $\bar{x}_k^- = x_k^-$. A conventional criterion is to check if some prior covariances do not satisfy inequality $\mathrm{P}_i^- < \epsilon\mathrm{I}$, for some predefined ϵ, where P_i^- is the covariance of the ith component [6, p.216]. Note that finding reasonable grid points $\xi_{\mathrm{g}}^{(i)}$ and an optimal constant $c_{\mathrm{g}} > 0$ usually requires some heavy computation.

C. Update, Step 3

The update Eq. (3) is usually computed approximately using a bank of EKFs. In this paper we use that approximation. It is possible to compute the update step using a bank of other Kalman-type filters [30] or a bank of PFs [31]. Using the bank of EKFs approximation we get

$$\begin{aligned} & p_{\bar{x}_k^+}(\xi) \propto p_{v_k}(y_k - h_k(\xi)) p_{\bar{x}_k^-}(\xi) \\ &= \sum_{j=1}^{n_{v_k}} \sum_{i=1}^{\bar{n}_k^-} \beta_{j,k}\, \mathrm{N}_{\mathrm{R}_{j,k}}^{\bar{v}_{j,k}}(y_k - h_k(\xi))\, \bar{\alpha}_{i,k}^-\, \mathrm{N}_{\bar{\Sigma}_{i,k}^-}^{\bar{\mu}_{i,k}^-}(\xi) \\ &\approx \sum_{j=1}^{n_{v_k}} \sum_{i=1}^{\bar{n}_k^-} \bar{\alpha}_{i,k}^- \beta_{j,k}\, \mathrm{N}_{\bar{\Sigma}_{i,k}^-}^{\bar{\mu}_{i,k}^-}(\xi) \cdots \\ &\quad \mathrm{N}_{\mathrm{R}_{j,k}}^{\bar{v}_{j,k}}\left(y_k - h_k(\bar{\mu}_{i,k}^-) - \mathrm{H}_{i,k}(\xi - \bar{\mu}_{i,k}^-)\right) \\ &= \sum_{j=1}^{n_{v_k}} \sum_{i=1}^{\bar{n}_k^-} \bar{\alpha}_{i,k}^- \beta_{j,k}\, \mathrm{N}_{\bar{\Sigma}_{i,k}^-}^{\bar{\mu}_{i,k}^-}(\xi) \cdots \\ &\quad \mathrm{N}_{\mathrm{R}_{j,k}}^{\mathrm{H}_{i,k}\xi}\left(y_k - h_k(\bar{\mu}_{i,k}^-) + \mathrm{H}_{i,k}\bar{\mu}_{i,k}^- - \bar{v}_{j,k}\right) \\ &\overset{\text{Thm. 25}}{=} \sum_{j=1}^{n_{v_k}} \sum_{i=1}^{\bar{n}_k^-} \bar{\alpha}_{i,k}^- \beta_{j,k}\, \mathrm{N}_{\bar{\Sigma}_{i*j,k}^+}^{\bar{\mu}_{i*j,k}^+}(\xi) \cdots \\ &\quad \mathrm{N}_{\mathrm{H}_{i,k}\bar{\Sigma}_{i,k}^-\mathrm{H}_{i,k}^T + \mathrm{R}_{j,k}}^{\mathrm{H}_{i,k}\bar{\mu}_{i,k}^-}\left(y_k - h_k(\bar{\mu}_{i,k}^-) + \mathrm{H}_{i,k}\bar{\mu}_{i,k}^- - \bar{v}_{j,k}\right), \end{aligned} \tag{10}$$

where $\mathrm{H}_{i,k} = \frac{\partial h_k(\xi)}{\partial \xi}\big|_{\xi=\bar{\mu}_{i,k}^-}$. So

$$\bar{x}_k^+ \sim \mathrm{M}(\bar{\alpha}_{i*j,k}^+, \bar{\mu}_{i*j,k}^+, \bar{\Sigma}_{i*j,k}^+)_{(i*j,\bar{n}_k^+)}, \tag{11}$$

where

$$\begin{aligned} \bar{n}_k^+ &= n_{v_k}\bar{n}_k^-, \\ \bar{\alpha}_{i*j,k}^+ &= \frac{\bar{\alpha}_{i,k}^- \beta_{j,k}\, \mathrm{N}_{\mathrm{H}_{i,k}\bar{\Sigma}_{i,k}^-\mathrm{H}_{i,k}^T + \mathrm{R}_{j,k}}^{h_k(\bar{\mu}_{i,k}^-) + \bar{v}_{j,k}}(y_k)}{\sum_{j=1}^{n_{v_k}} \sum_{i=1}^{\bar{n}_k^-} \bar{\alpha}_{i,k}^- \beta_{j,k}\, \mathrm{N}_{\mathrm{H}_{i,k}\bar{\Sigma}_{i,k}^-\mathrm{H}_{i,k}^T + \mathrm{R}_{j,k}}^{h_k(\bar{\mu}_{i,k}^-) + \bar{v}_{j,k}}(y_k)}, \\ \bar{\mu}_{i*j,k}^+ &= \bar{\mu}_{i,k}^- + \mathrm{K}_{i*j,k}(y_k - h_k(\bar{\mu}_{i,k}^-) - \bar{v}_{j,k}), \\ \bar{\Sigma}_{i*j,k}^+ &= (\mathrm{I} - \mathrm{K}_{i*j,k}\mathrm{H}_{i,k})\, \bar{\Sigma}_{i,k}^- \text{ and} \\ \mathrm{K}_{i*j,k} &= \bar{\Sigma}_{i,k}^- \mathrm{H}_{i,k}^T \left(\mathrm{H}_{i,k}\bar{\Sigma}_{i,k}^-\mathrm{H}_{i,k}^T + \mathrm{R}_{j,k}\right)^{-1}. \end{aligned}$$

D. Reduce the number of components, Step 4

One major challenge when using GMF efficiently is keeping the number of components as small as possible without losing significant information. There are many ways to do so. We use two different types of mixture reduction algorithms: forgetting and merging [21], [30], [32].

1) Forgetting components: We re-index the posterior approximation $\bar{x}_k^+$ Eq. (11) such that

$$\bar{x}_k^+ \sim \mathrm{M}(\bar{\alpha}_{i,k}^+, \bar{\mu}_{i,k}^+, \bar{\Sigma}_{i,k}^+)_{(i,\bar{n}_k^+)},$$

where $\bar{\alpha}_{i,k}^+ \geq \bar{\alpha}_{i+1,k}^+$. Let $\epsilon_{\mathrm{f}} = \frac{1}{2N}$ be the threshold value. Let $\bar{n}_{k,\mathrm{f}}^+$ be the index such that

$$\sum_{i=1}^{\bar{n}_{k,\mathrm{f}}^+} \bar{\alpha}_{i,k}^+ \geq 1 - \epsilon_{\mathrm{f}}$$

We forget all mixture components whose index $i > \bar{n}_{k,\mathrm{f}}^+$ and after normalization we get $\bar{x}_{k,\mathrm{f}}^+$. Now

$$\bar{x}_{k,\mathrm{f}}^+ \sim \mathrm{M}(\bar{\alpha}_{i,k,\mathrm{f}}^+, \bar{\mu}_{i,k,\mathrm{f}}^+, \bar{\Sigma}_{i,k,\mathrm{f}}^+)_{(i,\bar{n}_{k,\mathrm{f}}^+)}, \tag{12}$$

where

$$\bar{\alpha}^+_{i,k,\mathrm{f}} = \frac{\bar{\alpha}^+_{i,k}}{\sum_{j=1}^{\bar{n}^+_{k,\mathrm{f}}} \bar{\alpha}^+_{j,k}},\ \bar{\mu}^+_{i,k,\mathrm{f}} = \bar{\mu}^+_{i,k} \text{ and } \bar{\Sigma}^+_{i,k,\mathrm{f}} = \bar{\Sigma}^+_{i,k}.$$

2) Merging components: Our merging procedure is iterative. We merge two components, say the i_1th component and the i_2th component, into one component using moment matching method if they are sufficiently similar, that is if (for simplicity we suppress indices k and f) both

$$\|\bar{\mu}^+_{i_1} - \bar{\mu}^+_{i_2}\| \leq \epsilon_{\mathrm{m}_1} \text{ and} \tag{13a}$$
$$\|\bar{\Sigma}^+_{i_1} - \bar{\Sigma}^+_{i_2}\| \leq \epsilon_{\mathrm{m}_2} \tag{13b}$$

inequalities hold. Here we assume that the threshold values $\epsilon_{\mathrm{m}_1} \xrightarrow[N\to\infty]{} 0$ and $\epsilon_{\mathrm{m}_2} \xrightarrow[N\to\infty]{} 0$. The new component, which replaces components i_1 and i_2, is a component whose weight, mean and covariance matrix are

$$\begin{aligned}
\bar{\alpha}^+_{i_1,\mathrm{m}} &= \bar{\alpha}^+_{i_1} + \bar{\alpha}^+_{i_2} \\
\bar{\mu}^+_{i_1,\mathrm{m}} &= \frac{\bar{\alpha}^+_{i_1}}{\bar{\alpha}^+_{i_1,\mathrm{m}}}\bar{\mu}^+_{i_1} + \frac{\bar{\alpha}^+_{i_2}}{\bar{\alpha}^+_{i_1,\mathrm{m}}}\bar{\mu}^+_{i_2} \text{ and} \\
\bar{\Sigma}^+_{i_1,\mathrm{m}} &= \frac{\bar{\alpha}^+_{i_1}}{\bar{\alpha}^+_{i_1,\mathrm{m}}}\left(\bar{\Sigma}^+_{i_1} + \left(\bar{\mu}^+_{i_1} - \bar{\mu}^+_{i_1,\mathrm{m}}\right)\left(\bar{\mu}^+_{i_1} - \bar{\mu}^+_{i_1,\mathrm{m}}\right)^T\right) + \ldots \\
&\quad \frac{\bar{\alpha}^+_{i_2}}{\bar{\alpha}^+_{i_1,\mathrm{m}}}\left(\bar{\Sigma}^+_{i_2} + \left(\bar{\mu}^+_{i_2} - \bar{\mu}^+_{i_1,\mathrm{m}}\right)\left(\bar{\mu}^+_{i_2} - \bar{\mu}^+_{i_1,\mathrm{m}}\right)^T\right),
\end{aligned}$$

respectively. After re-indexing (forgetting component i_2) we merge iteratively more components until there are no sufficiently similar components, components that statisfy inequalities (13). Herewith, after re-indexing, we get

$$x^+_k \sim \mathrm{M}(\alpha^+_{i,k}, \mu^+_{i,k}, \Sigma^+_{i,k})_{(i,n_k)}.$$

IV. Convergence results of GMF

In this section, we present the convergence results of GMF. First we present some well know convergence results.

Definition 8 (Weak convergence): Let x and x_N, where $N \in \mathbb{N}$, be n-dimensional random variables. We say that x_N converges (weakly) to x if

$$F_{x_N}(\xi) \xrightarrow[N\to\infty]{} F_x(\xi),$$

for all points ξ for which the cumulative density function $F_x(\xi)$ is continuous. We use the abbreviation

$$x_N \xrightarrow[N\to\infty]{\mathrm{w}} x.$$

Theorem 9: The following conditions are equivalent

1) $x_N \xrightarrow[N\to\infty]{\mathrm{w}} x$.
2) $\mathrm{E}(g(x_N)) \xrightarrow[N\to\infty]{} \mathrm{E}(g(x))$ for all continuous functions g that vanish outside a compact set.
3) $\mathrm{E}(g(x_N)) \xrightarrow[N\to\infty]{} \mathrm{E}(g(x))$ for all continuous bounded functions g.
4) $\mathrm{E}(g(x_N)) \xrightarrow[N\to\infty]{} \mathrm{E}(g(x))$ for all bounded measurable functions g such that $\mathrm{P}(x \in C(g)) = 1$, where $C(g)$ is the continuity set of g.

Proof: See, for example, the book [33, p.13]. □

Theorem 10 (Slutsky Theorems): 1) If

$$x_N \xrightarrow[N\to\infty]{\mathrm{w}} x,$$

and if $f : \mathbb{R}^n \to \mathbb{R}^k$ is such that $\mathrm{P}(x \in C(f)) = 1$, where $C(f)$ is the continuity set of f, then

$$f(x_N) \xrightarrow[N\to\infty]{\mathrm{w}} f(x).$$

2) If $\{x_N\}$ and $\{y_N\}$ are independent, and if $x_N \xrightarrow[N\to\infty]{\mathrm{w}} x$ and $y_N \xrightarrow[N\to\infty]{\mathrm{w}} y$, then

$$\begin{bmatrix} x_N \\ y_N \end{bmatrix} \xrightarrow[N\to\infty]{\mathrm{w}} \begin{bmatrix} x \\ y \end{bmatrix},$$

where x and y are taken to be independent.

Proof: See, for example, the book [33, p.39, p.42]. □

Now we show the convergence results of GMF (Algorithm 2). The outline of the convergence results of GMF is given in Algorithm 3. The details of the convergence results are given in Sections IV-A–IV-D. The initial step of Algorithm 3 is self-evident because we assume that the initial state is a Gaussian mixture. Furthermore if our (exact) initial state has an arbitrary density function it is possible to approximate it as a Gaussian mixture such that the approximation weakly converges to the exact initial state (Sec. III-B).

Algorithm 3 Outline of showing the convergence results of the Gaussian mixture filter (Algorithm 2)

Initial state: Show that $x^+_0 \xrightarrow[N\to\infty]{\mathrm{w}} x^+_{0,\mathrm{exact}}$.
for $k = 1$ to n_{meas} **show**

1) Prediction, Sec. IV-A:
$$x^+_{k-1} \xrightarrow[N\to\infty]{\mathrm{w}} x^+_{k-1,\mathrm{exact}} \Longrightarrow x^-_k \xrightarrow[N\to\infty]{\mathrm{w}} x^-_{k,\mathrm{exact}}.$$
2) Approximation, Sec. IV-B:
$$x^-_k \xrightarrow[N\to\infty]{\mathrm{w}} x^-_{k,\mathrm{exact}} \Longrightarrow \bar{x}^-_k \xrightarrow[N\to\infty]{\mathrm{w}} x^-_{k,\mathrm{exact}}.$$
3) Update, Sec. IV-C:
$$\bar{x}^-_k \xrightarrow[N\to\infty]{\mathrm{w}} x^-_{k,\mathrm{exact}} \Longrightarrow \bar{x}^+_k \xrightarrow[N\to\infty]{\mathrm{w}} x^+_{k,\mathrm{exact}}.$$
4) Reduce the number of components, Sec. IV-D:
$$\bar{x}^+_k \xrightarrow[N\to\infty]{\mathrm{w}} x^+_{k,\mathrm{exact}} \Longrightarrow x^+_k \xrightarrow[N\to\infty]{\mathrm{w}} x^+_{k,\mathrm{exact}}.$$

end for

A. Convergence results of Step 1 (prediction)

Here we show that if $x^+_{k-1} \xrightarrow[N\to\infty]{\mathrm{w}} x^+_{k-1,\mathrm{exact}}$ then $x^-_k \xrightarrow[N\to\infty]{\mathrm{w}} x^-_{k,\mathrm{exact}}$ (Thm. 11).

Theorem 11 (Prediction convergence): If

$$x^+_{k-1} \xrightarrow[N\to\infty]{\mathrm{w}} x^+_{k-1,\mathrm{exact}},$$

w_{k-1} and $\{x^+_{k-1,N}|N\in\mathbb{N}\}$[5] are independent, and w_{k-1} and $x^+_{k-1,\text{exact}}$ are independent then

$$x^-_k \xrightarrow[N\to\infty]{\text{w}} x^-_{k,\text{exact}}.$$

Proof: Because w_{k-1} and $\{x^+_{k-1,N}|N\in\mathbb{N}\}$ are independent then w_{k-1} and $\{\mathrm{F}_{k-1}x^+_{k-1,N}|N\in\mathbb{N}\}$ are independent. From Thm. 10 we see that

$$\mathrm{F}_{k-1}x^+_{k-1} \xrightarrow[N\to\infty]{\text{w}} \mathrm{F}_{k-1}x^+_{k-1,\text{exact}}$$

and

$$\begin{bmatrix} \mathrm{F}_{k-1}x^+_{k-1} \\ w_{k-1}\end{bmatrix} \xrightarrow[N\to\infty]{\text{w}} \begin{bmatrix} \mathrm{F}_{k-1}x^+_{k-1,\text{exact}} \\ w_{k-1}\end{bmatrix}.$$

Because

$$x^-_k = \begin{bmatrix} \mathrm{I} & \mathrm{I}\end{bmatrix}\begin{bmatrix} \mathrm{F}_{k-1}x^+_{k-1} \\ w_{k-1}\end{bmatrix} \text{ and}$$
$$x^-_{k,\text{exact}} = \begin{bmatrix} \mathrm{I} & \mathrm{I}\end{bmatrix}\begin{bmatrix} \mathrm{F}_{k-1}x^+_{k-1,\text{exact}} \\ w_{k-1}\end{bmatrix}$$

it follows that

$$x^-_k \xrightarrow[N\to\infty]{\text{w}} x^-_{k,\text{exact}}.$$

□

B. Convergence results of Step 2 (approximation)

Here we show that if $x^-_k \xrightarrow[N\to\infty]{\text{w}} x^-_{k,\text{exact}}$ then $\bar{x}^-_k \xrightarrow[N\to\infty]{\text{w}} x^-_{k,\text{exact}}$. It is enough to show that

$$F_{x^-_k}(\xi) - F_{\bar{x}^-_k}(\xi) \xrightarrow[N\to\infty]{} 0,$$

for all ξ. If we use the conventional approximation method see Sec. III-B and if we use the new method (BGMA) see Thm. 21 and Corollary 22.

Furthermore, we require that the most of the covariance matrices $\bar{\Sigma}^-_{k,i,N}$ of the components of our GM approximation $\bar{x}^-_{k,N}$ are arbitrary small. That is if $\epsilon>0$ then there is N_0 such that for all $N>N_0$

$$\sum_{j=1}^{d}\left(\bar{\Sigma}^-_{k,i,N}\right)_{j,j} < \epsilon, \tag{14}$$

for almost all i. Both the conventional approximation (Sec. III-B) and BGMA (Sec. V and Corollary 20) satisfy this requirement.

C. Convergence results of Step 3 (update)

Here we show that if $\bar{x}^-_k \xrightarrow[N\to\infty]{\text{w}} x^-_{k,\text{exact}}$ then $\bar{x}^+_k \xrightarrow[N\to\infty]{\text{w}} x^+_{k,\text{exact}}$. The distribution $\bar{x}^+_k$ is computed from the prior approximation $\bar{x}^-_k$ using the bank of EKF approximations (Sec. III-C). We use the abbreviation $\bar{x}^{+,\text{Bayes}}_k$ for the distribution that is obtained from the prior approximation $\bar{x}^-_k$ using the exact update Eq. (3) (see also Eq. (15)). First we show that if $\bar{x}^-_k \xrightarrow[N\to\infty]{\text{w}} x^-_{k,\text{exact}}$ then $\bar{x}^{+,\text{Bayes}}_k \xrightarrow[N\to\infty]{\text{w}} x^+_{k,\text{exact}}$ (Thm. 12). After that it is enough to show that

$$F_{\bar{x}^{+,\text{Bayes}}_k}(\xi) - F_{\bar{x}^+_k}(\xi) \xrightarrow[N\to\infty]{} 0,$$

for all ξ (Thm. 13).

Theorem 12 (Correct posterior convergence): Assume that

$$\bar{x}^-_k \xrightarrow[N\to\infty]{\text{w}} x^-_{k,\text{exact}},$$

and the density functions of $\bar{x}^-_k$ and $x^-_{k,\text{exact}}$ are $p_{\bar{x}^-_k}(\xi)$ and $p_{x^-_{k,\text{exact}}}(\xi)$, respectively. Now

$$\bar{x}^{+,\text{Bayes}}_k \xrightarrow[N\to\infty]{\text{w}} x^+_{k,\text{exact}}.$$

Proof: Using the assumptions and Thm. 9 we get that

$$\int p(y_k|\xi)p_{\bar{x}^-_{k,N}}(\xi)\mathrm{d}\xi \xrightarrow[N\to\infty]{} \int p(y_k|\xi)p_{x^-_{k,\text{exact}}}(\xi)\mathrm{d}\xi,$$

where the likelihood $p(y_k|\xi)=p_{v_k}(y_k-h_k(\xi))$. Furthermore, all these integrals are positive because

$$p(y_k|\xi)p_{\bar{x}^-_{k,N}}(\xi)>0 \text{ and } p(y_k|\xi)p_{x^-_{k,\text{exact}}}(\xi)>0,$$

for all ξ. Respectively, because a set $\{x\,|\,x<z\}$ is open[6], we get that

$$\int_{-\infty}^{z} p(y_k|\xi)p_{\bar{x}^-_{k,N}}(\xi)\mathrm{d}\xi \xrightarrow[N\to\infty]{} \int_{-\infty}^{z} p(y_k|\xi)p_{x^-_{k,\text{exact}}}(\xi)\mathrm{d}\xi,$$

for all z. Combining these results we get that

$$F_{\bar{x}^{+,\text{Bayes}}_k}(z) \xrightarrow[N\to\infty]{} F_{x^+_{k,\text{exact}}}(z),$$

for all z, where

$$\begin{aligned} F_{\bar{x}^{+,\text{Bayes}}_k}(z) &= \frac{\int_{-\infty}^{z} p(y_k|\xi)p_{\bar{x}^-_k}(\xi)\mathrm{d}\xi}{\int p(y_k|\xi)p_{\bar{x}^-_k}(\xi)\mathrm{d}\xi} \text{ and} \\ F_{x^+_{k,\text{exact}}}(z) &= \frac{\int_{-\infty}^{z} p(y_k|\xi)p_{x^-_{k,\text{exact}}}(\xi)\mathrm{d}\xi}{\int p(y_k|\xi)p_{x^-_{k,\text{exact}}}(\xi)\mathrm{d}\xi}. \end{aligned} \tag{15}$$

□

Theorem 13 (Bank of EKFs convergence): Let

$$\begin{aligned} F_{\bar{x}^+_k}(z) &= \frac{\int_{-\infty}^{z} p_{\text{EKF}}(y_k|\xi)p_{\bar{x}^-_k}(\xi)\mathrm{d}\xi}{\int p_{\text{EKF}}(y_k|\xi)p_{\bar{x}^-_k}(\xi)\mathrm{d}\xi} \text{ and} \\ F_{\bar{x}^{+,\text{Bayes}}_k}(z) &= \frac{\int_{-\infty}^{z} p(y_k|\xi)p_{\bar{x}^-_k}(\xi)\mathrm{d}\xi}{\int p(y_k|\xi)p_{\bar{x}^-_k}(\xi)\mathrm{d}\xi}, \end{aligned}$$

where the likelihood

$$p(y_k|\xi)=p_{v_k}(y_k-h_k(\xi))$$

and the bank of EKF likelihood approximations[7] (see Eq. (10))

$$p_{\text{EKF}}(y_k|\xi)=p_{v_k}(y_k-h_k(\bar{\mu}^-_{i,k})-\mathrm{H}_{i,k}(\xi-\bar{\mu}^-_{i,k})).$$

Then

$$F_{\bar{x}^{+,\text{Bayes}}_k}(\xi) - F_{\bar{x}^+_k}(\xi) \xrightarrow[N\to\infty]{} 0.$$

[5] Usually we suppress the index N (parameter of GMF), that is $x^+_{k-1,N} \triangleq x^+_{k-1}$.

[6] Here sign $''<''$ is interpreted elementwise.

[7] Note that current approximation is also a function of index i (see Eq. (10)).

Proof: It is enough to show that

$$\int \left|p(y_k|\xi) - p_{\text{EKF}}(y_k|\xi)\right| p_{\bar{x}_k^-}(\xi)\mathrm{d}\xi \underset{N\to\infty}{\longrightarrow} 0.$$

Now

$$\begin{aligned}
&\int \left|p(y_k|\xi) - p_{\text{EKF}}(y_k|\xi)\right| p_{\bar{x}_k^-}(\xi)\mathrm{d}\xi \\
&\leq \sum_{j=1}^{n_{v_k}} \sum_{i=1}^{\bar{n}_k^-} \beta_{j,k}\bar{\alpha}_{i,k}^- \int \left|\mathrm{N}_{\mathrm{R}_{j,k}}^{\bar{v}_{j,k}}(z) - \mathrm{N}_{\mathrm{R}_{j,k}}^{\bar{v}_{j,k}}(\tilde{z}_i)\right| \mathrm{N}_{\bar{\Sigma}_{i,k}^-}^{\bar{\mu}_{i,k}^-}(\xi)\mathrm{d}\xi \\
&= \sum_{j=1}^{n_{v_k}} \sum_{i=1}^{\bar{n}_k^-} \frac{\beta_{j,k}\bar{\alpha}_{i,k}^-}{\sqrt{\det(2\pi \mathrm{R}_{j,k})}} \cdots \\
&\int \left|\exp\left(-\frac{1}{2}\|z\|_{\mathrm{R}_{j,k}^{-1}}^2\right) - \exp\left(-\frac{1}{2}\|\tilde{z}_i\|_{\mathrm{R}_{j,k}^{-1}}^2\right)\right| \mathrm{N}_{\bar{\Sigma}_{i,k}^-}^{\bar{\mu}_{i,k}^-}(\xi)\mathrm{d}\xi \\
&= \sum_{j=1}^{n_{v_k}} \sum_{i=1}^{\bar{n}_k^-} \frac{\beta_{j,k}\bar{\alpha}_{i,k}^-}{\sqrt{\det(2\pi \mathrm{R}_{j,k})}} \epsilon_{i,j},
\end{aligned}$$

where $z = y_k - h_k(\xi)$, $\tilde{z}_i = y_k - h_k(\bar{\mu}_{i,k}^-) - \mathrm{H}_{i,k}(\xi - \bar{\mu}_{i,k}^-)$ and $\epsilon_{i,j}$ is

$$\int \left|\exp\left(-\frac{1}{2}\|z\|_{\mathrm{R}_{j,k}^{-1}}^2\right) - \exp\left(-\frac{1}{2}\|\tilde{z}_i\|_{\mathrm{R}_{j,k}^{-1}}^2\right)\right| \mathrm{N}_{\bar{\Sigma}_{i,k}^-}^{\bar{\mu}_{i,k}^-}(\xi)\mathrm{d}\xi.$$

It is easy to see that

$$\epsilon_{i,j} < 1. \tag{16}$$

Based on the assumptions (see p. 4) we know that almost all $\bar{\mu}_{i,k}^-$ have a neighbourhood $\bar{C}_i$ such that

$$\left|\xi^T h_{k_j}''(x)\xi\right| \leq c_{\mathrm{H}}\xi^T \begin{bmatrix} \mathrm{I}_{d\times d} & 0 \\ 0 & 0 \end{bmatrix} \xi, \text{ for all } \xi \in \mathbb{R}^{n_x} \tag{17}$$

where c_{H} is some constant, $j \in \{1, \ldots, n_y\}$, n_y is the number of measurements (length of vector y), d see p. 4 and $x \in \bar{C}_i$. We select $\bar{C}_i$ such that it is as big as possible (union of all possible sets). Especially we see that if $x \in \bar{C}_i$ then $\begin{bmatrix} x_{1:d} \\ \bar{x} \end{bmatrix} \in \bar{C}_i$, where $\bar{x} \in \mathbb{R}^{n_x - d}$ is an arbitrary vector.

The index set I_1 contains the index i if both inequalities (14) and (17) hold, the rest of the indices belong to the index set I_2. Now

$$\begin{aligned}
&\int |p(y_k|\xi) - p_{\text{EKF}}(y_k|\xi)| p_{\bar{x}_k^-}(\xi)\mathrm{d}\xi \\
&= \sum_{j=1}^{n_{v_k}} \sum_{i=1}^{\bar{n}_k^-} \frac{\beta_{j,k}\bar{\alpha}_{i,k}^-}{\sqrt{\det(2\pi \mathrm{R}_{j,k})}} \epsilon_{i,j}, \\
&\overset{(16)}{\leq} \sum_{j=1}^{n_{v_k}} \sum_{i\in I_1} \frac{\beta_{j,k}\bar{\alpha}_{i,k}^-}{\sqrt{\det(2\pi \mathrm{R}_{j,k})}} \epsilon_{i,j} + \sum_{i\in I_2} \frac{\bar{\alpha}_{i,k}^-}{\sqrt{\det\left(2\pi \bar{\mathrm{R}}_k\right)}},
\end{aligned}$$

where $\det\left(2\pi\bar{\mathrm{R}}_k\right) = \min_j \det(2\pi \mathrm{R}_{j,k})$. Since almost all indices belong to the index set I_1,

$$\sum_{i\in I_2} \frac{\bar{\alpha}_{i,k}^-}{\sqrt{\det\left(2\pi\bar{\mathrm{R}}_k\right)}} \underset{N\to\infty}{\longrightarrow} 0.$$

Appendix C (Lemma 27) shows that $\epsilon_{i,j} \underset{N\to\infty}{\longrightarrow} 0$ when $i \in I_1$. □

D. Convergence results of Step 4 (reduce the number of components)

Here we show that if $\bar{x}_k^+ \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} x_{k,\text{exact}}^+$ then $x_k^+ \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} x_{k,\text{exact}}^+$. First we show that if $\bar{x}_k^+ \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} x_{k,\text{exact}}^+$ then $\bar{x}_{k,\mathrm{f}}^+ \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} x_{k,\text{exact}}^+$, (Thm. 14), see Eq. (12).

Theorem 14 (Forgetting components): If

$$\bar{x}_k^+ \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} x_{k,\text{exact}}^+$$

then

$$\bar{x}_{k,\mathrm{f}}^+ \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} x_{k,\text{exact}}^+.$$

(See Sec. III-D1.)

Proof: Take arbitrary $\epsilon > 0$, then there is an n_1 such that

$$|F_{x_{k,\text{exact}}^+}(\xi) - F_{\bar{x}_k^+}(\xi)| \leq \frac{\epsilon}{2},$$

for all ξ when $N > n_1$. Now

$$\begin{aligned}
&|F_{x_{k,\text{exact}}^+}(\xi) - F_{\bar{x}_{k,\mathrm{f}}^+}(\xi)| \\
&= |F_{x_{k,\text{exact}}^+}(\xi) - F_{\bar{x}_k^+}(\xi) + F_{\bar{x}_k^+}(\xi) - F_{\bar{x}_{k,\mathrm{f}}^+}(\xi)| \\
&\leq |F_{x_{k,\text{exact}}^+}(\xi) - F_{\bar{x}_k^+}(\xi)| + |F_{\bar{x}_k^+}(\xi) - F_{\bar{x}_{k,\mathrm{f}}^+}(\xi)| \\
&\leq \frac{\epsilon}{2} + \frac{1}{2N} \leq \epsilon,
\end{aligned}$$

for all $\xi \in \mathbb{R}^{n_x}$, when $N \geq \max(n_1, \frac{1}{\epsilon})$. This completes the proof (see Def. 8). □

Theorem 15 (Merging components): If

$$\bar{x}_{k,\mathrm{f}}^+ \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} x_{k,\text{exact}}^+$$

then

$$x_k^+ \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} x_{k,\text{exact}}^+.$$

(See Sec. III-D2.)

Proof: Based on Thm. 14 it is enough to show that

$$|F_{\bar{x}_{k,\mathrm{f}}^+}(\xi) - F_{x_k^+}(\xi)| \underset{N\to\infty}{\longrightarrow} 0,$$

for all ξ. Because all cumulative density functions are continuous and

$$\begin{aligned}
\|\bar{\mu}_{i_1}^+ - \bar{\mu}_{i_1,\mathrm{m}}^+\| &\underset{N\to\infty}{\longrightarrow} 0, \\
\|\bar{\mu}_{i_2}^+ - \bar{\mu}_{i_1,\mathrm{m}}^+\| &\underset{N\to\infty}{\longrightarrow} 0, \\
\|\bar{\Sigma}_{i_1}^+ - \bar{\Sigma}_{i_1,\mathrm{m}}^+\| &\underset{N\to\infty}{\longrightarrow} 0 \text{ and} \\
\|\bar{\Sigma}_{i_2}^+ - \bar{\Sigma}_{i_1,\mathrm{m}}^+\| &\underset{N\to\infty}{\longrightarrow} 0
\end{aligned}$$

then

$$|F_{\bar{x}_{k,\mathrm{f}}^+}(\xi) - F_{x_k^+}(\xi)| \underset{N\to\infty}{\longrightarrow} 0.$$

This completes the proof (see Def. 8). □

V. Box Gaussian Mixture Approximation

In this section, we define the Box Gaussian Mixture Approximation (BGMA) and present some of its properties. Finally, we show that BGMA converges weakly to the original distribution.

Definition 16 (BGMA): The Box Gaussian Mixture Approximation of $x \sim \mathrm{N}_n(\mu, \Sigma)$, note $n \triangleq n_x$, where $\Sigma > 0$, is

$$x_N \sim \mathrm{M}(\alpha_i, \mu_i, \Sigma_i)_{(i,(2N^2+1)^d)},$$

where the multi-index $i \in \mathbb{Z}^d$, with $d \leq n$ and $\|i\|_\infty \leq N^2$. The parameters are defined as

$$\begin{aligned} \alpha_i &= \int_{A_i} p_x(\xi)\mathrm{d}\xi, \\ \mu_i &= \int_{A_i} \xi \frac{p_x(\xi)}{\alpha_i}\mathrm{d}\xi, \text{ and} \\ \Sigma_i &= \int_{A_i} (\xi - \mu_i)(\xi - \mu_i)^T \frac{p_x(\xi)}{\alpha_i}\mathrm{d}\xi, \end{aligned} \tag{18}$$

where the sets

$$A_i = \left\{ x \middle| l(i) < \mathrm{A}(x - \mu) \leq u(i) \right\},$$

constitute a partition of $\mathbb{R}^n$. We assume that $\mathrm{A} = \left[\begin{array}{cc} \mathrm{A}_{11} & 0 \end{array} \right]$, $\mathrm{A}_{11} \in \mathbb{R}^{d \times d}$ and

$$\mathrm{A}\Sigma\mathrm{A}^T = \mathrm{I}. \tag{19}$$

Here the limits $l(i)$ and $u(i)$ are

$$\begin{aligned} l_j(i) &= \begin{cases} -\infty, & \text{if } i_j = -N^2 \\ \frac{i_j}{N} - \frac{1}{2N}, & \text{otherwise} \end{cases}, \\ u_j(i) &= \begin{cases} \infty, & \text{if } i_j = N^2 \\ \frac{i_j}{N} + \frac{1}{2N}, & \text{otherwise} \end{cases}. \end{aligned}$$

Now we show that the assumption Eq. (19) enables feasible computation time for the parameters of BGMA.

Theorem 17 (Parameters of BGMA): Let

$$x_N \sim \mathrm{M}(\alpha_i, \mu_i, \Sigma_i)_{(i,(2N^2+1)^d)},$$

be the BGMA of $x \sim \mathrm{N}_n(\mu, \Sigma)$, where $\Sigma > 0$ (see Def. 16). Then the parameters are

$$\begin{aligned} \alpha_i &= \prod_{j=1}^{d} \left(\Phi\left(u_j(i)\right) - \Phi\left(l_j(i)\right) \right), \\ \mu_i &= \mu + \Sigma\mathrm{A}^T \epsilon_i, \text{ and} \\ \Sigma_i &= \Sigma - \Sigma\mathrm{A}^T \Lambda_i \mathrm{A}\Sigma, \end{aligned}$$

where

$$\begin{aligned} \Phi(x) &= \int_{-\infty}^{x} N_1^0(\xi)\mathrm{d}\xi, \\ \Lambda_i &= \mathrm{diag}(\delta_i + \mathrm{diag}(\epsilon_i\epsilon_i^T)), \\ \epsilon_i &= \sum_{j=1}^{d} e_j \frac{e^{-\frac{1}{2}l_j(i)^2} - e^{-\frac{1}{2}u_j(i)^2}}{\sqrt{2\pi}\left(\Phi\left(u_j(i)\right) - \Phi\left(l_j(i)\right)\right)}, \text{ and} \\ \delta_i &= \sum_{j=1}^{d} e_j \frac{u_j(i)e^{-\frac{1}{2}u_j(i)^2} - l_j(i)e^{-\frac{1}{2}l_j(i)^2}}{\sqrt{2\pi}\left(\Phi\left(u_j(i)\right) - \Phi\left(l_j(i)\right)\right)}, \end{aligned}$$

where $e_j \in \mathbb{R}^d$ is the jth column of the identity matrix I. The sets A_i, and limits $l(i)$ and $u(i)$ are given in Def. 16.

Proof: We use the following block matrix notation

$$\Sigma = \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right], \quad \bar{\mathrm{A}} = \left[\begin{array}{cc} \mathrm{A}_{11} & 0 \\ -\mathrm{D}\Sigma_{21}\Sigma_{11}^{-1} & \mathrm{D} \end{array} \right],$$

where $\mathrm{D} = \left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)^{-\frac{1}{2}}$. Because $\Sigma > 0$ then $\mathrm{D} > 0$. We see that $\bar{\mathrm{A}}\Sigma\bar{\mathrm{A}}^T = \mathrm{I}$. We use the variable transformation

$$\bar{x} = \bar{\mathrm{A}}(x - \mu).$$

Because $x \sim \mathrm{N}(\mu, \Sigma)$ then $\bar{x} \sim \mathrm{N}(0, \mathrm{I})$, and if $x \in A_i$ then $\bar{x} \in B_i$ and vice versa. Here

$$B_i = \left\{ \bar{x} \middle| \left[\begin{array}{c} l(i) \\ -\infty \end{array} \right] < \bar{x} \leq \left[\begin{array}{c} u(i) \\ \infty \end{array} \right] \right\}.$$

Now we compute parameters Eq. (18)

$$\begin{aligned} \alpha_i &= \mathrm{P}(\bar{x} \in B_i) = \prod_{j=1}^{d} \left(\Phi\left(u_j(i)\right) - \Phi\left(l_j(i)\right) \right), \\ \mu_i &= \mu + \int_{B_i} \bar{\mathrm{A}}^{-1}\eta \frac{p_{\bar{x}}(\eta)}{\alpha_i}\mathrm{d}\eta = \mu + \Sigma\mathrm{A}^T\epsilon_i, \\ \Sigma_i &= \bar{\mathrm{A}}^{-1} \int_{B_i} (\eta - \epsilon_i)(\eta - \epsilon_i)^T \frac{p_{\bar{x}}(\eta)}{\alpha_i}\mathrm{d}\eta\bar{\mathrm{A}}^{-T} \\ &= \Sigma - \Sigma\mathrm{A}^T\Lambda_i\mathrm{A}\Sigma. \end{aligned}$$

Here we have used the knowledge that $\bar{\mathrm{A}}^{-1} = \Sigma\bar{\mathrm{A}}^T$. ◻

In Fig. 1, we compare the density function of the Gaussian distribution

$$x \sim \mathrm{N}_2\left(\left[\begin{array}{c} 0 \\ 0 \end{array} \right], \left[\begin{array}{cc} 13 & 12 \\ 12 & 13 \end{array} \right] \right)$$

and the density function of its BGMA with parameters $d = 2$, $N = 2$ and

$$\mathrm{A} = \left[\begin{array}{cc} \frac{1}{\sqrt{13}} & 0 \\ -\frac{12}{5\sqrt{13}} & \frac{\sqrt{13}}{5} \end{array} \right].$$

Fig. 1 shows the contour plots of the Gaussian and the BGMA density functions such that 50% of the probability mass is inside the innermost curve and 95% of the probability mass is inside the outermost curve.

Theorem 18 shows that BGMA has the same mean and covariance as the original distribution.

Theorem 18 (Mean and Covariance of BGMA): Let

$$x_N \sim \mathrm{M}(\alpha_i, \mu_i, \Sigma_i)_{(i,(2N^2+1)^d)},$$

be the BGMA of $x \sim \mathrm{N}_n(\mu, \Sigma)$, where $\Sigma > 0$ (see Def. 16). Then

$$\mathrm{E}(x_N) = \mu \quad \text{and} \quad \mathrm{V}(x_N) = \Sigma.$$

Proof: Now

$$\begin{aligned} \mathrm{E}(x_N) &\overset{\text{Thm. 5}}{=} \sum_i \alpha_i\mu_i \overset{\text{Def. 16}}{=} \sum_i \alpha_i \int_{A_i} \xi \frac{p_x(\xi)}{\alpha_i} d\xi \\ &= \sum_i \int_{A_i} \xi p_x(\xi) d\xi = \int \xi p_x(\xi)\mathrm{d}\xi = \mu, \end{aligned}$$

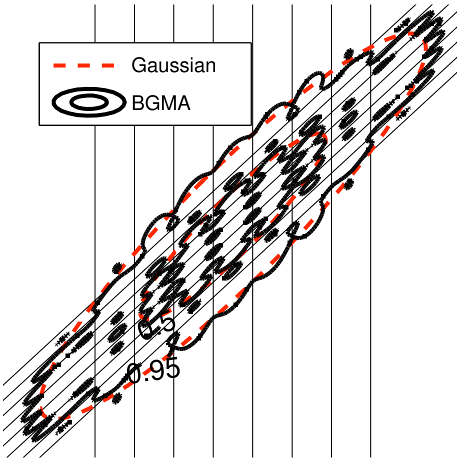

Fig. 1. Example of the BGMA

and

$$\begin{aligned}
\mathrm{V}(x_N) &\overset{\text{Thm. 5}}{=} \sum_i \alpha_i \left(\Sigma_i + (\mu_i - \mu)(\mu_i - \mu)^T\right) \\
&= \sum_i \alpha_i \left(\Sigma_i + \mu_i(\mu_i - \mu)^T\right) \\
&\overset{\text{Def. 16}}{=} \sum_i \alpha_i \left(\left(\int_{A_i} \xi(\xi - \mu_i)^T \frac{p_x(\xi)}{\alpha_i} \mathrm{d}\xi + \mu_i(\mu_i - \mu)^T \right)\right) \\
&\overset{\text{Def. 16}}{=} \sum_i \alpha_i \int_{A_i} \xi\xi^T \frac{p_x(\xi)}{\alpha_i} \mathrm{d}\xi - \mu\mu^T \\
&= \int \xi\xi^T p_x(\xi)\mathrm{d}\xi - \mu\mu^T = \Sigma.
\end{aligned}$$

□

Lemma 19 considers bounded boxes of Def. 16, i.e. boxes with parameters $d = n$ and $\|i\|_\infty < N^2$. Lemma 19 shows that a ball with radius $r_{\text{in}} = \frac{1}{2N\|\mathrm{A}\|}$ fits inside all boxes, and all boxes fit inside a ball whose radius is $r_{\text{out}} = \frac{\sqrt{n}}{2N}\|\mathrm{A}^{-1}\|$. Note that the proportion of these radiuses $\frac{r_{\text{in}}}{r_{\text{out}}} = \frac{1}{\sqrt{n}\kappa(\mathrm{A})}$ does not depend on the parameter N. Here $\kappa(\mathrm{A}) = \|\mathrm{A}\|\|\mathrm{A}^{-1}\|$ is the condition number of matrix A.

Lemma 19: Let

$$A = \left\{ x \middle| -\frac{1}{2N}\mathbf{1} < \mathrm{A}x \leq \frac{1}{2N}\mathbf{1} \right\},$$
$$R_{\text{in}} = \left\{ x \middle| \|x\| \leq r_{\text{in}} \right\} \text{ and } R_{\text{out}} = \left\{ x \middle| \|x\| \leq r_{\text{out}} \right\},$$

where $\mathbf{1}$ is a vector allof whose elements are ones,

$$r_{\text{in}} = \frac{1}{2N\|\mathrm{A}\|}, \quad r_{\text{out}} = \frac{\sqrt{n}}{2N}\|\mathrm{A}^{-1}\|$$

and A is non-singular. Now $R_{\text{in}} \subset A \subset R_{\text{out}}$.

Proof: If $x \in R_{\text{in}}$ then

$$\|\mathrm{A}x\| \leq \|\mathrm{A}\|\|x\| \leq \frac{1}{2N}.$$

So $R_{\text{in}} \subset A$. If $x \in A$ then

$$\|x\| \leq \|\mathrm{A}x\|\|\mathrm{A}^{-1}\| \leq \|\frac{1}{2N}\mathbf{1}\|\|\mathrm{A}^{-1}\| \leq \frac{\sqrt{n}}{2N}\|\mathrm{A}^{-1}\|.$$

So $A \subset R_{\text{out}}$. □

Corollary 20 considers the center boxes ($\|i\|_\infty < N^2$) of BGMA (Def. 16) and shows that the covariances of the first d dimensions converge to zero when N approaches infinity.

Corollary 20: Covariances Σ_i are the same as in Def. 16. Now

$$\sum_{j=1}^{d} \Sigma_{i_{j,j}} \underset{N\to\infty}{\longrightarrow} 0, \text{ when } \|i\|_\infty < N^2.$$

Proof: Because

$$\begin{aligned}
\sum_{j=1}^{d} \Sigma_{i_{j,j}} &= \int_{A_i} \|\xi_{1:d} - \mu_{i_{1:d}}\|^2 \frac{p_x(\xi)}{\alpha_i}\mathrm{d}\xi \\
&\overset{\text{Lem. 19}}{\leq} \int_{A_i} \frac{d}{N^2}\|\mathrm{A}_{11}^{-1}\|^2 \frac{p_x(\xi)}{\alpha_i}\mathrm{d}\xi \\
&= \frac{d}{N^2}\|\mathrm{A}_{11}^{-1}\|^2
\end{aligned}$$

then $\sum_{j=1}^{d} \Sigma_{i_{j,j}} \underset{N\to\infty}{\longrightarrow} 0$, for all $\|i\|_\infty < N^2$. □

Theorem 26 (see Appendix B) shows that the BGMA converges weakly to the original distribution when the center boxes are bounded. Theorem 21 uses this result to show that BGMA converges weakly to the original distribution even if all boxes are unbounded.

Theorem 21 (BGMA convergence, Gaussian case): Let

$$x_N \sim \mathrm{M}(\alpha_i, \mu_i, \Sigma_i)_{(i,(2N^2+1)^d)}$$

be BGMA of $x \sim \mathrm{N}_n(\mu, \Sigma)$, where $\Sigma > 0$ (see Def. 16). Now

$$x_N \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} x.$$

Proof: First we define new random variables

$$\begin{aligned}
\bar{x} &= \bar{\mathrm{A}}(x - \mu) \sim \mathrm{N}(0, \mathrm{I}) \quad \text{and} \\
\bar{x}_N &= \bar{\mathrm{A}}(x_N - \mu) \sim \mathrm{M}(\alpha_i, \bar{\mathrm{A}}(\mu_i - \mu), \bar{\mathrm{A}}\Sigma_i\bar{\mathrm{A}}^T)_{(i,(2N^2+1)^d)},
\end{aligned}$$

where $\bar{\mathrm{A}}$ is defined in Thm. 17. Note that $\bar{\mathrm{A}}\Sigma_i\bar{\mathrm{A}}^T$ are diagonal matrices. It is enough to show that (because of Slutsky's Theorem 10)

$$\bar{x}_N \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} \bar{x}.$$

Let F_N and F be the cumulative density functions corresponding to the random variables $\bar{x}_N$ and $\bar{x}$. We have to show that

$$F_N(\bar{x}) \underset{N\to\infty}{\longrightarrow} F(\bar{x}), \quad \forall \bar{x} \in \mathbb{R}^n. \tag{20}$$

Because

$$\begin{aligned}
F_N(\bar{x}) &= \sum_i \alpha_i \int_{-\infty}^{\bar{x}_1} \mathrm{N}_{\mathrm{I}-\Lambda_i}^{\epsilon_i}(\eta_1)\mathrm{d}\eta_1 \int_{-\infty}^{\bar{x}_2} \mathrm{N}_{\mathrm{I}}^{0}(\eta_2)\mathrm{d}\eta_2 \\
&= G_N(\bar{x}_1) \int_{-\infty}^{\bar{x}_2} \mathrm{N}_{\mathrm{I}}^{0}(\eta_2)\mathrm{d}\eta_2, \\
F(\bar{x}) &= \int_{-\infty}^{\bar{x}_1} \mathrm{N}_{\mathrm{I}}^{0}(\eta_1)\mathrm{d}\eta_1 \int_{-\infty}^{\bar{x}_2} \mathrm{N}_{\mathrm{I}}^{0}(\eta_2)\mathrm{d}\eta_2 \\
&= G(\bar{x}_1) \int_{-\infty}^{\bar{x}_2} \mathrm{N}_{\mathrm{I}}^{0}(\eta_2)\mathrm{d}\eta_2,
\end{aligned}$$

where $\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$, it is enough to show that

$$\bar{x}_{N_{1:d}} \underset{N\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} \bar{x}_{1:d} \tag{21}$$

Based on Thm. 26 (see Appendix B), Eq. (21) is true, which implies the theorem. □

Corollary 22 (BGMA convergence, GM case): Let

$$\bar{x}_{j,N} \sim \mathrm{M}(\bar{\alpha}_i, \bar{\mu}_i, \bar{\Sigma}_i)_{(i,(2N^2+1)^d)}$$

be the BGMA of $x_j \sim \mathrm{N}_n(\mu_j, \Sigma_j)$, where $\Sigma_j > 0$ (see Def. 16). Let x be the GM whose density function is

$$p_x(\xi) = \sum_{j=i}^{N_x} \alpha_j \, \mathrm{N}_{\Sigma_j}^{\mu_j}(\xi).$$

and x_N be the GM whose density function is

$$p_{x_N}(\xi) = \sum_{j=i}^{N_x} \alpha_j p_{\bar{x}_{j,N}}(\xi).$$

Now

$$x_N \xrightarrow[N\to\infty]{\mathrm{w}} x.$$

Proof: Take arbitrary $\epsilon > 0$, then there are n_j, $j = 1, \ldots, N_x$ such that (Thm. 21)

$$|F_{x_j}(\xi) - F_{\bar{x}_{j,N_j}}(\xi)| \leq \epsilon, \tag{22}$$

for all j and ξ, when $N_j > n_j$. Now

$$\begin{aligned} |F_x(\xi) - F_{x_N}(\xi)| &= \Big| \sum_{j=1}^{N_x} \alpha_j \left(F_{x_j}(\xi) - F_{x_{j,N}}(\xi) \right) \Big| \\ &\leq \sum_{j=1}^{N_x} \alpha_j |F_{x_j}(\xi) - F_{x_{j,N}}(\xi)| \\ &\overset{(22)}{\leq} \sum_{j=1}^{N_x} \alpha_j \epsilon = \epsilon, \end{aligned}$$

for all ξ, when $N > \max_j\{n_j\}$. □

VI. Box Gaussian Mixture Filter

The Box Gaussian Mixture Filter (BGMF) is a GMF (Sec. III) that approximates the prior x_k^- as a new GM (Step 2 in Algorithm 2) using BGMA (Sec. V) separately for each component of the prior. Section IV shows that BGMF converges weakly to the exact posterior distribution.

VII. Simulations

In the simulations we consider only the case of a single time step. Our state $x = \begin{bmatrix} r_{\mathrm{u}} \\ v_{\mathrm{u}} \end{bmatrix}$ consists of the 2D-position vector r_{u} and the 2D-velocity vector v_{u} of the user. The prior distribution is

$$x \sim \mathrm{N}\left(\begin{bmatrix} 100 \\ 10 \\ 10 \\ 10 \end{bmatrix}, \begin{bmatrix} 90000 & 0 & 0 & 0 \\ 0 & 7500 & 0 & 2500 \\ 0 & 0 & 1000 & 0 \\ 0 & 2500 & 0 & 7500 \end{bmatrix} \right),$$

and the current measurement (see Eq. (8)) is

$$\begin{bmatrix} 500 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \|r_{\mathrm{u}}\| \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x + v,$$

where v is independent of x and

$$v \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 10^4 & 0 & 0 & 0 \\ 0 & 10^3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right).$$

So now $d = 2$ and $n = 4$ (see Def. 16). The current posterior of the 2D-position is shown in Fig. 2. We see that the posterior distribution is multimodal.

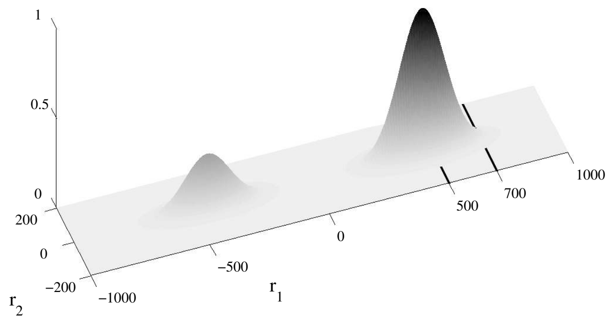

Fig. 2. The posterior of the position.

Now we compute the posterior approximations using BGMF (Algorithm 2 steps (2) and (3)) and a Particle Filter [18]. BGMF uses BGMA with parameters $N \in \{0, 1, 2, \ldots, 9\}$; the corresponding numbers of posterior mixture components are (Def. 16)

$$n_{\text{BGMF}} \in \{1, 9, 81, \ldots, 26569\}.$$

The numbers of particles in the Particle Filter are

$$n_{\text{PF}} \in \left\{2^2 \cdot 100, 2^3 \cdot 100, \ldots, 2^{17} \cdot 100\right\}.$$

We compute error statistics

$$|\mathrm{P}\left(x_{\text{true}} \in C\right) - \mathrm{P}\left(x_{\text{app.}} \in C\right)|,$$

where the set $C = \left\{ x \,\middle|\, |e_1^T x - 600| \leq 100 \right\}$ (see Fig. 2). We know that $\mathrm{P}\left(x_{\text{true}} \in C\right) \approx 0.239202$. These error statistics are shown as a function of CPU time in Fig. 3. Thm. 9 shows that these error statistics converge to zero when the posterior approximation converges weakly to the correct posterior.

Fig. 3 is consistent with the convergence results. It seems that the error statistics of both BGMF and PF converge to zero when the number of components or particles increase. We also see that in this case $2^{10} \cdot 100 \approx$ 1e5 particles in PF are definitely too few. However, BGMF gives promising results with only 81 components ($N = 2$) when CPU time is significantly less than one second, which is good considering real time implementations.

VIII. Conclusion

In this paper, we have presented the Box Gaussian Mixture Filter (BGMF), which is based on Box Gaussian Mixture Approximation (BGMA). We have presented the general form

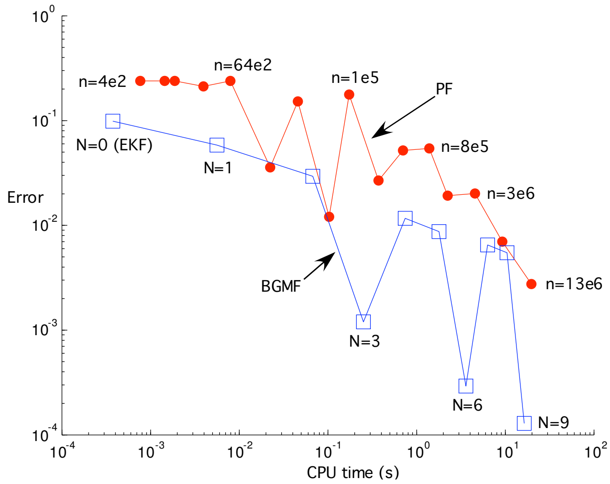

Fig. 3. Simulation results of BGMF and PF.

of Gaussian Mixture Filters (GMF) and we have shown that GMF converges weakly to the correct posterior at given time instant. BGMF is a GMF so we have shown that BGMF converges weakly to the correct posterior at given time instant. We have illustrated this convergence result with a tiny example in which BGMF outperforms a basic particle filter. The assumptions of BGMF fit very well in positioning and previous work [27] shows that BGMF is feasible for real time positioning implementations. BGMF also has smaller computational and memory requirements than conventional GMFs, because it splits only those dimensions where we get nonlinear measurements.

Appendix A
Product of two normal densities.

The aim of this appendix is to compute the product of two Gaussian densities (Theorem 25).

Lemma 23: If $\mathrm{A} > 0$ (s.p.d.) then

$$\|a \pm \mathrm{A}^{-1}b\|_{\mathrm{A}}^2 = a^T\mathrm{A}a \pm 2b^Ta + b^T\mathrm{A}^{-1}b.$$

Proof:

$$\begin{aligned}\|a \pm \mathrm{A}^{-1}b\|_{\mathrm{A}}^2 &= \left(a \pm \mathrm{A}^{-1}b\right)^T \mathrm{A}\left(a \pm \mathrm{A}^{-1}b\right)\\ &= a^T\mathrm{A}a \pm a^Tb \pm b^Ta + b^T\mathrm{A}^{-1}b\\ &= a^T\mathrm{A}a \pm 2b^Ta + b^T\mathrm{A}^{-1}b\end{aligned}$$

□

Lemma 24: If $\Sigma_1, \Sigma_2 > 0$ then

$$\|x-\mu\|_{\Sigma_1^{-1}}^2 + \|y - \mathrm{H}x\|_{\Sigma_2^{-1}}^2 = \|x-\bar{\mu}\|_{\Sigma_3^{-1}}^2 + \|y - \mathrm{H}\mu\|_{\Sigma_4^{-1}}^2,$$

where

$$\begin{aligned}\bar{\mu} &= \mu + \mathrm{K}(y - \mathrm{H}\mu),\\ \Sigma_3 &= (\mathrm{I}-\mathrm{KH})\Sigma_1,\\ \mathrm{K} &= \Sigma_1\mathrm{H}^T\Sigma_4^{-1}, \text{ and}\\ \Sigma_4 &= \mathrm{H}\Sigma_1\mathrm{H}^T + \Sigma_2.\end{aligned}$$

Proof:

$$\begin{aligned}&\|x-\mu\|_{\Sigma_1^{-1}}^2 + \|y-\mathrm{H}x\|_{\Sigma_2^{-1}}^2\\ &\overset{\text{Lem. 23}}{=} x^T(\Sigma_1^{-1} + \mathrm{H}^T\Sigma_2^{-1}\mathrm{H})x - 2(\mu^T\Sigma_1^{-1} + y^T\Sigma_2^{-1}\mathrm{H})x\\ &\quad + \mu^T\Sigma_1^{-1}\mu + y^T\Sigma_2^{-1}y\\ &\overset{\text{Lem. 23}}{=} \|x - \mathrm{C}c\|_{\mathrm{C}^{-1}}^2 - c^T\mathrm{C}c + \mu^T\Sigma_1^{-1}\mu + y^T\Sigma_2^{-1}y\\ &= \|x-\bar{\mu}\|_{\Sigma_3^{-1}}^2 + \|y-\mathrm{H}y\|_{\Sigma_4^{-1}}^2,\end{aligned}$$

where $c = \Sigma_1^{-1}\mu + \mathrm{H}^T\Sigma_2^{-1}y$,

$$\begin{aligned}\mathrm{C} &= (\Sigma_1^{-1} + \mathrm{H}^T\Sigma_2^{-1}\mathrm{H})^{-1}\\ &\overset{\star_1}{=} \Sigma_1 - \Sigma_1\mathrm{H}^T(\mathrm{H}\Sigma_1\mathrm{H}^T + \Sigma_2)^{-1}\mathrm{H}\Sigma_1\\ &= (\mathrm{I}-\mathrm{KH})\Sigma_1 = \Sigma_3,\end{aligned}$$

in step $\star_1$ we use the matrix inversion lemma [34, p.729].

$$\begin{aligned}\mathrm{C}c &= (\mathrm{I}-\mathrm{KH})\Sigma_1(\Sigma_1^{-1}\mu + \mathrm{H}^T\Sigma_2^{-1}y)\\ &= (\mathrm{I}-\mathrm{KH})\mu + (\mathrm{I}-\mathrm{KH})\Sigma_1\mathrm{H}^T\Sigma_2^{-1}y\\ &= (\mathrm{I}-\mathrm{KH})\mu + \Sigma_1\mathrm{H}^T\Sigma_2^{-1}y - \mathrm{K}(\Sigma_4 - \Sigma_2)\Sigma_2^{-1}y\\ &= (\mathrm{I}-\mathrm{KH})\mu + \Sigma_1\mathrm{H}^T\Sigma_2^{-1}y - \Sigma_1\mathrm{H}^T\Sigma_2^{-1}y + \mathrm{K}y\\ &= \mu + \mathrm{K}(y - \mathrm{H}\mu) = \bar{\mu}\end{aligned}$$

and

$$\begin{aligned}c^T\mathrm{C}c &= (\Sigma_1^{-1}\mu + \mathrm{H}^T\Sigma_2^{-1}y)^T((\mathrm{I}-\mathrm{KH})\mu + \mathrm{K}y)\\ &= (\mu^T\Sigma_1^{-1} + y^T\Sigma_2^{-1}\mathrm{H})((\mathrm{I}-\mathrm{KH})\mu + \mathrm{K}y)\\ &= \mu^T(\Sigma_1^{-1} - \mathrm{H}^T\Sigma_4^{-1}\mathrm{H})\mu + \mu^T\mathrm{H}^T\Sigma_4^{-1}y\ldots\\ &\quad + y^T(\Sigma_2^{-1}\mathrm{H} - \Sigma_2^{-1}\mathrm{HKH})\mu + y^T\Sigma_2^{-1}\mathrm{HK}y\\ &\overset{(23)}{=} \mu^T(\Sigma_1^{-1} - \mathrm{H}^T\Sigma_4^{-1}\mathrm{H})\mu + \mu^T\mathrm{H}^T\Sigma_4^{-1}y\ldots\\ &\quad + y^T\Sigma_4^{-1}\mathrm{H}\mu + y^T(\Sigma_2^{-1} - \Sigma_4^{-1})y\\ &= -(y^T\Sigma_4^{-1}y - 2\mu^T\mathrm{H}^T\Sigma_4^{-1}y + \mu^T\mathrm{H}^T\Sigma_4^{-1}\mathrm{H}\mu)\ldots\\ &\quad + y^T\Sigma_2^{-1}y + \mu^T\Sigma_1^{-1}\mu\\ &= -\|y-\mathrm{H}y\|_{\Sigma_4^{-1}}^2 + y^T\Sigma_2^{-1}y + \mu^T\Sigma_1^{-1}\mu.\end{aligned}$$

$$\Sigma_2^{-1}\mathrm{HK} = \Sigma_2^{-1} - \Sigma_4^{-1} \tag{23}$$

□

Theorem 25 (Product of two Gaussians): If $\Sigma_1, \Sigma_2 > 0$ then

$$\mathrm{N}_{\Sigma_1}^{\mu}(x)\,\mathrm{N}_{\Sigma_2}^{\mathrm{H}x}(y) = \mathrm{N}_{\Sigma_3}^{\bar{\mu}}(x)\,\mathrm{N}_{\Sigma_4}^{\mathrm{H}\mu}(y),$$

where

$$\begin{aligned}\bar{\mu} &= \mu + \mathrm{K}(y - \mathrm{H}\mu),\\ \Sigma_3 &= (\mathrm{I}-\mathrm{KH})\Sigma_1,\\ \mathrm{K} &= \Sigma_1\mathrm{H}^T\Sigma_4^{-1}, \text{ and}\\ \Sigma_4 &= \mathrm{H}\Sigma_1\mathrm{H}^T + \Sigma_2.\end{aligned}$$

Proof:

$$
\begin{aligned}
&\mathrm{N}^{\mu}_{\Sigma_1}(x)\,\mathrm{N}^{\mathrm{H}x}_{\Sigma_2}(y)\\
&= \frac{\exp\left(-\frac{1}{2}\|x-\mu\|^2_{\Sigma_1^{-1}}\right)}{(2\pi)^{\frac{n_x}{2}}\sqrt{\det(\Sigma_1)}}\frac{\exp\left(-\frac{1}{2}\|y-\mathrm{H}x\|^2_{\Sigma_2^{-1}}\right)}{(2\pi)^{\frac{n_y}{2}}\sqrt{\det(\Sigma_2)}}\\
&\overset{\text{Lem. 24}}{=} \frac{\exp\left(-\frac{1}{2}\|x-\bar{\mu}\|^2_{\Sigma_3^{-1}}\right)\exp\left(-\frac{1}{2}\|y-\mathrm{H}\mu\|^2_{\Sigma_4^{-1}}\right)}{(2\pi)^{\frac{n_x}{2}}(2\pi)^{\frac{n_y}{2}}\sqrt{\det(\Sigma_1)\det(\Sigma_2)}}\\
&\overset{(24)}{=} \frac{\exp\left(-\frac{1}{2}\|x-\bar{\mu}\|^2_{\Sigma_3^{-1}}\right)}{(2\pi)^{\frac{n_x}{2}}\sqrt{\det(\Sigma_3)}}\frac{\exp\left(-\frac{1}{2}\|y-\mathrm{H}\mu\|^2_{\Sigma_4^{-1}}\right)}{(2\pi)^{\frac{n_y}{2}}\sqrt{\det(\Sigma_4)}}\\
&= \mathrm{N}^{\bar{\mu}}_{\Sigma_3}(x)\,\mathrm{N}^{\mathrm{H}\mu}_{\Sigma_4}(y)
\end{aligned}
$$

where n_x and n_y are dimension of x and y, respectively.

$$
\begin{aligned}
&\det(\Sigma_1)\det(\Sigma_2)\\
&=\det\left(\begin{bmatrix}\Sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix}\right)\\
&=\det\left(\begin{bmatrix}\mathrm{I} & 0\\ \mathrm{H} & \mathrm{I}\end{bmatrix}\begin{bmatrix}\Sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix}\begin{bmatrix}\mathrm{I} & \mathrm{H}^T\\ 0 & \mathrm{I}\end{bmatrix}\right)\\
&=\det\left(\begin{bmatrix}\Sigma_1 & \Sigma_1\mathrm{H}^T\\ \mathrm{H}\Sigma_1 & \mathrm{H}\Sigma_1\mathrm{H}^T+\Sigma_2\end{bmatrix}\right)\\
&=\det\left(\begin{bmatrix}\mathrm{I} & \mathrm{K}\\ 0 & \mathrm{I}\end{bmatrix}\begin{bmatrix}\Sigma_3 & 0\\ \mathrm{H}\Sigma_1 & \Sigma_4\end{bmatrix}\right)\\
&=\det(\Sigma_3)\det(\Sigma_4)
\end{aligned}
\tag{24}
$$

□

Appendix B
BGMA convergence when $d = n$

Theorem 26 (BGMA convergence when $d = n$): Let

$$x_N \sim \mathrm{M}(\alpha_i, \mu_i, \Sigma_i)_{(i,(2N^2+1)^d)}$$

be the BGMA of $x \sim \mathrm{N}_n(\mu, \Sigma)$, where $\Sigma > 0$ (see Def. 16). We assume that $d = n$. Now

$$x_N \xrightarrow[N\to\infty]{\mathrm{w}} x.$$

Proof: Let F_N and F be the cumulative density functions corresponding to the random variables x_N and x. We have to show that

$$F_N(x) \xrightarrow[N\to\infty]{} F(x), \quad \forall x \in \mathbb{R}^n. \tag{25}$$

Let $x \in \mathbb{R}^n$ be an arbitrary vector whose components are x_j, and define the index sets

$$
\begin{aligned}
IN(x) &= \{i|\mu_i \le x \,\text{and}\, \|\mathrm{i}\|_\infty < \mathrm{N}^2\},\\
OUT(x) &= \{i|\mu_i \not\le x \,\text{and}\, \|\mathrm{i}\|_\infty < \mathrm{N}^2\},\\
OUT_j(x) &= \{i|(\mu_i)_j > x_j \,\text{and}\, \|\mathrm{i}\|_\infty < \mathrm{N}^2\}\,\text{and}\\
OUT_j^l(x) &=\\
&\{i|lr_{\text{in}} < (\mu_i)_j - x_j \le (l+1)r_{\text{in}} \,\text{and}\, \|\mathrm{i}\|_\infty < \mathrm{N}^2\},
\end{aligned}
$$

where $r_{\text{in}} = \frac{1}{2N\|\mathrm{A}\|}$ and less than or equal sign $'' \le''$ is interpreted elementwise. First we show that

$$\sum_{i\in IN(x)} \alpha_i \to F(x). \tag{26}$$

Now

$$F(x - 2r_{\text{out}}\mathbf{1}) - \epsilon_{\text{edge}} \le \sum_{j\in IN(x)} \alpha_j \le F(x + 2r_{\text{out}}\mathbf{1}),$$

where $r_{\text{out}} = \frac{\sqrt{n}}{2N}\|\mathrm{A}^{-1}\|$ and

$$\epsilon_{\text{edge}} = \mathrm{P}\left(x \in \bigcup_{\|i\|_\infty = N^2} A_i\right).$$

We see that $r_{\text{out}} \xrightarrow[N\to\infty]{} 0$ and $\epsilon_{\text{edge}} \xrightarrow[N\to\infty]{} 0$. Using these results and the continuity of the cumulative density function $F(x)$ we see that equation (26) holds. So equation (25) holds if

$$\epsilon_N(x) = F_N(x) - \sum_{i\in IN(x)} \alpha_i \xrightarrow[N\to\infty]{} 0, \quad \forall x \in \mathbb{R}^n. \tag{27}$$

Now we show that this equation (27) holds. We find upper and lower bounds of $\epsilon_N(x)$. The upper bound of $\epsilon_N(x)$ is

$$
\begin{aligned}
\epsilon_N(x) &= \int_{\xi\le x} p_{x_N}(\xi)\mathrm{d}\xi - \sum_{i\in IN(x)} \alpha_i\\
&\le \epsilon_{\text{edge}} + \int_{\xi\le x} \sum_{i\in OUT(x)} \alpha_i\,\mathrm{N}^{\mu_i}_{\Sigma_i}(\xi)\mathrm{d}\xi\\
&\le \epsilon_{\text{edge}} + \sum_{j=1}^{n}\int_{\xi_j\le x_j} \sum_{i\in OUT_j(x)} \alpha_i\,\mathrm{N}^{\mu_i}_{\Sigma_i}(\xi)\mathrm{d}\xi\\
&= \epsilon_{\text{edge}} + n\sum_{l=0}^{\infty}\int_{\xi_j\le x_j} \sum_{i\in OUT_j^l(x)} \alpha_i\,\mathrm{N}^{\mu_i}_{\Sigma_i}(\xi)\mathrm{d}\xi\\
&\le \epsilon_{\text{edge}} + n\sum_{l=0}^{\infty}\int_{\xi_j\le x_j} \sum_{i\in OUT_j^l(x)} \alpha_i\,\mathrm{N}^{x_j+lr_{\text{in}}}_{r^2_{\text{out}}}(\xi_j)\mathrm{d}\xi_j\\
&\le \epsilon_{\text{edge}} + n\sum_{l=0}^{\infty}\int_{y\le 0} \mathrm{N}_1^{l\frac{r_{\text{in}}}{r_{\text{out}}}}(y)\mathrm{d}y\,\alpha_{\max}\\
&\overset{(29)}{\le} \epsilon_{\text{edge}} + n\left(\frac{\kappa(\mathrm{A})\sqrt{2\pi n}+2}{4}\right)\alpha_{\max},
\end{aligned}
\tag{28}
$$

where

$$
\begin{aligned}
\alpha_{\max} &= \sup_{j,l}\left(\sum_{i\in OUT_j^l(x)} \alpha_i\right)\\
&\le \sup_{j,c\in\mathbb{R}}\left(\mathrm{P}\left(|x_j - c| \le 4r_{\text{out}}\right)\right).
\end{aligned}
$$

and

$$\sum_{l=0}^{\infty}\int_{y\le 0}\mathrm{N}_1^{l\frac{r_{\text{in}}}{r_{\text{out}}}}(y)\mathrm{d}y$$
$$= \sum_{l=0}^{\infty}\int_{y\le 0}\frac{\exp\left(-\frac{y^2}{2}+l\frac{r_{\text{in}}}{r_{\text{out}}}y-\frac{1}{2}\left(l\frac{r_{\text{in}}}{r_{\text{out}}}\right)^2\right)}{\sqrt{2\pi}}\mathrm{d}y$$
$$\le \sum_{l=0}^{\infty}\exp\left(-\frac{1}{2}\left(l\frac{r_{\text{in}}}{r_{\text{out}}}\right)^2\right)\int_{y\le 0}\frac{\exp\left(-\frac{y^2}{2}\right)}{\sqrt{2\pi}}\mathrm{d}y \tag{29}$$
$$= \frac{1}{2}\sum_{l=0}^{\infty}\exp\left(-\frac{1}{2}\left(l\frac{r_{\text{in}}}{r_{\text{out}}}\right)^2\right)$$
$$\le \frac{1}{2}\left(1+\int_0^{\infty}\exp\left(-\frac{1}{2}\left(l\frac{r_{\text{in}}}{r_{\text{out}}}\right)^2\right)\mathrm{d}l\right)$$
$$\overset{\text{Lem. 19}}{=}\left(\frac{\kappa(\mathrm{A})\sqrt{2\pi n}+2}{4}\right).$$

The lower bound of $\epsilon_N(x)$ is

$$\epsilon_N(x) = \int_{\xi\le x}p_{x_k}(\xi)\mathrm{d}\xi - \sum_{i\in IN(x)}\alpha_i$$
$$\ge -\int_{\xi\not\le x}\sum_{i\in IN(x)}\alpha_i\,\mathrm{N}_{\Sigma_i}^{\mu_i}(\xi)\mathrm{d}\xi$$
$$\ge -\sum_{j=1}^{n}\int_{\xi_j>x_j}\sum_{i\in IN(x)}\alpha_i\,\mathrm{N}_{\Sigma_i}^{\mu_i}(\xi)\mathrm{d}\xi \tag{30}$$
$$\ge -n\sum_{l=-\infty}^{-1}\int_{\xi_j>x_j}\mathrm{N}_{r_{\text{out}}^2}^{x_j+(l+1)r_{\text{in}}}(\xi_j)\mathrm{d}\xi_j\alpha_{\max}$$
$$\overset{(29)}{\ge} -n\left(\frac{\kappa(\mathrm{A})\sqrt{2\pi n}+2}{4}\right)\alpha_{\max}$$

So from equations (28) and (30) we get that

$$|\epsilon_N(x)| \le \epsilon_{\text{edge}} + n\left(\frac{\kappa(\mathrm{A})\sqrt{2\pi n}+2}{4}\right)\alpha_{\max},$$

because $\epsilon_{\text{edge}}\underset{N\to\infty}{\longrightarrow}0$ and $\alpha_{\max}\underset{N\to\infty}{\longrightarrow}0$ then $\epsilon_N(x)\underset{N\to\infty}{\longrightarrow}0$. Now using Eq. (26) we get

$$F_N(x)\to F(x),\quad \forall x\in\mathbb{R}^n.$$

□

APPENDIX C
LEMMA FOR UPDATE STEP

Lemma 27: Let $\epsilon_{i,j}=\int\star\,\mathrm{d}x$, where[8]

$$\star = \mathrm{N}_{\Sigma_i}^{\mu_i}(x)\left|\exp\left(-\frac{1}{2}\|z\|^2_{\mathrm{R}_j^{-1}}\right)-\exp\left(-\frac{1}{2}\|\tilde{z}_i\|^2_{\mathrm{R}_j^{-1}}\right)\right|,$$

$z=y-h(x)$, $\tilde{z}_i=y-h(\mu_i)-h'(\mu_i)(x-\mu_i)$, $i\in I_1$ (see Thm. 13). Now

$$\epsilon_{i,j}\underset{N\to\infty}{\longrightarrow}0.$$

[8]Here we simplify a little bit our notation.

Proof: First we define sets $C_{i,k}\subset\bar{C}_i$ which become smaller when N becomes larger.

$$C_{i,k}=\left\{x\,\middle|\,\left\|\left[\begin{array}{cc}\mathrm{I}_{d\times d} & 0\end{array}\right](x-\mu_i)\right\|\le\frac{1}{k}\right\},$$

where $i\in I_1$, $k>k_{\min}$ and

$$k_{\min}=\max_j\left(\sqrt{\frac{n_y c_{\mathrm{H}}(3\sqrt{\|\mathrm{R}_j\|}\|\mathrm{R}_j^{-1}\|+2)}{2}},\sqrt{\|\mathrm{R}_j^{-1}\|}\right). \tag{31}$$

Because $C_{i,k}\subset\bar{C}_i$ the Hessian matrices $h_{1j}''(x)$ are bounded when $x\in C_{i,k}$. So there is a constant c_{H} such that (17)

$$\left|\xi^T h_j''(x)\xi\right|\le c_{\mathrm{H}}\xi^T\left[\begin{array}{cc}\mathrm{I}_{d\times d} & 0\\ 0 & 0\end{array}\right]\xi, \tag{32}$$

where $j\in\{1,\ldots,n_y\}$.

Now

$$\epsilon_{i,j}=\int_{C_{i,k}}\star\,\mathrm{d}x+\int_{\complement C_{i,k}}\star\,\mathrm{d}x. \tag{33}$$

We show that $\int_{C_{i,k}}\star\,\mathrm{d}x\underset{k\to\infty}{\longrightarrow}0$ and $\int_{\complement C_{i,k}}\star\,\mathrm{d}x\underset{N\to\infty}{\longrightarrow}0$ for all k. We start to approximate integral $\int_{C_{i,k}}\star\,\mathrm{d}x$, and our goal is to show that $\int_{C_{i,k}}\star\,\mathrm{d}x\underset{k\to\infty}{\longrightarrow}0$. Now

$$\int_{C_{i,k}}\star\,\mathrm{d}x=\int_{C_{i,k}}\mathrm{N}_{\Sigma_i}^{\mu_i}(x)|f_{i,j}(x)|\mathrm{d}x, \tag{34}$$

where

$$f_{i,j}(x)=\exp\left(-\frac{1}{2}\|z\|^2_{\mathrm{R}_j^{-1}}\right)-\exp\left(-\frac{1}{2}\|z+\zeta_i\|^2_{\mathrm{R}_j^{-1}}\right),$$
$$\zeta_i=h(x)-h(\mu_i)-h'(\mu_i)(x-\mu_i). \tag{35}$$

Using Taylor's theorem we get

$$\zeta_i=\sum_{j=1}^{n_y}e_j\frac{1}{2}(x-\mu_i)^T h_j''(\bar{x}_j)(x-\mu_i),$$

where $\bar{x}_j\in C_{i,k}$ for all $j\in\{1,\ldots,n_y\}$ and e_j is the jth column of the identity matrix $\mathrm{I}_{n_y\times n_y}$. Now

$$\|\zeta_i\|\le\sum_{j=1}^{n_y}\left|\frac{1}{2}(x-\mu_i)^T h_j''(\bar{x}_j)(x-\mu_i)\right|$$
$$\overset{(32)}{\le}\sum_{j=1}^{n_y}\frac{c_{\mathrm{H}}}{2}\left\|\left[\begin{array}{cc}\mathrm{I}_{d\times d} & 0\end{array}\right](x-\mu_i)\right\|^2 \tag{36}$$
$$\overset{x\in C_{i,k}}{\le}\frac{n_y c_{\mathrm{H}}}{2k^2},$$

where $k>k_{\min}$. So $\|\zeta_i\|\underset{k\to\infty}{\longrightarrow}0$, for all $i\in I_1$ when $x\in C_{i,k}$. Now we start to approximate $f_{i,j}(x)$, our goal being Eq. (44). We divide the problem into two parts, namely, $\|z\|^2_{\mathrm{R}_j^{-1}}\ge k^2$ and $\|z\|^2_{\mathrm{R}_j^{-1}}<k^2$. First we assume that $\|z\|^2_{\mathrm{R}_j^{-1}}\ge k^2$. Now

$$\|z\|^2\ge\frac{\|z\|^2_{\mathrm{R}_j^{-1}}}{\|\mathrm{R}_j^{-1}\|}\ge\frac{k^2}{\|\mathrm{R}_j^{-1}\|}\overset{(31)}{>}1 \tag{37}$$

and using Eq. (31), Eq. (36) and Eq. (37) we get

$$\|\zeta_i\| \leq \min_j \left(1, \|z\|, \frac{1}{3\sqrt{\|\mathrm{R}_j\|\|\mathrm{R}_j^{-1}\|}}\right). \tag{38}$$

Now

$$\begin{aligned}
&\left| \|z+\zeta_i\|^2_{\mathrm{R}_j^{-1}} - \|z\|^2_{\mathrm{R}_j^{-1}} \right| \\
&= \left| 2z^T\mathrm{R}_j^{-1}\zeta_i + \|\zeta_i\|^2_{\mathrm{R}_j^{-1}} \right| \\
&\leq (2\|z\| + \|\zeta_i\|)\,\|\mathrm{R}_j^{-1}\|\|\zeta_i\| \\
&\overset{(38)}{\leq} 3\|z\|\|\mathrm{R}_j^{-1}\|\|\zeta_i\| \\
&\overset{(40)}{\leq} 3\sqrt{\|\mathrm{R}_j\|}\sqrt{\|z\|^2_{\mathrm{R}_j^{-1}}}\|\mathrm{R}_j^{-1}\|\|\zeta_i\| \\
&\overset{(38)}{\leq} \sqrt{\|z\|^2_{\mathrm{R}_j^{-1}}} \overset{(\|z\|^2_{\mathrm{R}_j^{-1}} \geq k^2 > 1)}{\leq} \frac{1}{k}\|z\|^2_{\mathrm{R}_j^{-1}}.
\end{aligned} \tag{39}$$

Here we used the inequality

$$\begin{aligned}
\|z\|^2 &= z^T\mathrm{R}_j^{-\frac{1}{2}}\mathrm{R}_j\mathrm{R}_j^{-\frac{1}{2}}z \\
&\leq \|\mathrm{R}_j\|\|\mathrm{R}_j^{-\frac{1}{2}}z\|^2 = \|\mathrm{R}_j\|\|z\|^2_{\mathrm{R}_j^{-1}}.
\end{aligned} \tag{40}$$

So when $\|z\|^2_{\mathrm{R}_j^{-1}} \geq k^2$ we can approximate Eq. (35) as follows

$$\begin{aligned}
&f_{i,j}(x) \\
&\leq \exp\left(-\frac{1}{2}\|z\|^2_{\mathrm{R}_j^{-1}}\right) + \exp\left(-\frac{1}{2}\|z+\zeta_i\|^2_{\mathrm{R}_j^{-1}}\right), \\
&\overset{(39)}{\leq} \exp\left(-\frac{1}{2}\|z\|^2_{\mathrm{R}_j^{-1}}\right) + \exp\left(-\frac{k-1}{2k}\|z\|^2_{\mathrm{R}_j^{-1}}\right) \\
&\overset{\|z\|^2_{\mathrm{R}_j^{-1}} \geq k^2}{\leq} 2\exp\left(-\frac{k^2-k}{2}\right).
\end{aligned} \tag{41}$$

Now we assume that $\|z\|^2_{\mathrm{R}_j^{-1}} < k^2$, then

$$\begin{aligned}
&\left| \|z+\zeta_i\|^2_{\mathrm{R}_j^{-1}} - \|z\|^2_{\mathrm{R}_j^{-1}} \right| \\
&= \left| 2z^T\mathrm{R}_j^{-1}\zeta_i + \|\zeta_i\|^2_{\mathrm{R}_j^{-1}} \right| \\
&\overset{(40)}{\leq} \left(2\sqrt{\|\mathrm{R}_j\|}k + \|\zeta_i\|\right)\|\mathrm{R}_j^{-1}\|\|\zeta_i\| \\
&\overset{(36)}{\leq} \left(2\sqrt{\|\mathrm{R}_j\|}k + \|\zeta_i\|\right)\|\mathrm{R}_j^{-1}\|\frac{n_y c_{\mathrm{H}}}{2k^2} \\
&\overset{(38)}{\leq} \left(2\sqrt{\|\mathrm{R}_j\|}k + 1\right)\|\mathrm{R}_j^{-1}\|\frac{n_y c_{\mathrm{H}}}{2k^2} \\
&\leq 2c_{\max}\frac{1}{k},
\end{aligned} \tag{42}$$

where $c_{\max} = \max_j \left(2\sqrt{\|\mathrm{R}_j\|}+1\right)\|\mathrm{R}_j^{-1}\|\frac{n_y c_{\mathrm{H}}}{4}$. So when $\|z\|^2_{\mathrm{R}_j^{-1}} < k^2$, we can approximate Eq. (35) as follows

$$\begin{aligned}
f_{\mathrm{i,j}}(x) &\leq \left| 1 - \exp\left(\frac{1}{2}\|z\|^2_{\mathrm{R}_j^{-1}} - \frac{1}{2}\|z+\zeta_i\|^2_{\mathrm{R}_j^{-1}}\right)\right| \\
&\overset{(42)}{\leq} \exp\left(c_{\max}\frac{1}{k}\right) - 1.
\end{aligned} \tag{43}$$

Combining these results we get that if $k > k_{\text{min}}$, then

$$\begin{aligned}
f_{i,j}(x) \leq \max\Bigg(&2\exp\left(-\frac{k^2-k}{2}\right), \ldots \\
&\exp\left(c_{\max}\frac{1}{k}\right) - 1\Bigg)
\end{aligned} \tag{44}$$

Using this result we get (Eq. (34))

$$\begin{aligned}
\int_{C_{i,k}} \star\, \mathrm{d}x &= \int_{C_{i,k}} \mathrm{N}^{\mu_i}_{\Sigma_i}(x)|f_{i,j}(x)|\mathrm{d}x, \\
&\leq \max\left(2\exp\left(-\frac{k^2-k}{2}\right), \exp\left(c_{\max}\frac{1}{k}\right) - 1\right)
\end{aligned} \tag{45}$$

and then $\int_{C_{i,k}} \star\, \mathrm{d}x \underset{k\to\infty}{\longrightarrow} 0$.[9]

Finally we approximate the second integral $\int_{\complement C_{i,k}} \star\, \mathrm{d}x$ of Eq. (33) and our goal is to show that $\int_{\complement C_{i,k}} \star\, \mathrm{d}x \underset{N\to\infty}{\longrightarrow} 0$, for all k. Now

$$\begin{aligned}
\int_{\complement C_{i,k}} \star\, \mathrm{d}x &\leq \int_{\complement C_{i,k}} \mathrm{N}^{\mu_i}_{\Sigma_i}(x)\mathrm{d}x \\
&= \mathrm{P}(x \in \complement C_{i,k}) \\
&= \mathrm{P}\left(\left\|\begin{bmatrix} \mathrm{I}_{d\times d} & 0 \end{bmatrix}(x-\mu_i)\right\| > \frac{1}{k}\right).
\end{aligned}$$

We know that (see Sec. IV-B, Sec. III-B (conventional approximation) and Corollary 20 (BGMA))

$$\sum_{j=1}^{d} (\Sigma_i)_{j,j} \underset{N\to\infty}{\longrightarrow} 0, \ \forall i \in I_2. \tag{46}$$

Using this information, Chebyshev's inequality and

$$\begin{bmatrix} \mathrm{I}_{d\times d} & 0 \end{bmatrix}(x-\mu_i) \sim \mathrm{N}\left(0, (\Sigma_i)_{(1:d,1:d)}\right),$$

we see that $\int_{\complement C_{i,k}} \star\, \mathrm{d}x \underset{N\to\infty}{\longrightarrow} 0$ for all k and $i \in I_1$. Collecting these results we get Eq. (33) $\epsilon_{i,j} \underset{N\to\infty}{\longrightarrow} 0$.

□

ACKNOWLEDGMENT

The author would like to thank Niilo Sirola, Tommi Perälä and Robert Piché for their comments and suggestions. This study was partly funded by Nokia Corporation. The author acknowledges the financial support of the Nokia Foundation and the Tampere Graduate School in Information Science and Engineering.

[9] Actually it is straightforward to see that this integral converges to zero because $f_{i,j}(\mu_i) = 0$ and function $f_{i,j}(x)$ is continuous. However based on Eq. (44) we have some idea of the speed of convergence.

REFERENCES

[1] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. Springer, 2001.

[2] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter, Particle Filters for Tracking Applications.* Boston, London: Artech House, 2004.

[3] Y. Bar-Shalom, R. X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation, Theory Algorithms and Software*. John Wiley & Sons, 2001.

[4] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in Science and Engineering. Academic Press, 1970, vol. 64.

[5] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, ser. Mathematics in Science and Engineering. Academic Press, 1982, vol. 141-2.

[6] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, ser. Prentice-Hall information and system sciences. Prentice-Hall, 1979.

[7] C. Ma, "Integration of GPS and cellular networks to improve wireless location performance," *Proceedings of ION GPS/GNSS 2003*, pp. 1585–1596, 2003.

[8] G. Heinrichs, F. Dovis, M. Gianola, and P. Mulassano, "Navigation and communication hybrid positioning with a common receiver architecture," *Proceedings of The European Navigation Conference GNSS 2004*, 2004.

[9] M. S. Grewal and A. P. Andrews, *Kalman Filtering Theory and Practice*, ser. Information and system sciences series. Prentice-Hall, 1993.

[10] D. Simon, *Optimal State Estimation Kalman, H_∞ and Nonlinear Approaches*. John Wiley & Sons, 2006.

[11] S. Ali-Löytty, N. Sirola, and R. Piché, "Consistency of three Kalman filter extensions in hybrid navigation," in *Proceedings of The European Navigation Conference GNSS 2005*, Munich, Germany, Jul. 2005.

[12] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *American Control Conference*, vol. 3, 1995, pp. 1628–1632.

[13] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, March 2004.

[14] R. S. Bucy and K. D. Senne, "Digital synthesis of non-linear filters," *Automatica*, vol. 7, no. 3, pp. 287–298, May 1971.

[15] S. C. Kramer and H. W. Sorenson, "Recursive Bayesian estimation using piece-wise constant approximations," *Automatica*, vol. 24, no. 6, pp. 789–801, 1988.

[16] N. Sirola and S. Ali-Löytty, "Local positioning with parallelepiped moving grid," in *Proceedings of 3rd Workshop on Positioning, Navigation and Communication 2006 (WPNC'06)*, Hannover, March 16th 2006, pp. 179–188.

[17] N. Sirola, "Nonlinear filtering with piecewise probability densities," Tampere University of Technology, Research report 87, 2007.

[18] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[19] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 736–746, March 2002.

[20] D. L. Alspach and H. W. Sorenson, "Nonlinear bayesian estimation using gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, Aug 1972.

[21] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, July 1971.

[22] T. Lefebvre, H. Bruyninckx, and J. De Schutter, "Kalman filters for nonlinear systems: a comparison of performance," *International Journal of Control*, vol. 77, no. 7, May 2004.

[23] N. Sirola, S. Ali-Löytty, and R. Piché, "Benchmarking nonlinear filters," in *Nonlinear Statistical Signal Processing Workshop*, Cambridge, September 2006.

[24] W. Cheney and W. Light, *A Course in Approximation Theory*, ser. The Brooks/Cole series in advanced mathematics. Brooks/Cole Publishing Company, 2000.

[25] W. Rudin, *Real and Complex Analysis*, 3rd ed., ser. Mathematics Series. McGraw-Hill Book Company, 1987.

[26] J. T.-H. Lo, "Finite-dimensional sensor orbits and optimal nonlinear filtering," *IEEE Transactions on Information Theory*, vol. 18, no. 5, pp. 583–588, September 1972.

[27] S. Ali-Löytty, "Efficient Gaussian mixture filter for hybrid positioning," in *Proceedings of PLANS 2008*, May 2008, (to be published).

[28] A. N. Shiryayev, *Probability*. Springer-Verlag, 1984.

[29] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, ser. Probability and mathematical statistics. London Academic Press, 1989.

[30] S. Ali-Löytty and N. Sirola, "Gaussian mixture filter in hybrid navigation," in *Proceedings of The European Navigation Conference GNSS 2007*, May 2007, pp. 831–837.

[31] J. Kotecha and P. Djuric, "Gaussian sum particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2602–2612, October 2003.

[32] D. J. Salmond, "Mixture reduction algorithms for target tracking," *State Estimation in Aerospace and Tracking Applications, IEE Colloquium on*, pp. 7/1–7/4, 1989.

[33] T. Ferguson, *A Course in Large Sample Theory*. Chapman & Hall, 1996.

[34] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, ser. Prentice-Hall information and system sciences. Prentice-Hall, 2000.