Truncated Unscented Kalman Filtering

Ángel F. García-Fernández, Mark R. Morelande, Jesús Grajal

***Abstract*—We devise a filtering algorithm to approximate the first two moments of the posterior probability density function (PDF). The novelties of the algorithm are in the update step. If the likelihood has a bounded support, we can use a modified prior distribution that meets Bayes' rule exactly. Applying a Kalman filter (KF) to the modified prior distribution, referred to as truncated Kalman filter (TKF), can vastly improve the performance of the conventional Kalman filter, particularly when the measurements are informative relative to the prior. The application of the TKF to practical problems in which the measurement noise PDF has unbounded support is achieved by imposing several approximating assumptions which are valid only when the measurements are informative. This implies that we adaptively choose between an approximation to the KF or the TKF according to the information provided by the measurement. The resulting algorithm based on the unscented transformation is referred to as truncated unscented KF.**

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

I. Introduction

Filtering refers to the estimation of the state of a process based on indirect measurements obtained of that process over time. The minimum mean square error (MMSE) estimator of the state requires knowledge of the posterior probability density function (PDF), i.e., the PDF of the state given the measurements [1]. For most problems of interest the posterior PDF is intractable so approximations are required. The posterior PDF is often approximated in a recursion which involves two steps: prediction and update [2]. In the prediction step, the PDF of the state at the current time given all previous measurements is approximated. This PDF is referred to as the prior PDF. The probabilistic relationship between the measurement and the state is encapsulated by the likelihood. The update step involves using the likelihood along with Bayes' rule and the prior PDF to approximate the posterior PDF.

We focus on the update step because this is where the main difficulties lie in nonlinear filtering approximation. It has been shown in [3] that the prediction of an approximation to the posterior does not increase the error in the resulting prior approximation. No such guarantee exists for the update step. In fact, in many cases the update step can potentially result in a significantly higher error. An important example in which the update can induce large errors is when the measurements are precise compared to the prior [3].

A large variety of nonlinear filtering approximations have been proposed. The following discussion is not intended to be an exhaustive account but rather concentrates on the most influential techniques. In recent times, the increased availability of computing power has led to a lot of interest in sequential Monte Carlo methods, or particle filters (PFs) [4], [5]. PFs obtain a weighted sample from the posterior by drawing from an importance density. An important virtue of PFs is that they provide an asymptotically exact approximation of the posterior PDF as the sample size tends to infinity [6]. However, the performance of a PF for a finite sample size may be poor. Similar comments apply to point-mass filters based on deterministic grid approximations [7], [8]. For this reason there is still considerable interest in computationally efficient Gaussian approximations.

Once a Gaussian approximation to a non-Gaussian posterior PDF is adopted, the possibility of optimal inference disappears. Even so, there are many problems for which a Gaussian approximation can be of sufficient fidelity to provide estimates of similar accuracy to PFs at a fraction of the computational expense. The mean and covariance matrix of the Gaussian approximation are often obtained by approximating the Kalman filter (KF) recursion, although other methods have been suggested [9], [10]. The most well-known of the KF approximations is the extended KF (EKF) which uses a linearised approximation to the measurement function [1]. More accurate approximation of the moments required for the KF recursion can be obtained using numerical integration. This is the approach adopted by the unscented KF (UKF) [11], [12], cubature KF (CKF) [13], [14], linear regression KF (LRKF) [15] and the Gauss-Hermite filter [16] among others [17], [18]. Some interesting relations among some of these KF-type algorithms are given in [19].

The aforementioned methods suppose that the KF provides a good approximation to the posterior and it is therefore desirable to approximate the KF equations as closely as possible. However, in certain conditions, e.g., if the measurements are sufficiently precise, the KF approximates the posterior poorly [20], [21]. When this happens the performance of a KF approximation cannot be greatly improved by simply using a more accurate approximation method, as in the UKF compared to the EKF. Instead, the approach used to find an accurate but computationally efficient Gaussian approximation to the posterior must be reconsidered.

Ángel F. García-Fernández and Jesús Grajal are with the Grupo de Microondas y Radar, Departamento de Señales, Sistemas y Radiocomunicaciones, ETSI de Telecomunicación, Universidad Politécnica Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain. email: agarcia@gmr.ssr.upm.es, jesus@gmr.ssr.upm.es. Mark R. Morelande is with the Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, Victoria 3010, Australia. email: mrmore@unimelb.edu.au.

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 TEC2008-02148, TEC2011-28683-C02-01 and Comonsens (Consolider-Ingenio 2010, CSD2008-00010). Part of this work was done when Ángel F. García-Fernández was a visiting researcher at The University of Melbourne in 2009 and 2010.

The key contributions of this paper are as follows. First we propose the notion that the performance of the KF can be improved by applying the KF update to a modified prior. This is formalised in the truncated Kalman filter (TKF) which applies to likelihoods with a bounded, connected support [22]. The TKF involves applying the usual KF update to a prior truncated by the likelihood. A theoretical analysis for the case of a cubic measurement is presented to demonstrate the potential of this idea. There are two barriers to an exact implementation of the TKF. First, although it can be argued that all real likelihoods are of bounded support since no real sensor can provide an infinitely large measurement, most models assume a likelihood with unbounded support. Second, the moments required in the TKF recursion are generally intractable. The second contribution is the development of a practical algorithm based on the TKF which addresses these issues. The resulting algorithm is of similar computational expense to other Gaussian approximations but is demonstrated to have much better performance in a number of examples.

The rest of the paper is organised as follows. In Section II, we review the update stage of the KF for nonlinear measurements. In Section III, we introduce the TKF. In Section IV, we address the TKF approximation for unbounded noise support and the TUKF. Numerical examples examining the filter performance are provided in Section V. Finally, conclusions are drawn in Section VI.

II. Kalman filter for nonlinear measurement models

In this section, we analyse the update phase of a Kalman filter for nonlinear measurement models. We assume the following measurement equation:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \boldsymbol{\eta} \tag{1}$$

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 a zero mean measurement noise with any PDF. One should note that we assume the measurement noise is additive for the clarity of presentation although this is not the general case [12]. We also denote by $p_0(\cdot)$ the prior distribution of $\mathbf{x}$. The MMSE estimator $\hat{\mathbf{x}}_{pos}$ of $\mathbf{x}$ given $\mathbf{z}$ is [1]:

$$\hat{\mathbf{x}}_{pos} = \mathsf{E}[\mathbf{x}|\mathbf{z}] \tag{2}$$

whose mean square error matrix is

$$\mathbf{P}_{pos} = \mathsf{E}\left[(\mathbf{x} - \hat{\mathbf{x}}_{pos})(\mathbf{x} - \hat{\mathbf{x}}_{pos})^T |\mathbf{z}\right] \tag{3}$$

which depends on the measurement and its trace is the MMSE given that specific measurement. Calculating (2) and (3) is our ultimate objective as they are the first two moments of the posterior. Yet, they are very difficult to calculate as we need to know the posterior.

Conversely, the updated mean of the KF provides the linear MMSE estimator of $\mathbf{x}$ given $\mathbf{z}$, $\hat{\mathbf{x}}_{u,0}$ [23]:

$$\hat{\mathbf{x}}_{u,0} = \hat{\mathbf{x}}_{p,0} + \boldsymbol{\Psi}_0 \mathbf{S}_0^{-1}(\mathbf{z} - \hat{\mathbf{z}}_0) \tag{4}$$

where $\hat{\mathbf{x}}_{p,0}$ is the mean of the prior distribution of $\mathbf{x}$ and

$$\hat{\mathbf{z}}_0 = \mathsf{E}[\mathbf{z}] = \int \mathsf{E}[\mathbf{z}|\mathbf{x}]\, p_0(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{5}$$

$$\mathbf{S}_0 = \mathsf{cov}[\mathbf{z}] = \int \mathsf{E}\left[(\mathbf{z} - \hat{\mathbf{z}}_0)(\mathbf{z} - \hat{\mathbf{z}}_0)^T |\mathbf{x}\right] p_0(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{6}$$

$$\boldsymbol{\Psi}_0 = \mathsf{cov}[\mathbf{x}, \mathbf{z}] = \int \mathsf{E}\left[(\mathbf{x} - \hat{\mathbf{x}}_{p,0})(\mathbf{z} - \hat{\mathbf{z}}_0)^T |\mathbf{x}\right] p_0(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{7}$$

The updated covariance matrix given by the KF, $\mathbf{P}_{u,0}$, is

$$\mathbf{P}_{u,0} = \mathbf{P}_{p,0} - \boldsymbol{\Psi}_0 \mathbf{S}_0^{-1} \boldsymbol{\Psi}_0^T \tag{8}$$

where $\mathbf{P}_{p,0}$ is the covariance of the prior distribution of $\mathbf{x}$. As shown in Chapter 12 in [23], $\mathbf{P}_{u,0}$ is the mean square error matrix averaged over all possible measurements even in a nonlinear set-up:

$$\mathbf{P}_{u,0} = \mathsf{E}\left[(\mathbf{x} - \hat{\mathbf{x}}_{u,0})(\mathbf{x} - \hat{\mathbf{x}}_{u,0})^T\right] \tag{9}$$

In filtering problems, $\hat{\mathbf{x}}_{u,0}$ and $\mathbf{P}_{u,0}$ are usually regarded as approximations of the first two moments of the posterior, $\hat{\mathbf{x}}_{pos}$ and $\mathbf{P}_{pos}$, as if one assumes that the variable $(\mathbf{x}, \mathbf{z})$ is jointly Gaussian, the KF update is equivalent to Bayes' rule [16]. However, when the measurement function is nonlinear, there is no guarantee that $\hat{\mathbf{x}}_{u,0}$ and $\mathbf{P}_{u,0}$ are close to $\hat{\mathbf{x}}_{pos}$ and $\mathbf{P}_{pos}$. In fact, using a second order Taylor approximation, it can be shown that measurement nonlinearities become more significant for smaller measurement noise variance and larger prior variance. This applies not only to the EKF [20], [24] but also more generally to the KF [21]. Then, in such conditions, all the algorithms that are based on Kalman filtering, such as the UKF, CKF, EKF or Monte Carlo KF (MCKF) are expected to approximate the posterior poorly. By MCKF we denote a KF in which the integrals (5), (6) and (7) are calculated via Monte Carlo simulation using samples from the prior.

A. Illustration of the KF's deficiencies

We think it is very important to illustrate the problems of Kalman filtering with nonlinear measurements models to understand the paper. Then, let us consider a range-bearing measurement model [5]:

$$z_r = \sqrt{x^2 + y^2} + \eta_r \tag{10}$$

$$z_\theta = \arctan\left(\frac{y}{x}\right) + \eta_\theta \tag{11}$$

where $\mathbf{z} = [z_r, z_\theta]^T$, $\mathbf{x} = [x, y]^T$ represents the position of a target, η_r is the measurement noise for the range with variance σ_r^2, η_θ is the measurement noise for the bearing with variance σ_θ^2 and these noises are zero-mean Gaussian distributed and independent. In this case, we use circular sums and subtractions when performing these operations on the angle measurement, z_θ [25]. The objective is to approximate the first two moments of the posterior as they correspond to the MMSE estimator and the covariance matrix of the estimation error, equations (2) and (3). We analyse the case when $p_0(\cdot)$ is Gaussian with mean $\hat{\mathbf{x}}_{p,0} = [5, 5]^T$ and covariance matrix $\mathbf{P}_{p,0} = 36\mathbf{I}_2$ where $\mathbf{I}_m$ is the $m \times m$ identity matrix. Besides, the noise parameters are $\sigma_r = 1\,\mathrm{m}$ and $\sigma_\theta = 2\pi/180\,\mathrm{rad}$. For the UKF, we use $N_s = 2n_x + 1$ sigma points and the weight of the sigma point located on the mean of $p_0(\cdot)$ is 1/3. For the MCKF, we use $2 \cdot 10^4$ samples of the prior to approximate (5), (6) and (7).

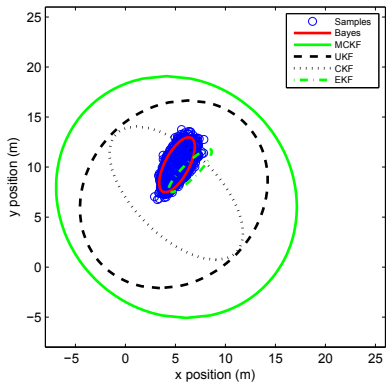


Figure 1: Samples from the posterior PDF and 3σ-ellipses for the true posterior and conventional KF-type algorithms. All conventional KF-type approximations are rather inaccurate.

In Fig. 1, we plot $2 \cdot 10^4$ samples of the posterior PDF, which are obtained using Bayes' rule. This is the PDF we want to approximate after receiving a certain measurement. The 3σ-ellipse of the posterior is shown in red colour. We use this ellipse as it represents the first two moments of the posterior [1]. For linear/Gaussian models, the KF's ellipse coincides with the ellipse of the posterior. However, in this nonlinear example, the ellipses given by the EKF, UKF, CKF or MCKF are rather inaccurate compared to the ellipse of the posterior. The results illustrate an important concept: improved approximation of the KF does not necessarily lead to improved approximation of the posterior. The MCKF with 20000 samples is by far the most accurate approximation of the KF. However it is arguably the worst approximation to the posterior. This suggests that any efforts to improve the accuracy of the KF moment approximations, as in the UKF and CKF, will not necessarily improve overall performance especially when nonlinearities are significant. In the rest of the paper, we develop an algorithm to approximate the first two moments of the posterior based on approximating the measurement noise PDF as one with bounded support.

III. Truncated Kalman Filter with Bounded Noise Support

In this section, we present the ideas behind the truncated Kalman filter (TKF). In subsection III-A, we justify applying the KF to a modified prior PDF instead of the real prior PDF so as to improve the performance of the estimation when some conditions are met. In subsection III-B, we provide an example in which the advantages of using the TKF over the KF are demonstrated.

A. Theory

We write the state vector as $\mathbf{x} = \left[\mathbf{a}^T, \mathbf{b}^T\right]^T$, where $\mathbf{a} \in \mathbb{R}^{n_a}$, $\mathbf{b} \in \mathbb{R}^{n_b}$ and $n_x = n_a + n_b$, such that the measurement equation is

$$\mathbf{z} = \mathbf{h}(\mathbf{a}) + \boldsymbol{\eta} \tag{12}$$

where $\mathbf{h}(\cdot)$ is a function of some elements of the state. The TKF is derived under the assumptions:

A1 The measurement function $\mathbf{h}(\cdot)$ is a continuous, bijective function.

A2 The PDF of the additive noise has a bounded, connected support:

$$p_{\boldsymbol{\eta}}(\boldsymbol{\eta}) = 0 \quad \text{if} \quad \boldsymbol{\eta} \notin I_{\boldsymbol{\eta}} \subset \mathbb{R}^{n_z} \tag{13}$$

where $I_{\boldsymbol{\eta}}$ is an n_z-dimensional connected region.

Due to Assumption A2, the PDF of the measurement conditioned on the state can be written as

$$p(\mathbf{z}|\mathbf{x}) = p(\mathbf{z}|\mathbf{a}) = p_{\boldsymbol{\eta}}(\mathbf{z} - \mathbf{h}(\mathbf{a}))\, \chi_{I_{\boldsymbol{\eta}}}(\mathbf{z} - \mathbf{h}(\mathbf{a})) \tag{14}$$

where $\chi_{I_{\boldsymbol{\eta}}}(\cdot)$ is the indicator function on the subset $I_{\boldsymbol{\eta}}$.

Due to Assumption A1, (14) can be written as:

$$p(\mathbf{z}|\mathbf{x}) = p_{\boldsymbol{\eta}}(\mathbf{z} - \mathbf{h}(\mathbf{a}))\, \chi_{I_{\mathbf{x}}(\mathbf{z})}(\mathbf{x}) \tag{15}$$

where

$$I_{\mathbf{x}}(\mathbf{z}) = \left\{ \mathbf{x} \,\middle|\, \mathbf{x} = \left[\left(\mathbf{h}^{-1}(\mathbf{z} - \boldsymbol{\eta})\right)^T, \mathbf{b}^T\right]^T, \boldsymbol{\eta} \in I_{\boldsymbol{\eta}}, \mathbf{b} \in \mathbb{R}^{n_b} \right\} \tag{16}$$

$$= I_{\mathbf{a}}(\mathbf{z}) \times \mathbb{R}^{n_b} \tag{17}$$

The set $I_{\mathbf{x}}(\mathbf{z})$ depends on the current measurement $\mathbf{z}$, the support of the measurement noise $I_{\boldsymbol{\eta}}$ and the inverse function of $\mathbf{h}(\cdot)$. The posterior PDF of $\mathbf{x}$ applying Bayes' rule and (15) is:

$$p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x}) \cdot p_0(\mathbf{x}) = p_{\boldsymbol{\eta}}(\mathbf{z} - \mathbf{h}(\mathbf{a}))\, \chi_{I_{\mathbf{x}}(\mathbf{z})}(\mathbf{x})\, p_0(\mathbf{x}) \tag{18}$$

A usual approach to approximate the first two moments of the posterior is to apply a Kalman filter using $p_0(\cdot)$ as the prior (KF). However, (18) can be rewritten as

$$p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})\, p_1(\mathbf{x}; \mathbf{z}) \tag{19}$$

where

$$p_1(\mathbf{x}; \mathbf{z}) = \frac{1}{\varepsilon_1} p_0(\mathbf{x})\, \chi_{I_{\mathbf{x}}(\mathbf{z})}(\mathbf{x}) \tag{20}$$

and ε_1 is a normalising constant and $p_1(\cdot)$ is a modified "prior" PDF, which is parameterised by $\mathbf{z}$. We should note that $p_1(\cdot)$ is a truncated version of $p_0(\cdot)$ and the result of applying Bayes' rule, equations (18) and (19), is the same if we use $p_0(\cdot)$ or $p_1(\cdot)$. This is illustrated in Fig. 2.

Equation (19) suggests an alternative approach to KF using $p_1(\cdot)$ as the prior, what we call the truncated KF (TKF). It should be noted that this approach does not calculate the conventional KF moments, i.e., (5), (6) and (7) as these are calculated with respect to $p_0(\cdot)$. The motivation behind this approach is that measurement nonlinearities are less significant when the prior variance is low [20], [21], [24]. Then, the use of $p_1(\cdot)$ instead of $p_0(\cdot)$ aims to reduce the variance of the prior PDF [26] so that the effect of nonlinearities is decreased. Therefore, the KF applied to $p_1(\cdot)$ is expected to have higher performance than the KF applied to $p_0(\cdot)$, especially, for informative measurements.

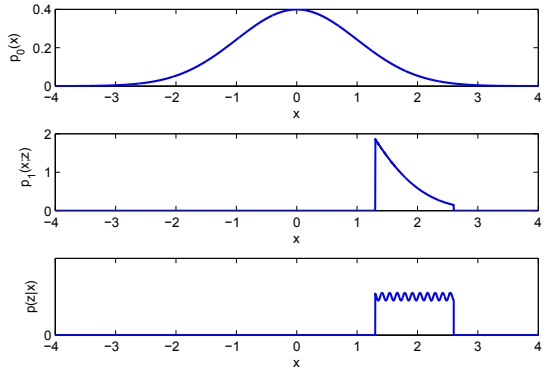


Figure 2: Graphical representation of $p_0(\mathbf{x})$, $p_1(\mathbf{x};\mathbf{z})$ and $p(\mathbf{z}|\mathbf{x})$ in one dimension. When the likelihood has a bounded support, Bayes' rule will exactly be met by the "prior" PDFs: $p_0(\mathbf{x})$ and $p_1(\mathbf{x};\mathbf{z})$.

B. Example

In this example we show the benefits of using $p_1(\cdot)$ rather than $p_0(\cdot)$ as the prior when the KF and TKF are tractable. That is, integrals (5), (6) and (7) can be calculated analytically for $p_0(\cdot)$ and $p_1(\cdot)$. Consider a random parameter x, whose prior is $p_0(x) = \mathcal{N}\left(x;\, \hat{x}_{p,0}, \sigma^2_{p,0}\right)$ (Gaussian PDF with mean $\hat{x}_{p,0}$ and variance $\sigma^2_{p,0}$ evaluated at x), which generates the measurement

$$z = ax^3 + \eta \tag{21}$$

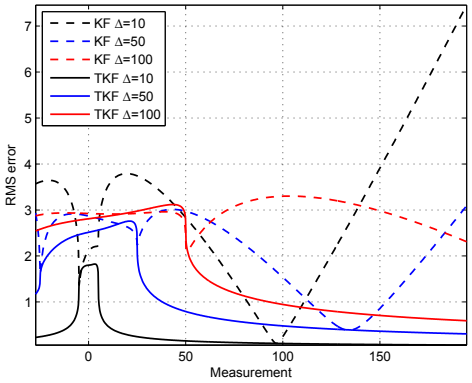
where $a \in \mathbb{R}$ is a known parameter and $\eta \sim U_{[-\Delta/2,\Delta/2]}$ (uniform PDF in the interval $[-\Delta/2, \Delta/2]$). We compare the RMS errors of estimating x for both methods noting that the information provided by the measurement about x is higher when z increases as the likelihood is narrower. Consequently, the KF is expected to provide high errors for high values of z.

In this case, $I_x(z) = \left[\sqrt[3]{\frac{z-\Delta/2}{a}},\ \sqrt[3]{\frac{z+\Delta/2}{a}}\right]$ and thus

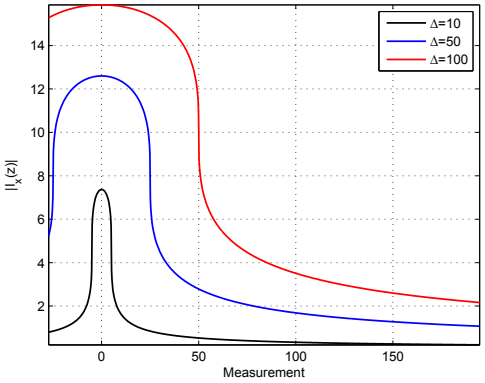
$$p_1(x;z) = \frac{1}{\varepsilon_1}\mathcal{N}\left(x;\, \hat{x}_{p,0}, \sigma^2_{p,0}\right)\chi_{I_x(z)}(x) \tag{22}$$

We can calculate integrals (5), (6) and (7) analytically for $p_0(\cdot)$ and $p_1(\cdot)$. The RMS errors of the KF and TKF estimators averaged over the posterior PDF for a particular measurement also admit a closed form expression. The analytical expressions can be found in [27].

We analyse the case where $a = 0.1$, $\hat{x}_{p,0} = 3$, $\sigma^2_{p,0} = 10$ and $\Delta \in \{10, 50, 100\}$. We plot the RMS error for an update step of the KF and TKF against the measurement values that occur for $x \in [\hat{x}_{p,0} - 3\sigma_{p,0},\ \hat{x}_{p,0} + 3\sigma_{p,0}]$ and the size of $I_x(z)$ in Fig. 3. When the measurement is lower than $\Delta/2$, i.e., the size of $I_x(z)$ is large, the KF and TKF perform similarly. When the measurement is higher than $\Delta/2$, $I_x(z)$ gets smaller and the RMS error of the KF increases considerably. There is one value of the measurement for which the KF's RMS error drops even for very informative measurements. This is because the estimate given by the KF is a linear function of the measurement, see (4), and for very informative measurements, the posterior mean is close to $x = \sqrt[3]{z/a}$ as the likelihood is narrow and concentrated around this point. Then, in our case, for $z > 0$, there is one point in which the KF estimate is close to the posterior mean for informative measurements producing a low RMS error but it increases rapidly.



(a)



(b)

Figure 3: (a) RMS errors of the KF and TKF and (b) size of $I_x(z)$ plotted against the measurement for different values of Δ. The TKF outperforms the KF especially when the size of $I_x(z)$ is small.

On the other hand, the TKF performs much better in general. In addition, the higher the measurement is or the lower Δ is, the smaller the size of $I_x(z)$ is, and the TKF improvement over the KF is more significant as the measurement is more informative. The reason for this is that the smaller the size of $I_x(z)$ is, the variance of $p_1(\cdot)$ gets smaller compared to the variance of $p_0(\cdot)$ [26] and, therefore, the KF applied to $p_1(\cdot)$ should work better.

IV. Truncated Kalman Filter with unbounded noise support

In this section, we apply the TKF ideas when the measurement noise support is unbounded. However, it should be noted that in reality the measurement noise support is always bounded as no noise can provide an infinitely large value. It is only the models which assume unbounded measurement noise support for convenience. Firstly, we reconsider the example of Section III-B and demonstrate that truncating the measurement noise PDF and applying the TKF can provide

a significant performance improvement over the KF with nonlinear measurement models. Note that this truncation puts more confidence into the measurement as the measurement noise assumed by the filter is actually lower than the real one. Secondly, we analyse the effect of some approximations that are made for general measurement models and introduce a variant of the TKF. Thirdly, we propose the truncated unscented Kalman filter (TUKF) as an approximation to the TKF for general problems.

A. Truncation of the measurement noise

One possible way of using the TKF's ideas when the measurement noise PDF support is unbounded is to truncate it and directly apply the method of Section III. The truncated noise PDF $p_{\eta_T}(\cdot)$ is calculated neglecting the areas where the original noise PDF $p_\eta(\cdot)$ is low according to a threshold Γ_η

$$p_{\eta_T}(\eta) = \frac{p_\eta(\eta)}{\varepsilon_\eta} \chi_{\frac{p_\eta(\eta)}{p_\eta(0)} > \Gamma_\eta}(\eta) \tag{23}$$

where ε_η is the normalising constant such that its integral is one.

Assuming the noise support is bounded when it is not and using the TKF provides a high performance gain in comparison with the KF for informative measurements. The implication is that it is a better approximation to Bayes' rule to truncate the noise and apply the TKF than approximating Bayes' rule using the KF. To demonstrate this, we reconsider the example of Section III-B for $\triangle = 100$ but using a zero-mean Gaussian noise with variance $\triangle^2/12$, i.e., a noise with the same mean and variance but with unbounded support. The RMS error is calculated via Monte Carlo simulation using 10000 samples from the prior and is shown in Fig. 4(a) for different values of Γ_η. Note that the main drop in the RMS error occurs for the measurements that make $I_x(z)$ decrease considerably as happened in Fig. 3. It is also important to notice that for low values of z (uninformative measurements), the KF and TKF have roughly the same performance. Therefore, Fig. 4(a) shows that it can be beneficial to truncate the noise PDF and apply the TKF if the measurement function is nonlinear. It should be noted that if the truncation threshold is too high, TKF performance decreases and does not outperform the KF for all measurement values. In this example, if $\Gamma_\eta < 0.4$, the TKF outperforms the KF.

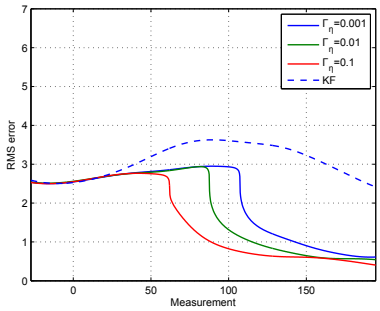
B. Practical approach: use of a mixture

In the example of Section IV-A, we set a truncation threshold Γ_η and applied the TKF steps: calculate $I_{\mathbf{x}}(\mathbf{z})$, the first two moments of $p_1(\cdot)$ and the KF moments for $p_1(\cdot)$. In a general scenario, neither of these steps is possible without several approximations. As explained in the next section, the approximations used to calculate $p_1(\cdot)$ are:

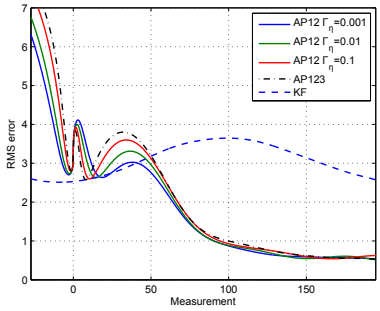
AP1 The measurement function $\mathbf{h}(\cdot)$ is locally linear.

AP2 The marginal prior of $\mathbf{a}$, $p_0(\mathbf{a})$, is constant over the region $I_{\mathbf{a}}(\mathbf{z})$.

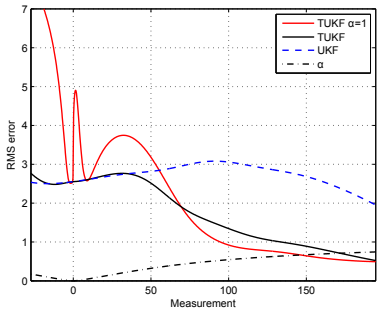
AP3 The measurement noise satisfies $\boldsymbol{\eta} \sim U_{I_{\boldsymbol{\eta}}}$ where $I_{\boldsymbol{\eta}}$ is such that the truncated noise has the same first two moments as the real noise $\mathsf{E}[\boldsymbol{\eta}] = 0$ and $\mathsf{cov}[\boldsymbol{\eta}] = \mathbf{R}$.



(a)



(b)



(c)

Figure 4: RMS error plotted against the measurement for the KF and the TKF for different values of Γ_η: (a) Without approximations (b) With different approximations (c) Proposed practical algorithm: TUKF with $\alpha = 1$, with adaptive α and UKF. The TUKF with adaptive α generally outperforms the UKF.

The implications of AP1 and AP2 are the following: for informative measurements, AP1 and AP2 are reasonable as the likelihood is narrow and the approximated TKF should work better than the KF but, for uninformative measurements, AP1 and AP2 are inaccurate as the likelihood is broad and the approximated TKF should work worse than the KF. This can be checked in Fig. 4(b) where the RMS error of the estimator under AP1 and AP2 to calculate $p_1(\cdot)$ is shown[1]. It can be seen that, as the measurement becomes more informative, i.e. as z increases, the RMSE for the TKF under AP1 and AP2 approaches the RMSE obtained by the TKF with truncated

[1]The linearisation of AP1 is done around the value that maximises the likelihood as explained in Section IV-C1. Once $p_1(\cdot)$ is approximated under AP1 and AP2, the TKF moments are calculated analytically.

measurement noise. However, for less informative measurements, the performance of the TKF deteriorates significantly compared to both the TKF with truncated measurement noise and the KF. These observations support the notion that AP1 and AP2 are reasonable for informative measurements but should not be applied for uninformative ones.

Considering only AP1 and AP2, we still need to set a truncation threshold, which becomes less important as the measurement is more informative. However, AP3 simplifies the algorithm as it approximates the measurement noise PDF by a uniform PDF and, therefore, the truncation threshold is not necessary. The exact form of $I_{\boldsymbol{\eta}}$ is not needed in the practical algorithm, see Section IV-C. The RMS error under AP1, AP2 and AP3 is also shown in Fig. 4(b). AP3 might seem restrictive but, as shown in Fig. 4(b), it does not have an important effect on performance compared to the RMS error under AP1 and AP2. This suggests that the error introduced by approximating the measurement noise PDF as uniform is low compared to the error introduced by approximations AP1 and AP2. It is clear from Fig. 4(b) that once AP1 and AP2 are applied, the TKF should only be used selectively, specifically when the measurements are informative. The way we propose to solve this problem is based on reconsidering Bayes' rule with bounded noise support, equation (18). It is shown in [22] that the prior

$$p_2(\mathbf{x};\mathbf{z}) = \alpha_0 p_1(\mathbf{x};\mathbf{z}) + (1-\alpha_0)\, p_0(\mathbf{x}) \tag{24}$$

where $\alpha_0 \in [0,1]$ also meets Bayes' rule exactly for bounded noise support. The parameter α_0 introduces a degree of freedom that allows several techniques to approximate the first two moments of the posterior. The KF uses $\alpha_0 = 0$ and the TKF $\alpha_0 = 1$. However, (24) suggests another equally valid option: apply a KF update to $p_0(\cdot)$ (conventional KF), another KF update to $p_1(\cdot)$ (TKF) and merge both results to form the final approximation to the posterior. Then, denoting the mean and covariance matrix of $p_1(\cdot)$ as $\hat{\mathbf{x}}_{p,1}$ and $\mathbf{P}_{p,1}$, the updated mean and covariance matrix are (see Chapter 2 in [5]):

$$\hat{\mathbf{x}}_{u,2} = \alpha \hat{\mathbf{x}}_{u,1} + (1-\alpha)\, \hat{\mathbf{x}}_{u,0} \tag{25}$$

$$\begin{aligned}\mathbf{P}_{u,2} =& \alpha \left[\mathbf{P}_{u,1} + (\hat{\mathbf{x}}_{u,1} - \hat{\mathbf{x}}_{u,2})(\hat{\mathbf{x}}_{u,1} - \hat{\mathbf{x}}_{u,2})^T\right] \\ &+ (1-\alpha)\left[\mathbf{P}_{u,0} + (\hat{\mathbf{x}}_{u,0} - \hat{\mathbf{x}}_{u,2})(\hat{\mathbf{x}}_{u,0} - \hat{\mathbf{x}}_{u,2})^T\right]\end{aligned} \tag{26}$$

where $\alpha \in [0,1]$ maintains the degree of freedom of α_0. It is shown in [27] that $\alpha = \alpha_0 / [\alpha_0 + (1-\alpha_0)\varepsilon_1]$ where ε_1 comes from (20). As $\alpha_0 \in [0,1]$ and $\varepsilon_1 \in [0,1]$, $\alpha \in [0,1]$ is another parameter that we can choose freely and, therefore, we can forget α_0 and ε_1. The results of Fig. 4(b) suggest that the TKF should be favoured when the measurements are informative and the KF should be emphasised for uninformative measurements. This corresponds to choosing α close to one for informative measurements and nearly zero for uninformative measurements. In the following section we propose an algorithm for approximating (25) and (26) and adaptively selecting α.

C. Truncated unscented Kalman filter

A closed form solution of the TKF is rarely tractable because of the calculation of the region $I_{\mathbf{x}}(\mathbf{z})$ and integrals (5)-(7) for $p_0(\cdot)$ and $p_1(\cdot)$. Then, we need to resort to approximations: the first one to approximate the first two moments of $p_1(\cdot)$ and the second one to approximate the moments (5), (6) and (7) with respect to $p_0(\cdot)$ and $p_1(\cdot)$. These issues and the selection of α are addressed in this section.

1) Approximation of the first two moments of $p_1(\cdot)$: The calculation of $p_1(\cdot)$, which requires the calculation of $I_{\mathbf{a}}(\mathbf{z})$, does not have an analytical expression in general. In this subsection, we explain how to approximate the first two moments of $p_1(\cdot)$ as they are everything we need to know about $p_1(\cdot)$ in the practical implementation of the filter, as will be explained in Section IV-C2.

Firstly, we need to approximate $I_{\mathbf{a}}(\mathbf{z})$. To this end, we use AP1 such that the measurement function $\mathbf{h}(\cdot)$ is approximated using a first-order Taylor series around $\mathbf{a} = \tilde{\mathbf{a}}(\mathbf{z})$. We choose $\tilde{\mathbf{a}}(\mathbf{z}) = \arg\max_{\mathbf{a}} p(\mathbf{z}|\mathbf{a})$ because it is the most probable observable state according to the measurement and $p_1(\cdot)$ is located around it. When A1 holds, the measurement noise is additive and zero-mean Gaussian distributed, the value of $\mathbf{a}$ that maximises the likelihood is $\tilde{\mathbf{a}}(\mathbf{z}) = \mathbf{h}^{-1}(\mathbf{z})$. Therefore:

$$\mathbf{h}(\mathbf{a}) \approx \mathbf{h}(\tilde{\mathbf{a}}(\mathbf{z})) + \tilde{\mathbf{H}}(\mathbf{a} - \tilde{\mathbf{a}}(\mathbf{z})) \tag{27}$$

where

$$\tilde{\mathbf{H}} = \left[\nabla_{\mathbf{a}} \mathbf{h}^T(\mathbf{a})\right]^T\Big|_{\mathbf{a}=\tilde{\mathbf{a}}(\mathbf{z})} \tag{28}$$

is the Jacobian of $\mathbf{h}(\mathbf{a})$ evaluated at $\tilde{\mathbf{a}}(\mathbf{z})$. Approximating $\mathbf{h}^{-1}(\cdot)$ using (27), $\tilde{\mathbf{a}}(\mathbf{z}) = \mathbf{h}^{-1}(\mathbf{z})$ and substituting them into (16), we get:

$$I_{\mathbf{a}}(\mathbf{z}) = \left\{\mathbf{a} \,\middle|\, \mathbf{a} = \tilde{\mathbf{a}}(\mathbf{z}) - \tilde{\mathbf{H}}^{-1}\boldsymbol{\eta},\, \boldsymbol{\eta} \in I_{\boldsymbol{\eta}}\right\} \tag{29}$$

Then, (20) becomes

$$p_1(\mathbf{x};\mathbf{z}) = p_1(\mathbf{a},\mathbf{b};\mathbf{z}) = \frac{1}{\varepsilon_1}\chi_{I_{\mathbf{a}}(\mathbf{z})}(\mathbf{a})\, p_0(\mathbf{b}|\mathbf{a})\, p_0(\mathbf{a}) \tag{30}$$

Under Approximation AP2, (30) can be written as

$$p_1(\mathbf{a},\mathbf{b};\mathbf{z}) = \frac{1}{\varepsilon_2}\chi_{I_{\mathbf{a}}(\mathbf{z})}(\mathbf{a})\, p_0(\mathbf{b}|\mathbf{a}) \tag{31}$$

where ε_2 is a normalising constant. Integrating out variable $\mathbf{b}$ in (31):

$$p_1(\mathbf{a};\mathbf{z}) = \frac{1}{\varepsilon_2}\chi_{I_{\mathbf{a}}(\mathbf{z})}(\mathbf{a}) \tag{32}$$

Under AP3, Appendix A shows that the mean $\boldsymbol{\mu}_{a,1}$ and covariance matrix $\boldsymbol{\Sigma}_{a,1}$ of $p_1(\mathbf{a};\mathbf{z})$ are

$$\boldsymbol{\mu}_{a,1} = \int \mathbf{a} p_1(\mathbf{a};\mathbf{z})\, \mathrm{d}\mathbf{a} = \tilde{\mathbf{a}}(\mathbf{z}) \tag{33}$$

$$\boldsymbol{\Sigma}_{a,1} = \int (\mathbf{a} - \boldsymbol{\mu}_{a,1})(\mathbf{a} - \boldsymbol{\mu}_{a,1})^T p_1(\mathbf{a};\mathbf{z})\, \mathrm{d}\mathbf{a} = \tilde{\mathbf{H}}^{-1}\mathbf{R}\left(\tilde{\mathbf{H}}^{-1}\right)^T \tag{34}$$

Noting that $p_0(\mathbf{x}) = p_0(\mathbf{a},\mathbf{b})$, the mean $\hat{\mathbf{x}}_{p,0}$ and covariance matrix $\mathbf{P}_{p,0}$ of $p_0(\cdot)$ is partitioned as

$$\hat{\mathbf{x}}_{p,0} = \begin{bmatrix} \mathsf{E}[\mathbf{a}] \\ \mathsf{E}[\mathbf{b}] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{a,0} \\ \boldsymbol{\mu}_{b,0} \end{bmatrix} \tag{35}$$

$$\mathbf{P}_{p,0} = \begin{bmatrix} \mathsf{cov}\left[\mathbf{a}\right] & \mathsf{cov}\left[\mathbf{a},\,\mathbf{b}\right] \\ \mathsf{cov}\left[\mathbf{b},\,\mathbf{a}\right] & \mathsf{cov}\left[\mathbf{b}\right] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{a,0} & \boldsymbol{\Sigma}_{ab,0} \\ \boldsymbol{\Sigma}_{ab,0}^T & \boldsymbol{\Sigma}_{b,0} \end{bmatrix} \tag{36}$$

Lastly, as shown in Appendix B, the mean $\hat{\mathbf{x}}_{p,1}$ and covariance matrix $\mathbf{P}_{p,1}$ of $p_1\left(\mathbf{a},\,\mathbf{b};\mathbf{z}\right)$ are:

$$\hat{\mathbf{x}}_{p,1} = \begin{bmatrix} \boldsymbol{\mu}_{a,1} \\ \boldsymbol{\mu}_{b,1} \end{bmatrix} \tag{37}$$

$$\mathbf{P}_{p,1} = \begin{bmatrix} \boldsymbol{\Sigma}_{a,1} & \boldsymbol{\Sigma}_{ab,1} \\ \left(\boldsymbol{\Sigma}_{ab,1}\right)^T & \boldsymbol{\Sigma}_{b,1} \end{bmatrix} \tag{38}$$

where

$$\boldsymbol{\mu}_{b,1} = \boldsymbol{\mu}_{b,0} + \boldsymbol{\Sigma}_{ab,0}^T \boldsymbol{\Sigma}_{a,0}^{-1}\left(\boldsymbol{\mu}_{a,1} - \boldsymbol{\mu}_{a,0}\right) \tag{39}$$

$$\boldsymbol{\Sigma}_{ab,1} = \boldsymbol{\Sigma}_{a,1}\left(\boldsymbol{\Sigma}_{a,0}^{-1}\right)^T \boldsymbol{\Sigma}_{ab,0} \tag{40}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{b,1} =& \boldsymbol{\Gamma} - \left(\boldsymbol{\mu}_{b,1} - \boldsymbol{\mu}_{b,0}\right)\left(\boldsymbol{\mu}_{b,1} - \boldsymbol{\mu}_{b,0}\right)^T \\ &+ \boldsymbol{\Sigma}_{ab,0}^T \boldsymbol{\Sigma}_{a,0}^{-1}\left[\boldsymbol{\Sigma}_{a,1} + \left(\boldsymbol{\mu}_{a,1} - \boldsymbol{\mu}_{a,0}\right)\left(\boldsymbol{\mu}_{a,1} - \boldsymbol{\mu}_{a,0}\right)^T\right] \\ &\times \left(\boldsymbol{\Sigma}_{a,0}^{-1}\right)^T \boldsymbol{\Sigma}_{ab,0} \end{aligned} \tag{41}$$

and

$$\boldsymbol{\Gamma} = \boldsymbol{\Sigma}_{b,0} - \boldsymbol{\Sigma}_{ab,0}^T \boldsymbol{\Sigma}_{a,0}^{-1} \boldsymbol{\Sigma}_{ab,0} \tag{42}$$

2) Approximation of the prior moments: This subsection describes how to approximate the moments (5), (6) and (7) with respect to $p_0\left(\cdot\right)$ and $p_1\left(\cdot\right)$. These approximations are based on the unscented transformation (UT) [11], or more precisely, on the UT for conditionally linear models [28]. The PDFs $p_0\left(\cdot\right)$ and $p_1\left(\cdot\right)$ can be written as:

$$p_j\left(\mathbf{a},\,\mathbf{b}\right) = p_0\left(\mathbf{b}\left|\mathbf{a}\right.\right) p_j\left(\mathbf{a}\right) \tag{43}$$

where $j \in \{0, 1\}$ and we highlight that the conditional PDF of $\mathbf{b}$ given $\mathbf{a}$ is the same for both PDFs because the truncation is applied to $\mathbf{a}$.

The first integral to approximate is (5). Substituting (12) and (43) into (5), (5) can be written as:

$$\hat{\mathbf{z}}_j = \int \mathbf{h}\left(\mathbf{a}\right) p_0\left(\mathbf{b}\left|\mathbf{a}\right.\right) p_j\left(\mathbf{a}\right) \mathrm{d}\mathbf{a}\mathrm{d}\mathbf{b} \tag{44}$$

Integrating out $\mathbf{b}$:

$$\hat{\mathbf{z}}_j = \int \mathbf{h}\left(\mathbf{a}\right) p_j\left(\mathbf{a}\right) \mathrm{d}\mathbf{a} \tag{45}$$

The UT approximation to (45) proceeds by selecting N_s sigma points $\mathcal{A}_j^1, ..., \mathcal{A}_j^{N_s}$ along with weights $w^1, ..., w^{N_s}$. These sigma points and weights can be found using any of the methods discussed in [12]. In all the implementations, we use $N_s = 2n_a + 1$ sigma points and the weight of the sigma point located on the mean is 1/3. The transformed sigma-points are calculated as

$$\mathcal{Z}_j^i = \mathbf{h}\left(\mathcal{A}_j^i\right), \quad i = 1, ..., N_s \tag{46}$$

Then, the UT approximation to (45) is

$$\hat{\mathbf{z}}_j = \sum_{i=1}^{N_s} w^i \mathcal{Z}_j^i \tag{47}$$

The second integral to approximate is (6):

$$\mathbf{S}_j = \int \mathsf{E}\left[\left(\mathbf{z} - \hat{\mathbf{z}}_j\right)\left(\mathbf{z} - \hat{\mathbf{z}}_j\right)^T \left|\mathbf{x}\right.\right] p_0\left(\mathbf{b}\left|\mathbf{a}\right.\right) p_j\left(\mathbf{a}\right) \mathrm{d}\mathbf{a}\mathrm{d}\mathbf{b} \tag{48}$$

Substituting (12) into (48), we get

$$\mathbf{S}_j = \mathbf{R} + \int \left(\mathbf{h}\left(\mathbf{a}\right) - \hat{\mathbf{z}}_j\right)\left(\mathbf{h}\left(\mathbf{a}\right) - \hat{\mathbf{z}}_j\right)^T p_j\left(\mathbf{a}\right) \mathrm{d}\mathbf{a} \tag{49}$$

The UT approximation to (49) can be calculated using (46):

$$\mathbf{S}_j = \mathbf{R} + \sum_{i=1}^{N_s} w^i \left(\mathcal{Z}_j^i - \hat{\mathbf{z}}_j\right)\left(\mathcal{Z}_j^i - \hat{\mathbf{z}}_j\right)^T \tag{50}$$

The third integral we need to approximate is (7):

$$\begin{aligned} \boldsymbol{\Psi}_j &= \int \mathsf{E}\left[\left(\mathbf{x} - \hat{\mathbf{x}}_{p,j}\right)\left(\mathbf{z} - \hat{\mathbf{z}}_j\right)^T \left|\mathbf{x}\right.\right] p_0\left(\mathbf{b}\left|\mathbf{a}\right.\right) p_j\left(\mathbf{a}\right) \mathrm{d}\mathbf{a}\mathrm{d}\mathbf{b} \\ &= \int \left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} - \hat{\mathbf{x}}_{p,j}\right)\left(\mathbf{h}\left(\mathbf{a}\right) - \hat{\mathbf{z}}_j\right)^T p_0\left(\mathbf{b}\left|\mathbf{a}\right.\right) p_j\left(\mathbf{a}\right) \mathrm{d}\mathbf{a}\mathrm{d}\mathbf{b} \end{aligned} \tag{51}$$

Integrating out $\mathbf{b}$ and taking into account that $\boldsymbol{\upsilon}_b\left(\mathbf{a}\right)$ is given by (75):

$$\boldsymbol{\Psi}_j = \int \left(\begin{bmatrix} \mathbf{a} \\ \boldsymbol{\upsilon}_b\left(\mathbf{a}\right) \end{bmatrix} - \hat{\mathbf{x}}_{p,j}\right)\left(\mathbf{h}\left(\mathbf{a}\right) - \hat{\mathbf{z}}_j\right)^T p_j\left(\mathbf{a}\right) \mathrm{d}\mathbf{a} \tag{52}$$

Let $\mathcal{X}_j^i = \left[\left(\mathcal{A}_j^i\right)^T, \left(\boldsymbol{\upsilon}_b\left(\mathcal{A}_j^i\right)\right)^T\right]^T$ for $i = 1, ..., N_s$. Then, (52) can be approximated as [28]

$$\boldsymbol{\Psi}_j = \sum_{i=1}^{N_s} w^i \left(\mathcal{X}_j^i - \hat{\mathbf{x}}_{p,j}\right)\left(\mathcal{Z}_j^i - \hat{\mathbf{z}}_j\right)^T \tag{53}$$

3) Selection of α: Approximations AP1 and AP2 are accurate for informative measurements but inaccurate for uninformative measurements. Therefore, the degree of freedom given by α, which was obtained from Bayes' rule as indicated in Section IV-B, should be chosen such that it favours $p_1\left(\cdot\right)$ when its variance is small (this indicates that the measurement is informative) compared to the variance of $p_0\left(\cdot\right)$. Then, we propose to use the following rule that meets these requirements:

$$\alpha = \frac{\gamma \mathrm{tr}\left(\boldsymbol{\Sigma}_{a,0}\right)}{\gamma \mathrm{tr}\left(\boldsymbol{\Sigma}_{a,0}\right) + \left(1 - \gamma\right)\mathrm{tr}\left(\boldsymbol{\Sigma}_{a,1}\right)} \tag{54}$$

where $\gamma \in [0,\,1]$ is a parameter that controls the weights of the traces of the covariance matrices to select α and $\mathrm{tr}\left(\mathbf{A}\right)$ denotes the trace of matrix $\mathbf{A}$. When the information about $\mathbf{a}$ measured from the data is high compared to the prior [29], which means that $\mathrm{tr}\left(\boldsymbol{\Sigma}_{a,1}\right)$ is low compared to $\mathrm{tr}\left(\boldsymbol{\Sigma}_{a,0}\right)$, $\alpha \to 1$. On the contrary, when the information about $\mathbf{a}$ measured from the data is low, which means that $\mathrm{tr}\left(\boldsymbol{\Sigma}_{a,1}\right)$ is high compared to $\mathrm{tr}\left(\boldsymbol{\Sigma}_{a,0}\right)$, $\alpha \to 0$, and the TUKF boils down to a UKF. The steps of the TUKF are shown in Table I and the RMS error for the example using $\gamma = 0.1$ is given in Fig. 4(c). Comparing this figure with Figs. 4(a) and (b), it is clear that the TUKF with $\alpha = 1$ approximates the TKF under AP1, AP2 and AP3, the UKF is approximating the KF and our method is approximating the TKF exploiting the degree of freedom given by α outperforming the UKF. The key to the close correspondence between the performances of the TUKF

Table I: Truncated Unscented Kalman Filter update phase steps

- Apply a UKF to $p_0(\cdot)$ (characterised by $\mathbf{x}_{p,0}$ and $\mathbf{P}_{p,0}$)
  - Select sigma points $\mathcal{A}_0^1, ..., \mathcal{A}_0^{N_s}$.
  - Compute the transformed sigma points $\mathcal{Z}_0^1, ..., \mathcal{Z}_0^{N_s}$ using (46).
  - Approximate $\hat{\mathbf{z}}_0$, $\mathbf{S}_0$ and $\mathbf{\Psi}_0$ using (47), (50) and (53).
  - Calculate $\hat{\mathbf{x}}_{u,0}$ and $\mathbf{P}_{u,0}$ using (4) and (8).
- Apply a UKF to $p_1(\cdot)$ (characterised by $\mathbf{x}_{p,1}$ and $\mathbf{P}_{p,1}$)
  - Calculate $\mathbf{x}_{p,1}$ and $\mathbf{P}_{p,1}$ using (37) and (38).
  - Select sigma points $\mathcal{A}_1^1, ..., \mathcal{A}_1^{N_s}$.
  - Compute the transformed sigma points $\mathcal{Z}_1^1, ..., \mathcal{Z}_1^{N_s}$ using (46).
  - Approximate $\hat{\mathbf{z}}_1$, $\mathbf{S}_1$ and $\mathbf{\Psi}_1$ using (47), (50) and (53).
  - Calculate $\hat{\mathbf{x}}_{u,1}$ and $\mathbf{P}_{u,1}$ using $\hat{\mathbf{z}}_1$, $\mathbf{S}_1$ and $\mathbf{\Psi}_1$ in (4) and (8).
- Calculate α using (54).
- Approximate the first two moments of the posterior using (25) and (26).

and the TKF is that the TUKF seeks to approximate the TKF only in conditions where the TKF performs significantly better than the KF, i.e., when the measurements are informative. When the measurements are uninformative, the TUKF seeks to approximate the KF which performs similarly to the TKF.

4) Single Point TUKF (SP-TUKF): In [22], we presented a modification of the TUKF which adds only one extra sigma-point to the collection of sigma-points given by the UKF. It should be noted that the TUKF really improves the estimate given by conventional KF-type algorithms when the information of the measurement is very high compared to the information of the prior. In that case, the region $I_{\mathbf{a}}(\mathbf{z})$ is very small. Then, another version of the algorithm uses a single sigma-point to represent $p_1(\mathbf{a};\mathbf{z})$ chosen such that it matches the mean of $p_1(\mathbf{a};\mathbf{z})$, given by (33). In this case, we approximate the first two moments of $p_2(\mathbf{a};\mathbf{z}) = \alpha p_1(\mathbf{a};\mathbf{z}) + (1-\alpha)\, p_0(\mathbf{a})$, given by the sigma-point representation, and then apply only one KF update to $p_2(\mathbf{a};\mathbf{z})$. The problem with this scheme is that if $\alpha \to 1$, the filter becomes unstable as the covariance matrix of $p_2(\cdot)$ tends to be singular. The way to solve it is to use α_{sp} rather than α such that

$$\alpha_{sp} = \alpha_{max} \cdot \alpha \tag{55}$$

where $\alpha_{max} < 1$ is the maximum value α_{sp} can take and α is given by (54). The steps of this version of the algorithm are shown in [22]. The computational burden of this version of the algorithm is lower than the TUKF as we only perform one KF update. However, it requires an extra parameter α_{max} and it does not attain the performance of the TUKF.

D. Discussion on the TUKF

It should be kept in mind that, despite the similarity in nomenclature, the proposed method differs in a fundamental way from other Gaussian approximations such as the EKF, UKF and CKF. These filters all seek to approximate the KF recursion by an accurate approximation of (5)-(7). As we have demonstrated in Section II this is not always a desirable goal. Instead, the proposed method is based on the approximation of the TKF, which applies the KF recursion to a truncated prior. The impossibility of applying the TKF in general leads to a practical algorithm for adaptively switching between TKF and KF approximations. The use of the TKF approximation when warranted, i.e., when measurements are sufficiently precise, results in considerable performance improvement compared to KF approximations, as will be seen in Section V.

Its main limitation though is that it is not such a general tool as conventional KF-type algorithms, which can always be used, as the current form of the algorithm requires $\mathbf{h}(\cdot)$ to be a bijective function of some elements of the state. Nevertheless, if $\mathbf{h}(\cdot)$ is not bijective the posterior could be multimodal and none of the conventional KF-type algorithms would provide an accurate representation of the posterior, as they represent a unimodal distribution.

How to generalise the algorithm when $\mathbf{h}(\cdot)$ is not bijective is well worth exploring and will be a topic for future research. In this case, there are several disconnected regions of the state space that are likely to produce a given measurement. Then, we would need to truncate $p_0(\cdot)$ in each one of these regions, apply a KF to each truncated distribution and represent the posterior as a Gaussian mixture.

In this paper, we approximate $p_0(\cdot)$ and $p_1(\cdot)$ using the sigma-points given by the UKF. However, if we used the CKF or other methods, the foundations of the algorithm would be the same and it would only imply minor changes. We have chosen the UKF as it is the method that is used more widely.

It should also be mentioned that there are some works that deal with Kalman filtering when the prior is truncated or there are inequality constraints [30], [31]. However, their approach and motivation have nothing to do with ours. In [30], the posterior is truncated because of the inequality constraints of the model. They apply a conventional KF with a linear/Gaussian model and, then, they calculate the first two moments of the truncated distribution. In [30], they provide an example with nonlinear measurement function but they linearise the measurement function like an EKF. In [31], the prior is truncated but they use linear/Gaussian models. Then, these methods do not try to address the problems of KF with nonlinear measurement models, they apply a conventional KF when either the prior or the posterior is truncated.

V. Numerical examples

In this section, we present three numerical examples to show the benefits of using the TUKF. Other examples can be found in [22], [32]. The first example is the range-bearing scenario introduced in Section II. We provide a detailed analysis on how the algorithms approximate the posterior in the update phase without accounting for filtering. The second example deals with a tracking scenario with the same measurement model and, lastly, we analyse a univariate nonstationary growth model. These two examples are included to show the performance benefits of the TUKF in filtering scenarios. In all the simulations, we use a TUKF with $\gamma = 0.1$ and an SP-TUKF with $\alpha_{max} = 0.8$. Besides, the RMS errors are calculated with respect to the true state, they do not refer to the RMS errors estimated by the filters.

A. Range-bearing case

Firstly, we go back to the range-bearing scenario introduced in Section II to highlight the shortcomings of Kalman filtering with nonlinear measurement models. It should be recalled that

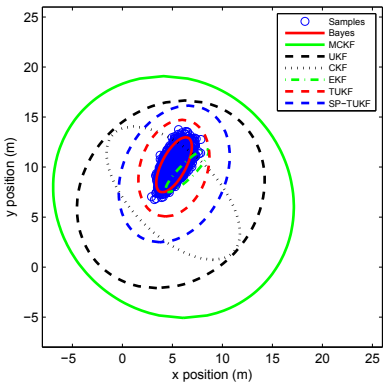


Figure 5: The TUKF and SP-TUKF approximate the first two moments of the posterior much better than conventional KF-type algorithms

the prior is perfectly known. For the sake of illustration, we redraw Fig. 1 in Fig. 5 but adding the ellipses corresponding to the TUKF and the SP-TUKF. It is clear that the TUKF approximation of the first two moments of the posterior is much more accurate than the rest of the algorithms.

Now we proceed to analyse this scenario more deeply plotting the Kullback-Leibler (KL) divergence assuming the statistics are Gaussian between the posterior PDF and the different algorithms. For the purposes of this comparison, an accurate approximation to the posterior is obtained by drawing $2 \cdot 10^4$ samples from the prior and weighting the samples according to the likelihood.

The KL divergence is obtained for several positions of the prior mean, $\hat{\mathbf{x}}_{p,0}$, and true locations of the target to analyse the influence of the prior mean and the true location of the target on performance. In order to plot the results, we define 9 points forming a square grid whose side is 5 meters around the prior mean each one labelled as in Fig. 6. We analyse three scenarios with a different prior means $\hat{\mathbf{x}}_{p,0}$: $[10,\, 10]^T$, $[20,\, 20]^T$ and $[30,\, 30]^T$. The rest of the parameters are those indicated in Section II.

In this case, $\tilde{\mathbf{a}}\left(\mathbf{z}\right) = \left[z_r \cos z_\theta,\, z_r \sin z_\theta\right]^T$. Also, the trace of the covariance matrix of $p_1\left(\cdot\right)$, which is needed to obtain α, can be calculated using (34):

$$\mathrm{tr}\left(\mathbf{\Sigma}_{a,1}\right) = \sigma_r^2 + \sigma_\theta^2 z_r^2 \tag{56}$$

Then, in this example, the amount of information we get from the measurement depends only on z_r^2. When z_r is low, i.e., the target is closer to the radar, conventional KF-type algorithms approximate the posterior worse than when the target is far afield.

The KL divergence for the true position of the target according to the labels in each scenario are shown in Figs. 7(a), (c) and (e), respectively, noting that the y-axis changes in the figures for the clarity of presentation. The TUKF and SP-TUKF have better performance than the MCKF, the UKF and the CKF in all the scenarios. EKF performance is quite variable, sometimes performs quite well while others its performance is the worst one. There is no doubt that this is not a desirable property of an algorithm. In the first scenario,

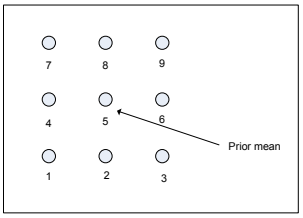


Figure 6: Labelling of the true target positions around the prior mean forming a square grid of side 5 m.

the MCKF has worse performance than the UKF and the CKF on the whole. Besides, the performance of these three filters is far from the performance attained by the TUKF and SP-TUKF as in this case the target is located close to the origin and the measurement carries a lot of information. In the second scenario, the TUKF works better than the rest of the algorithms in spite of the fact that the measurement carries less information than in the first scenario. Finally, in the third scenario, TUKF and SP-TUKF outperforms all the conventional KF-type algorithms but their differences have been reduced as the measurement does not carry so much information as before. In addition, the performance of TUKF is usually higher than the performance of SP-TUKF except for the third scenario.

It should also be pointed out that the MCKF is usually outperformed by the UKF and the CKF in these cases. When the number of samples of the MCKF tends to infinity, its updated mean and covariance matrix converge to the updated mean and covariance matrix of the KF (given by the integrals (5), (6) and (7)). However, in nonlinear systems a close approximation to the KF does not necessarily imply a close approximation to the posterior PDF. This is clearly evident in Fig. 5.

The RMSE for the true position of the target according to the labels in each scenario are shown in Figs.7 (b), (d) and (f), noting that the y-axis changes in the figures. The estimate to calculate the RMSE for Bayes' rule is obtained using the mean of the posterior PDF. As happened with the KL divergence, RMSE of the EKF is quite variable. In the first scenario, Figure 7(b), the RMSE for the TUKF and SP-TUKF are generally much lower than the RMSE of the rest of the filters as the information of the measurement is high and is quite close to the RMSE attained by Bayes' rule. In the second scenario, Figure 7(d), the RMSE for the TUKF and SP-TUKF are lower than the RMSE of the rest of the filters but the differences are closer as the measurement is less informative. Lastly, in scenario 3, Figure 7(f), where the measurement is much less informative, all the filters have roughly the same RMSE although the TUKF and SP-TUKF posteriors are closer to the real posterior as pointed out when analysing the KL divergence, Figure 7(e). It can be confusing that the posterior mean, which is given by Bayes' rule, is not always the best estimator for some values of the state. However, the posterior mean is the MMSE estimator conditioned on a value of the measurement not the true value of the state, see Section II, so its performance does not have to be the best one for a given value of the state. In fact, a blind estimator that estimates

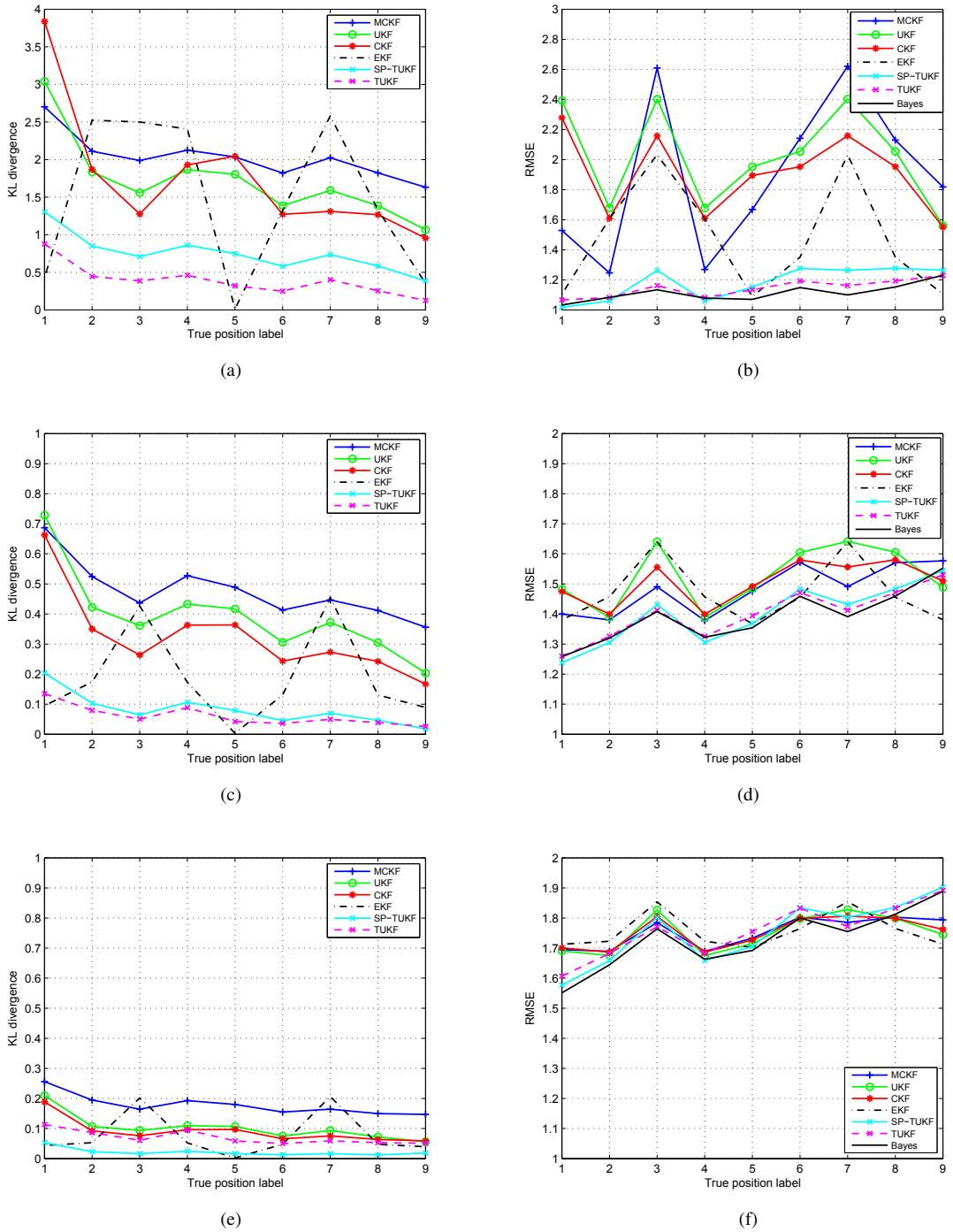


Figure 7: Range-bearing update: (a) KL divergence and (b) RMSE for scenario 1. (c) KL divergence and (d) RMSE for scenario 2. (e) KL divergence and (f) RMSE for scenario 3. The TUKF and SP-TUKF vastly outperform the conventional KF-type algorithms in scenario 1. In scenario 3, all the algorithms give roughly the same RMSE although the KL divergence is slightly lower for the TUKF and SP-TUKF as the measurement has much less information than in scenarios 1 and 2.

Table II: Averaged RMSE for range-bearing scenarios

	MCKF	UKF	CKF	EKF	SP-TUKF	TUKF	Bayes' rule
Scenario 1	3.85	3.89	3.91	4.46	1.48	1.25	1.14
Scenario 2	2.09	2.11	2.09	2.19	1.43	1.45	1.42
Scenario 3	2.01	2.02	2.01	2.05	1.78	1.79	1.77

always the same value regardless of the measurement will achieve zero error when the true value of the state coincides with this blind estimate although its performance is very poor in general. To clarify this, we show the averaged RMSE for all the algorithms in Table II. As expected, the lowest error is given by Bayes' rule followed by the TUKF and SP-TUKF and, in general, EKF error is the highest.

B. Target tracking with range-bearing measurements

In the previous example, we highlighted the improvement in performance of the TUKF over conventional KF-type algorithms only in the update equation. In this example, we show the performance benefits in a dynamic system. We analyse target tracking using range-bearing measurements from a radar located at the origin of the coordinate system. The state vector at time k is $\mathbf{x}^k = \left[\mathbf{a}^k, \mathbf{b}^k\right]^T$ where $\mathbf{a}^k$ is the position vector and $\mathbf{b}^k$ is the velocity vector. The measurement model is given by (10) and (11) but bearing in mind that the position vector in this example is $\mathbf{a}^k$.

The dynamic model of the target is the nearly-constant velocity model [1]:

$$p\left(\mathbf{x}^{k+1}\left|\mathbf{x}^k\right.\right) = \mathcal{N}\left(\mathbf{x}^{k+1};\, \mathbf{F}\cdot\mathbf{x}^k,\, \mathbf{Q}\right) \tag{57}$$

$$\mathbf{F} = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix} \otimes \mathbf{I}_2 \tag{58}$$

$$\mathbf{Q} = \sigma_u^2 \begin{pmatrix} \tau^3/3 & \tau^2/2 \\ \tau^2/2 & \tau \end{pmatrix} \otimes \mathbf{I}_2 \tag{59}$$

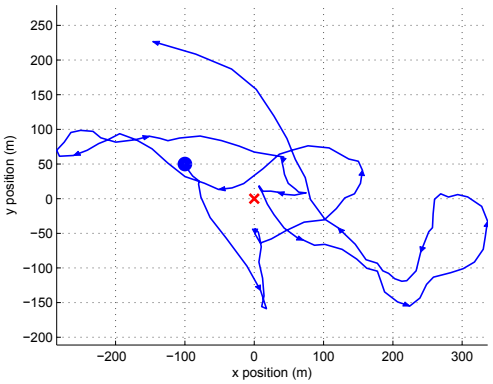
where $\mathcal{N}\left(\mathbf{x};\, \overline{\mathbf{x}},\, \mathbf{Q}\right)$ is the Gaussian PDF evaluated at $\mathbf{x}$ with mean $\overline{\mathbf{x}}$ and covariance matrix $\mathbf{Q}$, $\otimes$ is the Kronecker product, τ is the sampling period and σ_u^2 is the continuous-time process noise intensity [1].

The scenario we use to evaluate the performance of the algorithms is represented in Fig. 8(a). We also show the distance from the radar to the target at each time step in Fig. 8(b). This is important because conventional KF-type algorithms are expected to perform worse than the TUKF when the target gets close to the radar as explained in Section V-A. The sampling period of the trajectory is $\tau = 1\,\mathrm{s}$, $\sigma_u = 10\,\mathrm{m/s}^{3/2}$ and there are $l = 140$ time steps in the simulation. This trajectory corresponds to one realisation of the dynamic system described by (57). We evaluate the tracking performance calculating the RMS of the position error by a Monte Carlo simulation with $m = 200$ realisations.

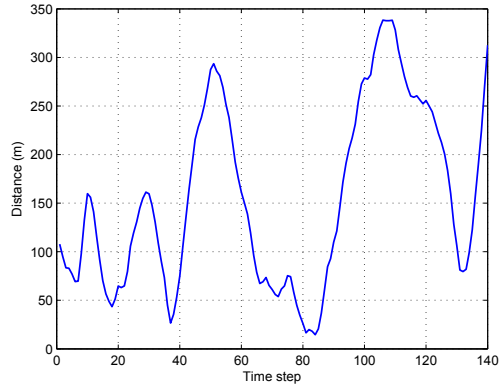
The prior PDF of the target's state is:

$$\mathbf{x}^0 \backsim \mathcal{N}\left(\mathbf{x}^0;\, \overline{\mathbf{x}}^0, \mathbf{Q}\right) \tag{60}$$

where, in each Monte Carlo run, $\overline{\mathbf{x}}^0$ is drawn from a Gaussian distribution whose mean is the real state of the target at time 0 and whose covariance matrix is $\mathbf{Q}$.



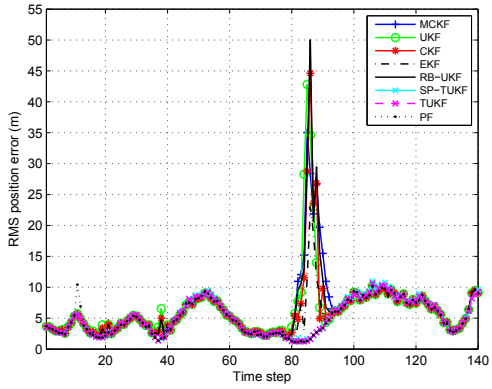
(a)



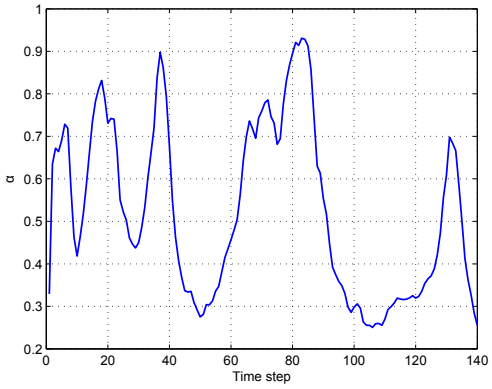
(b)

Figure 8: Scenario for range-bearing tracking: (a) Target trajectory: 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 radar location is represented by a red cross. (b) Distance from the target to the radar at each time step.

We are going to compare the proposed algorithms, TUKF and SP-TUKF, with six algorithms for tracking: UKF, CKF, EKF, MCKF ($2\cdot 10^4$ samples), Rao-Blackwellised UKF (RB-UKF) [28] and a PF based on sampling importance resampling with 10000 particles [2]. The parameters of the algorithms are the same as in the previous example. The RMS position plotted against time is shown in Fig. 9(a). As expected, EKF, UKF, CKF, MCKF and RB-UKF perform poorly when the target gets close to the radar at around time steps 40 and 80, see Fig. 8(b). This is due to the fact that the likelihood function is very narrow compared to the prior and conventional KF-type algorithms do not perform properly [20]. In a dynamic system, this can lead to filter divergence as happens at around time 80. In this case, conventional KF-type algorithms manage to recover the track around 10 time steps later. However, TUKF and SP-TUKF provide a much more accurate estimate of the trajectory as they can deal with situations when the information given by the measurement is very high. They do it by adaptively calculating the parameter α that allows us to identify when the measurement carries a lot of information and conventional KF-type algorithms fail. The value of α plotted



(a)



(b)

Figure 9: (a) RMS position error plotted against time. When the target is close to the radar (around time steps 40 and 80), SP-TUKF and TUKF vastly outperform conventional KF approximations. (b) Value of α plotted against time for the TUKF. When the target gets closer to the radar, α increases.

against time is shown in Fig. 9(b). When the distance from the radar to the target increases, α decreases, as the information provided by the measurement is lower and conventional KF-type algorithms perform well. As expected, the PF also attains a low error at all time steps but with a much higher computational expense.

We show the robustness of our algorithms with respect to γ in Table III. One should note that for the SP-TUKF, $\gamma = 1$ does not mean that we reject the information of the prior, as with the TUKF, because α_{sp} cannot be one, see (55). Recall that $\gamma \in [0, 1]$ is used in the calculation of the weight α given to the truncated prior. A low value of γ tends to reduce the weight given to the truncated prior while a high value of γ tends to place more emphasis on the truncated prior. The results of Table III show that the performances of the SP-TUKF and TUKF do not vary greatly for $\gamma > 0$ although slightly better performance is obtained for smaller values of γ. It should be noted that even with small values of γ the weight α given to the truncated prior can be large, provided the measurements are sufficiently precise. This can be seen in Figure 9(b) which shows that large values of α are selected even with $\gamma = 0.1$. Also, Table III shows that the performance of the TUKF with $\gamma = 0.1$ is very close to the PF performance.

We also include the averaged RMS position error over time for different measurement noise parameters in Table IV where we have underlined the cases in which the error is much higher than for the PF. We analyse cases in which the measurement is very informative and others in which it is very uninformative to show that our method is robust in all situations. Firstly, one should note that SP-TUKF and TUKF are very close to PF performance especially for informative measurements. They outperform the conventional KF-type algorithms in all cases, especially, when the measurement is informative. This demonstrates the power of both SP-TUKF and TUKF in nonlinear filtering problems. When the measurement is very uninformative, all filters behave similarly, cases: 2, 5 and 8, except for the EKF that gives higher errors. Moreover, among the conventional KF-type algorithms, MCKF has the most robust performance as it never diverges and EKF diverges more frequently than the rest.

C. Univariate nonstationary growth model (UNGM)

We analyse the univariate nonstationary growth model (UNGM) characterised by:

$$p\left(x^k \left| x^{k-1}\right.\right) = \mathcal{N}\left(x^k;\, f^k\left(x^{k-1}\right), Q\right) \tag{61}$$

$$f^k\left(x^{k-1}\right) = \frac{x^{k-1}}{2} + \frac{25x^{k-1}}{1+\left(x^{k-1}\right)^2} + 8\cos\left(1.2k\right) \tag{62}$$

$$p\left(z^k \left| x^k\right.\right) = \mathcal{N}\left(z^k;\, \frac{\left(x^k\right)^3}{20},\, R\right) \tag{63}$$

This model is analysed in [2], [9] but with a measurement function that depends on $\left(x^k\right)^3$ rather than $\left(x^k\right)^2$ so that it is bijective. We use the same parameters as in [9]. That is, we use $Q = 1$ and $R = 1$, data was generated using $x^0 = 0.1$ and the prior PDF at time step 0 is $x^0 \backsim \mathcal{N}\left(x^0;\, 0, 1\right)$.

The true state we use in the simulations with 100 time steps is plotted in Fig. 10. We analyse the performance of the algorithms using Monte Carlo simulation with 200 runs. We should realise that, in this case, the measurement function, given by (63), depends on the whole state so RB-UKF cannot be applied here. The parameters of the algorithms are the same as in the previous examples.

The RMS error plotted in log scale against time for the algorithms are shown in Fig. 11(a). We have represented the first twenty time steps for the clarity of representation. TUKF and SP-TUKF have a much higher performance than conventional KF-type algorithms and their performance is close to PF's. In this case, the error given by the EKF is the highest one. We also plot the value of α against time for the first twenty time steps for the TUKF in Fig. 11(b). Comparing Fig. 11(b) with Fig. 10, it can be seen that α is larger when $\left|x^k\right|$ increases as the measurement is more informative. We also show the averaged RMS error over time in Table V. Highest performance is provided by the PF closely followed by the TUKF. The EKF, MCKF, UKF and CKF are far from the performance of SP-TUKF and TUKF.

Table III: Averaged RMS position error over time for different values of γ (range-bearing)

Algorithm	γ 0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
SP-TUKF	8.48	5.85	5.93	5.98	6.01	6.03	6.05	6.07	6.09	6.10	6.12
TUKF	8.48	5.74	5.84	5.94	6.02	6.09	6.16	6.22	6.28	6.33	6.38
UKF						8.19					
CKF						8.02					
MCKF						8.04					
EKF						6.48					
RB-UKF						8.48					
PF						5.70					

Table IV: Averaged RMS position error over time for different measurement noise parameters

Case	σ_r(m)	σ_θ (rad)	SP-TUKF	TUKF	UKF	CKF	MCKF	EKF	RB-UKF	PF
1	1	$0.2\pi/180$	1.19	1.18	5.10	136.57	5.76	192.12	128.08	1.34
2	1	$20\pi/180$	38.29	37.62	38.33	37.98	38.28	50.15	37.93	33.21
3	$\sqrt{10}$	$0.2\pi/180$	3.19	3.18	161.11	222.31	7.35	99.47	251.3	3.26
4	$\sqrt{10}$	$2\pi/180$	6.55	6.45	11.26	9.86	9.36	7.35	10.21	6.34
5	$\sqrt{10}$	$20\pi/180$	38.58	37.82	38.71	38.24	38.51	44.51	38.35	33.31
6	$\sqrt{0.1}$	$0.2\pi/180$	0.71	0.70	4.70	3.19	5.27	128.31	2.78	0.96
7	$\sqrt{0.1}$	$2\pi/180$	5.77	5.66	7.27	7.93	8.24	5.92	6.54	5.86
8	$\sqrt{0.1}$	$20\pi/180$	38.25	37.61	38.31	37.86	38.24	51.32	37.91	34.22

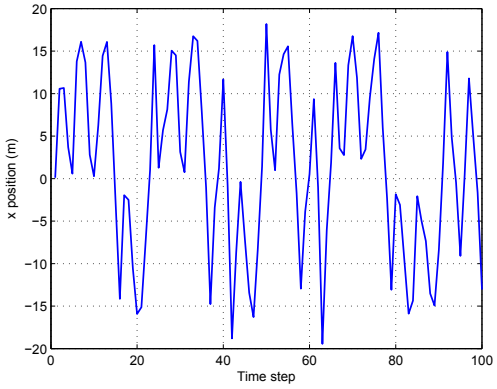


Figure 10: True state for the univariate nonstationary growth model

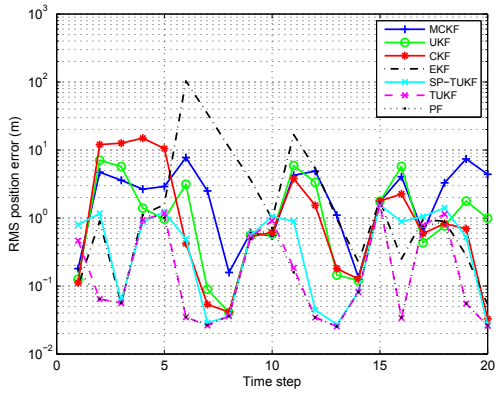
Table V: Averaged RMS error over time (UNGM)

Algorithm	SP-TUKF	TUKF	UKF	CKF	MCKF	EKF	PF
RMS	0.81	0.64	3.86	7.46	3.84	26.94	0.61

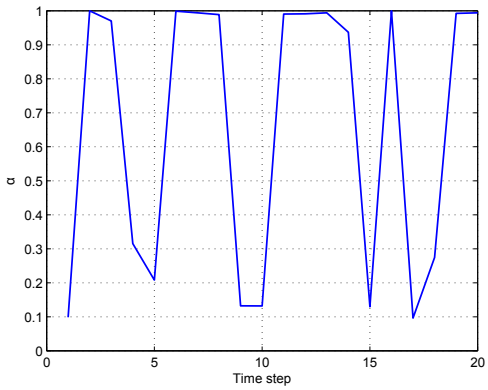
VI. Conclusions

We have developed a new approach to approximate the first two moments of the posterior in nonlinear systems with a bijective measurement function. We do it by approximating the PDF of the measurement noise by a truncated one and approximating Kalman filter equations applied to a mixture of the prior and a truncated version of the prior. On the whole, our approach should achieve better performance than any conventional Kalman-filter-type algorithms if the system sometimes provides very informative nonlinear measurements compared to the prior. As we have demonstrated in the paper, if we get a very informative measurement, conventional Kalman-filter-type algorithms do not work properly.

Its drawback is that it is not such a general tool as conventional Kalman-filter-type algorithms as it requires the measurement function to be bijective. However, our approach



(a)



(b)

Figure 11: UNGM scenario for the first 20 time steps: (a) RMS error plotted in log scale against time: The TUKF and SP-TUKF have a much higher performance compared to conventional KF-type algorithms. (b) Value of α for the TUKF: When $|x^k|$ increases, α increases as the measurement is more informative.

can be generalised when the measurement function is not bijective. In this case, there would be several disconnected regions of the state space that are likely to produce a given measurement. Then, we would need to truncate $p_0(\cdot)$ in each one of these regions, apply a Kalman filter to each truncated distribution and represent the posterior as a Gaussian mixture. How to select the weights of this mixture will be addressed in a future work. The study of the consistency of the proposed algorithm is also a topic for future research.

APPENDIX A

In this appendix, we calculate the mean $\boldsymbol{\mu}_{a,1}$ and covariance matrix $\boldsymbol{\Sigma}_{a,1}$ of $p_1(\mathbf{a};\mathbf{z})$, which is given by (32). Firstly, we calculate ε_2:

$$\varepsilon_2 = \int \chi_{I_{\mathbf{a}}(\mathbf{z})}(\mathbf{a})\,\mathrm{d}\mathbf{a} \tag{64}$$

Changing the variable inside the integral using

$$\mathbf{a} = \tilde{\mathbf{a}}(\mathbf{z}) - \tilde{\mathbf{H}}^{-1}\mathbf{u} \tag{65}$$

we get

$$\varepsilon_2 = \int \chi_{I_{\mathbf{a}}(\mathbf{z})}\left(\tilde{\mathbf{a}}(\mathbf{z}) - \tilde{\mathbf{H}}^{-1}\mathbf{u}\right)\left|\det\left(\tilde{\mathbf{H}}^{-1}\right)\right|\mathrm{d}\mathbf{u} \tag{66}$$

where $\left|\det\left(\tilde{\mathbf{H}}^{-1}\right)\right|$ is the absolute value of the determinant of $\tilde{\mathbf{H}}^{-1}$. Using (29), (66) becomes

$$\varepsilon_2 = \left|\det\left(\tilde{\mathbf{H}}^{-1}\right)\right|\int \chi_{I_{\boldsymbol{\eta}}}(\mathbf{u})\,\mathrm{d}\mathbf{u} = \left|\det\left(\tilde{\mathbf{H}}^{-1}\right)\right||I_{\boldsymbol{\eta}}| \tag{67}$$

where $|I_{\boldsymbol{\eta}}|$ indicates the area of the support of the measurement noise. The mean of $p_1(\mathbf{a};\mathbf{z})$ is

$$\boldsymbol{\mu}_{a,1} = \frac{1}{\left|\det\left(\tilde{\mathbf{H}}^{-1}\right)\right||I_{\boldsymbol{\eta}}|}\int \mathbf{a}\chi_{I_{\mathbf{a}}(\mathbf{z})}(\mathbf{a})\,\mathrm{d}\mathbf{a} \tag{68}$$

Using the change of variables of (65), we obtain

$$\boldsymbol{\mu}_{a,1} = \frac{1}{\left|\det\left(\tilde{\mathbf{H}}^{-1}\right)\right||I_{\boldsymbol{\eta}}|}\int \left(\tilde{\mathbf{a}}(\mathbf{z}) - \tilde{\mathbf{H}}^{-1}\mathbf{u}\right)\chi_{I_{\boldsymbol{\eta}}}(\mathbf{u})\left|\det\left(\tilde{\mathbf{H}}^{-1}\right)\right|\mathrm{d}\mathbf{u} \tag{69}$$

Using the fact that the mean of the measurement noise is zero and using Approximation AP3:

$$\frac{1}{|I_{\boldsymbol{\eta}}|}\int \mathbf{u}\chi_{I_{\boldsymbol{\eta}}}(\mathbf{u})\,\mathrm{d}\mathbf{u} = 0 \tag{70}$$

equation (69) becomes

$$\boldsymbol{\mu}_{a,1} = \tilde{\mathbf{a}}(\mathbf{z}) \tag{71}$$

Using the same procedure, the covariance matrix of $p_1(\mathbf{a};\mathbf{z})$ is

$$\begin{aligned}\boldsymbol{\Sigma}_{a,1} &= \frac{1}{\left|\det\left(\tilde{\mathbf{H}}^{-1}\right)\right||I_{\boldsymbol{\eta}}|}\int \left(\mathbf{a}-\tilde{\mathbf{a}}(\mathbf{z})\right)\left(\mathbf{a}-\tilde{\mathbf{a}}(\mathbf{z})\right)^T\chi_{I_{\mathbf{a}}(\mathbf{z})}(\mathbf{a})\,\mathrm{d}\mathbf{a}\\ &= \frac{1}{\left|\det\left(\tilde{\mathbf{H}}^{-1}\right)\right||I_{\boldsymbol{\eta}}|}\int \left(\tilde{\mathbf{H}}^{-1}\mathbf{u}\right)\left(\tilde{\mathbf{H}}^{-1}\mathbf{u}\right)^T\chi_{I_{\boldsymbol{\eta}}}(\mathbf{u})\left|\det\left(\tilde{\mathbf{H}}^{-1}\right)\right|\mathrm{d}\mathbf{u}\end{aligned} \tag{72}$$

In Section IV-C, the covariance matrix of the noise is $\mathbf{R}$. Then, (72) becomes

$$\boldsymbol{\Sigma}_{a,1} = \tilde{\mathbf{H}}^{-1}\mathbf{R}\left(\tilde{\mathbf{H}}^{-1}\right)^T \tag{73}$$

APPENDIX B

In this appendix, we calculate the first two moments of $p_1(\mathbf{a},\mathbf{b};\mathbf{z})$ considering we know the first two moments of $p_0(\mathbf{a},\mathbf{b})$ and $p_1(\mathbf{a};\mathbf{z})$, given by (35), (36), (33) and (34). The truncation only affects the part of the PDF that corresponds with $\mathbf{a}$, see (16). Then, this implies that the conditional PDF of $\mathbf{b}$ given $\mathbf{a}$ is not affected by the truncation. To emphasise this, we can write:

$$p_1(\mathbf{a},\mathbf{b};\mathbf{z}) = p_1(\mathbf{a};\mathbf{z})\,p_0(\mathbf{b}\,|\mathbf{a}) \tag{74}$$

Assuming that $p_0(\mathbf{a},\mathbf{b})$ is Gaussian [1]

$$\boldsymbol{\upsilon}_b(\mathbf{a}) \triangleq \mathsf{E}[\mathbf{b}\,|\mathbf{a}] = \boldsymbol{\mu}_{b,0} + \boldsymbol{\Sigma}_{ab,0}^T\boldsymbol{\Sigma}_{a,0}^{-1}\left(\mathbf{a}-\boldsymbol{\mu}_{a,0}\right) \tag{75}$$

$$\boldsymbol{\Gamma} \triangleq \mathsf{cov}[\mathbf{b}\,|\mathbf{a}] = \boldsymbol{\Sigma}_{b,0} - \boldsymbol{\Sigma}_{ab,0}^T\boldsymbol{\Sigma}_{a,0}^{-1}\boldsymbol{\Sigma}_{ab,0} \tag{76}$$

Now, we calculate the first two moments of $p_1(\mathbf{a},\mathbf{b};\mathbf{z})$ given by (37) and (38). The mean $\boldsymbol{\mu}_{b,1}$:

$$\begin{aligned}\boldsymbol{\mu}_{b,1} &= \int \mathbf{b}p_1(\mathbf{a},\mathbf{b};\mathbf{z})\,\mathrm{d}\mathbf{a}\mathrm{d}\mathbf{b} = \int \mathbf{b}p_0(\mathbf{b}\,|\mathbf{a})\,p_1(\mathbf{a};\mathbf{z})\,\mathrm{d}\mathbf{a}\mathrm{d}\mathbf{b}\\ &= \int \mathsf{E}[\mathbf{b}\,|\mathbf{a}]\,p_1(\mathbf{a};\mathbf{z})\,\mathrm{d}\mathbf{a}\end{aligned} \tag{77}$$

Using (75) and (33) in (77):

$$\boldsymbol{\mu}_{b,1} = \boldsymbol{\mu}_{b,0} + \boldsymbol{\Sigma}_{ab,0}^T\boldsymbol{\Sigma}_{a,0}^{-1}\left(\boldsymbol{\mu}_{a,1}-\boldsymbol{\mu}_{a,0}\right) \tag{78}$$

The cross-covariance $\boldsymbol{\Sigma}_{ab,1}$:

$$\boldsymbol{\Sigma}_{ab,1} = \int \left(\mathbf{a}-\boldsymbol{\mu}_{a,1}\right)\left(\mathbf{b}-\boldsymbol{\mu}_{b,1}\right)^T p_0(\mathbf{b}\,|\mathbf{a})\,p_1(\mathbf{a};\mathbf{z})\,\mathrm{d}\mathbf{a}\mathrm{d}\mathbf{b} \tag{79}$$

Using (75) in (79):

$$\begin{aligned}\boldsymbol{\Sigma}_{ab,1} = \int \left(\mathbf{a}-\boldsymbol{\mu}_{a,1}\right)&\left(\boldsymbol{\mu}_{b,0} + \boldsymbol{\Sigma}_{ab,0}^T\boldsymbol{\Sigma}_{a,0}^{-1}\left(\mathbf{a}-\boldsymbol{\mu}_{a,0}\right) - \boldsymbol{\mu}_{b,1}\right)^T\\ &\times p_1(\mathbf{a};\mathbf{z})\,\mathrm{d}\mathbf{a}\end{aligned} \tag{80}$$

Expanding the terms in (79) and using (34)

$$\boldsymbol{\Sigma}_{ab,1} = \boldsymbol{\Sigma}_{a,1}\left(\boldsymbol{\Sigma}_{a,0}^{-1}\right)^T\boldsymbol{\Sigma}_{ab,0} \tag{81}$$

Lastly, using a similar procedure, the covariance matrix $\boldsymbol{\Sigma}_{b,1}$ can be calculated and is given by (41).

REFERENCES

[1] Y. Bar-Shalom, T. Kirubarajan, and X. R. Li, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2002.

[2] 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.

[3] C. Musso, N. Oudjane, and F. Le Gland, "Improving regularised particle filters," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer-Verlag, 2001.

[4] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

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

[6] 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, Mar. 2002.

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

[8] M. Šimandl, J. Královec, and T. Söderström, "Advanced point-mass method for nonlinear state estimation," *Automatica*, vol. 42, pp. 1133–1145, 2006.

[9] J. H. Kotecha and P. M. Djuric, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2592–2601, Oct. 2003.

[10] F. Daum, "Nonlinear filters: beyond the Kalman filter," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, pp. 57–69, Aug. 2005.

[11] 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.

[12] 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.

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

[14] I. Arasaratnam, S. Haykin, and T. Hurd, "Cubature Kalman filtering for continuous-discrete systems: Theory and simulations," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 4977–4993, Oct. 2010.

[15] T. Lefebvre, H. Bruyninckx, and J. De Schuller, "Comment on "a new method for the nonlinear transformation of means and covariances in filters and estimators" [and authors' reply]," *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1406–1409, Aug. 2002.

[16] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[17] Y. Wu, D. Hu, M. Wu, and X. Hu, "A numerical-integration perspective on Gaussian filters," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2910–2921, Aug. 2006.

[18] T. Lefebvre, H. Bruyninckx, and J. D. Schutter, "Kalman filters for nonlinear systems: a comparison of performance," *International Journal of Control*, vol. 77, no. 7, pp. 639–653, May 2004.

[19] F. Gustafsson and G. Hendeby, "Some relations between extended and unscented Kalman filters," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 545–555, Feb. 2012.

[20] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[21] M. R. Morelande and A. F. García-Fernández, "Analysis of the Kalman filter for nonlinear measurements," *to be submitted*, 2012.

[22] 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.

[23] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.

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

[25] N. Nikolaidis and I. Pitas, "Nonlinear processing and analysis of angular signals," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3181–3194, Dec. 1998.

[26] K. Burdett, "Truncated means and variances," *Economics Letters*, vol. 52, pp. 263–267, 1996.

[27] 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/

[28] 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.

[29] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1386–1396, May 1998.

[30] D. Simon and D. L. Simon, "Constrained Kalman filtering via density function truncation for turbofan engine health estimation," *International Journal of Systems Science*, vol. 41, no. 2, pp. 159–171, Feb. 2010.

[31] M. Dueker, "Kalman filtering with truncated normal state variables for Bayesian estimation of macroeconomic models," *Economics Letters*, vol. 93, no. 1, pp. 141–149, Oct. 2006.

[32] A. F. García-Fernández, M. R. Morelande, and J. Grajal, "Multitarget simultaneous localization and mapping of a sensor network," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4544–4558, Oct. 2011.

Ángel F. García-Fernández received his degree in Telecommunication Engineering (with honours) and the Ph.D. from Universidad Politécnica de Madrid, Spain, in 2007 and 2011, respectively.
Since 2007 he has been a researcher at the Department of Signals, Systems and Radiocommunications of the same university. His research activities and interests are in the area of Bayesian filtering, multitarget tracking, wireless sensor networks, and radar signal processing.

Mark R. Morelande received the B.Eng. degree in aerospace avionics from Queensland University of Technology, Brisbane, Australia in 1997 and the Ph.D. in electrical engineering from Curtin University of Technology, Perth, Australia in 2001.
In 2001 he was a Postdoctoral Fellow at the Centre for Eye Research, Queensland University of Technology. From 2002-2005 he was a Research Fellow at the Cooperative Research Centre for Sensor, Signal and Information Processing, University of Melbourne. He is now a Senior Research Fellow in the Melbourne Systems Laboratory, also at The University of Melbourne. His research interests include non-stationary signal analysis and target tracking with particular emphasis on multiple target tracking and sequential Monte Carlo methods.

Jesús Grajal was born in Toral de los Guzmanes (León), Spain, in 1967. He received the Ingeniero de Telecomunicación and the Ph.D. degrees from the Technical University of Madrid, Madrid, Spain in 1992 and 1998, respectively.
Since 2001 he has been an Associate Professor at the Signals, Systems, and Radiocommunications Department of the Technical School of Telecommunication Engineering of the same University. His research activities are in the area of hardware-design for radar systems, radar signal processing and broadband digital receivers for radar and spectrum surveillance applications.