Mixture truncated unscented Kalman filtering

Ángel F. García-Fernández*, Mark R. Morelande†, Jesús Grajal*
*Dpto. Señales, Sistemas y Radiocomunicaciones, Universidad Politécnica de Madrid, Spain
†Melbourne Systems Laboratory, The University of Melbourne, Australia
Emails: agarcia@gmr.ssr.upm.es, mrmore@unimelb.edu.au, jesus@gmr.ssr.upm.es

***Abstract*—This paper proposes a computationally efficient nonlinear filter that approximates the posterior probability density function (PDF) as a Gaussian mixture. The novelty of this filter lies in the update step. If the likelihood has a bounded support made up of different regions, we can use a modified prior PDF, which is a mixture, that meets Bayes' rule exactly. The central idea of this paper is that a Kalman filter applied to each component of the modified prior mixture can improve the approximation to the posterior provided by the Kalman filter. In practice, bounded support is not necessary.**

***Index Terms*—Bayes' rule, Kalman filter, nonlinear filtering**

I. Introduction

Bayesian filtering aims to estimate the current state of a process based on previous measurements up to the current time. In this framework, all the information about the current state is contained in the posterior probability density function (PDF), i.e., the PDF of the current state given the current and previous measurements. It is of great interest to approximate the posterior as it enables us to apply widely used estimators such as the minimum mean square error (MMSE) estimator [1]. Theoretically, the posterior can be calculated recursively every time step in two phases: prediction and update. The prediction step provides an expression for the prior PDF, which refers to the PDF of the current state given all previous measurements, based on the posterior at the previous time step. The update step establishes the relationship between the posterior (at the current time step) and the prior, which is given by Bayes' rule.

The main difficulty in nonlinear/non-Gaussian systems is that the posterior cannot be calculated analytically and must be approximated. A common family of algorithms used to approximate the posterior are particle filters (PFs) [2]. PFs provide an accurate approximation to the posterior given a sufficiently high number of particles but can perform poorly otherwise. Another usual option is to use Kalman-filter-type algorithms such as the extended Kalman filter (EKF) [1] or unscented Kalman filter (UKF) [3]. KF-type algorithms approximate the posterior as a Gaussian by approximating the KF recursion. Their computational burden is very low compared to PFs but the possibility of optimal inference disappears although if the posterior is unimodal they can provide an accurate enough approximation.

Gaussian mixture approximations to the posterior can also be used. These are useful when the posterior is multimodal. In [4], the prior at the initial time step is approximated as a Gaussian mixture and an EKF is applied to each component. However, the number of components is fixed with time and after several time steps all the components can be the same resulting in a Gaussian rather than a Gaussian mixture approximation. This is not the case for the box Gaussian mixture filter (BGMF) [5] as it adaptively selects the number of components depending on the measurement nonlinearity. Yet, its computational complexity scales exponentially with the state dimension so it suffers from the curse of dimensionality. This makes it unsuitable for moderately high-dimensional problems.

This papers focuses on the update phase with nonlinear measurements. In this case, KF approximations do not perform well if the measurement is informative compared to the prior [6]–[8]. In [6]–[8], the truncated Kalman filter (TKF) is proposed to provide a more accurate Gaussian approximation to the posterior when the measurement function is bijective and the measurement noise PDF support is bounded. The idea is to apply a KF to a truncated prior that meets Bayes' rule exactly rather than the original prior. In practice, most models assume measurement noises with unbounded support. In such cases, the measurement noise PDF can be approximated by one with bounded support and the truncated unscented Kalman filter (TUKF) can be applied.

This paper introduces the mixture TKF (MTKF). It is a generalisation of the TKF when the likelihood support is composed of several regions. As the likelihood is multimodal, the prior and the posterior might be multimodal and, therefore, the MTKF approximates them as Gaussian mixtures rather than single Gaussians. As the MTKF is intractable in general, we also introduce the mixture TUKF (MTUKF) as a practical approximation to the MTKF. The MTUKF is a generalisation of the TUKF for this kind of likelihood. Importantly, its computational complexity is low and it does not suffer from the curse of dimensionality as in the method of [5].

The rest of this paper is organised as follows. In Section II, we review the update step of KF approximations with multimodal priors. In Section III, we present the MTKF. In Section IV, we develop the MTUKF. Numerical examples are provided in Section V. Finally, conclusions are given in Section VI.

II. KF update step with multimodal prior

Regardless of the form of the prior, the KF update as defined in [3] can always be applied. However, if the prior is multimodal, the posterior might be multimodal but the KF always provides a unimodal approximation. Therefore, this strategy does not seem to be appropriate. Nevertheless, the

static multiple-model estimator filter (SMMEF) can be applied using the KF update and preserving the multimodality of the prior [2]. This algorithm is reviewed in this section.

The state is represented by $\mathbf{x} \in \mathbb{R}^{n_x}$. For ease of explanation, we do not include the time index and the conditioning on past measurements. Therefore, the prior PDF is denoted by $p_0(\cdot)$ and is assumed to be a mixture

$$p_0(\mathbf{x}) = \sum_{j=1}^{n} w_j p_{0,j}(\mathbf{x}) \tag{1}$$

where n is the number of components, $p_{0,j}(\cdot)$ is the jth component and w_j is its weight.

We assume the following measurement equation:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \boldsymbol{\eta} \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is the state, $\mathbf{z} \in \mathbb{R}^{n_z}$ is the measurement, $\mathbf{h}(\cdot)$ is a nonlinear function and $\boldsymbol{\eta}$ is the zero mean measurement noise with any PDF and covariance matrix $\mathbf{R}$.

The aim is to calculate a Gaussian mixture approximation to the posterior $q(\cdot)$, which is given by Bayes' rule

$$q(\mathbf{x}) \propto \ell(\mathbf{x};\mathbf{z})\, p_0(\mathbf{x}) \tag{3}$$

where $\ell(\mathbf{x};\mathbf{z})$ is the likelihood of $\mathbf{x}$ after observing $\mathbf{z}$. It should be noted that $\ell(\mathbf{x};\mathbf{z})$ is characterised by (2).

Substituting (1) into (3), the posterior can be written as

$$q(\mathbf{x}) \propto \sum_{j=1}^{n} w_j \ell(\mathbf{x};\mathbf{z})\, p_{0,j}(\mathbf{x}) \tag{4}$$

$$= \sum_{j=1}^{n} w_j \rho_{0,j} q_{0,j}(\mathbf{x}) \tag{5}$$

where

$$q_{0,j}(\mathbf{x}) = \frac{\ell(\mathbf{x};\mathbf{z})\, p_{0,j}(\mathbf{x})}{\rho_{0,j}} \tag{6}$$

$$\rho_{0,j} = \int \ell(\mathbf{x};\mathbf{z})\, p_{0,j}(\mathbf{x})\, d\mathbf{x} \tag{7}$$

where $q_{0,j}(\cdot)$ denotes the posterior for the jth prior component. A typical implementation of the SMMEF consists of approximating (6) and (7) using the KF update, see [2, Sec. 2.3.1]. This is optimal in the linear/Gaussian case. In the nonlinear case, KF-type algorithms such as the EKF [4] or the UKF can be used instead. However, this approach has the following drawbacks:

- D1: If the likelihood is multimodal, $q_{0,j}(\cdot)$ is in general multimodal. However, KF-type approximations always provide a unimodal approximation.
- D2: If the likelihood is unimodal but the measurement is very informative compared to the prior, KF-type approximations do not work properly [6].

Both drawbacks can result in an important loss in performance. TKF approaches were proposed in [6]–[8] as an alternative to conventional KF-type approximations for bijective measurements functions. TKF approaches are able to deal properly with informative measurements (Drawback D2). However, the TKF as presented in [6]–[8] can only be applied to unimodal likelihoods. In the next section, the TKF is generalised so that it is able to overcome Drawbacks D1 and D2.

III. Mixture truncated Kalman filter

The MTKF assumes that the state vector can be written as $\mathbf{x} = \left[\mathbf{a}^T, \mathbf{b}^T\right]^T$, where T denotes transpose, $\mathbf{a} \in \mathbb{R}^{n_a}$, $\mathbf{b} \in \mathbb{R}^{n_b}$ and $n_x = n_a + n_b$, such that the measurement equation, see (2), becomes

$$\mathbf{z} = \mathbf{h}(\mathbf{a}) + \boldsymbol{\eta} \tag{8}$$

The MTKF is derived under the following assumptions:

- A1: The measurement function is such that the equation

$$\mathbf{z} = \mathbf{h}(\mathbf{a}) \tag{9}$$

  admits m real solutions $\tilde{\mathbf{a}}_i(\mathbf{z}) \in \mathbb{R}$ $i = 1, ..., m$ where the number of solutions might change depending on $\mathbf{z}$.
- A2: The PDF of the measurement noise has a bounded, connected support $I_{\boldsymbol{\eta}}$.

Assumptions A1 and A2 imply that the support of the likelihood $I = \{\mathbf{a} : \mathbf{z} - \mathbf{h}(\mathbf{a}) \in I_{\boldsymbol{\eta}}\}$ can be split into m disjoint regions $\{I_1, ..., I_m : I_i \cap I_j = \oslash \text{ if } i \neq j\}$, where I_i contains $\tilde{\mathbf{a}}_i(\mathbf{z})$, if the size $|I_{\boldsymbol{\eta}}|$ of $I_{\boldsymbol{\eta}}$ is sufficiently small. These regions depend on $\mathbf{z}$ but explicit dependence has been removed for the sake of clarity. Then, the likelihood can be written as

$$\ell(\mathbf{x};\mathbf{z}) = \ell(\mathbf{x};\mathbf{z}) \sum_{i=1}^{m} \chi_i(\mathbf{x}) \tag{10}$$

where $\chi_i(\mathbf{x}) = 1$ if $\mathbf{x} \in I_i \times \mathbb{R}^{n_b}$ and zero otherwise.

Substituting (10) into (4), Bayes' rule becomes

$$q(\mathbf{x}) \propto \ell(\mathbf{x};\mathbf{z}) \sum_{i=1}^{m} \chi_i(\mathbf{x}) \sum_{j=1}^{n} w_j p_{0,j}(\mathbf{x}) \tag{11}$$

As with the TKF [6]–[8], Bayes' rule is also met by the truncated prior

$$p_T(\mathbf{x}) \propto \sum_{i=1}^{m} \chi_i(\mathbf{x}) \sum_{j=1}^{n} w_j p_{0,j}(\mathbf{x}) \tag{12}$$

and we can write the posterior as

$$q(\mathbf{x}) \propto \ell(\mathbf{x};\mathbf{z})\, p_T(\mathbf{x}) \tag{13}$$

The truncated prior can be written as

$$p_T(\mathbf{x}) = \frac{\sum_{j=1}^{n} w_j p_{0,j}(\mathbf{x}) \sum_{i=1}^{m} \chi_i(\mathbf{x})}{\sum_{j=1}^{n} w_j \sum_{i=1}^{m} \int_{I_i} p_{0,j}(\boldsymbol{\xi})\, d\boldsymbol{\xi}} = \frac{\sum_{j=1}^{n} w_j \sum_{i=1}^{m} \varepsilon_{i,j} p_{i,j}(\mathbf{x})}{\sum_{j=1}^{n} w_j \sum_{i=1}^{m} \varepsilon_{i,j}} \tag{14}$$

where

$$p_{i,j}(\mathbf{x}) = \frac{p_{0,j}(\mathbf{x})\, \chi_i(\mathbf{x})}{\varepsilon_{i,j}} \tag{15}$$

is the truncated PDF of the jth component of the prior $p_{0,j}(\cdot)$ in region I_i and

$$\varepsilon_{i,j} = \int p_{0,j}(\mathbf{x})\, \chi_i(\mathbf{x})\, d\mathbf{x} \tag{16}$$

For $j = 1, ..., n$, $i = 1, ..., m$, let

$$q_{i,j}(\mathbf{x}) = \frac{\ell(\mathbf{x};\mathbf{z})\, p_{i,j}(\mathbf{x})}{\rho_{i,j}} \tag{17}$$

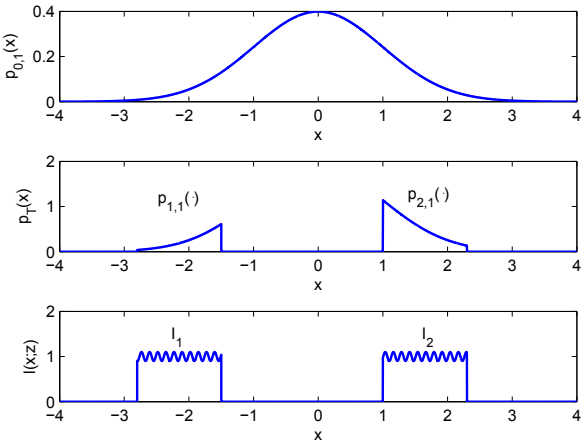


Figure 1: Graphical representation of the "priors" $p_{0,1}(\mathbf{x})$, $p_T(\mathbf{x})$ and the likelihood $\ell(\mathbf{x};\mathbf{z})$ in one dimension. When the likelihood has a bounded support, Bayes' rule is met by the "prior" PDFs $p_{0,1}(\mathbf{x})$ and $p_T(\mathbf{x})$.

where

$$\rho_{i,j} = \int \ell(\mathbf{x};\mathbf{z})\, p_{i,j}(\mathbf{x})\, d\mathbf{x} \tag{18}$$

Substituting (14) into (13)

$$q(\mathbf{x}) \propto \sum_{j=1}^{n} w_j \sum_{i=1}^{m} \varepsilon_{i,j} \ell(\mathbf{x};\mathbf{z})\, p_{i,j}(\mathbf{x}) \tag{19}$$

Using (17) and (18) into (19), we get

$$q(\mathbf{x}) \propto \sum_{j=1}^{n} \sum_{i=1}^{m} w_j \varepsilon_{i,j} \rho_{i,j} q_{i,j}(\mathbf{x}) \tag{20}$$

As indicated in Section II, the usual approach to approximate the posterior as a Gaussian mixture is to apply a KF to each $p_{0,j}(\cdot)$ to approximate (6) and (7). However, equation (20) suggests an alternative approach. A KF can be applied to each $p_{i,j}(\cdot)$ to approximate (17) and (18). This way we can handle multimodal likelihoods (Drawback D1) and informative measurements (Drawback D2) as it uses the same idea as the TKF and the TKF works well for informative measurements.

An exemplar situation with $n = 1$ and $m = 2$ is illustrated in Figure 1. The SMMEF approximates the posterior applying a KF to the prior $p_{0,1}(\cdot)$ and the posterior would be represented as a single Gaussian. As the posterior is multimodal, this approximation is rather poor. On the contrary, the MTKF applies a KF to the truncated priors $p_{1,1}(\cdot)$ and $p_{2,1}(\cdot)$ in regions I_1 and I_2 and represents the posterior as a Gaussian mixture providing a more accurate approximation to the posterior.

IV. Mixture truncated unscented Kalman filter

If the measurement noise support is not bounded, we can still apply the MTKF by approximating the noise PDF as one with bounded support. The reasons why this is useful are thoroughly explained in [8].

We assume the components of the prior mixture are Gaussian

$$p_{0,j}(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}_{p,0,j}, \mathbf{P}_{p,0,j}) \tag{21}$$

where

$$\hat{\mathbf{x}}_{p,0,j} = \begin{bmatrix} \boldsymbol{\mu}_{a,0,j} \\ \boldsymbol{\mu}_{b,0,j} \end{bmatrix} \tag{22}$$

$$\mathbf{P}_{p,0,j} = \begin{bmatrix} \boldsymbol{\Sigma}_{a,0,j} & \boldsymbol{\Sigma}_{ab,0,j} \\ \boldsymbol{\Sigma}_{ab,0,j}^T & \boldsymbol{\Sigma}_{b,0,j} \end{bmatrix} \tag{23}$$

A closed form solution of the MTKF is rarely tractable because of the calculation of the regions $I_1, ..., I_m$ and the KF moments for $p_{i,j}(\cdot)$. As in the TUKF [6], we will use the unscented transformation [3] to approximate the KF moments for $p_{i,j}(\cdot)$. Then we show how the UKF can be used to approximate the densities $q_{i,j}(\cdot)$ and the weights $\varepsilon_{i,j}$. The assumptions used in the approximation of the priors, posteriors and weights mean the MTUKF should only be applied in certain situations. We therefore develop a rule for adaptively switching between the MTUKF and the SMMEF with UKFs.

A. Approximation of the first two moments of $p_{i,j}(\cdot)$

The same approximations as in the TUKF are used [6]

- AP1: The measurement function $\mathbf{h}(\cdot)$ is locally linear.
- AP2: The marginal prior of $\mathbf{a}$ of the jth component, $p_{0,j}^a(\cdot)$, is constant over the region I_i.
- AP3: The measurement noise satisfies $\boldsymbol{\eta} \sim U_{I_{\boldsymbol{\eta}}}$ where $I_{\boldsymbol{\eta}}$ is such that the truncated noise has the same first two moments as the real noise $\mathsf{E}[\boldsymbol{\eta}] = 0$ and $\mathsf{cov}[\boldsymbol{\eta}] = \mathbf{R}$.

Using AP1, the measurement function in region I_i can be approximated as

$$\mathbf{h}(\mathbf{a}) \approx \mathbf{h}(\tilde{\mathbf{a}}_i(\mathbf{z})) + \tilde{\mathbf{H}}_i(\mathbf{a} - \tilde{\mathbf{a}}_i(\mathbf{z})) \tag{24}$$

where

$$\tilde{\mathbf{H}}_i = \left[\nabla_{\mathbf{a}} \mathbf{h}^T(\mathbf{a})\right]^T \Big|_{\mathbf{a}=\tilde{\mathbf{a}}_i(\mathbf{z})} \tag{25}$$

is the Jacobian of $\mathbf{h}(\cdot)$ evaluated at $\tilde{\mathbf{a}}_i(\mathbf{z})$ (recall that $\tilde{\mathbf{a}}_i(\mathbf{z})$ is the solution to (9) in region I_i).

Following the same procedure as in the TUKF (see [6, Sec. 3.4.3]), it can be shown that the mean $\hat{\mathbf{x}}_{p,i,j}$ and covariance matrix $\mathbf{P}_{p,i,j}$ of $p_{i,j}(\cdot)$ under Approximations AP1, AP2 and AP3 are

$$\hat{\mathbf{x}}_{p,i,j} = \begin{bmatrix} \boldsymbol{\mu}_{a,i} \\ \boldsymbol{\mu}_{b,i,j} \end{bmatrix} \tag{26}$$

$$\mathbf{P}_{p,i,j} = \begin{bmatrix} \boldsymbol{\Sigma}_{a,i} & \boldsymbol{\Sigma}_{ab,i,j} \\ \boldsymbol{\Sigma}_{ab,i,j}^T & \boldsymbol{\Sigma}_{b,i,j} \end{bmatrix} \tag{27}$$

where

$$\boldsymbol{\mu}_{a,i} = \tilde{\mathbf{a}}_i(\mathbf{z}) \tag{28}$$

$$\boldsymbol{\Sigma}_{a,i} = \tilde{\mathbf{H}}_i^{-1} \mathbf{R} \left(\tilde{\mathbf{H}}_i^{-1}\right)^T \tag{29}$$

$$\boldsymbol{\mu}_{b,i,j} = \boldsymbol{\mu}_{b,0,j} + \boldsymbol{\Sigma}_{ab,0,j}^T \boldsymbol{\Sigma}_{a,0,j}^{-1} \left(\boldsymbol{\mu}_{a,i} - \boldsymbol{\mu}_{a,0,j}\right) \tag{30}$$

$$\boldsymbol{\Sigma}_{ab,i,j} = \boldsymbol{\Sigma}_{a,i} \left(\boldsymbol{\Sigma}_{a,0,j}^{-1}\right)^T \boldsymbol{\Sigma}_{ab,0,j} \tag{31}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{b,i,j} =&\boldsymbol{\Gamma}_j - \left(\boldsymbol{\mu}_{b,i,j} - \boldsymbol{\mu}_{b,0,j}\right)\left(\boldsymbol{\mu}_{b,i,j} - \boldsymbol{\mu}_{b,0,j}\right)^T \\ &+ \boldsymbol{\Sigma}_{ab,0,j}^T \boldsymbol{\Sigma}_{a,0,j}^{-1}\left[\boldsymbol{\Sigma}_{a,i,j} + \left(\boldsymbol{\mu}_{a,i,j} - \boldsymbol{\mu}_{a,0,j}\right)\right. \\ &\left.\left(\boldsymbol{\mu}_{a,i,j} - \boldsymbol{\mu}_{a,0,j}\right)^T\right]\left(\boldsymbol{\Sigma}_{a,0,j}^{-1}\right)^T \boldsymbol{\Sigma}_{ab,0,j} \end{aligned} \tag{32}$$

and

$$\boldsymbol{\Gamma}_j = \boldsymbol{\Sigma}_{b,0,j} - \boldsymbol{\Sigma}_{ab,0,j}^T \boldsymbol{\Sigma}_{a,0,j}^{-1} \boldsymbol{\Sigma}_{ab,0,j} \tag{33}$$

B. UKF approximations to $q_{i,j}(\cdot)$ and $\rho_{i,j}$

The KF approximations to $q_{i,j}(\cdot)$ and $\rho_{i,j}$ are

$$q_{i,j}(\mathbf{x}) \approx \mathcal{N}\left(\mathbf{x}; \hat{\mathbf{x}}_{u,i,j}, \mathbf{P}_{u,i,j}\right) \tag{34}$$

$$\rho_{i,j} \approx \mathcal{N}\left(\mathbf{z}; \hat{\mathbf{z}}_{i,j}, \mathbf{S}_{i,j}\right) \tag{35}$$

where

$$\hat{\mathbf{x}}_{u,i,j} = \hat{\mathbf{x}}_{p,i,j} + \boldsymbol{\Psi}_{i,j}\mathbf{S}_{i,j}^{-1}\left(\mathbf{z} - \hat{\mathbf{z}}_{i,j}\right) \tag{36}$$

$$\mathbf{P}_{u,i,j} = \mathbf{P}_{p,i,j} - \boldsymbol{\Psi}_{i,j}\mathbf{S}_{i,j}^{-1}\boldsymbol{\Psi}_{i,j}^T \tag{37}$$

and

$$\hat{\mathbf{z}}_{i,j} = \int \mathsf{E}\left[\mathbf{z}\,|\mathbf{x}\right] p_{i,j}(\mathbf{x})\, d\mathbf{x} \tag{38}$$

$$\mathbf{S}_{i,j} = \int \mathsf{E}\left[\left(\mathbf{z} - \hat{\mathbf{z}}_{i,j}\right)\left(\mathbf{z} - \hat{\mathbf{z}}_{i,j}\right)^T |\mathbf{x}\right] p_{i,j}(\mathbf{x})\, d\mathbf{x} \tag{39}$$

$$\boldsymbol{\Psi}_{i,j} = \int \mathsf{E}\left[\left(\mathbf{x} - \hat{\mathbf{x}}_{p,i,j}\right)\left(\mathbf{z} - \hat{\mathbf{z}}_{i,j}\right)^T |\mathbf{x}\right] p_{i,j}(\mathbf{x})\, d\mathbf{x} \tag{40}$$

Therefore, the KF approximation to $q_{i,j}(\cdot)$ amounts to calculating the moments (38)-(40). As these cannot be calculated analytically in general, we resort to the unscented transformation (UT) [9], or more precisely, the UT for conditionally linear models [10].

The UKF approximation proceeds by selecting N_s sigma points $\mathcal{A}_{i,j}^1, ..., \mathcal{A}_{i,j}^{N_s}$ along with weights $w^1, ..., w^{N_s}$. These sigma points and weights match the first two moments of $p_{i,j}(\mathbf{a})$ and can be found using any of the methods discussed in [3]. The transformed sigma-points are calculated as

$$\mathcal{Z}_{i,j}^l = \mathbf{h}\left(\mathcal{A}_{i,j}^l\right), \quad l = 1, ..., N_s \tag{41}$$

Then, the UT approximation to (38) is

$$\hat{\mathbf{z}}_{i,j} = \sum_{l=1}^{N_s} w^l \mathcal{Z}_{i,j}^l \tag{42}$$

The second integral to approximate is (39)

$$\mathbf{S}_{i,j} = \mathbf{R} + \sum_{l=1}^{N_s} w^l \left(\mathcal{Z}_{i,j}^l - \hat{\mathbf{z}}_{i,j}\right)\left(\mathcal{Z}_{i,j}^l - \hat{\mathbf{z}}_{i,j}\right)^T \tag{43}$$

Let $\mathcal{X}_{i,j}^l = \left[\left(\mathcal{A}_{i,j}^l\right)^T, \left(\boldsymbol{\upsilon}_j\left(\mathcal{A}_{i,j}^l\right)\right)^T\right]^T$ for $l = 1, ..., N_s$ where

$$\boldsymbol{\upsilon}_j(\mathbf{a}) = \boldsymbol{\mu}_{b,0,j} + \boldsymbol{\Sigma}_{ab,0,j}^T \boldsymbol{\Sigma}_{a,0,j}^{-1}\left(\mathbf{a} - \boldsymbol{\mu}_{a,0,j}\right) \tag{44}$$

Then, (40) can be approximated as [10]

$$\boldsymbol{\Psi}_{i,j} = \sum_{l=1}^{N_s} w^l \left(\mathcal{X}_{i,j}^l - \hat{\mathbf{x}}_{p,i,j}\right)\left(\mathcal{Z}_j^l - \hat{\mathbf{z}}_{i,j}\right)^T \tag{45}$$

It should be noted that the approximated first two moments of $p_{i,j}(\mathbf{a})$, $\boldsymbol{\mu}_{a,i}$ and $\boldsymbol{\Sigma}_{a,i}$, which are given by (28) and (29), do not depend on j because of AP2. Thus, $\hat{\mathbf{z}}_{i,j}$ and $\mathbf{S}_{i,j}$ do not depend on j either although $\boldsymbol{\Psi}_{i,j}$ does depend on j.

C. Approximation of $\varepsilon_{i,j}$

Using Approximation AP2, $\varepsilon_{i,j}$ becomes

$$\begin{aligned}\varepsilon_{i,j} =& \int p_{0,j}(\mathbf{x})\, \chi_i(\mathbf{x})\, d\mathbf{x} \\ \approx& p_{0,j}^a\left(\tilde{\mathbf{a}}_i(\mathbf{z})\right) \int \chi_i(\mathbf{x})\, d\mathbf{x}\end{aligned} \tag{46}$$

where $\tilde{\mathbf{a}}_i(\mathbf{z})$ is the point that maximises the likelihood in region I_i. This approximation is accurate if I_i is small enough, i.e., the measurement is sufficiently informative.

Using Approximation AP1 and (24), (46) can be approximated as

$$\varepsilon_{i,j} \approx p_{0,j}^a\left(\tilde{\mathbf{a}}_i(\mathbf{z})\right) \left|\det\left(\tilde{\mathbf{H}}_i^{-1}\right)\right| |I_{\boldsymbol{\eta}}| \tag{47}$$

where $|I_{\boldsymbol{\eta}}|$ is the size of the support of the measurement (see [6, Appendix A.5] for a detailed analogue calculation). Note that because of the normalisation of the final weights of the mixture, see (20), the result is independent of $|I_{\boldsymbol{\eta}}|$ so it does not have to be approximated.

D. Adaptive selection

As discussed in [6], [8], Approximations AP1 and AP2 are accurate if the measurement is informative compared to the prior but they are not otherwise. Therefore, the procedure explained so far would have a good performance if the measurement is informative but bad otherwise. In [6]–[8], we proposed an adaptive way so that the UKF approximation is favoured when the measurement is not informative and the TUKF approximation is favoured when the measurement is informative. In the same way, here, we can favour the SMMEF for non-informative measurements or the MTUKF for informative measurements. This adaptive scheme approximates the posterior as

$$\begin{aligned}q(\mathbf{x}) =& \alpha \frac{\sum_{j=1}^n \sum_{i=1}^m w_j \varepsilon_{i,j} \rho_{i,j} q_{i,j}(\mathbf{x})}{\sum_{j=1}^n \sum_{i=1}^m w_j \varepsilon_{i,j} \rho_{i,j}} \\ &+ (1-\alpha)\frac{\sum_{j=1}^n w_j \rho_{0,j} q_{0,j}(\mathbf{x})}{\sum_{j=1}^n w_j \rho_{0,j}}\end{aligned} \tag{48}$$

where the mixture that accompanies α is the output of the UKFs applied to the mixture of truncated priors and the mixture that accompanies $1-\alpha$ is the output of the SMMEF that uses UKF approximations to $q_{0,j}(\cdot)$. The parameter $\alpha \in [0, 1]$ and should tend to 1 for informative measurements such that we favour the truncated prior and should tend to 0 for non-informative measurements such that the true prior is favoured. The following rule that meets these requirements is proposed:

$$\alpha = \frac{\gamma \sum_{j=1}^n w_j \mathrm{tr}\left(\boldsymbol{\Sigma}_{a,0,j}\right)}{\gamma \sum_{j=1}^n w_j \mathrm{tr}\left(\boldsymbol{\Sigma}_{a,0,j}\right) + \frac{(1-\gamma)}{m}\sum_{i=1}^m \mathrm{tr}\left(\boldsymbol{\Sigma}_{a,i}\right)} \tag{49}$$

where $\gamma \in [0, 1]$ is a parameter that controls the weights of the traces of the covariance matrices to select α and $\mathrm{tr}(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$. Note that the number of components in the posterior mixture is higher than in the prior mixture. This leads to an ever increasing number of components in successive iterations of the filter. Therefore, in practice, the number of components must be controlled by merging the

Table I: Steps of the update phase of the MTUKF

- For $i = 1, ..., m$
  - Calculate $\boldsymbol{\mu}_{a,i}$ and $\boldsymbol{\Sigma}_{a,i}$ using (28) and (29).
  - Calculate the sigma points $\mathcal{A}'^{1}_{i}, ..., \mathcal{A}'^{N_s}_{i}$ matching the moments $\boldsymbol{\mu}_{a,i}$ and $\boldsymbol{\Sigma}_{a,i}$ given by (28) and (29).
- For $j = 1, ..., n$
  - Calculate the sigma points $\mathcal{A}^{1}_{0,j}, ..., \mathcal{A}^{N_s}_{0,j}$ of $p^{a}_{0,j}(\cdot)$, which match the moments $\boldsymbol{\mu}_{a,0,j}$ and $\boldsymbol{\Sigma}_{a,0,j}$ given by (22) and (23) [10].
  - Compute the transformed sigma points $\mathcal{Z}^{1}_{0,j}, ..., \mathcal{Z}^{N_s}_{0,j}$ using (41).
  - Approximate $\hat{\mathbf{z}}_{0,j}, \mathbf{S}_{0,j}$ and $\boldsymbol{\Psi}_{0,j}$ using (42), (43) and (45).
  - Approximate $q_{0,j}(\cdot)$ by its first two moments $\hat{\mathbf{x}}_{u,0,1}$ and $\mathbf{P}_{u,0,1}$ using (36) and (37).
  - Approximate $\rho_{0,j}$ using (35).
  - For $i = 1, ..., m$
    - Assign the sigma points $\left(\mathcal{A}^{1}_{i,j}, ..., \mathcal{A}^{N_s}_{i,j}\right) = \left(\mathcal{A}'^{1}_{i}, ..., \mathcal{A}'^{N_s}_{i}\right)$ of $p_{i,j}(\cdot)$.
    - Compute the transformed sigma points $\mathcal{Z}^{1}_{i,j}, ..., \mathcal{Z}^{N_s}_{i,j}$ using (41).
    - Approximate $\hat{\mathbf{z}}_{i,j}, \mathbf{S}_{i,j}$ and $\boldsymbol{\Psi}_{i,j}$ using (42), (43) and (45).
    - Approximate $q_{i,j}(\cdot)$ by its first two moments $\hat{\mathbf{x}}_{u,i,j}$ and $\mathbf{P}_{u,i,j}$ using (36) and (37).
    - Approximate $\rho_{i,j}$ using (35).
    - Approximate $\varepsilon_{i,j}$ using (47).
- Calculate α using (49).
- Calculate the final weights of the mixture of the posterior given by (48).
- Reduce the number of components of the posterior [11].

components that are sufficiently alike and setting an upper limit on the number of components. In the implementations, we use the joining algorithm in [11]. Finally, the steps of the algorithm are shown in Table I.

V. Numerical examples

In this section, the performances of several algorithms are analysed in two examples. The parameter γ of the MTUKF, which is needed to calculate (49), is set to 0.1 as in other TUKF approaches [6]–[8] and the maximum number of components of the posterior mixture is 10. The following algorithms are analysed: UKF, cubature KF (CKF) [12], Monte Carlo KF (MCKF) [6], [8], EKF, BGMF [5], SMMEF and a particle filter (PF) based on sampling importance resampling [13], in which the importance density is the prior. The UKF uses $N_s = 2n_a + 1$ sigma points and the weight of the sigma point located on the mean is 1/3. The MCKF uses 20000 samples from the prior and the PF uses 10000 particles. The BGMF has been implemented with parameters $N = 1$ and $c_\Sigma = 1$, see [5]. The SMMEF has been implemented using a bank of EKFs [4] in which the prior at the initial time step is approximated using the box Gaussian mixture approximation with parameters $N = 1$ and $c_\Sigma = 1$, see [5]. The performances of the algorithms are analysed using a Monte Carlo simulation with 200 runs.

A. Univariate nonstationary growth model

1) Model description: The univariate nonstationary growth model (UNGM) is characterised by:

$$p\left(x^k \left| x^{k-1}\right.\right) = \mathcal{N}\left(x^k;\, f^k\left(x^{k-1}\right), Q\right) \tag{50}$$

$$f^k\left(x^{k-1}\right) = \frac{x^{k-1}}{2} + \frac{25x^{k-1}}{1 + \left(x^{k-1}\right)^2} + 8\cos\left(1.2k\right) \tag{51}$$

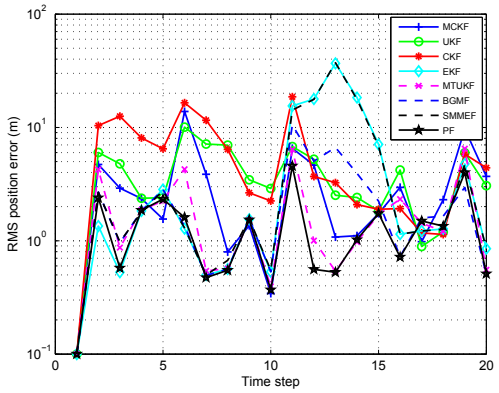


Figure 2: RMS error plotted in log scale against time for the first 20 time steps

$$p\left(z^k \left| x^k\right.\right) = \mathcal{N}\left(z^k;\, \frac{\left(x^k\right)^2}{20},\, R\right) \tag{52}$$

where $\mathcal{N}(x;\, \overline{x},\, Q)$ is the Gaussian PDF evaluated at x with mean $\overline{x}$ and variance Q. The same parameters as in [6] are used: $Q = 1$, $R = 1$, data was generated using $x^0 = 0.1$, the prior PDF at time step 0 is $x^0 \backsim \mathcal{N}\left(x^0;\, 0, 1\right)$ and the simulation has 100 time steps. The MTUKF prediction step uses a UKF prediction step for each component of the mixture.

2) MTUKF update: In this case, (9) admits $m = 2$ solutions if $z^k > 0$

$$\tilde{a}_1\left(z^k\right) = \sqrt{20z^k} \tag{53}$$

$$\tilde{a}_2\left(z^k\right) = -\sqrt{20z^k} \tag{54}$$

or $m = 0$ if $z^k < 0$ (if $z^k = 0$, (9) admits one solution but as it has probability zero, it is not considered). Therefore, if $z^k < 0$ the MTUKF cannot be applied and the SMMEF with UKFs is used instead (note that this would be equivalent to setting $\alpha = 0$ in the usual MTUKF update, see Table I).

3) Algorithms' performances: The RMS error plotted in log scale against time for the algorithms is shown in Figure 2. Only the first twenty time steps have been represented for clarity of representation. The PF reaches the lowest error among the algorithms followed by the MTUKF. Conventional KF-type algorithms are far from the performance of the PF and MTUKF because of Drawbacks D1 and D2 explained in Section II. After around 7 time steps, the SMMEF and the EKF have the same error as the SMMEF usually tends to converge to the EKF as time goes on. Therefore, even though the SMMEF approximates the posterior as a Gaussian mixture, in practice, it performs as the EKF in the long run.

Table II shows the averaged RMS error over time against the measurement noise variance. In brackets, the maximum number of components of the MTUKF is shown. PF is the best algorithm for the analysed noise variances followed by the MTUKF. The BGMF also has a very good performance although it does not perform very well for $R \in \{0.1, 1, 10\}$. Increasing the parameter N of this filter improves performance but also increases computational burden. KF-type algorithms perform much worse than the MTUKF and the PF because

the posterior is multimodal at some time steps especially for low R. KF-type algorithms' performances get worse than PF and MTUKF's as the measurement gets more informative. In addition, the error of the MTUKF is not affected much by the maximum number of components as, in this example, the posterior can be properly represented by two components at all time steps. The relatively good performance of the MCKF is due to the fact that it uses an MC approximation in the prediction step, which is very nonlinear, see (50).

B. Tracking with two sensors

1) Model description: The state vector of the target at time k is $\mathbf{x}^k = \left[\mathbf{a}^k, \mathbf{b}^k\right]^T$ where $\mathbf{a}^k = \left[a_x^k, a_y^k\right]^T$ is the position vector and $\mathbf{b}^k$ is the velocity vector of the target. The dynamic model of the target is the nearly-constant velocity model [1]:

$$p\left(\mathbf{x}^{k+1}\left|\mathbf{x}^k\right.\right) = \mathcal{N}\left(\mathbf{x}^{k+1};\, \mathbf{F}\mathbf{x}^k,\, \mathbf{Q}\right) \tag{55}$$

$$\mathbf{F} = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix} \otimes \mathbf{I}_2 \tag{56}$$

$$\mathbf{Q} = \sigma_u^2 \begin{pmatrix} \tau^3/3 & \tau^2/2 \\ \tau^2/2 & \tau \end{pmatrix} \otimes \mathbf{I}_2 \tag{57}$$

where $\otimes$ is the Kronecker product, τ is the sampling period and σ_u^2 is the continuous-time process noise intensity [1].

The prior PDF of the target's state is:

$$\mathbf{x}^0 \backsim \mathcal{N}\left(\mathbf{x}^0;\, \overline{\mathbf{x}}^0, \mathbf{Q}\right) \tag{58}$$

where we have used the prior covariance matrix equal to the process noise covariance. There are two sensors located at $[m_{1,x}, m_{1,y}]^T = [0,0]^T$ and $[m_{2,x}, m_{2,y}]^T = [m_{2,x}, 0]^T$. These sensors measure the target's range at times $\mathrm{mod}\,(k, N_b) \neq 0$ and the target's bearing at times $\mathrm{mod}\,(k, N_b) = 0$ where N_b is a parameter of the sensors such that

$$\mathbf{z}^k = \begin{cases} \left[h_r\left(a_x^k, a_y^k, m_{1,x}, m_{1,y}\right), h_r\left(a_x^k, a_y^k, m_{2,x}, m_{2,y}\right)\right]^T \\ \quad + \left[\eta_{r,1}, \eta_{r,2}\right]^T \quad \text{if } \mathrm{mod}\,(k, N_b) \neq 0 \\ \left[h_\theta\left(a_x^k, a_y^k, m_{1,x}, m_{1,y}\right), h_\theta\left(a_x^k, a_y^k, m_{2,x}, m_{2,y}\right)\right]^T \\ \quad + \left[\eta_{\theta,1}, \eta_{\theta,2}\right]^T \quad \text{if } \mathrm{mod}\,(k, N_b) = 0 \end{cases} \tag{59}$$

$$h_r\left(a_x^k, a_y^k, m_{1,x}, m_{1,y}\right) = \sqrt{\left(a_x^k - m_{1,x}\right)^2 + \left(a_y^k - m_{1,y}\right)^2} \tag{60}$$

$$h_\theta\left(a_x^k, a_y^k, m_{1,x}, m_{1,y}\right) = \mathrm{atan2}\left(a_y^k - m_{1,y}, a_x^k - m_{1,x}\right) \tag{61}$$

where $\mathrm{atan2}\,(\cdot,\cdot)$ is the four-quadrant inverse tangent, $\mathbf{z}^k = \left[z_1^k, z_2^k\right]^T$ is the measurement vector, $\eta_{r,i}$ is the range noise for the ith sensor and $\eta_{\theta,i}$ is the bearing noise for the ith sensor. All the measurement noises are independent. The range noises have mean zero and covariance matrix R_r. The bearing noises have mean zero and covariance matrix R_θ. Note that if the sensors always measure only range, after the target crosses the axis $y = 0$ several times (this axis is where the sensors lie) the posterior becomes bimodal with one node situated at a position (a_x, a_y) and another $(a_x, -a_y)$ with even weights. In this case, the MMSE estimator will lie in the axis $y = 0$ and therefore it is not useful for tracking purposes. To avoid this, bearing measurements are used every N_b steps. These measurements make the posterior unimodal and, then, the MMSE estimator becomes useful.

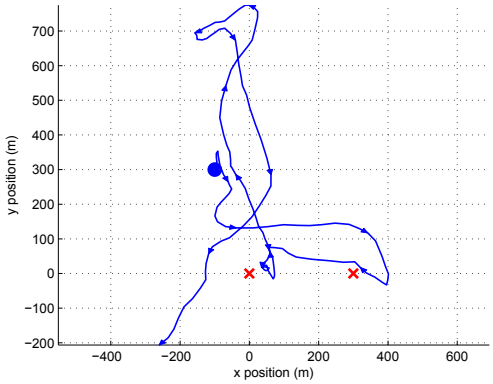


Figure 3: Scenario for the tracking example: The initial target position is represented by a blue circle. The target position and direction of movement every 10 time steps are represented by arrows. The two sensor locations are represented by red crosses $m_{2,x} = 300\,\mathrm{m}$.

The scenario of the simulation is shown in Figure 3. The sampling period of the trajectory is $\tau = 1\,\mathrm{s}$, $\sigma_u = 10\,\mathrm{m/s^{3/2}}$ and there are 149 time steps. This trajectory corresponds to one realisation of the dynamic system described by (55). In each Monte Carlo run, the filter prior mean $\overline{\mathbf{x}}^0$ is drawn from a Gaussian distribution whose mean is the real state of the target at time 0 and whose covariance matrix is $\mathbf{Q}$.

The variances of the range and bearing noises are $R_r = 1\,\mathrm{m}^2$ and $R_\theta = (\pi/180)^2\,\mathrm{rad}^2$, respectively. Bearing measurements are taken every $N_b = 40$ steps.

2) MTUKF update: Let us assume we use range measurements at time k and let

$$p = \left(z_1^k\right)^2 - \frac{1}{4\left(m_{2,x}\right)^2}\left[\left(\left(z_1^k\right)^2 - \left(z_2^k\right)^2\right)^2 + \left(m_{2,x}\right)^4 + 2\left(\left(z_1^k\right)^2 - \left(z_2^k\right)^2\right)\left(m_{2,x}\right)^2\right] \tag{62}$$

It can be shown that (9) has $m = 2$ solutions if $p > 0$

$$\tilde{a}_{1,x}\left(\mathbf{z}^k\right) = \frac{\left(\left(z_1^k\right)^2 - \left(z_2^k\right)^2\right) + \left(m_{2,x}\right)^2}{2m_{2,x}} \tag{63}$$

$$\tilde{a}_{1,y}\left(\mathbf{z}^k\right) = \sqrt{p} \tag{64}$$

$$\tilde{a}_{2,x}\left(\mathbf{z}^k\right) = \tilde{a}_{1,x}\left(\mathbf{z}^k\right) \tag{65}$$

$$\tilde{a}_{2,y}\left(\mathbf{z}^k\right) = -\sqrt{p} \tag{66}$$

where $\tilde{\mathbf{a}}_i\left(\mathbf{z}\right) = \left[\tilde{a}_{i,x}\left(\mathbf{z}^k\right), \tilde{a}_{i,y}\left(\mathbf{z}^k\right)\right]^T$ and we have assumed that $[m_{1,x}, m_{1,y}]^T = [0,0]^T$ and $[m_{2,x}, m_{2,y}]^T = [m_{2,x}, 0]^T$. If $p < 0$, there is no real solution to (9), then, the MTUKF cannot be applied and the SMMEF is used instead $(\alpha = 0)$.

Let us assume we use bearing measurements at time k and let

$$p_x = \frac{m_{2,x}\tan\left(z_2^k\right)}{\tan\left(z_2^k\right) - \tan\left(z_1^k\right)} \tag{67}$$

$$p_y = \tan\left(z_1^k\right) p_x \tag{68}$$

Table II: Averaged RMS error over time for different measurement noise parameters for the UNGM

R	MTUKF (10)	MTUKF (3)	MTUKF (2)	UKF	CKF	MCKF	EKF	BGMF	SMMEF	PF
0.001	3.61	3.61	3.58	9.09	30.53	4.53	26.66	3.34	25.32	3.17
0.01	3.96	3.96	3.94	6.16	14.74	4.99	20.27	3.71	19.98	3.29
0.1	4.10	4.10	4.12	4.86	9.71	4.78	11.51	7.75	11.20	3.54
1	4.45	4.49	4.59	5.04	7.37	4.86	11.34	7.09	11.33	3.99
10	4.90	4.96	5.12	5.70	6.27	5.42	9.52	6.60	9.46	4.74
100	7.34	7.51	7.62	8.04	7.90	7.64	10.93	7.36	10.18	6.31

If $[p_x, p_y]^T$ belongs to the same quadrant as the one indicated by the measurement vector $\mathbf{z}^k$, i.e.,

$$\mathbf{z}^k = \begin{bmatrix} \operatorname{atan2}(p_y, p_x) \\ \operatorname{atan2}(p_y, p_x - m_{2,x}) \end{bmatrix} \tag{69}$$

then, (9) has $m = 1$ solution $\tilde{\mathbf{a}}_1(\mathbf{z}) = [p_x, p_y]^T$. Otherwise, $m = 0$, the MTUKF cannot be applied and the SMMEF is used instead $(\alpha = 0)$.

3) Algorithms' performances: In this example, the number of components in the update phase for the BGMF is very high because the state is 4-dimensional. Therefore, for each highly nonlinear component [5], $\left(2N^2+1\right)^{n_x} = 81$ new components are generated (recall that $N = 1$). It should be noted that Rao-Blackwellisation [10] can be applied such that the number of new components is $\left(2N^2+1\right)^{n_a}$ rather than $\left(2N^2+1\right)^{n_x}$ but this approach is not the one proposed in [5]. Besides, the computational complexity of the merging algorithm [11] is $O(N_c^3)$ where N_c is the number of components. Therefore, to make the simulations faster in this example for the BGMF, before running the merging algorithm, we use pruning such that the components whose weight are 100 times lower than the maximum weight are removed. After [11], a maximum number of three components is considered for the BGMF. Even with this simplification, the running time of the BGMF is around 100 times the MTUKF (with 10 mixture components) and this gets worse as the state dimension increases.

The RMS position error plotted against time for the algorithms is shown in Figure 4. SMMEF error is not shown as it is roughly the same as the EKF. The MTUKF, BGMF and PF outperform the rest although MTUKF requires a fraction of the computational time of the PF and BGMF. It should be noted that the RMS position error increases from time 60 to 80 for all the algorithms. At time 80, it drops for the MTUKF, BGMF and the PF. The reason why the RMS position error increases during this time interval is that the posterior is bimodal (with approximately equal weights) and the MMSE lies in a region where there is no target. The posterior is bimodal because at around time 60, the target is close to the axis $y = 0$ and range measurements cannot determine if a target is located in $y > 0$ or $y < 0$. In Figure 5, we plot the MTUKF approximations to the posterior PDFs at time steps 79 and 80 for an exemplar run. At time step 79, the posterior is bimodal and the MMSE estimator lies around $[-40, 0]$ while the target is around $[-40, 286]$ and therefore, the position error is around 286 m. At time step 80, the sensors take a bearing measurement, the mode located in $y < 0$ disappears and the error slumps. The same happens for the PF. On the contrary, as conventional KF-type algorithms cannot approximate a multimodal posterior properly, this situation does not happen and, when the bearing measurements are taken, the error does not decrease so much.

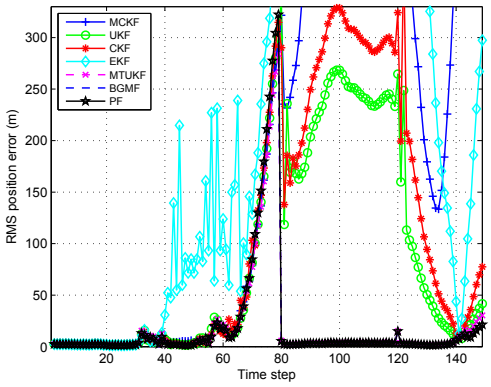
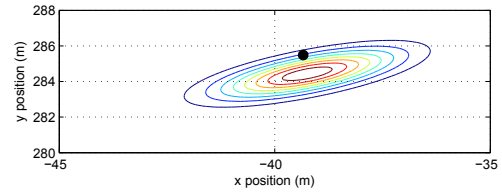


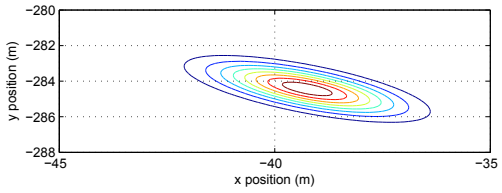
Figure 4: RMS position error plotted against time for the tracking scenario. MTUKF, BGMF and PF outperform the rest.

The averaged RMS position error over time for different measurement noise parameters are shown in Table III. Because conventional KF-type algorithms cannot approximate multimodal PDFs, their performance is very low compared to the MTUKF. Also, note that PF has a lower performance than the MTUKF when the measurement noise is low enough. The main reason is that PFs that sample from the prior do not work well for informative measurements [14]. This implies that a PF with measurement-directed sampling may be necessary in these cases. Nevertheless, since they usually use the EKF or UKF [2], they will not do well if the process noise variance is large, as it is in this scenario. Therefore, we can always find a low enough R_r such that the performance of PF is low compared to the MTUKF as the MTUKF can handle informative measurements properly. In addition, it should be taken into account that PF cannot keep multimodality for long times [15] and, therefore, we can always increase N_b such that any "conventional" PF with a given number of particles (even PFs based on the so-called optimal importance density [16]) will have a low performance compared to the MTUKF, which is able to keep multimodality forever. Only PFs that incorporate "mirror" particles adapted to the problem could be used to keep multimodality forever [15]. PF performance is low for $R_r = 0.1\ \mathrm{m}^2$. If we increase the number of particles to 10^5, the RMS error (52.91 m) is roughly equivalent to MTUKF's. However, for $R_r = 0.01\ \mathrm{m}^2$, even with 10^5 particles, PF's RMS error (295.85 m) is much higher than MTUKF's. BGMF performs well for most values of R_r but not so well for $R_r = 0.1\ \mathrm{m}^2$. It should be recalled that its computational burden is much higher than the MTUKF.

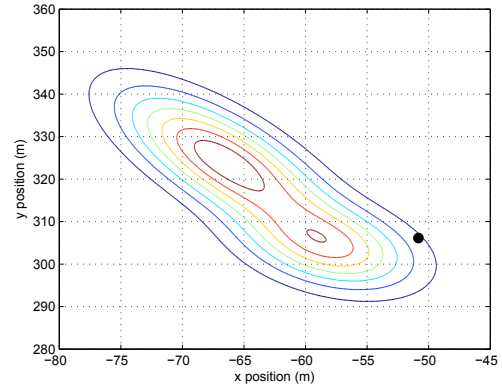
Table III: Averaged RMS position error (m) over time for different measurement noise parameters

R_r (m^2)	MTUKF (10)	MTUKF (3)	MTUKF (2)	UKF	CKF	MCKF	EKF	BGMF	SMMEF	PF
0.001	16.79	10.76	22.43	$1.44 \cdot 10^4$	$2.96 \cdot 10^3$	$1.38 \cdot 10^3$	$2.25 \cdot 10^4$	14.79	$1.78 \cdot 10^4$	498.11
0.01	25.14	25.73	33.40	$3.67 \cdot 10^3$	$1.48 \cdot 10^3$	665.94	$5.72 \cdot 10^3$	23.20	$4.74 \cdot 10^3$	426.76
0.1	47.52	49.12	53.39	684.10	334.29	374.90	700.61	62.61	845.17	445.43
1	50.72	50.98	53.35	137.48	165.00	317.96	117.83	52.59	151.32	57.02
10	52.68	55.02	60.62	220.93	132.30	340.04	82.49	52.03	86.16	51.50
100	58.63	63.45	65.78	495.13	346.95	324.87	94.80	56.47	111.06	55.80





(a)



(b)

Figure 5: Contour plots of the posterior PDF approximation using the MTUKF: (a) Time step $k = 79$ (b) Time step $k = 80$. The black circle represents the true target position at time k. In (a), the posterior is bimodal. At time 80, the sensors measure the target's bearing and the posterior becomes does not have a mode in region $y < 0$.

VI. Conclusions

We have proposed the MTUKF as a generalisation of the TUKF when the likelihood has a support made up of several regions. This algorithm has a much higher performance than conventional KF-type algorithms for two reasons. It can handle multimodal PDFs and approximates the posterior properly when the measurement is informative. In addition, it can keep multimodality forever so it can beat PFs in problems where multimodal posteriors stand for long periods of time. Importantly, all these improvements have been achieved with low computational burden.

VII. Acknowledgements

Ángel F. García-Fernández is supported by an FPU fellowship from Spanish MEC. This work was supported in part by the Spanish national research and development program under projects TEC2011-28683-C02-01 and Comonsens (Consolider-Ingenio 2010, CSD2008-00010).

References

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

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

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

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

[5] S. S. Ali-Löytty, "Box Gaussian mixture filter," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2165–2169, Sept. 2010.

[6] A. F. García-Fernández, "Detection and tracking of multiple targets using wireless sensor networks," Ph.D. dissertation, Universidad Politécnica de Madrid, 2011. [Online]. Available: http://oa.upm.es/9823/

[7] A. F. García-Fernández, M. R. Morelande, and J. Grajal, "Nonlinear filtering update phase via the single point truncated unscented Kalman filter," in *14th International Conference on Information Fusion*, 2011, pp. 17–24.

[8] ——, "Truncated unscented Kalman filtering," *to be published in IEEE Transactions on Signal Processing*, 2012.

[9] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.

[10] M. R. Morelande and B. Moran, "An unscented transformation for conditionally linear models," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, April 2007, pp. 1417–1420.

[11] D. Salmond, "Mixture reduction algorithms for point and extended object tracking in clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 2, pp. 667–686, April 2009.

[12] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, June 2009.

[13] M. 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, Feb. 2002.

[14] F. Gustafsson, S. Saha, and U. Orguner, "The benefits of down-sampling in the particle filter," in *14th International Conference on Information Fusion*, 2011, pp. 277–282.

[15] A. F. García-Fernández, M. R. Morelande, and J. Grajal, "Particle filter for extracting target label information when targets move in close proximity," in *14th International Conference on Information Fusion*, 2011, pp. 795–802.

[16] M. R. Morelande and A. Zhang, "Uniform sampling for multiple target tracking," in *14th International Conference on Information Fusion*, 2011, pp. 283–289.