Adaptive unscented Gaussian likelihood approximation filter

(article starts on next page)

Adaptive unscented Gaussian likelihood approximation filter

Ángel F. García-Fernández⋆, Mark R. Morelande†, Jesús Grajal⋆, Lennart Svensson°

⋆Dpto. de Señales, Sistemas y Radiocomunicaciones, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain

†Dept. of Electrical and Electronic Engineering, The University of Melbourne, Parkville, Victoria 3010, Australia

°Dept. of Signals and Systems, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

Abstract

This paper focuses on the update step of Bayesian nonlinear filtering. We first derive the unscented Gaussian likelihood approximation filter (UGLAF), which provides a Gaussian approximation to the likelihood by applying the unscented transformation to the inverse of the measurement function. The UGLAF approximation is accurate in the cases where the unscented Kalman filter (UKF) is not and the other way round. As a result, we propose the adaptive UGLAF (AUGLAF), which selects the best approximation to the posterior (UKF or UGLAF) based on the Kullback-Leibler divergence. This enables AUGLAF to outperform both the UKF and UGLAF.

Keywords: Bayes' rule, Kalman filter, Gaussian approximation, nonlinear filtering

1. Introduction

The objective of Bayesian filtering is to estimate the current state of a process based on previous measurements up to the current time. Estimators of interest such as the minimum mean square error (MMSE) estimator require knowledge of the probability density function (PDF) of the current state given the current and previous measurements. This PDF is referred to as the posterior PDF. In principle, the posterior can be calculated in a recursive procedure that involves two phases: prediction and update.

If the system is linear and Gaussian or the state space is discrete, the posterior PDF can be calculated in closed-form. In other cases, the posterior PDF is intractable in general so approximations are required. Particle filters (PFs) [1] obtain weighted samples from the posterior by sequentially drawing from an importance density. PFs give an asymptotically exact approximation of the posterior as the sample size tends to infinity although their performance for a finite sample size may be poor. Thus, it is of interest to develop computationally efficient Gaussian approximations, especially, if the posterior is unimodal.

Gaussian approximations are often based on the Kalman filter (KF). Even though the KF is most commonly known in the literature as the solution to the linear/Gaussian filtering recursion, the KF can also be applied in the nonlinear case [2, Sec. II.A]. In this case, the KF does not provide the true posterior PDF, it is an approximation. Particularly, the update step of the KF consists of approximating the first two moments of the posterior by the linear minimum mean square error (LMMSE) estimator with its mean square error matrix [2, Sec. II.A]. Calculating the LMMSE estimator requires knowledge of the mean and covariance matrix of the current measurement and cross-covariance between the current state and the current measurement. However, these moments (KF moments) cannot be calculated in closed-form so we require approximations. In this paper, we refer to these approximations as approximations to the KF. For instance, the extended Kalman filter (EKF) [3, 4] approximates the KF moments using analytical linearisation while the unscented KF (UKF) [2, 5] and cubature KF (CKF) [6, 7] use sigma-point methods. Other approaches, such as the iterated EKF (IEKF), use the maximum a posteriori (MAP) estimate in the update step to approximate the posterior [8]. However, the IEKF is not ensured to converge and requires the calculation of Jacobians.

In the update step with a nonlinear measurement function, analysis has shown that the KF and its approximations tend to perform well when the measurements are imprecise but deteriorate as the measurements become more precise [9, Sec. II.D]. Motivated by this, we focus on the update step and seek a computationally efficient Gaussian posterior density approximation which can be used in combination with the KF to provide good performance for precise and imprecise measurements. To this end we propose the following: 1) a filtering algorithm which is accurate for precise measurements and 2) a method to select between the proposed filter and the UKF as suggested by the available measurements. These contributions are discussed below.

We first develop a derivative-free filter that approxi-

Email address: angel.garciafernandez@curtin.edu.au ()

mates the likelihood as a Gaussian, which can be used to find a Gaussian approximation to the posterior. This method is based on the unscented transformation (UT) [2] and we refer to it as the unscented Gaussian likelihood approximation filter (UGLAF). The UGLAF is shown to be optimal for low enough measurement noise variance.

Motivated by the fact that the KF approximation is accurate for high measurement noise variance while the UGLAF approximation is accurate for low measurement noise variance, we propose an adaptive approach in which, depending on the situation, one of the two Gaussian posterior density approximations is selected. The choice between the candidate approximations is made using the Kullback-Leibler divergence [10]. The resulting algorithm, the adaptive UGLAF (AUGLAF), switches between the UGLAF and the UKF. The idea of using two approximations to the posterior depending on their suitability was also used in the truncated UKF (TUKF). The TUKF was proposed in [11] but it has two main drawbacks. First, the rule to select between the truncated KF and KF approximations is ad-hoc. Second, the TUKF is an optimistic filter for precise measurements.

The rest of the paper is organised as follows. In Section 2, the UGLAF is presented. We discuss the relationships between UGLAF and existing filters in the literature in Section 3. AUGLAF is introduced in Section 4. Numerical simulations in which the performances of these filters are analysed are provided in Section 5. Finally, conclusions are drawn in Section 6.

2. UGLAF

This paper focuses on the update phase of Bayesian filtering. Let $\mathbf{x} \in \mathbb{R}^{n_x}$ and $\mathbf{z} \in \mathbb{R}^{n_z}$ be the state and the measurement, respectively. We write the state vector as $\mathbf{x} = \left[\mathbf{a}^T, \mathbf{b}^T\right]^T$, where T stands for 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 is

$$\mathbf{z} = \mathbf{h}\left(\mathbf{a}\right) + \boldsymbol{\eta} \tag{1}$$

where $\mathbf{h}\left(\cdot\right)$ is the measurement function and $\boldsymbol{\eta}$ is the measurement noise with PDF $p_{\boldsymbol{\eta}}\left(\cdot\right)$ which is a zero-mean Gaussian[1] with covariance matrix $\mathbf{R}$. Using (1), the conditional PDF $p\left(\mathbf{z}\,|\mathbf{a}\right)$ of the measurement $\mathbf{z}$ given the state $\mathbf{a}$ becomes

$$p\left(\mathbf{z}\,|\mathbf{a}\right) = p_{\boldsymbol{\eta}}\left(\mathbf{z} - \mathbf{h}\left(\mathbf{a}\right)\right) \tag{2}$$

Given a measurement $\mathbf{z}$, the likelihood function $l\left(\cdot\right)$ is the conditional PDF (2) viewed as a function of the state

$$l\left(\mathbf{a}\right) = p_{\boldsymbol{\eta}}\left(\mathbf{z} - \mathbf{h}\left(\mathbf{a}\right)\right) \tag{3}$$

where we have dropped the dependence on $\mathbf{z}$ for the sake of notational simplicity. Given a prior $p\left(\cdot\right)$, the aim is to

calculate the posterior $q\left(\cdot\right)$, which is given by Bayes' rule

$$q\left(\mathbf{x}\right) \propto l\left(\mathbf{a}\right)p\left(\mathbf{x}\right) \tag{4}$$

where $\propto$ denotes proportionality. The UGLAF provides a Gaussian approximation to $q\left(\cdot\right)$ under the assumptions:

- A1 The measurement function $\mathbf{h}\left(\cdot\right)$ is a continuous, bijective function whose inverse is continuously differentiable (which implies $n_z = n_a$).

- A2 The likelihood is integrable.

- A3 The prior is Gaussian with mean $\overline{\mathbf{x}}$ and covariance matrix $\boldsymbol{\Sigma}$:

$$\overline{\mathbf{x}} = \left[\overline{\mathbf{a}}^T, \overline{\mathbf{b}}^T\right]^T \tag{5}$$

$$\boldsymbol{\Sigma} = \left[\begin{array}{cc} \boldsymbol{\Sigma}_a & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ab}^T & \boldsymbol{\Sigma}_b \end{array}\right] \tag{6}$$

In Section 2.1, we develop a Gaussian likelihood approximation. Then, in Section 2.2, we use the Gaussian likelihood approximation to obtain a Gaussian posterior approximation. We prove the optimality of UGLAF as the measurement noise variance tends to zero in Section 2.3. Finally, how to generalise UGLAF for nonbijective measurement functions is explained in Section 2.4.

2.1. Unscented Gaussian approximation to the likelihood

In general, a likelihood function is not a PDF because it does not integrate to one. However, Bayes' rule remains unaltered if we replace the likelihood by a scaled version of the likelihood. Then, under Assumption A2, we can define a normalised likelihood $\tilde{l}\left(\cdot\right)$ that integrates to one

$$\tilde{l}\left(\mathbf{a}\right) = \frac{l\left(\mathbf{a}\right)}{\int l\left(\tilde{\mathbf{a}}\right)d\tilde{\mathbf{a}}} \tag{7}$$

The posterior, which is given by (4), is then

$$q\left(\mathbf{x}\right) = \frac{\tilde{l}\left(\mathbf{a}\right)p\left(\mathbf{x}\right)}{\rho} \tag{8}$$

where $\rho = \int \tilde{l}\left(\mathbf{a}\right)p\left(\mathbf{x}\right)d\mathbf{x}$ is the normalising constant.

The objective of this section is to design a derivative-free method to obtain a Gaussian approximation to the normalised likelihood:

$$\tilde{l}\left(\mathbf{a}\right) \approx \mathcal{N}\left(\mathbf{a}; \overline{\mathbf{l}}, \mathbf{L}\right) \tag{9}$$

where $\mathcal{N}\left(\mathbf{a}; \overline{\mathbf{l}}, \mathbf{L}\right)$ is the Gaussian PDF evaluated at $\mathbf{a}$ with mean $\overline{\mathbf{l}}$ and covariance matrix $\mathbf{L}$. It should be noted that $\overline{\mathbf{l}}$ and $\mathbf{L}$ depend on $\mathbf{z}$ in general but this is omitted in the notation for the sake of simplicity.

Using (7) and (3), the mean of the normalised likelihood is given by

$$\overline{\mathbf{l}} = \int \mathbf{a}\tilde{l}\left(\mathbf{a}\right)d\mathbf{a} \tag{10}$$

---

[1] If the measurement noise is not Gaussian, it can be approximated as Gaussian using its first two moments.

$$= \frac{\int \mathbf{a} p_{\boldsymbol{\eta}} \left(\mathbf{z} - \mathbf{h}\left(\mathbf{a}\right)\right) d\mathbf{a}}{\int p_{\boldsymbol{\eta}} \left(\mathbf{z} - \mathbf{h}\left(\mathbf{a}\right)\right) d\mathbf{a}} \quad (11)$$

$$= \frac{\int \mathbf{a} \mathcal{N} \left(\mathbf{z}; \mathbf{h}\left(\mathbf{a}\right), \mathbf{R}\right) d\mathbf{a}}{\int \mathcal{N} \left(\mathbf{z}; \mathbf{h}\left(\mathbf{a}\right), \mathbf{R}\right) d\mathbf{a}} \quad (12)$$

where $\mathbf{z}$ is known. Equivalently, the covariance matrix of the normalised likelihood is

$$\mathbf{L} = \int \left(\mathbf{a} - \bar{\mathbf{l}}\right) \left(\mathbf{a} - \bar{\mathbf{l}}\right)^T \tilde{l}\left(\mathbf{a}\right) d\mathbf{a}$$

$$= \frac{\int \left(\mathbf{a} - \bar{\mathbf{l}}\right) \left(\mathbf{a} - \bar{\mathbf{l}}\right)^T \mathcal{N} \left(\mathbf{z}; \mathbf{h}\left(\mathbf{a}\right), \mathbf{R}\right) d\mathbf{a}}{\int \mathcal{N} \left(\mathbf{z}; \mathbf{h}\left(\mathbf{a}\right), \mathbf{R}\right) d\mathbf{a}} \quad (13)$$

Because of A1, we can make the change of variable $\mathbf{a} = \mathbf{h}^{-1}\left(\mathbf{y}\right)$ in (12) and (13) and write

$$\bar{\mathbf{l}} = \frac{\int \mathbf{h}^{-1}\left(\mathbf{y}\right) \left|\det \mathbf{h}'_I\left(\mathbf{y}\right)\right| \mathcal{N}\left(\mathbf{y}; \mathbf{z}, \mathbf{R}\right) d\mathbf{y}}{\int \left|\det \mathbf{h}'_I\left(\mathbf{y}\right)\right| \mathcal{N}\left(\mathbf{y}; \mathbf{z}, \mathbf{R}\right) d\mathbf{y}} \quad (14)$$

$$\mathbf{L} = \frac{\int \left(\mathbf{h}^{-1}\left(\mathbf{y}\right) - \bar{\mathbf{l}}\right) \left(\mathbf{h}^{-1}\left(\mathbf{y}\right) - \bar{\mathbf{l}}\right)^T \left|\det \mathbf{h}'_I\left(\mathbf{y}\right)\right| \mathcal{N}\left(\mathbf{y}; \mathbf{z}, \mathbf{R}\right) d\mathbf{y}}{\int \left|\det \mathbf{h}'_I\left(\mathbf{y}\right)\right| \mathcal{N}\left(\mathbf{y}; \mathbf{z}, \mathbf{R}\right) d\mathbf{y}} \quad (15)$$

where $\mathbf{h}'_I\left(\cdot\right)$ represents the Jacobian of the inverse of the measurement function and $\det \mathbf{A}$ denotes the determinant of matrix $\mathbf{A}$.

The integrals in (14)-(15) can be approximated using the unscented transformation (UT) [2]:

$$\bar{\mathbf{l}} \approx \bar{\mathbf{l}}_1 = \frac{\sum_{j=1}^{N_s} \omega^j \mathcal{A}^j \left|\det \mathbf{h}'_I\left(\mathcal{Y}^j\right)\right|}{\sum_{j=1}^{N_s} \omega^j \left|\det \mathbf{h}'_I\left(\mathcal{Y}^j\right)\right|} \quad (16)$$

$$\mathbf{L} \approx \mathbf{L}_1 = \frac{\sum_{j=1}^{N_s} \omega^j \left(\mathcal{A}^j - \bar{\mathbf{l}}_1\right) \left(\mathcal{A}^j - \bar{\mathbf{l}}_1\right)^T \left|\det \mathbf{h}'_I\left(\mathcal{Y}^j\right)\right|}{\sum_{j=1}^{N_s} \omega^j \left|\det \mathbf{h}'_I\left(\mathcal{Y}^j\right)\right|} \quad (17)$$

where

$$\mathcal{A}^j = \mathbf{h}^{-1}\left(\mathcal{Y}^j\right), \quad j = 1, ..., N_s \quad (18)$$

and the N_s sigma points $\mathcal{Y}^1, ..., \mathcal{Y}^{N_s}$ and weights $\omega^1, ..., \omega^{N_s}$ match the mean $\mathbf{z}$ and covariance matrix $\mathbf{R}$. How to select these sigma points and weights is explained in [2].

The UT approximation is not very convenient in this case because it requires the calculation of the Jacobian of the inverse of the measurement function. Therefore, we use the following approximation to calculate $\bar{\mathbf{l}}_1$ and $\mathbf{L}_1$:

• AP1

$$\bar{\mathbf{l}}_1 \approx \bar{\mathbf{l}}_2 = \sum_{j=1}^{N_s} \omega^j \mathcal{A}^j \quad (19)$$

$$\mathbf{L}_1 \approx \mathbf{L}_2 = \sum_{j=1}^{N_s} \omega^j \left(\mathcal{A}^j - \bar{\mathbf{l}}_2\right) \left(\mathcal{A}^j - \bar{\mathbf{l}}_2\right)^T \quad (20)$$

The calculations of $\bar{\mathbf{l}}_1$ and $\mathbf{L}_1$ under Approximation AP1 are much simpler than without AP1, see (16)-(17), as they

Table 1: Steps of the update phase of the UGLAF

- Select sigma points $\mathcal{Y}^1, ..., \mathcal{Y}^{N_s}$ matching the moments $\mathbf{z}$ and $\mathbf{R}$.
- Compute the transformed sigma points $\mathcal{A}^1, ..., \mathcal{A}^{N_s}$ using (18).
- Compute the Gaussian approximation to the likelihood, whose moments are $\bar{\mathbf{l}}$ and $\mathbf{L}$, using (19) and (20).
- Compute the UGLAF approximation to the posterior, whose moments are $\bar{\mathbf{u}}_1$ and $\mathbf{U}_1$, using (24) and (25).

do not require the calculation of the Jacobian. We are especially interested in UGLAF if the measurement noise is low as otherwise the KF is expected to work well [9]. It is proved in AppendixA that $\bar{\mathbf{l}}_1$ and $\bar{\mathbf{l}}_2$ and, $\mathbf{L}_1$ and $\mathbf{L}_2$ are alike as the measurement noise tends to zero. In addition, it is shown in Section 2.3 that both likelihood approximations tend to be optimal for low enough measurement noise, which justifies the use of AP1. One possible interpretation of AP1 is that we are approximating the Jacobian of $\mathbf{h}^{-1}\left(\cdot\right)$ as constant for all sigma points.

2.2. Gaussian approximation to the posterior

The Gaussian likelihood approximation, which is given by (9), can also be written as

$$\tilde{l}\left(\mathbf{a}\right) \approx \mathcal{N}\left(\mathbf{A}\mathbf{x}; \bar{\mathbf{l}}, \mathbf{L}\right) = \mathcal{N}\left(\bar{\mathbf{l}}; \mathbf{A}\mathbf{x}, \mathbf{L}\right) \quad (21)$$

where

$$\mathbf{A} = [\mathbf{I}_{n_a}, \mathbf{0}_{n_b}] \quad (22)$$

and $\mathbf{I}_n$ is the identity matrix of size n and $\mathbf{0}_n$ is the square zero matrix of size n. Then, substituting (21) into (8) and using the Gaussian product formula [12], the posterior approximation $\hat{q}^1\left(\cdot\right)$ becomes

$$q\left(\mathbf{x}\right) \approx \hat{q}^1\left(\mathbf{x}\right) = \mathcal{N}\left(\mathbf{x}; \bar{\mathbf{u}}_1, \mathbf{U}_1\right) \quad (23)$$

where

$$\bar{\mathbf{u}}_1 = \bar{\mathbf{x}} + \boldsymbol{\Psi} \mathbf{S}^{-1} \left(\bar{\mathbf{l}} - \hat{\mathbf{l}}\right) \quad (24)$$

$$\mathbf{U}_1 = \boldsymbol{\Sigma} - \boldsymbol{\Psi} \mathbf{S}^{-1} \boldsymbol{\Psi}^T \quad (25)$$

$$\hat{\mathbf{l}} = \mathbf{A}\bar{\mathbf{x}} \quad (26)$$

$$\mathbf{S} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T + \mathbf{L} \quad (27)$$

$$\boldsymbol{\Psi} = \boldsymbol{\Sigma}\mathbf{A}^T \quad (28)$$

Finally, the algorithm is summarised in Table 1.

2.3. Asymptotic convergence

Let us write the measurement noise covariance matrix as $\mathbf{R} = \tilde{\mathbf{R}}/\tilde{n}$. In this section, we show the convergence of UGLAF as $\tilde{n} \to \infty$. For simplicity, we write the proof for the case that $\mathbf{x} = \mathbf{a}$ although it can be generalised to the case $\mathbf{x} = \left[\mathbf{a}^T, \mathbf{b}^T\right]^T$. The UGLAF approximation to the normalised likelihood is denoted as

$$\hat{l}\left(\mathbf{a}\right) = \mathcal{N}\left(\mathbf{a}; \bar{\mathbf{l}}_{i,\tilde{n}}, \mathbf{L}_{i,\tilde{n}}\right) \quad (29)$$

where $i = 1$ if we use (16) and (17), and $i = 2$ if we use (19) and (20). As proved in Appendix A, regardless of whether we use AP1 or not, we can write

$$\bar{\mathbf{l}}_{i,\tilde{n}} = \mathbf{a}^{\star} + O\left(\tilde{n}^{-1}\right)$$
$$\mathbf{L}_{i,\tilde{n}} = \tilde{n}^{-1}\tilde{\mathbf{L}} + O\left(\tilde{n}^{-3/2}\right)$$

where $\tilde{\mathbf{L}} = \mathbf{H}_I\tilde{\mathbf{R}}\mathbf{H}_I^T$. We can then obtain (30) and (31). Using (B.1) in Appendix B, we get

$$\hat{l}\left(\mathbf{a}\right) = \frac{\exp\left(-\tilde{n}\left(\mathbf{a} - \mathbf{a}^{\star}\right)^T \tilde{\mathbf{L}}^{-1}\left(\mathbf{a} - \mathbf{a}^{\star}\right)\left[1 + O\left(\tilde{n}^{-1/2}\right)\right]/2\right)}{\left|2\pi\tilde{\mathbf{L}}/\tilde{n}\right|^{1/2}\left[1 + O\left(\tilde{n}^{-1/2}\right)\right]}$$
$$(32)$$

Applying (B.2) and (B.3) in Appendix B to the numerator and denominator, respectively, of (32) gives

$$\hat{l}\left(\mathbf{a}\right) = \mathcal{N}\left(\mathbf{a}; \mathbf{a}^{\star}, \tilde{\mathbf{L}}/\tilde{n}\right)\left[1 + O\left(\tilde{n}^{-1/2}\right)\right] \qquad (33)$$

The posterior approximation can be written as

$$\hat{q}^1\left(\mathbf{a}\right) \propto \hat{l}\left(\mathbf{a}\right)p\left(\mathbf{a}\right)$$
$$= \mathcal{N}\left(\mathbf{a}; \mathbf{a}^{\star}, \tilde{\mathbf{L}}/\tilde{n}\right)\mathcal{N}\left(\mathbf{a}; \bar{\mathbf{a}}, \boldsymbol{\Sigma}_a\right)\left[1 + O\left(\tilde{n}^{-1/2}\right)\right]$$
$$\propto \mathcal{N}\left(\mathbf{a}; \mathbf{a}^{\star} + \mathbf{K}_{\tilde{n}}\left(\bar{\mathbf{a}} - \mathbf{a}^{\star}\right), \tilde{\mathbf{L}}/\tilde{n} - \mathbf{K}_{\tilde{n}}\tilde{\mathbf{L}}/\tilde{n}\right)$$
$$\left[1 + O\left(\tilde{n}^{-1/2}\right)\right]$$

where $\mathbf{K}_{\tilde{n}} = \tilde{\mathbf{L}}\left(\tilde{\mathbf{L}} + \tilde{n}\boldsymbol{\Sigma}_a\right)^{-1}$. Since $\mathbf{K}_{\tilde{n}} = O\left(\tilde{n}^{-1}\right)$ we can argue as in the proof of (33) that

$$\hat{q}^1\left(\mathbf{a}\right) = \mathcal{N}\left(\mathbf{a}; \mathbf{a}^{\star}, \tilde{\mathbf{L}}/\tilde{n}\right)\left[1 + O\left(\tilde{n}^{-1/2}\right)\right] \qquad (34)$$

In [13], it is indicated that the true posterior density converges to (34). That is, the true posterior is asymptotically normal distributed with mean $\mathbf{a}^{\star}$ and covariance matrix $\tilde{\mathbf{L}}/\tilde{n}$ with a remainder $O\left(\tilde{n}^{-1/2}\right)$ if $\tilde{n} \rightarrow \infty$. A proof of this result is given in [14]. This shows that the posterior density obtained by the UGLAF asymptotically converges to the true posterior as $\tilde{n} \rightarrow \infty$, i.e., as the measurement noise variance decreases to zero. As a result, if the measurement noise is low enough, the UGLAF approximation to the posterior is accurate.

2.4. Generalisations of UGLAF

The UGLAF has been presented for bijective measurement functions $\mathbf{h}\left(\cdot\right) : \mathbb{R}^{n_a} \rightarrow \mathbb{R}^{n_a}$. We did this for the sake of clarity as in this case the inverse function $\mathbf{h}^{-1}\left(\cdot\right)$ exists for all possible values of the measurement. Nonetheless, UGLAF can be generalised for other types of measurement functions. The most direct generalisation is for injective measurement functions, which is described in Section 2.4.1. Other generalisations are beyond the scope of this paper but we want to briefly comment on them in the next paragraph.

If we perform the update step in several phases, UGLAF can be applied more generally. For example, 2-D target tracking using a sensor that measures range, bearings and Doppler shift can be handled by updating first the range and bearings measurements using UGLAF and then the Doppler shift using a UKF. As with the mixture TUKF [15], if the likelihood is multimodal, a Gaussian mixture approximation based on UGLAF is possible. In this case, the measurement function must meet some conditions, e.g., its domain and codomain can be partitioned such that the measurement function is injective in some regions.

2.4.1. Injective measurement function

In this section, we use the fact that an injective function induces a bijection by constraining its codomain by its image to apply UGLAF. Let us assume that the domain $\mathbb{R}^{n_a}$ of $\mathbf{h}\left(\cdot\right)$ is partitioned as $I_I \cup I_N = \mathbb{R}^{n_a}$, where I_I and I_N denote the regions where $\mathbf{h}\left(\cdot\right)$ is and is not injective, respectively. If we replace the domain and codomain of $\mathbf{h}\left(\cdot\right)$ by I_I and its image $\mathbf{h}\left(I_I\right)$, the resulting function $\mathbf{h}\left(\cdot\right) : I_I \rightarrow \mathbf{h}\left(I_I\right)$ is bijective and its inverse function $\mathbf{h}^{-1}\left(\cdot\right) : \mathbf{h}\left(I_I\right) \rightarrow I_I$ exists. For example, if $h(x) = x^2$ $x > 0$ and $h(x) = 0$ $x \leq 0$, we get $I_I = \{x : x > 0\}$ and $h(I_I) = \{y : y > 0\}$. The idea developed in the rest of the section is that we can apply UGLAF if all the sigma points belong to region $\mathbf{h}\left(I_I\right)$.

If the measurement noise PDF were a truncated Gaussian with bounded support $I_{\boldsymbol{\eta}}$, the mean of the normalised likelihood, which is given by (11), would be written as

$$\bar{\mathbf{l}} = \frac{\int_{\mathbf{z} - \mathbf{h}(\mathbf{a}) \in I_{\boldsymbol{\eta}}} \mathbf{a}p_{\boldsymbol{\eta}}\left(\mathbf{z} - \mathbf{h}\left(\mathbf{a}\right)\right)d\mathbf{a}}{\int_{\mathbf{z} - \mathbf{h}(\mathbf{a}) \in I_{\boldsymbol{\eta}}} p_{\boldsymbol{\eta}}\left(\mathbf{z} - \mathbf{h}\left(\mathbf{a}\right)\right)d\mathbf{a}} \qquad (35)$$

The equation for the covariance matrix of the normalised likelihood is analogous and is not written here. We can apply UGLAF to approximate (35) if we can perform the change of variable $\mathbf{a} = \mathbf{h}^{-1}\left(\mathbf{y}\right)$ in the integrals, see Section 2.1. This can be done if the image I_I of $\mathbf{h}^{-1}\left(\cdot\right)$ contains the region of integration $\{\mathbf{a} : \mathbf{z} - \mathbf{h}\left(\mathbf{a}\right) \in I_{\boldsymbol{\eta}}\}$. Using the fact that $\mathbf{y} = \mathbf{h}\left(\mathbf{a}\right)$, we can equivalently say that we can apply UGLAF if $\mathbf{h}\left(I_I\right)$ contains the set $\mathbf{z} - I_{\boldsymbol{\eta}} \triangleq \{\mathbf{y} : \mathbf{z} - \mathbf{y} \in I_{\boldsymbol{\eta}}\}$, i.e., $\mathbf{z} - I_{\boldsymbol{\eta}} \subset \mathbf{h}\left(I_I\right)$ or $I_{\boldsymbol{\eta}} \subset \mathbf{z} - \mathbf{h}\left(I_I\right)$. That is, the support of the measurement noise PDF must be contained in the set $\mathbf{z} - \mathbf{h}\left(I_I\right)$.

In practice, the support of the measurement noise is not bounded but if the inverse function exists for the values of the PDF of the measurement noise that are non-negligible, we can still use the UGLAF approximation. This condition can be written as:

$$I = \int \chi_{\mathbf{z} - \mathbf{h}(I_I)}\left(\boldsymbol{\eta}\right)p_{\boldsymbol{\eta}}\left(\boldsymbol{\eta}\right)d\boldsymbol{\eta} \approx 1$$
$$= \int \chi_{\mathbf{z} - \mathbf{h}(I_I)}\left(\boldsymbol{\eta}\right)\mathcal{N}\left(\boldsymbol{\eta}; \mathbf{0}, \mathbf{R}\right)d\boldsymbol{\eta} \approx 1 \qquad (36)$$

where $\chi_A\left(\cdot\right)$ is the indicator function on the subset A.

$$\hat{l}\left(\mathbf{a}\right) = \frac{\exp\left(-\left[\mathbf{a} - \mathbf{a}^\star + O\left(\tilde{n}^{-1/2}\right)\right]^T \left[\tilde{n}^{-1}\tilde{\mathbf{L}} + O\left(\tilde{n}^{-3/2}\right)\right]^{-1} \left[\mathbf{a} - \mathbf{a}^\star + O\left(\tilde{n}^{-1/2}\right)\right]/2\right)}{\left|2\pi\left[\tilde{n}^{-1}\tilde{\mathbf{L}} + O\left(\tilde{n}^{-3/2}\right)\right]\right|^{1/2}} \quad (30)$$

$$= \frac{\exp\left(-\tilde{n}\left[\mathbf{a} - \mathbf{a}^\star + O\left(\tilde{n}^{-1/2}\right)\right]^T \left[\tilde{\mathbf{L}}^{-1} + O\left(\tilde{n}^{-1/2}\right)\right] \left[\mathbf{a} - \mathbf{a}^\star + O\left(\tilde{n}^{-1/2}\right)\right]/2\right)}{\left|2\pi\left[\tilde{\mathbf{L}} + O\left(\tilde{n}^{-1/2}\right)\right]/\tilde{n}\right|^{1/2}} \quad (31)$$

Using the change of variable $\mathbf{y} = \mathbf{z} - \boldsymbol{\eta}$ in (36), we get

$$I = \int \chi_{\mathbf{h}(I_I)}\left(\mathbf{y}\right) \mathcal{N}\left(\mathbf{y}; \mathbf{z}, \mathbf{R}\right) d\boldsymbol{\eta} \approx 1 \quad (37)$$

The UT can also be used to integrate discontinuous functions [2, Sec. VII]. Therefore, as in Section 2.1, we can use the UT with sigma points $\mathcal{Y}^1, ..., \mathcal{Y}^{N_s}$, which are drawn with mean $\mathbf{z}$ and covariance matrix $\mathbf{R}$ to approximate (37)

$$I \approx \sum_{j=1}^{N_s} \omega^j \chi_{\mathbf{h}(I_I)}\left(\mathcal{Y}^j\right) \quad (38)$$

Consequently, if $\mathcal{Y}^j \in \mathbf{h}\left(I_I\right) : j \in \{1, ..., N_s\}$ then $I \approx 1$, we meet condition (36) and UGLAF can be applied. Otherwise, UGLAF cannot be applied but a UKF approximation can be used instead. Approximation (38) is quite useful as it does not require extra computational burden in UGLAF. Drawing sigma points $\mathcal{Y}^j$, $j \in \{1, ..., N_s\}$ comes at no extra computational burden as we already need them in the first step of UGLAF, see Table 1. If the inverse function exists for all of them (this is what Eq. (38) checks), we continue with the steps of UGLAF.

3. Comparison with existing filters

In this section, we compare the methodology and conceptual motivation of UGLAF with other Gaussian approximations.

3.1. UGLAF vs KF

In this section, we discuss the differences between the UGLAF and the KF. We find it convenient to use the following example to illustrate our reasoning. Consider a prior $p\left(x\right) = \mathcal{N}\left(x; 3, 4\right)$ and the measurement equation

$$z = 0.01x^3 + \eta \quad (39)$$

where η is zero-mean Gaussian with variance 0.1.

In order to obtain an approximation to the posterior, the KF approximates the joint PDF of the measurement and the state as Gaussian

$$p\left(\mathbf{z}\,|\mathbf{a}\right) p\left(\mathbf{x}\right) \approx \mathcal{N}\left(\left[\mathbf{x}^T, \mathbf{z}^T\right]^T; \left[\bar{\mathbf{x}}^T, \hat{\mathbf{z}}_0^T\right]^T, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Psi}_0 \\ \boldsymbol{\Psi}_0^T & \mathbf{S}_0 \end{bmatrix}\right) \quad (40)$$

where $\hat{\mathbf{z}}_0 = \mathrm{E}\left[\mathbf{z}\right]$, $\boldsymbol{\Psi}_0 = \mathrm{cov}\left[\mathbf{x}, \mathbf{z}\right]$ and $\mathbf{S}_0 = \mathrm{cov}\left[\mathbf{z}\right]$ are the KF moments. In general, these moments cannot be calculated exactly so the KF must be approximated. The approximations to the KF use linearisation (EKF) or sigma points matching the moments of the prior (UKF, CKF) to approximate these moments [11]. Once the joint PDF is approximated as Gaussian, the posterior can be calculated analytically for any value of the measurement. Therefore, the only enabling approximation to have a Gaussian posterior is (40). This is illustrated in Figures 1 (a) and (b), where we show the contour plots of the true joint posterior and its KF approximation, respectively. In our example, the KF moments are calculated exactly, see [16] for details. The following discussion applies to the KF and its approximations so the term KF encompasses the KF and its approximations for simplicity.

If the KF moments are calculated exactly, (40) is, in a sense, the best Gaussian approximation to the joint PDF as it matches its first two moments[2]. However, the problem at hand is not to calculate a Gaussian approximation to the joint PDF but to approximate the posterior. As the KF approximates the joint PDF by (40), it does not take into account the value of $\mathbf{z}$ (which is known) in performing the enabling approximation to obtain a Gaussian posterior. This is unimportant if the measurement noise is large enough as the Gaussian approximation to the joint PDF is accurate [9], recalling that both the noise and prior are Gaussian. However, it becomes important for low measurement noise in nonlinear scenarios as the Gaussian approximation to the joint PDF is not accurate in such cases [9]. As a result, there will be values of $\mathbf{z}$ for which the posterior approximation is extremely poor. This is evident in Figures 1 (a) and (b). Because the joint PDF is not accurately approximated by (40), there are clearly measurement values for which the posterior approximation provided by the KF is poor, particularly those which are far from the prior mean of the measurement.

Let us assume we measure the value $z = 1.5$ in our example. Once we know the measurement, we have a likelihood and we can calculate the posterior using (4). Unlike the KF, the only enabling approximation UGLAF makes is to represent the normalised likelihood as a Gaussian. As

---

[2]It is best in the sense that it minimises the KLD $D\left(p \parallel \hat{p}\right)$ where $p\left(\cdot\right)$ represents a PDF and $\hat{p}\left(\cdot\right)$ its approximation [10].
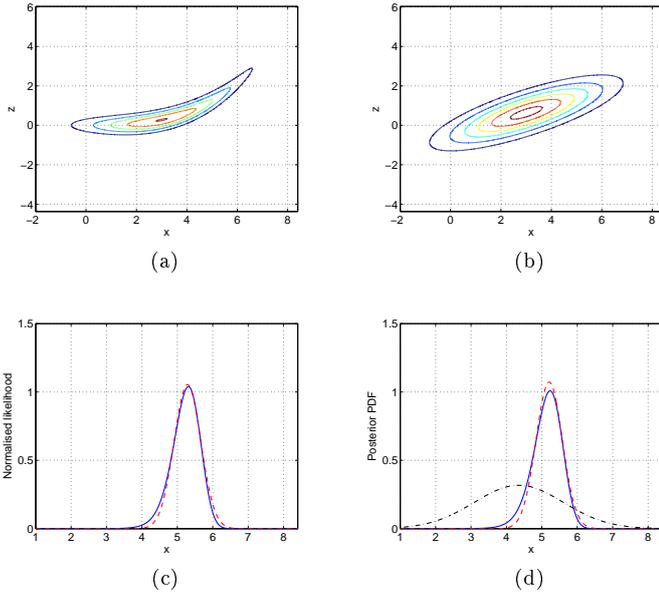
Figure 1: Comparison between the UGLAF and KF (a) True joint PDF (b) KF approximation (c) True normalised likelihood (solid) and UGLAF approximation (dashed). (d) True posterior (solid) and UGLAF (dashed) and KF (dot-dashed) outputs.

shown before, sigma points should be chosen matching the moments $\mathbf{z}$ and $\mathbf{R}$ so the UGLAF uses knowledge of $\mathbf{z}$ to construct an enabling approximation to obtain a Gaussian posterior. The true normalised likelihood and its UGLAF approximation are plotted in Figure 1 (c). We can see that the UGLAF approximates the likelihood very accurately. This results in a posterior approximation which is much closer to the true posterior than the KF approximation shown in Figure 1 (d).

3.2. Use of analytical linearisation instead of the UT

UGLAF approximates the first two moments of the likelihood using the UT. However, these moments can also be approximated by analytical linearisation of the measurement function around the point $\tilde{\mathbf{a}}(\mathbf{z})$ that maximises the likelihood. This method is referred to as linearised GLAF (LGLAF).

Using a first order Taylor series expansion of $\mathbf{h}(\cdot)$ around $\tilde{\mathbf{a}}(\mathbf{z}) = \mathbf{h}^{-1}(\mathbf{z})$, it can be shown that

$$\bar{\mathbf{l}} = \tilde{\mathbf{a}}(\mathbf{z}) \tag{41}$$

$$\mathbf{L}^{-1} = \tilde{\mathbf{H}}^T \mathbf{R}^{-1} \tilde{\mathbf{H}} \tag{42}$$

where $\tilde{\mathbf{H}}$ is the Jacobian of $\mathbf{h}(\cdot)$ evaluated at $\tilde{\mathbf{a}}(\mathbf{z})$. The rest of the algorithm would remain unaltered. Under the assumptions stated in this paper, the LGLAF coincides with the method[3] described in [17]. However, using the UT to calculate the first two moments of the likelihood

---

[3]The name of LGLAF in [17] is Gaussian mixture filter new1.

is more accurate than using linearisation [2]. In addition, the adaptive selection between UKF or UGLAF approximation to the posterior, which is explained in Section 4 and has a major effect on performance, was not considered in [17]. The better performance of UGLAF over LGLAF is demonstrated by simulations in Section 5.

4. Adaptive UGLAF

In this section, we propose the adaptive UGLAF (AUGLAF). As proved in Section 2.3, the UGLAF has high performance if the measurement noise covariance $\mathbf{R}$ is low. The KF and its approximations have the opposite behaviour. Their approximations to the true posterior PDF are accurate for large $\mathbf{R}$ and inaccurate for small $\mathbf{R}$ if the measurement function is not linear [9]. Therefore, AUGLAF uses the KF or UGLAF approximation to the posterior in an adaptive fashion depending on the accuracy of the approximations.

The Kullback–Leibler divergence (KLD) is a common way to measure the closeness between two PDFs [10]. We propose a method that estimates the KLD between several approximations to the posterior and the true posterior. The approximation that has the lowest KLD is expected to be the most accurate representation of the posterior. The procedure presented here is general but we use it specifically to choose between the UKF and UGLAF approximations, here denoted as $\hat{q}^0(\cdot) = \mathcal{N}(\cdot; \overline{\mathbf{u}}_0, \mathbf{U}_0)$ and $\hat{q}^1(\cdot)$, respectively.

It is shown in AppendixC that the KLD between the posterior approximation $\hat{q}^i(\cdot)$ and the posterior $q(\cdot)$ is

$$D\left(\hat{q}^i \parallel q\right) = c + D_i^l + D_i^n \tag{43}$$

$$c = \log \int l(\mathbf{a}) p(\mathbf{x}) d\mathbf{x} - \frac{n_x}{2} + \frac{1}{2}\log|\mathbf{\Sigma}|$$
$$+ \log(2\pi)^{n_z/2} + \frac{1}{2}\log|\mathbf{R}| \tag{44}$$

$$D_i^l = -\frac{1}{2}\log|\mathbf{U}_i|$$
$$+ \frac{1}{2}\mathrm{tr}\left[\mathbf{\Sigma}^{-1}\left(\mathbf{U}_i + (\overline{\mathbf{u}}_i - \overline{\mathbf{x}})(\overline{\mathbf{u}}_i - \overline{\mathbf{x}})^T\right)\right] \tag{45}$$

$$D_i^n = \frac{1}{2}\int (\mathbf{z} - \mathbf{h}(\mathbf{a}))^T \mathbf{R}^{-1}(\mathbf{z} - \mathbf{h}(\mathbf{a}))\hat{q}^i(\mathbf{x}) d\mathbf{x} \tag{46}$$

$$= \frac{1}{2}\int \mathrm{tr}\left[(\mathbf{z} - \mathbf{h}(\mathbf{a}))^T \mathbf{R}^{-1}(\mathbf{z} - \mathbf{h}(\mathbf{a}))\right]\hat{q}^i(\mathbf{x}) d\mathbf{x}$$

$$= \frac{1}{2}\mathrm{tr}\left[\mathbf{R}^{-1}\int (\mathbf{z} - \mathbf{h}(\mathbf{a}))(\mathbf{z} - \mathbf{h}(\mathbf{a}))^T \hat{q}^i(\mathbf{x}) d\mathbf{x}\right] \tag{47}$$

There are some important aspects to highlight in (43). First, c does not depend on $\hat{q}^i(\cdot)$. Thus, it does not have to be calculated to select the posterior approximation that minimises the KLD. Second, D_i^l can be calculated analytically so it poses no problem. Third, in general, D_i^n can

Table 2: Steps of the update phase of the AUGLAF

- Calculate the UKF approximation to the posterior, whose moments are $\overline{\mathbf{u}}_0$ and $\mathbf{U}_0$ [2]. - Calculate the UGLAF approximation to the posterior, whose moments are $\overline{\mathbf{u}}_1$ and $\mathbf{U}_1$, as indicated in Table 1. - Calculate D_0^l using (45). - Draw sigma-points $\mathcal{U}_0^1,...,\mathcal{U}_0^{N_s}$ matching the moments $\overline{\mathbf{u}}_0^a$ and $\mathbf{U}_0^a$, see (48) and (49). - Approximate D_0^n using (50). - Calculate D_1^l using (45). - Draw sigma-points $\mathcal{U}_1^1,...,\mathcal{U}_1^{N_s}$ matching the moments $\overline{\mathbf{u}}_1^a$ and $\mathbf{U}_1^a$, see (48) and (49). - Approximate D_1^n using (50) - Calculate the AUGLAF approximation to the posterior, whose moments are $\overline{\mathbf{u}}_2$ and $\mathbf{U}_2$, using (51).

be well approximated by the UT due to the fact that UT approximations of integrals of the type (47) appear in the UKF update step. For usual measurement functions, the UT provides reasonable approximations to these integrals with low computational burden [2]. We want to mention that we do not consider the KLD $D\left(q \parallel \hat{q}^i\right)$ instead of $D\left(\hat{q}^i \parallel q\right)$ because it cannot be computed accurately using the UT as we require the calculation of an integral w.r.t. the true posterior, which is unknown and not Gaussian in general.

As the function to integrate with respect to $\hat{q}^i\left(\cdot\right)$ in (46) only depends on $\mathbf{a}$, we only need to draw sigma points in this variable. If the first two moments are written as

$$\overline{\mathbf{u}}_i = \left[\left(\overline{\mathbf{u}}_i^a\right)^T, \left(\overline{\mathbf{u}}_i^b\right)^T\right]^T \tag{48}$$

$$\mathbf{U}_i = \left[\begin{array}{cc} \mathbf{U}_i^a & \mathbf{U}_i^{ab} \\ \left(\mathbf{U}_i^{ab}\right)^T & \mathbf{U}_i^b \end{array}\right] \tag{49}$$

where $\overline{\mathbf{u}}_i^a \in \mathbb{R}^{n_a}$, $\overline{\mathbf{u}}_i^b \in \mathbb{R}^{n_b}$ and the sigma points $\mathcal{U}_i^1,...,\mathcal{U}_i^{N_s}$ are selected to match the moments $\overline{\mathbf{u}}_i^a$ and $\mathbf{U}_i^a$ [2], the UT approximation to D_i^n is

$$D_i^n \approx \frac{1}{2}\sum_{j=1}^{N_s}\omega^j \left(\mathbf{z} - \mathbf{h}\left(\mathcal{U}_i^j\right)\right)^T \mathbf{R}^{-1}\left(\mathbf{z} - \mathbf{h}\left(\mathcal{U}_i^j\right)\right) \tag{50}$$

In short, the first two moments of the AUGLAF approximation to the posterior are

$$\left(\overline{\mathbf{u}}_2, \mathbf{U}_2\right) = \begin{cases} \left(\overline{\mathbf{u}}_0, \mathbf{U}_0\right) & \text{if } D_0^l + D_0^n < D_1^l + D_1^n \\ \left(\overline{\mathbf{u}}_1, \mathbf{U}_1\right) & \text{otherwise} \end{cases} \tag{51}$$

The steps of the algorithm are given in Table 2. We also want to remark that if the measurement equation is linear and the measurement noise is Gaussian, the UGLAF and AUGLAF provide the first two moments of the posterior exactly.

5. Simulations

In this section, the performances of UGLAF and AUGLAF are analysed in a ballistic target tracking case using range-bearing measurements obtained from a radar as in [2]. We compare them with: UKF, CKF, EKF, TUKF, LGLAF and a sampling importance resampling (SIR) PF [1], in which the importance density is the prior. The prediction step of LGLAF is performed using the UT. All these filters have a low computational burden except the PF.

We use the following parameters for the algorithms. The UKF uses $N_s = 2n_x + 1$ sigma points and the weight of the sigma point located on the mean is $1/3$. The PF uses 10000 particles. The parameter γ of the TUKF is set to 0.1 [16, 15]. The performances of the algorithms are analysed using Monte Carlo simulations with 20000 runs and common random numbers to generate the measurements.

The state vector at time step k is $\mathbf{x}^k = \left[\mathbf{a}^k, \mathbf{b}^k\right]^T$ where $\mathbf{a}^k = \left[p_x^k, p_y^k\right]^T$ is the position vector with respect to the center of the Earth and $\mathbf{b}^k = \left[v_x^k, v_y^k, \gamma^k\right]^T$ includes the velocity vector of the target and ballistic coefficient [2]. The measurement model is:

$$z_r^k = \sqrt{\left(p_x^k - \xi_x\right)^2 + \left(p_y^k - \xi_y\right)^2} + \eta_r \tag{52}$$

$$z_\theta^k = \text{atan2}\left(p_y^k - \xi_y, p_x^k - \xi_x\right) + \eta_\theta \tag{53}$$

where $\text{atan2}\left(\cdot, \cdot\right)$ is the four-quadrant inverse tangent, $\left[\xi_x, \xi_y\right]^T$ is the position vector of the radar, $\mathbf{z}^k = \left[z_r^k, z_\theta^k\right]^T$, 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 considered zero-mean Gaussian distributed and independent. The measurements are taken with a sampling period T. The measurement function is injective so UGLAF can be applied checking condition (37) using (38). That is, sigma points $\mathcal{Y}^j$, $j = 1,...,N_s$ must have the first component (distance) greater than 0 as the function is bijective if the codomain is constrained to distances greater than 0.

The dynamic model is given by a set of nonlinear differential equations [2]. We discretise the differential equations using the Euler approximation as in [2]. Thus, each sampling period is split into J subintervals and the state at the jth subinterval at time step k is denoted as $\mathbf{x}^{k,j} = \left[p_x^{k,j}, p_y^{k,j}, v_x^{k,j}, v_y^{k,j}, \gamma^{k,j}\right]^T$ where $\mathbf{x}^{k,0} = \mathbf{x}^k$ and $\mathbf{x}^{k,J} = \mathbf{x}^{k+1}$. The nonlinear process function that relates $\mathbf{x}^{k,j+1}$ to $\mathbf{x}^{k,j}$ is

$$p_x^{k,j+1} = p_x^{k,j} + \frac{T}{J}v_x^{k,j} \tag{54}$$

$$p_y^{k,j+1} = p_y^{k,j} + \frac{T}{J}v_y^{k,j} \tag{55}$$

$$v_x^{k,j+1} = v_x^{k,j} + \frac{T}{J}\left(D^{k,j}v_x^{k,j} + G^{k,j}p_x^{k,j}\right) \tag{56}$$

$$v_y^{k,j+1} = v_y^{k,j} + \frac{T}{J}\left(D^{k,j}v_y^{k,j} + G^{k,j}p_y^{k,j}\right) \tag{57}$$

7

Table 3: Dynamic model parameters

Parameter	Value
T	5 s
J	100
β_0	0.59783 s^{-1}
H_0	13.406 km
Gm_0	$3.986 \cdot 10^5$ km^3/s^2
R_0	6374 km
q	10^{-6} km^2/s^2

Table 4: Scenario 1 and 2 parameters

	Scenario 1	Scenario 2
Σ_x^0 (km^2)	1600	160
Σ_y^0 (km^2)	300	30
$\Sigma_{v_x}^0$ (km^2/s^2)	0.1	0.1
$\Sigma_{v_y}^0$ (km^2/s^2)	0.01	0.01
Σ_γ^0 (s^{-2})	$8 \cdot 10^{-3}$	$8 \cdot 10^{-3}$
σ_r^2 (km^2)	1	10
σ_θ^2 (rad^2)	$\left(\pi/180 \cdot 10^{-5}\right)^2$	$(20 \cdot \pi/180)^2$

$$\gamma^{k,j+1} = \gamma^{k,j} \tag{58}$$

where

$$D^{k,j} = -\beta_0 \exp\left(\gamma^{k,j} + \frac{R_0 - R^{k,j}}{H_0}\right) V^{k,j}$$

$$R^{k,j} = \sqrt{\left(p_x^{k,j}\right)^2 + \left(p_y^{k,j}\right)^2}$$

$$V^{k,j} = \sqrt{\left(v_x^{k,j}\right)^2 + \left(v_y^{k,j}\right)^2}$$

$$G^{k,j} = -\frac{Gm_0}{\left(R^{k,j}\right)^3}$$

where R_0 is the radius of the Earth and β_0, Gm_0 and H_0 are parameters that reflect environmental and target characteristics [2]. Let $\mathbf{f}(\cdot)$ denote the nonlinear function that relates $\mathbf{x}^{k+1}$ and $\mathbf{x}^k$ using (54)-(58) recursively. Then, the dynamic equation is

$$\mathbf{x}^{k+1} = \mathbf{f}\left(\mathbf{x}^k\right) + \mathbf{v}^{k+1} \tag{59}$$

where we have introduced a process noise $\mathbf{v}^{k+1}$, which is zero-mean Gaussian noise with covariance matrix $\mathbf{Q} = \text{diag}\left([0,0,q,q,0]\right)$, that accounts for random accelerations. The prior at time zero is Gaussian

$$p\left(\mathbf{x}^0\right) = \mathcal{N}\left(\mathbf{x}^0; \overline{\mathbf{x}}^0, \mathbf{\Sigma}^0\right) \tag{60}$$

We consider two scenarios to evaluate the performance of the algorithms. In both scenarios the radar is located at $[\xi_x, \xi_y]^T = [R_0, 0]^T$ and we use the same dynamic model parameters, which are given in Table 3. The number of time steps is 80. The target trajectory is shown in Figure 2. Its initial state and covariance matrix are $\mathbf{x}^0 = [6600\,\text{km}, 15\,\text{km}, 0\,\text{km/s}, 0.2\,\text{km/s}, 0.6932]^T$ and $\mathbf{\Sigma}^0 = \text{diag}\left(\left[\Sigma_x^0, \Sigma_y^0, \Sigma_{v_x}^0, \Sigma_{v_y}^0, \Sigma_\gamma^0\right]\right)$. The measurement model and prior at time 0 parameters for both scenarios are shown in Table 4.

In each Monte Carlo run, $\overline{\mathbf{x}}^0$ is drawn from a Gaussian PDF whose mean is the true state and covariance matrix $\mathbf{\Sigma}^0$. In both scenarios, the EKF performs quite badly and diverges in all the Monte Carlo runs so the EKF is not considered in the rest of the section.

In Scenario 1, the PF does not work well because the measurement noise is low and therefore requires a larger
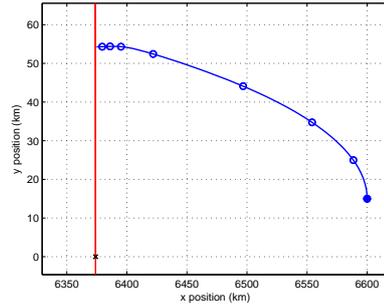

Figure 2: The trajectory of the target is represented as a blue line. The position of the target every ten time steps is represented as a circle. Its initial position is represented as a filled blue circle. The radar position is represented as a black cross. The surface of the Earth is represented as a red line.

sample size. The root mean square (RMS) position errors of the rest of the algorithms for Scenario 1 are shown in Figure 3. The filters with highest performance are UGLAF, AUGLAF, TUKF and LGLAF. The UKF and CKF fare much worse due to the low measurement noise and high prior uncertainty. Here, AUGLAF selects UGLAF with probability one up to time 30. This probability decreases to 0.2 at around time step 50 to slowly increase again to 1, which is reached at time step 70. Nevertheless, the improvement of UGLAF over the UKF approximation in the AUGLAF effectively happens at the initial time steps, in which the prior uncertainty is large.

The execution times in milliseconds of our Matlab implementation of the algorithms are AUGLAF (98), UGLAF (61), LGLAF (56), TUKF (89), UKF (65), CKF (65) and PF (60000). It can be clearly seen that the big improvement of AUGLAF only implies a 50% increase in the computational burden with respect to the UKF.

The RMS position errors for Scenario 2 are shown in Figure 4. LGLAF is the filter with worst performance followed by UGLAF and the TUKF. The filters with highest performance are: AUGLAF, UKF and CKF. The PF works as well as these three filters at the beginning but its performance deteriorates slightly later on. AUGLAF selects the UKF at all time steps. In Scenario 2, the prior uncertainty is reasonably low and measurement noise high. This is the reason why KF-type algorithms work well [9] and the AUGLAF selects the UKF at all time steps.

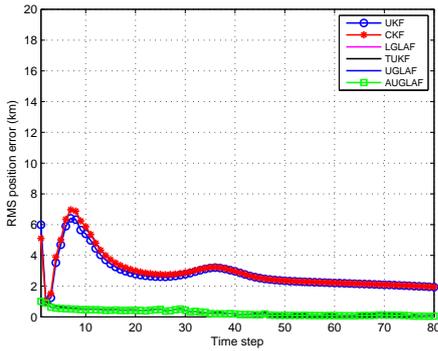To sum up, these two examples show the importance

Figure 3: Scenario 1 RMS position error: The lines of UGLAF, AUGLAF, LGLAF and TUKF are indistinguishable and they obtain highest performance among the filters.
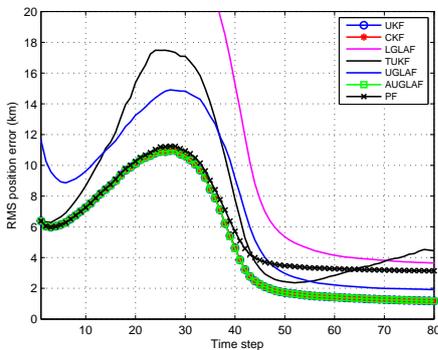

Figure 4: Scenario 2 RMS error in position: The lines of AUGLAF, UKF and CKF are indistinguishable. These filters have the highest performance.

of using an adaptive rule to select between the UKF and UGLAF approximations. As mentioned before, these filters perform well in different situations so AUGLAF selects the best one according to the KLD. The adaptive rule enables AUGLAF to be the algorithm with highest performance in both scenarios.

6. Conclusions

We have proposed the UGLAF, which is based on approximating the normalised likelihood by a Gaussian PDF using the unscented transformation. An important property of UGLAF is that it is asymptotically optimal as the measurement tends to zero. As a result, the direct application of UGLAF is expected to outperform the KF and its approximations, such as the EKF, UKF and CKF, if the measurement noise is low enough and the measurement function is nonlinear.

In order to provide good performance in a wider variety of situations, we have also devised the AUGLAF, which chooses the best approximation to the posterior between UKF or UGLAF. This selection algorithm follows a principled approach based on the KLD and can be applied

whenever there are several approximations to the posterior.

The drawback of the developed methods is that they require the measurement function to be bijective or injective. Nevertheless, more complex measurement models can also be handled following the same principles. In this respect, an interesting line of future research is to provide the generalisation of AUGLAF to multimodal likelihoods. In principle, this can be done following similar ideas as in the mixture TUKF [15].

Acknowledgements

AppendixA.

Let us write the measurement noise covariance matrix as $\mathbf{R} = \tilde{\mathbf{R}}/\tilde{n}$. In this appendix, we prove that Approximation AP1 is increasingly accurate as $\tilde{n} \to \infty$. The UGLAF approximation to the normalised likelihood is denoted as

$$\hat{l}\left(\mathbf{a}\right) = \mathcal{N}\left(\mathbf{a}; \bar{\mathbf{l}}_{i,\tilde{n}}, \mathbf{L}_{i,\tilde{n}}\right) \tag{A.1}$$

where $i = 1$ if we use (16) and (17), and $i = 2$ if we use (19) and (20). In the following, we show that both approximations are alike as $\tilde{n} \to \infty$.

With AP1

We consider the Taylor expansion of the inverse of the measurement function around $\mathbf{z}$

$$\mathbf{h}^{-1}\left(\mathcal{Y}^j\right) = \mathbf{a}^\star + \mathbf{H}_I\left(\mathcal{Y}^j - \mathbf{z}\right) + \mathbf{J}\left(\mathcal{Y}^j\right) \tag{A.2}$$

where $\mathbf{a}^\star = \mathbf{h}^{-1}\left(\mathbf{z}\right)$, $\mathbf{H}_I$ is the Jacobian of $\mathbf{h}^{-1}\left(\cdot\right)$ evaluated at $\mathbf{z}$, and $\mathbf{J}\left(\mathcal{Y}^j\right)$ is the remainder. We recall that $\mathbf{R} = \tilde{\mathbf{R}}/\tilde{n}$ and sigma points $\mathcal{Y}^j$ $j = 1, ..., N_s$ match the moments $\mathbf{z}$ and $\mathbf{R}$. Therefore, according to how sigma points are selected [2], $\mathcal{Y}^j - \mathbf{z}$ is $O\left(\tilde{n}^{-1/2}\right)$ and $\mathbf{J}\left(\mathcal{Y}^j\right)$ is $O\left(\tilde{n}^{-1}\right)$. Using (A.2) in (18) and (19), we get

$$\bar{\mathbf{l}}_{2,\tilde{n}} = \sum_{j=1}^{N_s} \omega^j \left[\mathbf{a}^\star + \mathbf{H}_I\left(\mathcal{Y}^j - \mathbf{z}\right) + \mathbf{J}\left(\mathcal{Y}^j\right)\right]$$
$$= \mathbf{a}^\star + O\left(\tilde{n}^{-1}\right) \tag{A.3}$$

as $\sum_{j=1}^{N_s} \omega^j \mathcal{Y}^j = \mathbf{z}$. Using a similar derivation for (20), we have

$$\mathbf{L}_{2,\tilde{n}} = \tilde{n}^{-1}\tilde{\mathbf{L}} + O\left(\tilde{n}^{-3/2}\right) \tag{A.4}$$

where $\tilde{\mathbf{L}} = \mathbf{H}_I\tilde{\mathbf{R}}\mathbf{H}_I^T$.

9

Without AP1

We consider the Taylor expansion of the determinant of the Jacobian of the measurement function around $\mathbf{z}$

$$\left|\det \mathbf{h}'_I\left(\mathcal{Y}^j\right)\right| = \left|\det \mathbf{h}'_I\left(\mathbf{z}\right)\right| + \boldsymbol{\Theta}\left(\mathcal{Y}^j - \mathbf{z}\right) + O\left(\tilde{n}^{-1}\right) \tag{A.5}$$

where $\boldsymbol{\Theta}$ is the Jacobian of $\left|\det \mathbf{h}'_I\left(\cdot\right)\right|$ evaluated at $\mathbf{z}$.

Substituting (A.2) and (A.5) into (16), we have

$$\begin{aligned}\bar{\mathbf{l}}_{1,n} &= \frac{\mathbf{a}^\star \left|\det \mathbf{h}'_I\left(\mathbf{z}\right)\right| + O\left(\tilde{n}^{-1}\right)}{\left|\det \mathbf{h}'_I\left(\mathbf{z}\right)\right| + O\left(\tilde{n}^{-1}\right)}\\ &= \mathbf{a}^\star + O\left(\tilde{n}^{-1}\right)\end{aligned} \tag{A.6}$$

where we have used that $\mathcal{Y}^j - \mathbf{z}$ is $O\left(\tilde{n}^{-1/2}\right)$ and (B.3) in AppendixB.

Substituting (A.2) and (A.5) into (17), we have

$$\begin{aligned}\mathbf{L}_{1,\tilde{n}} &= \frac{\tilde{n}^{-1}\mathbf{H}_I\tilde{\mathbf{R}}\mathbf{H}_I^T\left|\det \mathbf{h}'_I\left(\mathbf{z}\right)\right| + O\left(\tilde{n}^{-3/2}\right)}{\left|\det \mathbf{h}'_I\left(\mathbf{z}\right)\right| + O\left(\tilde{n}^{-1}\right)}\\ &= \tilde{n}^{-1}\tilde{\mathbf{L}} + O\left(\tilde{n}^{-3/2}\right)\end{aligned} \tag{A.7}$$

Equations (A.3), (A.4), (A.6) and (A.7) prove that the effect of Approximation AP1 becomes negligible for a low enough measurement noise.

AppendixB.

In this appendix, we include some relationships used in the derivation of UGLAF optimality. From matrices $\mathbf{P}$ and $\mathbf{C}$ with $\left|\mathbf{P}\right| \neq 0$ it can be shown that [18]

$$\begin{aligned}\left|\mathbf{P} + \mathbf{C}/\sqrt{\tilde{n}}\right| &= \left|\mathbf{P}\right|\left(1 + \text{tr}\left(\mathbf{C}\mathbf{P}^{-1}\right)/\sqrt{\tilde{n}}\right) + O\left(\tilde{n}^{-1}\right)\\ &= \left|\mathbf{P}\right| + O\left(\tilde{n}^{-1/2}\right)\end{aligned} \tag{B.1}$$

For any $b \in \mathbb{R}$, $c \in \mathbb{R}$ and $\tilde{n} \in \mathbb{N}$, using a Taylor series expansion about b we have

$$\begin{aligned}\exp\left(b + c/\sqrt{\tilde{n}}\right) &= \exp\left(b\right)\left[1 + c/\sqrt{\tilde{n}} + O\left(\tilde{n}^{-1}\right)\right]\\ &= \exp\left(b\right)\left[1 + O\left(\tilde{n}^{-1/2}\right)\right]\end{aligned} \tag{B.2}$$

$$\begin{aligned}\left(b + c/\sqrt{\tilde{n}}\right)^{-1} &= b^{-1} - cb^{-2}/\sqrt{\tilde{n}} + O\left(\tilde{n}^{-1}\right)\\ &= b^{-1} + O\left(\tilde{n}^{-1/2}\right)\end{aligned} \tag{B.3}$$

AppendixC.

The KLD between the posterior approximation $\hat{q}^i\left(\cdot\right) = \mathcal{N}\left(\cdot; \bar{\mathbf{u}}_i, \mathbf{U}_i\right)$ and the true posterior $q\left(\cdot\right)$ is derived. We make use of (4).

$$D\left(\hat{q}^i \parallel q\right) = \int \hat{q}^i\left(\mathbf{x}\right)\log \frac{\hat{q}^i\left(\mathbf{x}\right)}{q\left(\mathbf{x}\right)}d\mathbf{x}$$

$$\begin{aligned}&= \log\int l\left(\mathbf{a}\right)p\left(\mathbf{x}\right)d\mathbf{x} + \int \hat{q}^i\left(\mathbf{x}\right)\left[\log\mathcal{N}\left(\mathbf{x}; \bar{\mathbf{u}}_i, \mathbf{U}_i\right)\right.\\ &\quad \left.- \log l\left(\mathbf{a}\right) - \log p\left(\mathbf{x}\right)\right]d\mathbf{x}\\ &= \log\int l\left(\mathbf{a}\right)p\left(\mathbf{x}\right)d\mathbf{x} - \frac{n_x}{2} - \frac{1}{2}\log\left|\mathbf{U}_i\right| + \frac{1}{2}\log\left|\boldsymbol{\Sigma}\right|\\ &\quad + \frac{1}{2}\text{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\mathbf{U}_i + \left(\bar{\mathbf{u}}_i - \bar{\mathbf{x}}\right)\left(\bar{\mathbf{u}}_i - \bar{\mathbf{x}}\right)^T\right)\right]\\ &\quad + \log\left(2\pi\right)^{n_z/2} + \frac{1}{2}\log\left|\mathbf{R}\right|\\ &\quad + \frac{1}{2}\int\left(\mathbf{z} - \mathbf{h}\left(\mathbf{a}\right)\right)^T\mathbf{R}^{-1}\left(\mathbf{z} - \mathbf{h}\left(\mathbf{a}\right)\right)\hat{q}^i\left(\mathbf{x}\right)d\mathbf{x}\end{aligned}$$

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

[2] 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.

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

[4] R. J. Elliott and S. Haykin, "A Zakai equation derivation of the extended Kalman filter," *Automatica*, vol. 46, no. 3, pp. 620–624, 2010.

[5] J. Dunik, M. Simandl, and O. Straka, "Unscented Kalman filter: Aspects and adaptive setting of scaling parameter," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2411–2416, Sept. 2012.

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

[7] B. Jia, M. Xin, and Y. Cheng, "High-degree cubature Kalman filter," *Automatica*, vol. 49, no. 2, pp. 510–518, 2013.

[8] B. Bell and F. Cathey, "The iterated Kalman filter update as a Gauss-Newton method," *IEEE Transactions on Automatic Control*, vol. 38, no. 2, pp. 294–297, Feb. 1993.

[9] M. R. Morelande and A. F. García-Fernández, "Analysis of Kalman filter approximations for nonlinear measurements," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5477–5484, Nov. 2013.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer Science + Business Media, 2006.

[11] A. F. García-Fernández, M. R. Morelande, and J. Grajal, "Truncated unscented Kalman filtering," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3372–3386, July 2012.

[12] Y. Ho and R. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 333–339, Oct. 1964.

[13] N. Reid, "Likelihood and Bayesian approximations methods," in *Bayesian Statistics 5*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Eds. Oxford University Press, 1995.

[14] A. M. Walker, "On the asymptotic behaviour of posterior distributions," *Journal of the Royal Statistical Society. Series B (Methodological).*, vol. 31, no. 1, pp. 80–88, 1969.

[15] A. F. García-Fernández, M. R. Morelande, and J. Grajal, "Mixture truncated unscented Kalman filtering," in *15th International Conference on Information Fusion*, 2012, pp. 479–486.

[16] 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/

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

[18] M. Fiedler, "Bounds for the determinant of the sum of Hermitian matrices," *Proceedings of the American Mathematical Society*, vol. 30, no. 1, pp. 27–31, Sep. 1971.