



HAL
open science

A Gaussian Uniform Mixture Model for Robust Kalman Filtering

Mathieu Brunot

► **To cite this version:**

Mathieu Brunot. A Gaussian Uniform Mixture Model for Robust Kalman Filtering. *IEEE Transactions on Aerospace and Electronic Systems*, 2019, 14 (8), pp.1-9. 10.1109/TAES.2019.2953414 . hal-02475697

HAL Id: hal-02475697

<https://hal.science/hal-02475697>

Submitted on 12 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Gaussian Uniform Mixture Model for Robust Kalman Filtering

Mathieu Brunot

Abstract—This paper presents a Kalman-type recursive estimator for discrete time systems with a measurement noise modeled by a Gaussian-uniform mixture. The objective is to deal with data containing outliers that degrade the performances of the regular Kalman filter. The introduced non-Gaussian noise model takes into account the reliability of the measurement to be robust with respect to erroneous data. The Kalman-type estimator is based on the Masreliez's formulation that copes with non-Gaussian noise models. Results in different simulated conditions are displayed to evaluate the performances of the introduced algorithm and to compare it to state-of-art alternatives.

Index Terms—Kalman filters, Mixture models, Noise robustness, Bayesian estimation.

I. INTRODUCTION

SINCE the seminal work of R. Kalman and R. Bucy [1]–[2], the estimation theory has received a large attention from researchers. Consequently, a great number of algorithms have been developed for prediction, filtering and smoothing, for either linear or nonlinear systems. Those algorithms can also be categorized in terms of noise assumption. If the regular Kalman Filter (KF) assumes a Gaussian noise model, many physical systems deviate from this assumption and follow distributions with heavier tails leading to high intensity noise realizations known as outliers.

Many robust approaches have been suggested to deal with non-Gaussian measurements. An intuitive approach is to use a regular KF by adding a detection of the outliers to reduce their weights in the corresponding correction terms [3]–[4]. Several methods directly deal with non-Gaussian distributions. One of them consists in using the Student's t -distribution and performing the statistical inference with an Expectation-Maximization iteration [5]. That has been extended to sequential filtering and smoothing processes [6]. Other works consider the variational Bayes approximation to estimate non-Gaussian posterior distribution [7]–[8]. Without being exhaustive, one can also cite the Particle Filter (PF) and Gaussian Mixture Filter (GMF). Unlike the previously cited methods, they do not assume that the state distribution is Gaussian after each time step. The PF utilizes a set of weighted points, *i.e.* particles, to approximate the distributions [9]. The GMF approximates of the noise distribution by a finite sum of Gaussian terms [10]. The drawback of those two methods is that number of particles or Gaussian components have to be large enough to insure a good estimation accuracy, which may lead to unacceptable computing time.

In the case of model uncertainties, coming from model errors, unknown input process or noise properties, the risk-sensitive filter also represents an interesting alternative. Contrary to the regular KF, it minimizes an exponential cost function, tuned by the so called risk-sensitivity parameter; see *e.g.* [11]. This parameter can be constant or time-varying. Recent works have extended that filter to a family of robust KF parametrized by the τ -divergence family [12]–[13]. In [14], the authors have highlighted the link between the risk-sensitive filter and the \mathcal{H}_∞ filtering algorithm that may also be appealing. It indeed makes no statistical assumptions on the noise, except that it has finite energy [15]–[16].

Grandin and Marques [17] introduced a link between the theory of Probability and the theory of Evidence. The introduction of the reliability of the source in the measurement model led to a robust fusion process. In the specific case where the reliability is not a random variable but a known constant probability, the model developed in [17] becomes a Gaussian-Uniform (GU) mixture model. In [18], the authors have shown that it is a worthwhile alternative to Gaussian mixture in case of outlier data. Such a modelling has been considered with PF for multiple target tracking in [19] where the uniform distribution represents "clutter" or false alarm measurements. This PF method is referred to as Rao-Blackwellized Monte Carlo Data Association (MCDA) and was designed for multiple target tracking. Few years before, the Probabilistic Data Association Filter (PDAF) have been conceived with the same purpose and noise assumptions [20].

In this paper, we consider the GU measurement noise to provide a robust Kalman-type estimator. Masreliez [21] developed a generic filter formulation to deal with non-Gaussian observation noise. Although the formulation has the reputation for being complex [4]–[22], our goal is to present the calculation of this filter and to illustrate its interest with two examples.

The paper is organized as follows. The formal definition of the GU noise model and Masreliez filter are first given in Section II. Section III is devoted to the development of the filter equations by considering the GU mixture for the measurement noise. A simple test case is firstly considered in Section IV to illustrate the filter characteristics while Section V tackles a more involved multidimensional case. Finally, concluding remarks are provided in Section VI.

Notations

General notations

a, b, c = Scalars

$\mathbf{a}, \mathbf{b}, \mathbf{c}$ = Vectors

M. Brunot is with ONERA, Toulouse, France e-mail: mathieu.brunot@onera.fr.

$\mathbf{A}, \mathbf{B}, \mathbf{C}$ = Matrices

Conventions

\mathbf{A}^{-1} = Inverse of matrix \mathbf{A}

\mathbf{A}^T = Transpose of matrix \mathbf{A}

$|\mathbf{A}|$ = Determinant of matrix \mathbf{A}

$\hat{\mathbf{x}}$ = Estimated vector

$x_{i,k}$ = i^{th} component of vector \mathbf{x}_k

$[\mathbf{A}]_i$ = i^{th} row of matrix \mathbf{A}

$[\mathbf{A}]_{ij}$ = Element at i^{th} row and j^{th} column of \mathbf{A}

$\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}}$ = Jacobian matrix of function \mathbf{g}

II. PRELIMINARIES

A. Mixture noise model

To obtain a robust Kalman filter, a non-Gaussian measurement noise is studied. The principle is to take into account outliers by assuming a noise distribution with heavy tails. In one dimension, the following distribution is considered:

$$\mathcal{G}(\theta, \mu, \sigma, a) = F(\theta)\mathcal{N}(\mu, \sigma) + (1 - F(\theta))\mathcal{U}(\mu, a) \quad (1)$$

where F is the probability of a correct measurement parametrized by θ , \mathcal{N} is a Gaussian distribution (mean μ and standard deviation σ) and \mathcal{U} is a centered uniform law (mean μ and half span a). As explained in [17], the idea of such a model is that the Gaussian term expresses the precision of the instrument, whereas $F(\theta)$ assesses if the instrument senses what it is supposed to measure, *i.e.* the reliability. In case of invalid measurement, a uniform distribution is considered to model the error although a more complex modelling could be applied for some sensors. In this article, the author limits himself to the case where the instrument reliability is known in advance to reduce the complexity of the model. In the sequel, the reliability is a constant known parameter θ . The resulting Probability Density Function (p.d.f.) is a one dimensional GU mixture:

$$f_{\mathcal{G}}^{1D}(n) = \theta f_{\mathcal{N}}^{1D}(n) + (1 - \theta) f_{\mathcal{U}}^{1D}(n), \quad (2)$$

where

$$f_{\mathcal{N}}^{1D}(n) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{n - \mu}{\sigma}\right)^2\right\}$$

$$f_{\mathcal{U}}^{1D}(n) = \begin{cases} 0 & n \leq \mu - a \\ \frac{1}{2a} & \mu - a \leq n \leq \mu + a \\ 0 & \mu + a \leq n \end{cases}$$

In m dimensions, a multivariate distribution could be defined as follows

$$f_{\mathcal{G}}^m(\mathbf{n}) = \prod_{i=1}^m f_{\mathcal{G}_i}^{1D}(n_i)$$

$$= \prod_{i=1}^m [\theta_i f_{\mathcal{N}_i}^{1D}(n_i) + (1 - \theta_i) f_{\mathcal{U}_i}^{1D}(n_i)].$$

Nevertheless, such a distribution assumes that the m components are independent which may be a too strong assumption in practice. Thus, we define the m dimensional probability distribution \mathcal{G} :

$$\mathcal{G}(\theta, \boldsymbol{\mu}, \mathbf{R}, \mathbf{a}) = F(\theta)\mathcal{N}(\boldsymbol{\mu}, \mathbf{R}) + (1 - F(\theta))\mathcal{U}(\boldsymbol{\mu}, \mathbf{a}) \quad (3)$$

Therefore, the p.d.f. is given by

$$f_{\mathcal{G}}(\mathbf{n}) = \theta f_{\mathcal{N}}(\mathbf{n}) + (1 - \theta) f_{\mathcal{U}}(\mathbf{n}), \quad (4)$$

with

$$f_{\mathcal{N}}(\mathbf{n}) = \frac{1}{(2\pi)^{m/2} |\mathbf{R}|} \exp\left\{-\frac{1}{2}(\mathbf{n} - \boldsymbol{\mu})^T \mathbf{R}^{-1}(\mathbf{n} - \boldsymbol{\mu})\right\}$$

$$f_{\mathcal{U}}(\mathbf{n}) = \begin{cases} \frac{1}{A} & \mathbf{n} \in P(\boldsymbol{\mu}, \mathbf{a}) \\ 0 & \text{otherwise} \end{cases}$$

For the uniform part, $P(\boldsymbol{\mu}, \mathbf{a})$ is the m dimensional parallelepiped centered in $\boldsymbol{\mu}$ with faces defined by \mathbf{a} and volume A . By assuming that P is a rectangular parallelepiped:

$$A = \int_{\mu_1 - a_1}^{\mu_1 + a_1} \dots \int_{\mu_m - a_m}^{\mu_m + a_m} dx_1 \dots dx_m = 2^m \prod_{i=1}^m a_i. \quad (5)$$

With (4), the covariance matrix \mathbf{R} enables the model to take into account correlated measurements. The price to pay is that the probability of correct measurement $F(\theta)$ is the same for each component. However, each component can have a different tail for outliers incorporation by selecting appropriate a_i coefficients.

B. Masreliez filter

Like in [21], in this article, the dynamic system considered are modelled by a linear discrete-time state-space defined as follows

$$\mathbf{x}_{k+1} = \boldsymbol{\phi}_k \mathbf{x}_k + \mathbf{q}_k \quad (6)$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k \quad (7)$$

The plant noise \mathbf{q}_k is assumed white and Gaussian with covariance matrix \mathbf{Q}_k . In the following, $p_Z(\mathbf{z}_k | \mathbf{Z}^{k-1})$ denotes the observation prediction density, *i.e.* the density of the observation \mathbf{z}_k conditioned on prior observations $\mathbf{Z}^{k-1} = \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}$. The predicted state is assumed to be Gaussian with p.d.f. $p_X(\mathbf{x}_k | \mathbf{Z}^{k-1})$. According to [21], this assumption should be satisfied in many practical situations if the state noise is Gaussian or if the state prediction variance is small with respect to the observation variance. Its mean and covariance are given by

$$\bar{\mathbf{x}}_k \triangleq \mathbb{E}[(\mathbf{x}_k | \mathbf{Z}^{k-1})] = \boldsymbol{\phi}_{k-1} \hat{\mathbf{x}}_{k-1} \quad (8)$$

$$\mathbf{M}_k \triangleq \mathbb{E}[(\bar{\mathbf{x}}_k - \mathbf{x}_k)(\bar{\mathbf{x}}_k - \mathbf{x}_k)^T | \mathbf{Z}^{k-1}]$$

$$= \boldsymbol{\phi}_{k-1} \mathbf{P}_{k-1} \boldsymbol{\phi}_{k-1}^T + \mathbf{Q}_{k-1}$$

where $\hat{\mathbf{x}}_{k-1} \triangleq \mathbb{E}[(\mathbf{x}_{k-1} | \mathbf{Z}^{k-1})]$ is the estimated state vector at time $k-1$ and \mathbf{P}_{k-1} is its conditional covariance:

$$\mathbf{P}_{k-1} = \mathbb{E}[(\hat{\mathbf{x}}_{k-1} - \mathbf{x}_{k-1})(\hat{\mathbf{x}}_{k-1} - \mathbf{x}_{k-1})^T | \mathbf{Z}^{k-1}]. \quad (10)$$

The correction step satisfies:

$$\hat{\mathbf{x}}_k = \bar{\mathbf{x}}_k + \mathbf{M}_k \mathbf{H}_k^T \mathbf{g}_k(\mathbf{z}_k) \quad (11)$$

$$\mathbf{P}_k = \mathbf{M}_k - \mathbf{M}_k \mathbf{H}_k^T \mathbf{G}_k(\mathbf{z}_k) \mathbf{H}_k \mathbf{M}_k \quad (12)$$

$$\mathbf{M}_{k+1} = \boldsymbol{\phi}_k \mathbf{P}_k \boldsymbol{\phi}_k^T + \mathbf{Q}_k \quad (13)$$

where $\mathbf{g}_k(\mathbf{z}_k)$ is a vector valued function given by

$$g_{i,k}(\mathbf{z}_k) = - \left[\frac{\partial p_Z(\mathbf{z}_k | \mathbf{Z}^{k-1})}{\partial z_{i,k}} \right] [p_Z(\mathbf{z}_k | \mathbf{Z}^{k-1})]^{-1}. \quad (14)$$

That function is called the *score* function in [21]. The matrix $\mathbf{G}_k(\mathbf{z}_k)$ is such as

$$[\mathbf{G}_k(\mathbf{z}_k)]_{ij} = \frac{\partial g_{i,k}(\mathbf{z}_k)}{\partial z_{j,k}} \quad (15)$$

In our case, we assume that $\mathbf{n}_k \sim \mathcal{G}(\theta_k, \mathbf{0}, \mathbf{R}_k, \mathbf{a}_k)$ without loss of generality. The mean value $\boldsymbol{\mu}$ could indeed be nonzero and estimated in an augmented state vector as a measurement bias. The next section is devoted to the development of \mathbf{g}_k and \mathbf{G}_k in closed form.

III. ROBUST KALMAN FILTERING

A. Observation prediction density and its derivative

The objective is to obtain the expression of the observation prediction density $p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1})$. From (7), as the sum of two random variables: $\mathbf{H}_k \mathbf{x}_k$ and \mathbf{n}_k , the p.d.f. of \mathbf{z}_k is the convolution of their two p.d.f.s:

$$p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1}) = \int_{\mathbb{R}^m} f_{X_k|\mathbf{Z}^{k-1}}(\mathbf{z}_k - \mathbf{t}|\mathbf{Z}^{k-1}) f_{\mathcal{G}}(\mathbf{t}|\mathbf{Z}^{k-1}) dt$$

where the p.d.f. of the first term $f_{X_k|\mathbf{Z}^{k-1}}(\mathbf{t}|\mathbf{Z}^{k-1})$ is Gaussian with mean $\mathbf{H}_k \bar{\mathbf{x}}_k$ and covariance $\mathbf{L}_k = \mathbf{H}_k \mathbf{M}_k \mathbf{H}_k^T$. The second p.d.f. comes from our assumptions in the previous section. Therefore, it comes:

$$\begin{aligned} p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1}) &= \int_{\mathbb{R}^m} \left[(2\pi)^{m/2} |\mathbf{L}_k|^{1/2} \right]^{-1} \exp\{T_1^k(\mathbf{z}_k, \mathbf{t})\} \cdot \\ &\left[\theta_k \left[(2\pi)^{m/2} |\mathbf{R}_k|^{1/2} \right]^{-1} \exp\{T_2^k(\mathbf{t})\} + (1 - \theta_k) f_{\mathcal{U}}(\mathbf{t}) \right] dt \\ &= \theta_k \left[(2\pi)^{m/2} |\mathbf{S}_k|^{1/2} \right]^{-1} \exp\{T_3^k(\mathbf{z}_k)\} \\ &+ \frac{1 - \theta_k}{A_k} \int_{-\mathbf{a}_k}^{\mathbf{a}_k} \left[(2\pi)^{m/2} |\mathbf{L}_k|^{1/2} \right]^{-1} \exp\{T_1^k(\mathbf{z}_k, \mathbf{t})\} dt \\ &= \theta_k f_k^1 + \frac{1 - \theta_k}{A_k} f_k^2 \end{aligned} \quad (16)$$

where the terms in exponential are

$$\begin{aligned} T_1^k(\mathbf{z}_k, \mathbf{t}) &= -\frac{1}{2} [\boldsymbol{\epsilon}_k - \mathbf{t}]^T \mathbf{L}_k^{-1} [\boldsymbol{\epsilon}_k - \mathbf{t}] \\ T_2^k(\mathbf{t}) &= -\frac{1}{2} \mathbf{t}^T \mathbf{R}_k^{-1} \mathbf{t} \\ T_3^k(\mathbf{z}_k) &= -\frac{1}{2} \boldsymbol{\epsilon}_k^T \mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k \end{aligned}$$

with $A_k = 2^m \prod_{l=1}^m a_{l,k}$, $\mathbf{S}_k = \mathbf{H}_k \mathbf{M}_k \mathbf{H}_k^T + \mathbf{R}_k$ and $\boldsymbol{\epsilon}_k = \mathbf{z}_k - \mathbf{H}_k \bar{\mathbf{x}}_k$. The first term f_k^1 is a Gaussian p.d.f. and given by

$$f_k^1 = \left[(2\pi)^{m/2} |\mathbf{S}_k|^{1/2} \right]^{-1} \exp\{T_3^k(\mathbf{z}_k)\} \quad (17)$$

The second term f_k^2 is a little more complex:

$$\begin{aligned} f_k^2 &= \left[(2\pi)^{m/2} |\mathbf{L}_k|^{1/2} \right]^{-1} \int_{-\mathbf{a}_k}^{\mathbf{a}_k} \exp\{T_1^k(\mathbf{z}_k, \mathbf{t})\} dt \\ &= \left[(2\pi)^{m/2} |\mathbf{L}_k|^{1/2} \right]^{-1} \cdot J_k \end{aligned} \quad (18)$$

The term J_k is developed in Appendix A. It can be expressed as $J_k = \prod_{l=1}^m I_{l,k}$ and its derivative is $\frac{\partial J_k}{\partial z_{j,k}} = J_k \cdot \gamma_{j,k}$, with the intermediate variables $I_{l,k}$ and $\gamma_{j,k}$ defined by (27) and (31) respectively.

The calculation of the derivative is divided into two terms such as

$$\begin{aligned} \frac{\partial p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1})}{\partial z_{i,k}} &= \theta_k \frac{\partial f_k^1}{\partial z_{i,k}} + \frac{(1 - \theta_k)}{A_k} \frac{\partial f_k^2}{\partial z_{i,k}} \\ &= \theta_k f_k^{1,i} + \frac{(1 - \theta_k)}{A_k} f_k^{2,i} \end{aligned} \quad (19)$$

The first element of the derivative is:

$$\begin{aligned} f_k^{1,i} &= \left[(2\pi)^{m/2} |\mathbf{S}_k|^{1/2} \right]^{-1} \exp\{T_3^k(\mathbf{z}_k)\} \frac{\partial T_3^k(\mathbf{z}_k)}{\partial z_{i,k}} \\ &= -f_k^1 [\mathbf{S}_k^{-1} [\mathbf{z}_k - \mathbf{H}_k \bar{\mathbf{x}}_k]]_i. \end{aligned} \quad (20)$$

Using the results of Appendix A, the second element of the derivative is:

$$\begin{aligned} f_k^{2,i} &= \left[(2\pi)^{m/2} |\mathbf{L}_k|^{1/2} \right]^{-1} \frac{\partial J_k}{\partial z_{i,k}} \\ &= \left[(2\pi)^{m/2} |\mathbf{L}_k|^{1/2} \right]^{-1} J_k \cdot \gamma_{i,k} = f_k^2 \cdot \gamma_{i,k}. \end{aligned} \quad (21)$$

In this part, the expression of $p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1})$ and its derivative have been developed in closed-form so that the score function can be expressed.

B. Score function and its derivative

From the previous derivations, the score function comes out:

$$\begin{aligned} \mathbf{g}_{i,k}(\mathbf{z}_k) &= - \left[\frac{\partial p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1})}{\partial z_{i,k}} \right] [p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1})]^{-1} \\ &= - \frac{\theta_k f_k^{1,i} + \frac{1 - \theta_k}{A_k} f_k^{2,i}}{\theta_k f_k^1 + \frac{1 - \theta_k}{A_k} f_k^2} \\ &= \frac{[\theta_k f_k^1 \mathbf{S}_k^{-1} [\mathbf{z}_k - \mathbf{H}_k \bar{\mathbf{x}}_k]]_i - \frac{1 - \theta_k}{A_k} f_k^2 \gamma_{i,k}}{\theta_k f_k^1 + \frac{1 - \theta_k}{A_k} f_k^2} \\ &= \frac{\theta_k f_k^1 [\mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k]_i - \frac{1 - \theta_k}{A_k} f_k^2 \gamma_{i,k}}{p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1})}. \end{aligned} \quad (22)$$

Then, the matrix \mathbf{G}_k is expressed:

$$\begin{aligned} [\mathbf{G}_k(\mathbf{z}_k)]_{ij} &= \frac{\partial g_{i,k}(\mathbf{z}_k)}{\partial z_{j,k}} = \frac{\partial}{\partial z_{j,k}} \left(\frac{\theta_k f_k^{1,i} + \frac{1 - \theta_k}{A_k} f_k^{2,i}}{p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1})} \right) \\ &= \frac{\partial \left(\theta_k f_k^1 [\mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k]_i - \frac{1 - \theta_k}{A_k} f_k^2 \gamma_{i,k} \right)}{\partial z_{j,k}} \frac{1}{p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1})} \\ &\quad - \frac{\theta_k f_k^{1,i} + \frac{1 - \theta_k}{A_k} f_k^{2,i}}{p_Z^2(\mathbf{z}_k|\mathbf{Z}^{k-1})} \frac{\partial p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1})}{\partial z_{j,k}} \\ &= \frac{1}{p_Z(\mathbf{z}_k|\mathbf{Z}^{k-1})} \left(\theta_k \frac{\partial f_k^1 [\mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k]_i}{\partial z_{j,k}} - \frac{1 - \theta_k}{A_k} \frac{\partial f_k^2 \gamma_{i,k}}{\partial z_{j,k}} \right) \\ &\quad + g_{i,k}(\mathbf{z}_k) g_{j,k}(\mathbf{z}_k) \end{aligned} \quad (23)$$

where, the remaining terms to be estimated are

$$\begin{aligned} \frac{\partial f_k^1 [\mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k]_i}{\partial z_{j,k}} &= f_k^{1,j} [\mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k]_i + f_k^1 [\mathbf{S}_k^{-1}]_{ij} \\ &= f_k^1 \left([\mathbf{S}_k^{-1}]_{ij} - [\mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k]_i [\mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k]_j \right) \\ \frac{\partial f_k^2 \gamma_{i,k}}{\partial z_{j,k}} &= f_k^{2,j} \gamma_{i,k} + f_k^2 \frac{\partial \gamma_{i,k}}{\partial z_{j,k}} \\ &= f_k^2 \left(\gamma_{j,k} \gamma_{i,k} + \frac{\partial \gamma_{i,k}}{\partial z_{j,k}} \right) \end{aligned}$$

C. Corner cases

If the previous development is correct in theory, numerical issues may appear in practice in three specific cases, referred to as "Corner Cases" (CCs). Those CCs are:

- 1) the reliability coefficient is set to one;
- 2) the Gaussian component, f_k^1 , is too weak;
- 3) both components, f_k^1 and f_k^2 , are too weak.

The first CC would appear in the case where the designer sets $\theta_k = 1$. In this case, the filter is equivalent to the regular KF and it is more efficient to apply: $\mathbf{g}_k(\mathbf{z}_k) \leftarrow \mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k$ and $\mathbf{G}_k(\mathbf{z}_k) \leftarrow \mathbf{S}_k^{-1}$, setting $\theta_k = 1$ in (22) and (23).

The second CC appears in the case where the observation is an outlier for the Gaussian term but explained by the uniform distribution. With such a configuration, the p.d.f. of the Gaussian component may be too low with respect to the floating-point relative accuracy leading to numerical instability. In fact, if the Gaussian component f_k^1 is neglected, (22) and (23) can be reduced to $\mathbf{g}_k(\mathbf{z}_k) \leftarrow \gamma_k$ and $\mathbf{G}_k(\mathbf{z}_k) \leftarrow \frac{\partial \gamma_k}{\partial \mathbf{z}_k}$. The challenge is now to define when f_k^1 can be neglected. As long as the measurement error \mathbf{n}_k is Gaussian, the term $\boldsymbol{\epsilon}_k^T \mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k$ follows a chi-square distribution and the limit is evaluated for the χ^2 value with m degrees-of-freedom, such as

$$f_k^1 > \left[(2\pi)^{m/2} |\mathbf{S}_k|^{1/2} \right]^{-1} \exp\left\{-\frac{1}{2} x_{lim}\right\}, \quad (24)$$

where x_{lim} is the inverse of the chi-square cumulative distribution function, selected with a p -value of $p = 0.001$. One should note that x_{lim} is similar to the gate threshold of the PDAF method; see *e.g.* [23].

The remaining CC is when the outlier is not explained by the noise model. The uniform component f_k^2 is then too small and a source of instability. Two solutions can be thought to deal with such a situation. First, the regular KF step can be used. Second, the measurement can be discarded and no correction is performed. These two different solutions are referred to as the *regular* (reg.) and *prediction* (pred.) variants respectively. The main question left is the detection of an observation that does not follow the model. In practice, the value of $p_Z(\mathbf{z}_k | \mathbf{Z}^{k-1})$ is compared with the χ^2 value, such as

$$f_k^2 > \left[(2\pi)^{m/2} |\mathbf{L}_k|^{1/2} \right]^{-1} A_k \cdot \exp\left\{-\frac{1}{2} x_{lim}\right\}. \quad (25)$$

Although $[\boldsymbol{\epsilon}_k - \mathbf{t}]^T \mathbf{L}_k^{-1} [\boldsymbol{\epsilon}_k - \mathbf{t}]$ is not supposed to follow exactly a chi-square distribution, this criterion gives good results in practice as it is illustrated in the following sections. The pseudo-code of the score function and its derivative calculation is given in Algorithm 1.

Algorithm 1 Kalman-Masrelez Filter with GU Noise

Require: $\mathbf{L}_k, \mathbf{S}_k, \boldsymbol{\epsilon}_k, \theta_k, \mathbf{a}_k, A_k, x_{lim}$

- 1: **if** $\theta_k == 1$ **then**
- 2: $\mathbf{g}_k \leftarrow \mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k$
- 3: $\mathbf{G}_k \leftarrow \mathbf{S}_k^{-1}$
- 4: **else**
- 5: $\mathbf{C}_k \leftarrow \text{chol}(\mathbf{L}_k^{-1}, 'upper')$
- 6: Compute \mathbf{b}_k and \mathbf{e}_k
- 7: Compute \mathbf{I}_k with (27)
- 8: Compute f_k^2 with (18)
- 9: Compute f_k^1 with (17)
- 10: Compute p_Z with (16)
- 11: $f_k^{1,lim} \leftarrow \left[(2\pi)^{m/2} |\mathbf{L}_k|^{1/2} \right]^{-1} A_k \cdot \exp\left\{-\frac{1}{2} x_{lim}\right\}$
- 12: **if** $p_Z < (1 - \theta_k) f_k^{2,lim} / A_k$ **then**
- 13: *Prediction variant:*
- 14: $\mathbf{g}_k \leftarrow \text{zeros}(m, 1)$
- 15: $\mathbf{G}_k \leftarrow \text{zeros}(m, m)$
- 16: *Regular variant:*
- 17: $\mathbf{g}_k \leftarrow \mathbf{S}_k^{-1} \boldsymbol{\epsilon}_k$
- 18: $\mathbf{G}_k \leftarrow \mathbf{S}_k^{-1}$
- 19: **else**
- 20: Compute γ_k with (31)
- 21: Compute $\frac{\partial \gamma_k}{\partial \mathbf{z}_k}$ with (34)
- 22: $f_k^{1,lim} \leftarrow \left[(2\pi)^{m/2} |\mathbf{S}_k|^{1/2} \right]^{-1} \exp\left\{-\frac{1}{2} x_{lim}\right\}$
- 23: **if** $f_k^1 \geq f_k^{1,lim}$ **then**
- 24: Compute \mathbf{g}_k with (22)
- 25: Compute \mathbf{G}_k with (23)
- 26: **else**
- 27: $\mathbf{g}_k \leftarrow \gamma_k$
- 28: $\mathbf{G}_k \leftarrow \frac{\partial \gamma_k}{\partial \mathbf{z}_k}$
- 29: **return** $\mathbf{g}_k, \mathbf{G}_k$

IV. PRELIMINARY TESTS

A. Single state model

For the first evaluation of the introduced filter, a simple one dimensional test case is considered. The constant distance of an object is measured. In practice, one can think about a laser rangefinder used in a dusty environment, *i.e.* there is probability $1 - \theta$ that a particle positions itself between the object and the sensor. The system properties are as follows

- $x^* = 100$ m, true distance of the target;
- $x_0 = 50$ m, initial distance of the estimators;
- $P_0 = 20^2 m^2$, initial variance of the state;
- $Q = 1 m^2$, variance of the state noise;
- x , estimated state;
- $y = x + n$, measured distance with measurement noise n : zero mean and variance σ_n^2 .

The state model is therefore a random walk. The data are generated with the p.d.f. (2) where the noise parameters are constant: $\theta = 0.9$, $\mu = 0$, $\sigma = 5$ and $a = 100$.

Table I provides three criteria: the Root Mean Square (RMS) error, the Execution Time (ET) and the Average Normalized Equation Estimation Error Squared (ANEES) over 10000 Monte Carlo Simulations (MCS) of a regular KF, the two variants of the introduced Robust KF (RKF), the MCDA and

Table I
 NOMINAL CASE – 10000 MCS – 1D MODEL

	Crit.	KF	RKF (reg.)	RKF (pred.)	MCDA	PDAF
Gauss.	RMS	1.74	1.77	1.77	1.77	1.75
	ET [ms]	0.14	4.31	4.22	48.61	1.26
	ANEES	0.59	0.56	0.56	0.58	0.74
Mixt.	RMS	6.24	3.84	4.10	3.36	6.02
	ET [ms]	0.14	4.37	4.31	48.55	4.19
	ANEES	8.21	3.16	5.00	2.81	56.06

PDAF algorithms. Both MCDA and PDAF were designed to take into account clutter measurements and to track multiple targets. In our case, we consider only one target and use 10 particles for the MCDA. That number was chosen to have similar performances with the mixture noise case; see Table I. The MCDA software can be downloaded via <http://becs.aalto.fi/en/research/bayes/rbmcd/>. For the PDAF algorithm, the gate probability is based on x_{lim} defined in section III-C and the clutter density at time k is $\lambda_k = (1 - \theta)/V_k$, with the V_k the volume of the gate.

For the first example, the measurement noise is generated as Gaussian with the same standard deviation as the noise models of the KF and RKF filters: $\sigma_n = \sigma = 5$. The second example is called "Mixture" because the noise model of the RKF is exactly the one used to generate the data. In both cases, the situation is described as "Nominal" because there is no error introduced in the RKF noise model.

For the Gaussian case, it appears that both RKF variants are equivalent but slightly less precise than the regular KF. That can be explained by a slower convergence due to the noise model. The GU mixture indeed tends to explain more easily the measurement noise and thus corrects less quickly the state. This phenomenon can be observed in Fig. 1. This figure depicts the four estimations for the last Monte Carlo iteration. The MCDA and PDAF algorithms provide equivalent mean estimates in this introductory case. Regarding the ANEES, the PDAF has the closest to 1, which means that its covariance estimate best represents the underlying estimate error [24].

When the measurement noise is generated with the GU mixture, the KF has difficulty as expected. It seems that the *regular* variant is more robust than the *prediction* one. However, this is not so clear by looking at Fig. 2 (last MCS iteration) and hence requires more investigation. In this case, the MCDA algorithm gives a mean RMS error as low as the *regular* variant. Nonetheless, Table I shows that it requires more execution time. It may be observed that the ratio is more or less the number of particles considered. The difficulties of the PDAF can be explained by the case where the first observation is an outlier, which prevents the estimator to converge and causes large errors due to the coarse initialization.

B. Influence of the outliers model and the reliability coefficient

If the previous subsection illustrated the interest of the RKF in presence of outliers, the current one is devoted to the limit case where the outliers' model is erroneous. The measurement noise is still generated with the same GU mixture but both RKF variants and MCDA have a different range of outliers: a .

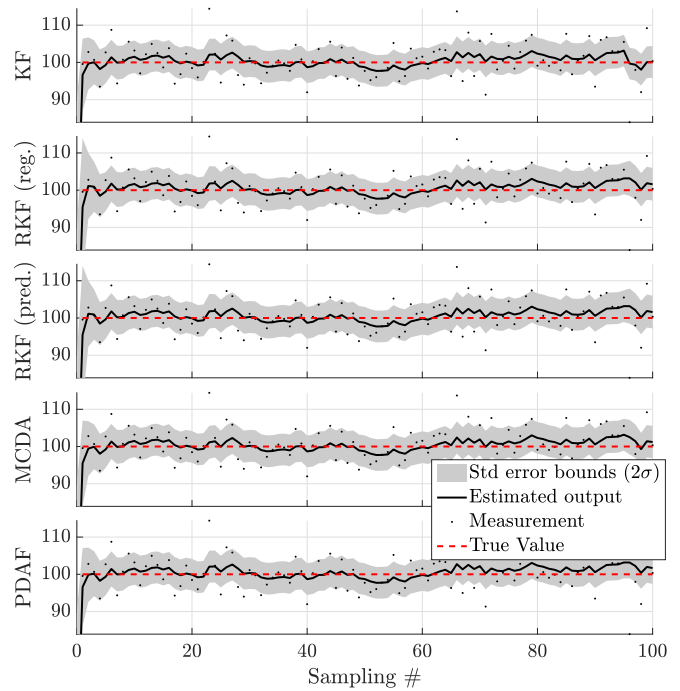


Figure 1. Nominal case – Gaussian noise – 1D model

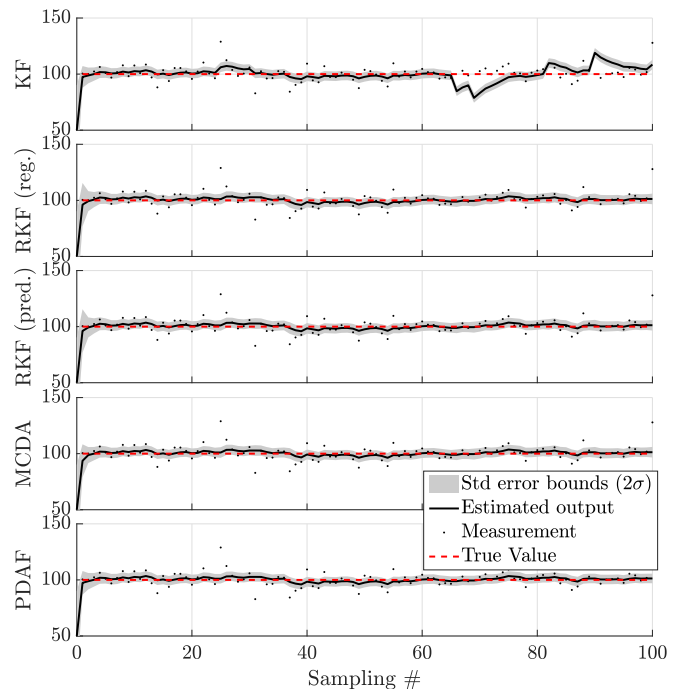


Figure 2. Nominal case – Mixture noise – 1D model

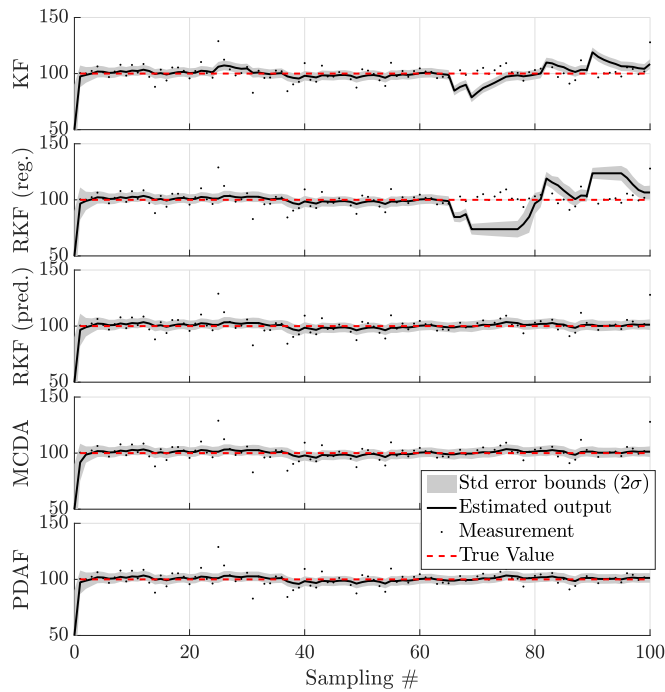


Figure 3. Errors in the error model – $\Delta a = -50\%$ – 1D model

Table II summarizes the results for 10000 MCS with outliers range errors $\Delta a = \pm 50\%$. As it could be expected, if the outliers' range is larger than the 'reality', the RKF variants do not suffer. In the opposite, when the range is narrower, the effect is disastrous for the *regular* variant whereas the *prediction* one is insensitive. This is also depicted in Fig. 3 that gives the estimation results of the last MCS iteration for $\Delta a = -50\%$. Here also, the MCDA algorithm proves to be robust by providing low RMS estimates, although it has a larger computational cost. Obviously, the results of the PDAF are insensitive since the method does not rely on the range information. With respect to the ANEES, the most credible estimator is the MCDA followed by the prediction variant of the RKF.

The second limitation of the GU mixture lies in the reliability coefficient. To test its influence on the estimation results, the RKF variants are considered with reliability errors: $\Delta\theta = \pm 10\%$. The results are provided in Table III. It appears that both variants similarly behave. Here also, the estimations are less accurate in the case where the reliability coefficient is not large enough; *i.e.* $\Delta\theta = -10\%$. Nevertheless, this is not disastrous as the case $\Delta a = -50\%$, which appears to be the limit case. In particular, the RMS errors stay far below the one of the KF. Although the clutter density depends on θ , the PDAF is not really prone to its miscalculation. Therefore, the time histories of the last MCS iteration are not presented because do not provide more information. Finally, by putting aside the regular variant, the same observation can be made for the ANEES criterion as the case $\Delta a = \pm 50\%$.

Table II
 ERRORS IN THE RELIABILITY COEFFICIENT – 10000 MCS – 1D MODEL

$\Delta\theta$	Crit.	KF	RKF (reg.)	RKF (pred.)	MCDA	PDAF
+10%	RMS	6.24	3.76	4.06	4.21	6.01
	ET [ms]	0.15	4.47	4.38	50.02	4.45
	ANEES	8.21	3.45	5.73	5.28	56.14
-10%	RMS	6.24	3.95	4.16	3.14	6.03
	ET [ms]	0.14	4.46	4.36	49.15	4.33
	ANEES	8.21	3.03	4.58	1.98	55.94

Table III
 ERRORS IN THE OUTLIERS RANGE – 10000 MCS – 1D MODEL

Δa	Crit.	KF	RKF (reg.)	RKF (pred.)	MCDA	PDAF
+50%	RMS	6.24	3.76	4.16	3.42	6.02
	ET [ms]	0.15	4.57	4.49	51.74	4.55
	ANEES	8.21	3.44	4.48	3.05	56.06
-50%	RMS	6.24	8.98	3.51	3.15	6.02
	ET [ms]	0.14	4.37	4.28	48.92	4.34
	ANEES	8.21	10.86	3.09	2.05	56.06

V. COMPLEMENTARY TESTS

A. Simplified car model

For further testing of the introduced RKF, a 2-dimensional car model is considered; see Examples 3.6 and 4.3 in [25]. Compared with the previous example, this one has a real dynamics (*i.e.* there is a movement) and illustrates a multi-dimensional estimation. In this case, the outliers could model multipath errors of the GPS receiver for instance. For the sake of clarity, the linear state space model is recalled here:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}\mathbf{x}_{k-1} + \mathbf{q}_{k-1} & \mathbf{q}_{k-1} &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \\ \mathbf{y}_k &= \mathbf{H}\mathbf{x}_k + \mathbf{n}_k & \mathbf{n}_k &\sim \mathcal{G}(\theta, \mathbf{0}, \mathbf{R}, \mathbf{a}), \end{aligned}$$

where $\mathbf{x} = [x_1, x_2, x_3, x_4]^T$ with the position of the car (x_1, x_2) and the corresponding velocities (x_3, x_4) . The matrices are

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & \mathbf{Q} &= \sigma_q^2 \begin{bmatrix} \frac{\Delta t^3}{3} & 0 & \frac{\Delta t^2}{2} & 0 \\ 0 & \frac{\Delta t^3}{3} & 0 & \frac{\Delta t^2}{2} \\ \frac{\Delta t^2}{2} & 0 & \Delta t & 0 \\ 0 & \frac{\Delta t^2}{2} & 0 & \Delta t \end{bmatrix}, \\ \mathbf{H} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} & \mathbf{R} &= \sigma_n^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \end{aligned}$$

where $\Delta t = 0.1$ s in the sampling period, $\sigma_q = 1$ m.s⁻¹/√Hz and $\sigma_n = 0.5$ m. The other parameters of the mixture model are $\theta = 0.9$ and $\mathbf{a} = [10, 10]^T$. The initial state is vector $\mathbf{x}_0 = [0, 0, 1, -1]^T$ and its covariance matrix \mathbf{P}_0 is the (4×4) identity matrix.

B. Results

Table IV provides the results over 10000 MCS, for the nominal case. It should be noticed that the RMS_x and RMS_y errors are computed for the whole state vector and only the positions respectively. As expected, with a Gaussian noise, the RKF does

not improve the estimation. That is confirmed by Fig. 4. This figure illustrates the vehicle position, the measurements and the estimates for the last MCS iteration. With the GU mixture noise, the *prediction* variant outperforms the KF whereas the *regular* one has major difficulties. When the outlier is not explained by the noise model, the application of the regular KF correction step appears to be counterproductive. Hence, the corresponding trajectory is not depicted in Fig. 5, where it is visible that the RKF provides a more reliable estimated trajectory compared to the regular filter. The MCDA algorithm gives results (RMS_x , RMS_y and ANEES) similar to the *prediction* variant but has the advantage of providing narrower confidence ellipsoids, *i.e.* smaller covariances, although the computational cost is still larger. The PDAF algorithm has also RMS_x and RMS_y equivalent to the *prediction* variant with a lowest computational cost, however, it is more optimistic according to the ANEES criterion. One should note here that there is no initialization issue.

Tables V and VI give the criteria values in the case of errors in the error model. Those results definitively disqualify the *regular* variant which gives worse results than the regular KF. On the contrary, the *prediction* variant appears to be robust to the errors and always provides mean RMS errors lower than the regular KF. Furthermore, its estimation is similar to the reference MCDA algorithm, though its confidence ellipses are larger. Here also, the PDAF provides estimates with RMS errors comparable to those of *prediction* variant and MCDA, with a lowest computational cost. Nonetheless, with a slightly larger amount of time, the *prediction* variant appears to be credible with respect to the ANESS criterion.

From those different observations, the following comments can be drawn

- the *regular* variant should be dropped;
- compared to the regular KF, the RKF is more robust to outliers but has a slower convergence and provides larger states covariances;
- compared to the MCDA, the RKF has a lower computational cost, provides as accurate estimates but is less optimal (larger covariances);
- compared to the PDAF, the RKF provides as accurate estimates with more credibility (ANEES closer to 1), is less sensitive to rough initialisation, but has a slightly larger computational cost;
- the worst case to test is the narrower outlier range ($\Delta a < 0$);
- the RKF is not too sensitive to the reliability coefficient.

VI. CONCLUSION

This paper suggests the use of a Gaussian-uniform mixture for the observation noise model in a Kalman-type recursive estimator. Such a noise model encompasses the outliers by expressing the instrument reliability. Based on the Masreliez's work, a closed-form formulation of the filter is drawn. Although this formulation is rigorously established, that may lead to numerical instability in several cases: (i) the filter is used as regular KF, (ii) the Gaussian component of the error model is too weak, (iii) both components of the error model are too

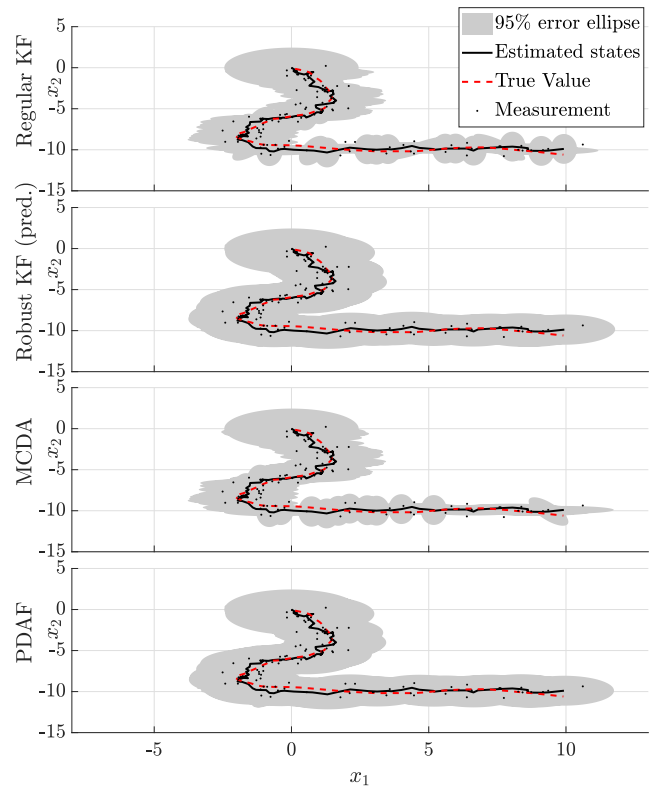


Figure 4. Nominal case – Gaussian noise – 2D model

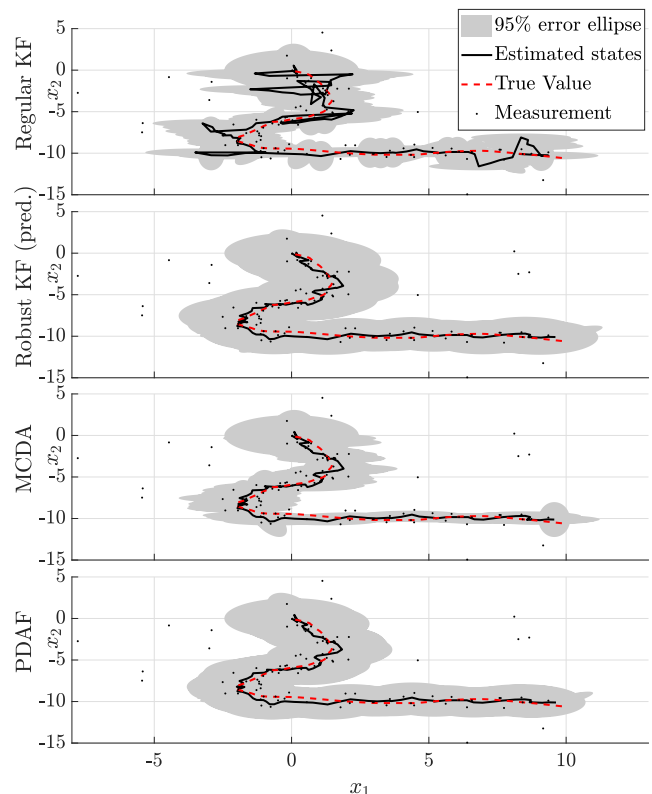


Figure 5. Nominal case – Mixture noise – 2D model

Table IV
NOMINAL CASE – 10000 MCS – 2D MODEL

	Crit.	KF	RKF (reg.)	RKF (pred.)	MCDA	PDAF
Gauss.	RMS _x	0.57	0.59*	0.58	0.58	0.58
	RMS _y	0.28	0.29*	0.28	0.28	0.28
	ET [ms]	1.09	11.66*	11.40	71.34	2.80
	ANEES	1.07	1.07*	1.03	1.07	1.11
Mixt.	RMS _x	1.30	1.41*	0.61	0.62	0.65
	RMS _y	0.93	1.07*	0.33	0.33	0.38
	ET [ms]	1.07	10.32*	11.20	69.93	2.79
	ANEES	8.00	8.70*	1.07	1.17	4.54

*: x_{lim} selected for a p -value $p = 0.02$

Table V
ERRORS IN THE OUTLIERS RANGE – 10000 MCS – 2D MODEL

Δa	Crit.	KF	RKF (reg.)	RKF (pred.)	MCDA	PDAF
+25%	RMS _x	1.30	1.37*	0.61	0.62	0.65
	RMS _y	0.93	1.02*	0.33	0.33	0.38
	ET [ms]	1.05	10.18*	11.11	69.04	2.77
	ANEES	8.00	8.45*	1.09	1.18	4.54
-25%	RMS _x	1.30	2.27*	0.62	0.61	0.65
	RMS _y	0.93	2.21*	0.34	0.33	0.38
	ET [ms]	1.06	10.95*	11.00	69.28	2.77
	ANEES	8.00	12.08*	1.06	1.16	4.54

*: x_{lim} selected for a p -value $p = 0.02$

weak. The resulting algorithm is able to detect those cases and adapts itself to provide an accurate estimation. Though case (iii) led to two variants, the simulation results indicate that the best solution is to perform only the prediction when the measurement is not explained at all by the GU mixture.

Regarding the prospects, the next step is to undertake a real-time implementation in order to evaluate the solution in realistic experimental situations.

APPENDIX A

DEFINITION AND DERIVATIVES OF J_k

From (18), the term J_k is defined as follows:

$$\begin{aligned}
 J_k &= \int_{-\mathbf{a}_k}^{\mathbf{a}_k} \exp\left\{-\frac{1}{2}\boldsymbol{\epsilon}_k^T \mathbf{L}_k^{-1} \boldsymbol{\epsilon}_k - \frac{1}{2}\mathbf{t}^T \mathbf{L}_k^{-1} \mathbf{t} - \boldsymbol{\epsilon}_k^T \mathbf{L}_k^{-1} \mathbf{t}\right\} dt \\
 &= \int_{-\mathbf{C}_k \mathbf{a}}^{\mathbf{C}_k \mathbf{a}_k} \exp\left\{-\frac{1}{2}\mathbf{e}_k^T \mathbf{e}_k - \mathbf{e}_k^T \mathbf{u} - \frac{1}{2}\mathbf{u}^T \mathbf{u}\right\} du \quad (26) \\
 &= \prod_{l=1}^m \int_{-b_{l,k}}^{b_{l,k}} \exp\left\{-\frac{1}{2}e_{l,k}^2 - \frac{1}{2}u_l^2 - e_{l,k}u_l\right\} du_l = \prod_{l=1}^m I_{l,k}
 \end{aligned}$$

We define the integration variable $\mathbf{u} = \mathbf{C}_k \mathbf{t}$ where \mathbf{C}_k is the upper triangular matrix coming from the Cholesky

Table VI
ERRORS IN THE RELIABILITY COEFFICIENT – 10000 MCS – 2D MODEL

$\Delta \theta$	Crit.	KF	RKF (reg.)	RKF (pred.)	MCDA	PDAF
+10%	RMS _x	1.30	1.50*	0.62	0.62	0.68
	RMS _y	0.93	1.18*	0.33	0.34	0.43
	ET [ms]	1.11	10.77*	11.57	72.45	2.85
	ANEES	8.00	9.46*	1.15	1.22	7.97
-10%	RMS _x	1.30	1.39*	0.61	0.61	0.64
	RMS _y	0.93	1.05*	0.33	0.33	0.37
	ET [ms]	1.08	10.35*	11.32	71.06	2.83
	ANEES	8.00	8.53*	1.04	1.15	3.47

*: x_{lim} selected for a p -value $p = 0.02$

factorization of $\mathbf{L}_k^{-1} = \mathbf{C}_k^T \mathbf{C}_k$. Therefore, the exponential term in the integral can be decomposed axis per axis, by writing $\mathbf{b}_k = \mathbf{C}_k \mathbf{a}_k$ and $\mathbf{e}_k = \mathbf{C}_k \boldsymbol{\epsilon}_k$. By assumption the covariance of the predicted state \mathbf{M}_k is symmetric positive definite. In addition, \mathbf{H}^T is assumed to be full column rank; i.e. $\text{rank}(\mathbf{H}^T) = m$. Consequently, \mathbf{L}_k and its inverse are symmetric positive definite, which is the requirement for the Cholesky factorization. Each component of the $(m \times 1)$ vector \mathbf{I}_k can be evaluated separately such as

$$\begin{aligned}
 I_{l,k} &= \int_{-b_{l,k}}^{b_{l,k}} \exp\left\{-\frac{1}{2}(u_l + e_{l,k})^2\right\} du_l \quad (27) \\
 &= \int_{e_{l,k}-b_{l,k}}^{e_{l,k}+b_{l,k}} \exp\left\{-\frac{1}{2}v_l^2\right\} dv_l \\
 &= \frac{\sqrt{2\pi}}{2} \left[\text{erf}\left(\left(b_{l,k} + e_{l,k}\right)\sqrt{\frac{1}{2}}\right) \right. \\
 &\quad \left. - \text{erf}\left(\left(-b_{l,k} + e_{l,k}\right)\sqrt{\frac{1}{2}}\right) \right]
 \end{aligned}$$

where erf is the error function.

In order to obtain the derivatives of J_k , one should noticed that it is not independent of \mathbf{z}_k . Thus, the derivative of $I_{l,k}$ is first developed:

$$\begin{aligned}
 \frac{\partial I_{l,k}}{\partial z_{j,k}} &= \frac{\sqrt{2\pi}}{2} \sqrt{\frac{1}{2}} \frac{\partial e_{l,k}}{\partial z_{j,k}} \frac{2}{\sqrt{\pi}} \left(\exp\left\{-\frac{1}{2}(b_{l,k} + e_{l,k})^2\right\} \right. \\
 &\quad \left. - \exp\left\{-\frac{1}{2}(b_{l,k} - e_{l,k})^2\right\} \right) \\
 &= \frac{\partial e_{l,k}}{\partial z_{j,k}} \alpha_{l,k} = [\mathbf{C}_k]_{lj} \alpha_{l,k}, \quad (28)
 \end{aligned}$$

by defining the $(m \times 1)$ vector $\boldsymbol{\alpha}_k$ such as

$$\alpha_{l,k} = \exp\left\{-\frac{1}{2}(b_{l,k} + e_{l,k})^2\right\} - \exp\left\{-\frac{1}{2}(b_{l,k} - e_{l,k})^2\right\}. \quad (29)$$

The first derivative is thus defined by:

$$\begin{aligned}
 \frac{\partial J_k}{\partial z_{j,k}} &= \sum_{l=1}^m \frac{\partial I_{l,k}}{\partial z_{j,k}} \prod_{i=1, i \neq l}^m I_{i,k} = \sum_{l=1}^m \frac{J_k}{I_{l,k}} \frac{\partial I_{l,k}}{\partial z_{j,k}} \quad (30) \\
 &= J_k \gamma_{j,k},
 \end{aligned}$$

by assuming $I_{l,k} \neq 0$ and defining the $(m \times 1)$ vector $\boldsymbol{\gamma}_k$ such as

$$\gamma_{j,k} = \sum_{l=1}^m \frac{1}{I_{l,k}} \frac{\partial I_{l,k}}{\partial z_{j,k}}. \quad (31)$$

With respect to (27), $I_{l,k} = 0$ implies that $b_{l,k} = 0$ since erf is a strictly monotonic function. Therefore, one should be cautious to ensure $\alpha_{l,k} \neq 0, \forall l, k$. To derive the second order derivatives, it appears in (30) that only $\partial \gamma_{j,k} / \partial z_{i,k}$ is missing. It first comes

$$\begin{aligned}
 \frac{\partial^2 I_{l,k}}{\partial z_{i,k} \partial z_{j,k}} &= \frac{\partial}{\partial z_{i,k}} ([\mathbf{C}_k]_{lj} \alpha_{l,k}) = [\mathbf{C}_k]_{lj} \frac{\partial \alpha_{l,k}}{\partial z_{i,k}} \\
 &= -[\mathbf{C}_k]_{lj} \frac{\partial e_{l,k}}{\partial z_{i,k}} \beta_{l,k} = -[\mathbf{C}_k]_{lj} [\mathbf{C}_k]_{li} \beta_{l,k} \quad (32)
 \end{aligned}$$

by defining the $(m \times 1)$ vector β_k such as

$$\beta_{l,k} = (b_{l,k} + e_{l,k}) \exp\left\{-\frac{1}{2}(b_{l,k} + e_{l,k})^2\right\} + (b_{l,k} - e_{l,k}) \exp\left\{-\frac{1}{2}(b_{l,k} - e_{l,k})^2\right\} \quad (33)$$

The derivatives of γ_k components are thus defined by:

$$\begin{aligned} \frac{\partial \gamma_{j,k}}{\partial z_{i,k}} &= \sum_{l=1}^m \left[\frac{\partial \frac{1}{I_{l,k}}}{\partial z_{i,k}} \frac{\partial I_{l,k}}{\partial z_{j,k}} + \frac{1}{I_{l,k}} \frac{\partial^2 I_{l,k}}{\partial z_{i,k} \partial z_{j,k}} \right] \\ &= \sum_{l=1}^m \left[-\frac{1}{I_{l,k}^2} \frac{\partial I_{l,k}}{\partial z_{i,k}} \frac{\partial I_{l,k}}{\partial z_{j,k}} + \frac{1}{I_{l,k}} \frac{\partial^2 I_{l,k}}{\partial z_{i,k} \partial z_{j,k}} \right] \end{aligned} \quad (34)$$

REFERENCES

- [1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Journal of basic engineering*, vol. 83, no. 1, pp. 95–108, 1961.
- [3] K. D. Rao, M. Swamy, and E. Plotkin, "Gps navigation with increased immunity to modeling errors," *IEEE transactions on aerospace and electronic systems*, vol. 40, no. 1, pp. 2–11, 2004.
- [4] J.-A. Ting, E. Theodorou, and S. Schaal, "A kalman filter for robust outlier detection." in *IROS*, 2007, pp. 1514–1519.
- [5] K. L. Lange, R. J. Little, and J. M. Taylor, "Robust statistical modeling using the t distribution," *Journal of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, 1989.
- [6] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "An outlier-robust kalman filter," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1551–1558.
- [7] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-t distribution," in *International Workshop on Machine Learning for Signal Processing*. IEEE, 2012, pp. 1–6.
- [8] Y. Huang, Y. Zhang, N. Li, Z. Wu, and J. A. Chambers, "A novel robust student-t-based kalman filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1545–1554, 2017.
- [9] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [10] H. W. Sorenson and D. L. Alspach, "Recursive bayesian estimation using gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, 1971.
- [11] R. N. Banavar and J. L. Speyer, "Properties of risk-sensitive filters/estimators," *IEE Proceedings-Control Theory and Applications*, vol. 145, no. 1, pp. 106–112, 1998.
- [12] M. Zorzi, "Robust kalman filtering under model perturbations," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2902–2907, 2016.
- [13] —, "Convergence analysis of a family of robust kalman filters based on the contraction principle," *SIAM Journal on Control and Optimization*, vol. 55, no. 5, pp. 3116–3131, 2017.
- [14] K. Glover and J. C. Doyle, "State-space formulae for all stabilizing controllers that satisfy an h-norm bound and relations to relations to risk sensitivity," *Systems & Control Letters*, vol. 11, no. 3, pp. 167–172, 1988.
- [15] P. Rawicz, P. Kalata, K. Murphy, and T. Chmielewski, "Explicit formulae for two state kalman, h/sub 2/and h/sub spl infin/target tracking filters," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 1, pp. 53–69, 2003.
- [16] W. Li and Y. Jia, "Nonlinear gaussian mixture phd filter with an h criterion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 4, pp. 2004–2016, 2016.
- [17] J.-F. Grandin and M. Marques, "Robust fusion with reliabilities weights," in *Sensor Fusion: Architectures, Algorithms, and Applications VI*, vol. 4731. International Society for Optics and Photonics, 2002, pp. 176–188.
- [18] P. Coretto and C. Hennig, "Maximum likelihood estimation of heterogeneous mixtures of gaussian and uniform distributions," *Journal of Statistical Planning and inference*, vol. 141, no. 1, pp. 462–473, 2011.

- [19] S. Särkkä, A. Vehtari, and J. Lampinen, "Rao-blackwellized particle filter for multiple target tracking," *Information Fusion*, vol. 8, no. 1, pp. 2–15, 2007.
- [20] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica*, vol. 11, no. 5, pp. 451–460, 1975.
- [21] C. Masreliez, "Approximate non-gaussian filtering with linear state and observation relations," *IEEE Transactions on Automatic Control*, vol. 20, no. 1, pp. 107–110, 1975.
- [22] B. Kovačević, v. Đurović, and S. Glavaški, "On robust kalman filtering," *International Journal of Control*, vol. 56, no. 3, pp. 547–562, 1992.
- [23] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, no. 6, pp. 82–100, Dec 2009.
- [24] X. Rong Li, Z. Zhao, and V. P. Jilkov, "Practical measures and test for credibility of an estimator," in *Proc. Workshop on Estimation, Tracking, and Fusion A Tribute to Yaakov Bar-Shalom*, 2001, pp. 481–495.
- [25] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013, vol. 3.

Mathieu Brunot received the Engineer Degree in aeronautics and automatic control (2013) from the ISAE-ENSICA, France, the M.Sc. in flight dynamics and flight control (2014) from Cranfield Univ., U.K., and the Ph.D. (2017) from the National Polytechnic Institute of Toulouse, France. He is currently a research engineer with ONERA. His research interests include air vehicle identification and estimation as well as navigation systems.