Kalman Filter and Joint Tracking and Classification based on Belief Functions in the TBM Framework.

Philippe Smets† and **Branko Ristic**‡

Abstract

The paper presents an approach to joint tracking and classification based on belief functions as understood in the transferable belief model (TBM). The TBM model is identical to the classical model except all probability functions are replaced by belief functions, which are more flexible for representing uncertainty. It is felt that the tracking phase is well handled by the classical Kalman filter but that the classification phase deserves amelioration. For the tracking phase, we derive a minimal set of assumptions needed in the TBM approach in order to recover the classical relations. For the classification phase, we distinguish between the observed target behaviors and the underlying target classes which are usually not in one-to-one correspondence. We feel the results obtained with the TBM approach are more reasonable than those obtained with the corresponding Bayesian classifiers.

Keywords: Transferable belief model, target classification, Kalman filter, belief function theory.

A short version of this study was presented at the Seventh International Conference on Information Fusion (FUSION 2004) in Stockholm, Sweden [1].

February 25, 2005

† IRIDIA, Université libre de Bruxelles, Av. F. Roosevelt 50, CP 194/6, 1050 Bruxelles, Belgium; tel: +32 (2) 344 82 96; email: psmets@ulb.ac.be

‡ DSTO, ISRD - 200 Labs, PO Box 1500, Edinburgh SA 5111, Australia, tel: +61 8 8259 7181, fax: +61 8 8259 7184, email: branko.ristic@dsto.defence.gov.au.

1 Introduction

The paper is devoted to joint tracking and classification (JTC) of targets based on kinematic data. The optimal Bayesian estimator for this problem [2, 3, 4, 5] consists of a bank of filters that match the expected dynamic behavior of each class (class-matched filters). In some numerical examples, however, this type of classifier may produce unsatisfactory performance due to inadequate mapping between target behavior and target class. This was a motivation to consider the classification phase of the JTC problem in the framework of the transferable belief model (TBM) [6, 7, 8, 9]. It was hoped - and in fact observed - that the use of belief functions, which are more flexible than the probability functions, could produce better classification results.

In order to use the TBM for the classification phase, we need also to formulate the class-matched filtering within the TBM framework. Indeed using a probabilistic approach for the tracking phased and then switching to the TBM for the classification phase would be inconsistent.

The TBM overall model for JTC is identical to the classical JTC model except all probability functions are replaced by belief functions. It is felt that the tracking phase is well handled by the classical Kalman filter (KF) but that the classification phase deserves amelioration.

For the tracking phase, we determine under which 'minimal' assumptions the TBM approach produces the classical KF relations (as a typical building block of class-matched filters) [10]. For this derivation we assume:

- for the dynamic equations, that the uncertainty in the additive noise is represented by belief functions whose pignistic transformations are Gaussian probability distribution functions, and the prior belief used at time 0 is a vacuous belief function,

- for the measurement equations, that the uncertainty in the additive noise is represented by Bayesian belief functions which are Gaussian (as in the classical case).

For the classification phase, we will distinguish between the observed target behaviors and the underlying target classes which are usually not in one-to-one correspondence. Within the TBM, the classification phase is based on the General Bayesian Theorem [11, 12]. The classification results differ significantly from those derived within the classical framework. It is due to the fact that the TBM provides more flexible ways to represent adequately the prior beliefs and the target behavior - target class relations. We argue that the results obtained with the TBM approach are more reasonable than those obtained with the corresponding Bayesian classifiers.

Historically, the KF has already been described as an example of evidential network by [13]. In fact the special nature of the belief functions we assume for the KF make it possible to solve easily the filtering without using the whole machinery of the TBM. Nevertheless for the more general models, we would use these evidential networks. Note that Kohlas and Monney [14, Ch.10] also present a formal solution for handling dynamic uncertainty as in the KF. Their solution is similar to the one presented here but it is derived within the hint model framework which is still based on probability theory, whereas the TBM does not assume the existence of any underlying probabilities.

Belief functions used in this papers are defined on continuous spaces, a topic presented in [15], used in [16] and summarized in appendix A.1.

The paper is organized as follows. In section 2, we present the classical JTC and an example that motivated this work as its classification results seem inappropriate to us. In section 3, we present some background material about the TBM. In section 5, we present the KF within the TBM framework. In section 6, we present the classification phase of the JTC within the TBM framework. Section 7 presents the conclusions.

2 JTC in the probabilistic framework

We first present the JTC problem in the classical probabilistic framework.

2.1 Problem formulation

The target state at time $t = 0, 1, 2, \ldots$ is represented by n-dimensional vector $\boldsymbol{x}_t \in R^n$. Target class is a time-invariant attribute which takes values from a discrete set: $C = \{c_1, \ldots, c_s\}$. For simplicity let us assume that the state vector evolves according to the linear target motion model:

$$\boldsymbol{x}_t = \boldsymbol{F}(c_i)\boldsymbol{x}_{t-1} + \boldsymbol{v}_{t-1} \tag{1}$$

where $\boldsymbol{F}(c_i)$ is $n \times n$ class dependent transition matrix and $\boldsymbol{v}_{t-1} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{Q})$ is an i.i.d. process noise[1]. It is assumed that 1) prior class probabilities $P(c_i)$ are known, 2) prior conditional state is independent of process noise v_{t-1} and measurement noise w_t, and 3) prior conditional state densities $p(\boldsymbol{x}_0|c_i)$ are known and Gaussian for $i = 1, \ldots, s$, i.e.

$$\boldsymbol{x}_0|c_i \sim \mathcal{N}_n(\boldsymbol{\mu}_0^i, \boldsymbol{\Sigma}_0^i)$$

Again for simplicity let us assume that kinematic measurements $\boldsymbol{z} \in R^m$ are linearly related to the target state:

$$\boldsymbol{z}_t = \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{w}_t \tag{2}$$

where $\boldsymbol{H}$ is $m \times n$ measurement matrix and $\boldsymbol{w}_t \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{R})$ is an i.i.d. measurement noise. Furthermore, process and measurement noises are mutually independent. The optimal joint state and class estimator in the Bayesian sense requires to construct at time t the posterior density $p(\boldsymbol{x}_t, c_i|\boldsymbol{Z}_t)$, where $\boldsymbol{Z}_t = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t\}$ is the sequence of observations up to time t.

[1]Let $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the n-dimensional normal distribution of mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The $\sim$ in expressions like $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that the variable $\boldsymbol{Y}$ is a n-dimensional random variable with a Gaussian probability density function.

The optimal Bayesian solution for JTC [4, 5] consists of a bank of class-matched filters as shown in figure 1. Class-matched filters are typically built as Kalman filters or a weighted sum of Kalman filters, as described in the example in section 2.2. The terms $p(\boldsymbol{x}_t|\boldsymbol{Z}_t, c_i)$ are the outputs used for tracking (state estimation). Class probabilities are computed recursively as follows:

$$P(c_i|\boldsymbol{Z}_t) = \alpha\, p(\boldsymbol{z}_t|\boldsymbol{Z}_{t-1}, c_i)\, P(c_i|\boldsymbol{Z}_{t-1}) \tag{3}$$

where $p(\boldsymbol{z}_t|\boldsymbol{Z}_{t-1}, c_i)$ is the likelihood of class i at time t and α is a normalizing constant.

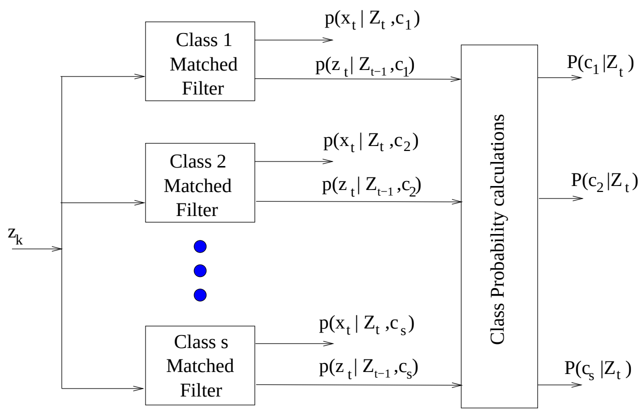


Figure 1: JTC scheme

2.2 A motivating example on classification

Suppose we have $s = 3$ classes of targets, and for classification we use their acceleration (that is maneuvering) capabilities. Class 1 (c_1) are for example commercial planes with modest acceleration, class 2 (c_2) are large military aircraft (e.g. bomber) which can perform medium level acceleration and class 3 (c_3) are light and agile military aircraft (fighter planes) which are able to attain very large levels of acceleration. In the steady state all platform classes fly with no acceleration in order to minimise the fuel consimption.

To simplify the analysis let us consider 1D geometry with state vector $\boldsymbol{x} = [x \quad \dot{x}]'$, and with a linear target motion model (for all three classes):

$$\boldsymbol{x}_t = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \boldsymbol{x}_{t-1} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} a + \boldsymbol{w}_{t-1}$$

where T is the sampling interval and a represents the input acceleration with limits $|a| \leq L_i$, where $L_i = 1$g, 3g and 5g for class c_1, c_2, c_3, respectively (g$= 9.81$ m/s^2 is the acceleration due to gravity). The measurement equation is given by (2) where $\boldsymbol{H} = [1 \quad 0]$. The motion with small acceleration ($|a| < 1$g) we will refer to as nearly constant velocity (CV) motion.

The question is how to implement the JTC scheme shown in figure 1 for this example. The bank of filters in JTC scheme can be more easily tuned to detect the *behavior* of targets, rather than their class. Let $B = \{b_1, \dots, b_n\}$ be the set of possible behaviors and $p(b_j|\boldsymbol{Z}_t)$, $j = 1, \dots, n$, be a vector of behavior probabilities. In our example $C = \{$commercial, bomber, fighter$\}$, while B can be for example $B = \{$nearly CV, slow turn, sharp turn$\}$. A reasonable mapping rule between B and C is:

$$\begin{array}{rll} \text{`sharp turn} & \text{implies} & \text{fighter',} \\ \text{`slow turn} & \text{implies} & \text{fighter or bomber', and} \\ \text{`nearly CV} & \text{implies} & \text{fighter or bomber or commercial'.} \end{array} \tag{4}$$

In general, behaviors are related to the classes by a matrix $\boldsymbol{M} = [M_{ij}]$, where $M_{ij} = p(c_i|b_j)$. The beliefs on C are constructed using $p(c_i|\boldsymbol{Z}_t) = \sum_j M_{ij} \cdot p(b_j|\boldsymbol{Z}_t)$, for all $i = 1, \dots, s$.

How does one choose the matrix $\boldsymbol{M}$? When behaviors and classes are in one-to-one correspondence, indexes can be organized so that $b_i \equiv c_i$, in which case $\boldsymbol{M} = \boldsymbol{I}$ and $p(c_i|\boldsymbol{Z}_t) = p(b_i|\boldsymbol{Z}_t)$. In our example, however, it may be better to use the following matrix:

$$\boldsymbol{M} = \begin{bmatrix} 1/3 & 0 & 0 \\ 1/3 & 1/2 & 0 \\ 1/3 & 1/2 & 1 \end{bmatrix} \tag{5}$$

which corresponds to rule (4), that is: if $p(b_1) = 1$, all three target classes will be of equal probability; if $p(b_2) = 1$, target classes 2 and 3 will be of equal probability; if $p(b_3) = 1$, target class 3 is certain.

Next we analyze the classification results obtained by Monte Carlo simulations using the JTC scheme of figure 1. Matrix $\boldsymbol{M}$ is the identity matrix or the matrix given by equation (5). The target is moving with CV in the first 36 scans, followed by 4 scans of acceleration with 2g (which is not compatible with a commercial plane), and then another 40 scans of CV motion. The sampling interval is $T = 3$ s.

The filters in JTC scheme are tuned to the behavior set B: a KF is used for behavior b_1; an interactive multiple-model (IMM) filter [10] with three modes (corresponding to $a \in \{-2a, 0, 2a\}$) for behavior b_2; an IMM with 5 modes (corresponding to $a \in \{-4a, -2a, 0, 2a, 4a\}$) for behavior b_3.

We fell that the outcomes before medium acceleration is observed should not support any of the three classes more than any other as the data are perfectly consistent with the three classes, and that after medium acceleration is observed, class c_1 should become impossible and the other two classes should become equi-probable.

Figure 2 shows the classification results averaged over 20 Monte Carlo runs.

- Case (a) uses $\boldsymbol{M} = \boldsymbol{I}$ and equi-priors on the 3 classes. Even though the CV is a steady state motion for all three classes, c_1 becomes rapidly the best supported class. This is due to the fact that commercial planes can only exhibit nearly CV, whereas the other two can exhibit other behaviors. We feel this conclusion is inadequate as none of the three classes should become better supported by a nearly CV behavior. Furthermore, after observing a medium acceleration, the classifier supports class c_2, which is as unjustified for the same argument as above.

- Case (b) uses $\boldsymbol{M}$ given by (5) and equi-priors on the 3 classes. We could hope the better representation of the relationship between the behavior and class would improve the results. It turns out that now class c_3 becomes better supported both before and after the medium acceleration is observed. This result is also unsatisfactory.

- Case (c) uses $\boldsymbol{M}$ given by (5) and priors $p_0^B(b_1) = 0.9999$, $p_0^B(b_2) = p_0^B(b_3) = 0.00005$. By adopting these priors, we can obtain desirable results before the medium acceleration maneuvre (all three classes are equally probable). But once the medium acceleration is observed, results deteriorate again as class c_3 is better supported than class c_2.

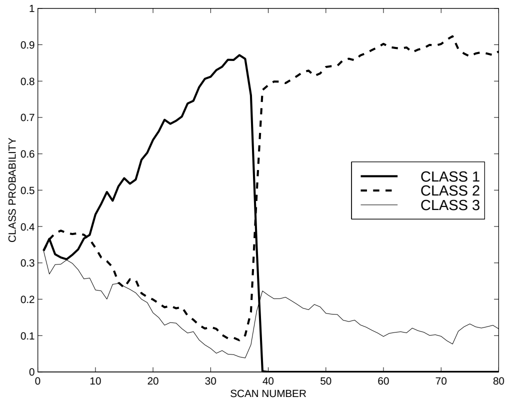
Other choices for matrix $\boldsymbol{M}$ and for the prior do not improve the classification. This example serves as a motivation to consider the classification problem using the transferable belief model. However, we need to look at the JTC problem in its entirety. Indeed creating a JTC where the tracking phase would be done within the probabilistic framework and the classification phase within the TBM framework would be easily criticized for being an opportunist patch work. If we want to apply the TBM in the classification phase, we should also use it in the tracking phase, which is what we do in section 5. It turns out that the classical KF relations developed in the probabilistic framework can also be derived in the TBM framework, with the advantage that many hypotheses underlying the probabilistic KF can be relaxed.

2.3 Background probabilistic relations

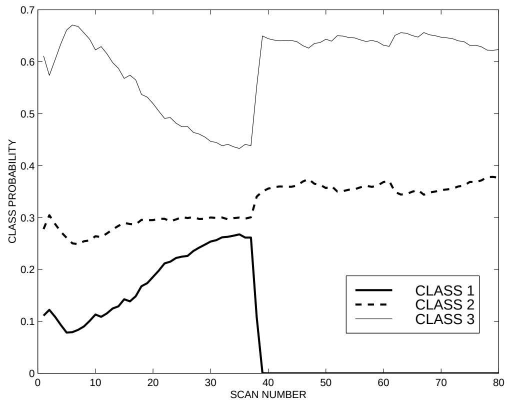
We present first a few well known properties of Gaussian random variables [10].

Lemma 2.1 *Let* $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and let* $\boldsymbol{F}$ *be a* $m \times n$ *matrix. Then* $\boldsymbol{FY} \sim \mathcal{N}_m(\boldsymbol{F\mu}, \boldsymbol{F\Sigma F}')$.

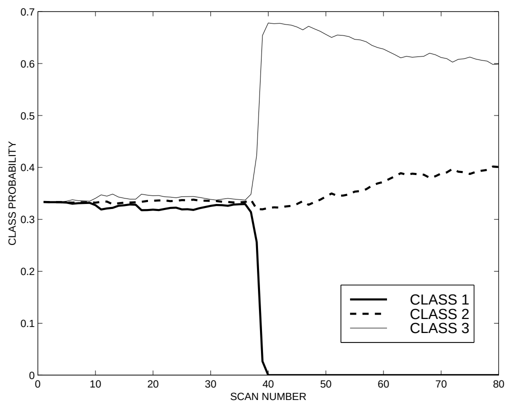
Lemma 2.2 *Let* $\boldsymbol{Y}_i \sim \mathcal{N}_n(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) : i = 1, 2$ *and let* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be two* $m \times n$ *matrices. If* $\boldsymbol{Y}_1$ *and* $\boldsymbol{Y}_2$ *are stochastically independent, then* $\boldsymbol{AY}_1 + \boldsymbol{BY}_2 \sim \mathcal{N}_m(\boldsymbol{A\mu}_1 + \boldsymbol{B\mu}_2, \boldsymbol{A\Sigma}_1\boldsymbol{A}' + \boldsymbol{B\Sigma}_2\boldsymbol{B}')$.



(a)



(b)



(c)

Figure 2: Classification results using JTC scheme in the probabilistic framework. The target is moving with CV in the first 36 scans, followed by 4 scans of acceleration with 2g, and then another 40 scans of CV motion.

Lemma 2.3 *Let* $\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Z} \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} \boldsymbol{\mu_X} \\ \boldsymbol{\mu_Z} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma_{XX}} & \boldsymbol{\Sigma_{XZ}} \\ \boldsymbol{\Sigma_{ZX}} & \boldsymbol{\Sigma_{ZZ}} \end{bmatrix})$. *Then the probability density function about* $\boldsymbol{X}$ *given* $\boldsymbol{Z} = \boldsymbol{z}$ *is given by* $\boldsymbol{X}|\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu_{X|z}}, \boldsymbol{\Sigma_{X|z}})$ *where*

$$\boldsymbol{\mu_{X|z}} = \boldsymbol{\mu_X} + \boldsymbol{\Sigma_{XZ}}\boldsymbol{\Sigma_{ZZ}^{-1}}(\boldsymbol{z} - \boldsymbol{\mu_Z})$$

and

$$\boldsymbol{\Sigma_{X|z}} = \boldsymbol{\Sigma_{XX}} - \boldsymbol{\Sigma_{XZ}}\boldsymbol{\Sigma_{ZZ}^{-1}}\boldsymbol{\Sigma_{ZX}}.$$

2.4 Kalman Filter Equations

Relations and notations for the Kalman filter are strongly influenced by [10]. For simplicity sake, we assume that the noise components and the parameters of the system are time independent (the $\boldsymbol{R}$, $\boldsymbol{Q}$, $\boldsymbol{F}$ and $\boldsymbol{H}$ matrices of section 2.1). Generalization to the time-dependent case is very simple, but the notation becomes so cumbersome that the major underlying ideas get hidden.

Consider the dynamic and measurement equations introduced by (1) and (2), respectively. Suppose temporarily that there is only one class and the prior state is Gaussian distributed, i.e.

$$\boldsymbol{x}_0 \sim N_n(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \tag{6}$$

The KF performs filtering in two phases (update and prediction):

- the updated state where one computes the beliefs about $\boldsymbol{x}_t$ given beliefs relative to $\hat{\boldsymbol{x}}_{t|t-1}$ induced by the past data and given the present measurement $\boldsymbol{z}_t$. It results in a probability function about $\boldsymbol{x}_{t|t}$.
- the predicted state where one computes the beliefs about the next state $\boldsymbol{x}_{t+1}$ given the beliefs built for the updated phase. It results in a probability function about $\boldsymbol{x}_{t+1|t}$.

In Fig: 3, we present the different components of the Kalman filtering process for what concerns the tracking phase.

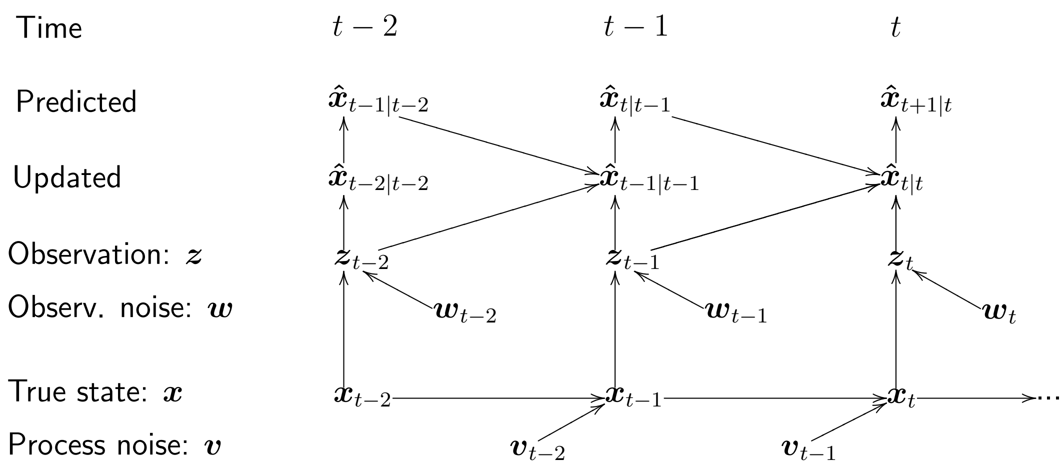


Figure 3: The Kalman filter for the tracking phase

These beliefs are computed using the following relations.

Definition 2.1 *We use the next notations:*

$$\hat{\boldsymbol{x}}_{j|t} \triangleq E(\boldsymbol{x}_j|\boldsymbol{z}_t) \tag{7}$$

$$\boldsymbol{\Sigma}_{j|t} \triangleq E((\boldsymbol{x}_j - \hat{\boldsymbol{x}}_{j|t})(\boldsymbol{x}_j - \hat{\boldsymbol{x}}_{j|t})'|\boldsymbol{z}_t) \tag{8}$$

The relations at $t = 0$ are:

$$\hat{\boldsymbol{x}}_{0|0} = \boldsymbol{\mu}_0, \tag{9}$$

$$\boldsymbol{\Sigma}_{0|0} = \boldsymbol{\Sigma}_0 \tag{10}$$

For $t = 1, 2, \ldots$, the relations are:

Predicted state

$$\boldsymbol{x}_{t|t-1} \sim \mathcal{N}(\hat{\boldsymbol{x}}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}) \tag{11}$$

$$\hat{\boldsymbol{x}}_{t|t-1} = \boldsymbol{F}\hat{\boldsymbol{x}}_{t-1|t-1} \tag{12}$$

$$\boldsymbol{\Sigma}_{t|t-1} = \boldsymbol{F}\boldsymbol{\Sigma}_{t-1|t-1}\boldsymbol{F}' + \boldsymbol{Q} \tag{13}$$

Update state

$$\hat{\boldsymbol{z}}_{t|t-1} = \boldsymbol{H}\hat{\boldsymbol{x}}_{t|t-1} \tag{14}$$

$$\boldsymbol{\nu}_t = \boldsymbol{z}_t - \hat{\boldsymbol{z}}_{t|t-1} \tag{15}$$

$$\boldsymbol{S}_t = \boldsymbol{H}\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}' + \boldsymbol{R} \tag{16}$$

$$\boldsymbol{W}_t = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}'\boldsymbol{S}_t^{-1} \tag{17}$$

$$\boldsymbol{x}_{t|t} \sim \mathcal{N}(\hat{\boldsymbol{x}}_{t|t}, \boldsymbol{\Sigma}_{t|t}) \tag{18}$$

$$\hat{\boldsymbol{x}}_{t|t} = \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{W}_t\boldsymbol{\nu}_t \tag{19}$$

$$\boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{W}_t\boldsymbol{S}_t\boldsymbol{W}_t' \tag{20}$$

The likelihood of the measurement sequence is required for classification. It is given by:

$$p(\boldsymbol{Z}_t) = \prod_{i=1,\ldots,t} p(\boldsymbol{z}_i|\boldsymbol{Z}_{i-1}) \tag{21}$$

where $p(\boldsymbol{z}_i|\boldsymbol{Z}_{i-1}) = \mathcal{N}(\boldsymbol{z}_i; \hat{\boldsymbol{z}}_{i|i-1}, \boldsymbol{S}_i)$. Thus:

$$p(\boldsymbol{Z}_t) = p(\boldsymbol{Z}_{t-1})\mathcal{N}(\boldsymbol{z}_t; \hat{\boldsymbol{z}}_{t|t-1}, \boldsymbol{S}_t).$$

3 TBM Background

We introduce some preliminary concepts related to the TBM.

The transferable belief model (TBM) [9, 17, 6] is a model to represent quantified beliefs based on the belief function theory developed by Shafer [18], but completely unrelated to any underlying probabilistic constraints as it is the case with the model of Dempster [19] and with the hint model [14]. These differences are not important here.

The essential tool is the basic belief assignment (bba) m^Ω which maps subsets of its domain Ω to $[0, 1]$. Its value $m^\Omega(A)$ for $A \subseteq \Omega$ is called the basic belief mass (bbm) when Ω in countable and the basic belief density (bbd) when Ω is the set of reals. It represents the amount - or density - of belief that specifically supports that the actual value of the variable on which beliefs are expressed belongs to A, and that supports nothing more specific due to a lack of information, but that might support any strict subset of A if further information justifies it.

3.1 Notation

We use the next notation to express the bbas. Suppose a variable V which possible values belong to the frame of discernment Ω. The bba relative to the the actual value ω_0 of the variable V and based on the conditioning facts $cond$ is denoted as $m^\Omega\{V\}[cond]$. The expression $m^\Omega\{V\}[cond](X)$ denoted the value taken by the bba at $X \subseteq \Omega$. The same notation is used for the functions related to the bba, like bel and pl. Some indices are neglected when the context makes them obvious.

We recall the definitions of a refinement, coarsening and vacuous extension.

Definition 3.1 (The refinement) *Let X be a frame of discernment. We say that Y is a refinement of X iff there exists a mapping $R : X \to Y$ such that $R(x) \subseteq Y, R(x) \neq \emptyset$ for all $x \in X$, and $R(x_1) \cap R(x_2) = \emptyset, x_1, x_2 \in X, x_1 \neq x_2$.*

Definition 3.2 (The coarsening) *Let Y be a frame of discernment. We say that X is a coarsening of Y iff Y is a refinement of X.*

Definition 3.3 (The vacuous extension: $\uparrow$) *Let X be a frame of discernment and let Y be a refinement of X based on R. Let m^X be a bba on X. Its vacuous extension $m^{X\uparrow Y}$ on Y is given by:*

$$m^{X\uparrow Y}(W) = \begin{cases} m^X(A), & \text{if } W = R(A), A \subseteq X \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

3.2 Credal variables

A credal variable is a variable for which we have defined a bba. The symbol $m^{R^n}\{\boldsymbol{x}\}$ denotes the bba of the credal variable $\boldsymbol{x}$ whose domain is R^n. The superscript R^n is often omitted in notation.

A credal space is a triple $(\Omega, \mathcal{A}, m)$ where Ω is a set, $\mathcal{A}$ an algebra on Ω (closed under union, intersection, complement, with $\emptyset$ and Ω), and m is a bba on Ω. If the credal variable $\boldsymbol{x}$ is defined on a credal space $(\Omega, \mathcal{A}, m)$, we write $\boldsymbol{x} \sim m$.

A probability space is the credal space where m satisfies $m(A) = 0,\ \forall A : |A| \neq 1.$.

If Ω is finite or countable, $\mathcal{A} = 2^{\Omega}$, the power set of $\Omega.$.

If $\Omega = R$, $\mathcal{A}$ is the Borel sigma-algebra on the set of real numbers [15, 16].

If $\Omega = R^n$, $\mathcal{A}$ is the cross product of n Borel sigma-algebra as just defined.

A vacuous credal variable $\boldsymbol{x}$ on Ω is a credal variable $\boldsymbol{x}$ which $bel^{\Omega}\{\boldsymbol{x}\}$ function is a vacuous belief function. Thus its related plausibility function $pl^{\Omega}\{\boldsymbol{x}\}$ satisfies $pl^{\Omega}\{\boldsymbol{x}\}(A) = 1, \forall A \subseteq \Omega, A \neq \emptyset$. We denote it $\boldsymbol{x} \sim VBF$.

3.3 The Gaussian Bayesian Belief Functions

To differentiate between a Gaussian pdf and a Gaussian Bayesian belief function, we use notation $\mathcal{N}$ for the former and NB for the latter. Mathematically, they are the same functions, the difference lies in their semantics.

Definition 3.4 $\boldsymbol{x} \sim NB_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *means that $\boldsymbol{x}$ is a n-dimensional credal variable on R^n whose basic belief density allocates non-zero densities only to the singletons of R^n and these densities are those of the n-dimensional Gaussian distribution of mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.*

Let $m\{\boldsymbol{x}\}$ denote the basic belief density (bbd) that represents the basic belief assignment *(bba) relative to $\boldsymbol{x}$. Then:*

$$m\{\boldsymbol{x}\}(A) = \begin{cases} \mathcal{N}(\boldsymbol{a};\boldsymbol{\mu},\boldsymbol{\Sigma}) & \textit{if } A = \{\boldsymbol{a}\}, \boldsymbol{a} \in R^n \\ 0 & \textit{otherwise.} \end{cases} \tag{23}$$

We also use the notation $m\{\boldsymbol{x}\} = NB(\boldsymbol{\mu},\boldsymbol{\Sigma})$ and we refer to $m\{\boldsymbol{x}\}$ as the Gaussian bbd (meaning that $\boldsymbol{x}$ is Gaussian).

The index n is omitted when the state dimension is clear from the context.

We use $\textcircled{\scriptsize$\cap$}$ and $\oplus$ to denote the conjunctive combination rule and Dempster's rule of combination (the normalized conjunctive combination rule), respectively.

We present a few useful lemmas. Lemma 3.1 to lemma 3.3 are just rephrasing equivalent theorems of probability theory.

Lemma 3.1 *If $\boldsymbol{x} \sim NB(\boldsymbol{\mu},\boldsymbol{\Sigma})$, then $\boldsymbol{F}\boldsymbol{x} \sim NB(\boldsymbol{F}\boldsymbol{\mu},\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F'})$.*

Lemma 3.2 *Let $\boldsymbol{x} \sim NB(\boldsymbol{\mu}_X,\boldsymbol{\Sigma}_X)$ and $\boldsymbol{y} \sim NB(\boldsymbol{\mu}_Y,\boldsymbol{\Sigma}_Y)$. Let $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}$. Then $\boldsymbol{z} \sim NB(\boldsymbol{A}\boldsymbol{\mu}_{\boldsymbol{X}} + \boldsymbol{B}\boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{A'} + \boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{B'})$.*

Lemma 3.3 *Let $\boldsymbol{x} \sim NB(\boldsymbol{\mu}_{\boldsymbol{X}},\boldsymbol{\Sigma}_{\boldsymbol{X}})$ and $\boldsymbol{w} \sim NB(\boldsymbol{0},\ \boldsymbol{R})$. Let $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}$. Then the conditional bbd on $\boldsymbol{x}$ given $\boldsymbol{z}$ is given by*

$$\boldsymbol{x}|\boldsymbol{z} \sim NB(\boldsymbol{A}^{-1}\boldsymbol{z}, \boldsymbol{A}^{-1}\boldsymbol{R}\boldsymbol{A}^{-1\prime}).$$

Lemma 3.4 *Let $\boldsymbol{x} \in R^n$. Let f_1 and f_2 are two pdf on R^n. For $i = 1, 2$, let $m_i\{\boldsymbol{x}\}$ be two bbds relative to $\boldsymbol{x}$ with*

$$m_i\{\boldsymbol{x}\}(A) = \begin{cases} f_i(\boldsymbol{a}) & \textit{if } A = \{\boldsymbol{a}\}, \boldsymbol{a} \in R^n \\ 0 & \textit{otherwise.} \end{cases} \tag{24}$$

Then

$$m_{1 \textcircled{\cap} 2}\{\boldsymbol{x}\}(A) \triangleq m_1\{\boldsymbol{x}\} \textcircled{\cap} m_2\{\boldsymbol{x}\}(A) = \begin{cases} f_1(\boldsymbol{a}) f_2(\boldsymbol{a}) & \textit{if } A = \{\boldsymbol{a}\}, \boldsymbol{a} \in R^n \\ 0 & \textit{otherwise.} \end{cases} \tag{25}$$

Proof. The only non null densities are given to the singletons. For the combination, the product $f_1(\boldsymbol{a}_1) f_2(\boldsymbol{a}_2)$ is allocated to $\boldsymbol{a}$ if $\boldsymbol{a}_1 = \boldsymbol{a}_2 = \boldsymbol{a}$, else to the empty set. □

Lemma 3.5 *Let $\boldsymbol{x} \in R^n$. Let $m_1\{\boldsymbol{x}\} = NB(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $m_2\{\boldsymbol{x}\} = NB(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ be two bbds relative to $\boldsymbol{x}$. Then (after normalization)*

$$m_{1 \oplus 2}\{\boldsymbol{x}\} = m_1\{\boldsymbol{x}\} \oplus m_2\{\boldsymbol{x}\} = NB(\boldsymbol{\nu}, \boldsymbol{S})$$

where $\boldsymbol{S}^{-1} = \boldsymbol{\Sigma_1}^{-1} + \boldsymbol{\Sigma_2}^{-1}$ and $\boldsymbol{\nu} = \boldsymbol{S}(\boldsymbol{\Sigma_1}^{-1}\boldsymbol{\mu_1} + \boldsymbol{\Sigma_2}^{-1}\boldsymbol{\mu_2})$.

Proof. By lemma 3.4, the result of the combination is a bbf and its value is computed from the pointwise product of the two underlying Gaussian pdfs. □

3.4 Decision making in TBM

When a decision must be made and uncertainty is represented by the TBM, the decision maker derives the so-called pignistic probability function using the pignistic transformation. The result is just a probability function that is used to make decisions using the classical expected utility theory. Given a bba m^Ω on a finite Ω, the pignistic probability function (denoted $BetP$) is defined as :

$$BetP(Y) = \sum_{X \subseteq \Omega} \frac{|Y \cap X|}{|X|} \frac{m^\Omega(X)}{1 - m^\Omega(\emptyset)}, \ \forall Y \subseteq \Omega \tag{26}$$

The nature of the pignistic transformation given by (26) is presented in [20, 9]. Its detailed justification is presented in [21].

When the credal variable X is defined on R, we end up with a pignistic probability density function $Betf$ which is defined as:

$$Betf(a) = \lim_{\varepsilon \to 0} \int_{x=-\infty}^{x=a} \int_{y=a+\varepsilon}^{y=\infty} \frac{m^R(X \in [x, y])}{y - x} dydx. \tag{27}$$

Details can be found in [15].

3.5 Doxastic independence

The concept of doxastic independence is the extension of the concept of stochastic independence in the TBM framework [22, 23]. Its syntactical definition is given by:

Definition 3.5 *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two credal variables defined on R^{n_1} and R^{n_2}, respectively. Let $m\{\boldsymbol{X}\}$ and $m\{\boldsymbol{Y}\}$ be their bbds. The credal variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are doxastically independent iff*

$$m \propto m\{\boldsymbol{X}\} \cap m\{\boldsymbol{Y}\}.$$

This definition generalizes the definition $P(A \cap B) = P(A)P(B)$ encountered in probability theory.

The major property used here is that stochastic independence also holds between their pignistic probabilities [22]. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two doxastically independent credal variables. Then

$$Betf\{\boldsymbol{X}, \boldsymbol{Y}\}(\boldsymbol{x}, \boldsymbol{y}) = Betf\{\boldsymbol{X}\}(\boldsymbol{x}) \cdot Betf\{\boldsymbol{Y}\}(\boldsymbol{y}).$$

Hence the pignistic transformation of two doxastically independent credal variables is the product of the pignistic transformation of the individual variables, a very satisfactory property.

4 JTC in the TBM framework

The overall model for JTC is represented by relations (1) and (2), where $\boldsymbol{v}$ and $\boldsymbol{w}$ are credal variables. The difference between this model and its classical version resides in the fact that belief functions replace the probability functions.

In a nutshell, we will:

1. keep the assumptions about the dynamic of the state vectors (1) and about the kinematic measurements (2),
2. relax the assumptions about the initial state vector (6) and replace it by $\boldsymbol{x}_0 \sim VBF$,
3. accept that the additive process noise $\boldsymbol{v}_t$ in (1) is a credal variable; its bbd is unknown but its pignistic transformation $Betf\{\boldsymbol{v}\}$ is a Gaussian pdf: $Betf\{\boldsymbol{v}\} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q})$,
4. accept that the additive measurement noise $\boldsymbol{w}_t$ in (2) is Gaussian credal variable: $\boldsymbol{w}_t \sim NB(\mathbf{0}, \boldsymbol{R})$,
5. select the most committed admissible bbd for the credal variable $\boldsymbol{x}_{t+1|t}$
6. use the General Bayesian Theorem for the final classification phase with a vacuous *a priori* belief on B and a $\boldsymbol{M}$ matrix that better translates the implication rules.

The whole prediction phase will be essentially the same as with the probability approach. The gain obtained by relaxing (6) about $\boldsymbol{x}_0$ is really not that essential. All it does is avoiding the "quarrel" about the choice of an adequate prior on $\boldsymbol{x}_0$. Relaxing constraints on $\boldsymbol{v}_t$ is surely more impressive, as we do not even require knowing its underlying bbd, all we use is its pignistic transformation. Selecting the most committed admissible bbd is arguable, but our purpose is to show what are the assumptions underlying the Kalman filter relations within the TBM, not to build a new set of Kalman filter relations.

Given the relaxed assumptions, in particular the third one about $\boldsymbol{v}_t$, the bbds of $\boldsymbol{x}_{t+1|t}$ and $\boldsymbol{x}_{t+1|t+1}$ cannot be derived. Still we can derive their related pignistic probabilities, and as far as this happens to be all we need in practice, ignoring the bbds is not a real issue.

What is interesting here is that the classical relations used in the Kalman filter can be justified within the TBM.

Major discrepancies appear in the classification phase. For the previous example of section 2.2, we will use conditional belief functions that represent exactly the rules between classes and behaviors. Such a representation is not achievable in probability theory.

5 Kalman filter in the TBM framework

We derive first the KF relations in the TBM framework.

As the initial assumptions on $\boldsymbol{x}_0$ are not those of the classical Kalman filters, we must derive the properties of $\boldsymbol{x}_{1|0}$, $\boldsymbol{z}_1$ and $\boldsymbol{x}_{1|1}$.

We then proceed with time $t = 2$ and derive the properties of $\boldsymbol{x}_{2|1}$, $\boldsymbol{z}_2$ and $\boldsymbol{x}_{2|2}$ which turn out to be those of the classical KF for what concerns their pignistic probabilities. We can just proceed then as with the classical KF relations.

5.1 Predicted state for t_1: bba on $\boldsymbol{x}_1$ induced by $\boldsymbol{x}_0$

Let $\boldsymbol{x}_0$ be a vacuous credal variable on R^n: $\boldsymbol{x}_0 \sim VBF$. Let $\boldsymbol{v}_0 \in R^n$ be a credal variable whose bbd is unknown but which pignistic transformation is Gaussian: $Betf\{\boldsymbol{v}\} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q})$. Let $\boldsymbol{w}_1 \in R^m$ satisfies $\boldsymbol{w}_1 \sim NB(\mathbf{0}, \boldsymbol{R})$. Assume $\boldsymbol{v}_0$ and $\boldsymbol{w}_1$ are doxastically independent credal variables. From (1) we have:

$$\boldsymbol{x}_1 = \boldsymbol{F}\boldsymbol{x}_0 + \boldsymbol{v}_0.$$

Since $\boldsymbol{x}_0 \sim VBF$, then $\boldsymbol{F}\boldsymbol{x}_0 \sim VBF$.

We ignore $m^{R^n}\{\boldsymbol{v}_0\}$. Still whatever $m^{R^n}\{\boldsymbol{v}_0\}$, $\boldsymbol{x}_1 \sim VBF$.

Formally, we combine the two bbds after vacuously extending them on their product space.

$$m^{R^{2n}}\{\boldsymbol{F}\boldsymbol{x}_0, \boldsymbol{v}_0\} = m^{R^n}\{\boldsymbol{F}\boldsymbol{x}_0\}^{\uparrow R^{2n}} \circledcap m^{R^n}\{\boldsymbol{v}_0\}^{\uparrow R^{2n}}$$

The bbd satisfies:

$$m^{R^{2n}}\{\boldsymbol{F}\boldsymbol{x}_0, \boldsymbol{v}_1\}(\boldsymbol{x}, \boldsymbol{v}) = \begin{cases} m^{R^n}\{\boldsymbol{v}_1\}(\boldsymbol{v}) & \text{if } \boldsymbol{x} = R^n \text{ and } \boldsymbol{v} \subseteq R^n \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

The bba $m^{R^n}\{\boldsymbol{x}_1\}$ for $\boldsymbol{x}_1 = \boldsymbol{F}\boldsymbol{x}_0 + \boldsymbol{v}_0$ is the result of a coarsening of $m^{R^{2n}}\{\boldsymbol{F}\boldsymbol{x}_0, \boldsymbol{v}_0\}$ which is a vacuous belief function ($\boldsymbol{x}_1 \sim VBF$).

This result reflects the natural rule that if one adds two terms, the value of one of them being completely unknown, the result is also completely unknown.

5.2 Updated state for t_1: bba on $\boldsymbol{x}_1$ induced by $\boldsymbol{x}_0$ and $\boldsymbol{z}_1$

For simplicity we assume that $m = n$. Generalization to $m \neq n$ is feasible, but equations become more complex, hiding the message of the paper.

From (2) we have:

$$\boldsymbol{z}_1 = \boldsymbol{H}\boldsymbol{x}_1 + \boldsymbol{w}_1$$

We can now use two different approaches with identical outcomes:

1. We can use the just derived bba $m^{R^n}\{\boldsymbol{x}_1\}$, vacuously extend it on the R^{2n} space, vacuously extend the a priori belief about $\boldsymbol{w}_1$ on the R^{2n} space and combine them with the conjunctive combination rule. Then we condition the result on the observation $\boldsymbol{z}_1$, and marginalize the result on $\boldsymbol{x}_1$ domain. This last bba would be the final bba on $\boldsymbol{x}_1$ induced by $\boldsymbol{x}_0$ and $\boldsymbol{z}_1$.

2. We can consider the conditional belief over $\boldsymbol{z}_1$ given $\boldsymbol{x}_1$. The conditional bbds are given by $m^{R^n}\{\boldsymbol{z}_1\}[\boldsymbol{x}_1] = NB(\boldsymbol{H}\boldsymbol{x}_1,\ \boldsymbol{R})$. We apply credal inference to determine the bbd on $\boldsymbol{x}_1$ given the observation $\boldsymbol{z}_1$. The result is $m^{R^n}\{\boldsymbol{x}_1\}[\boldsymbol{z}_1] \sim NB(\boldsymbol{H^{-1}}\boldsymbol{z}_1, \boldsymbol{H^{-1}RH^{-1\prime}})$. We conjunctively combine this bbd with the just derived bba $m^{R^n}\{\boldsymbol{x}_1\}$.

The argument used to show $\boldsymbol{x}_1 \sim VBF$ leads similarly to $\boldsymbol{z}_1 \sim VBF$.

For $\boldsymbol{x}_{1|1}$, we have $\boldsymbol{x}_{1|1} = \boldsymbol{H}^{-1}(\boldsymbol{z}_1 - \boldsymbol{w}_1)$ where $\boldsymbol{z}_1$ is the observed value and $\boldsymbol{w}_1 \sim NB(\boldsymbol{0}, \boldsymbol{R})$. Therefore $\boldsymbol{x}_{1|1} \sim NB(\hat{\boldsymbol{x}}_{1|1}, \boldsymbol{\Sigma}_{1|1})$ where $\hat{\boldsymbol{x}}_{1|1} = \boldsymbol{H}^{-1}\boldsymbol{z}_1$ and $\boldsymbol{\Sigma}_{1|1} = \boldsymbol{H}^{-1}\boldsymbol{R}\boldsymbol{H}'^{-1}$.

The left half of table 1 summarizes the results at time t_1.

	Time t_1			Time t_2		
	Variable	bbd	$Betf$	Variable	bbd	$Betf$
prior	$\boldsymbol{x}_0$	VBF	-	$\boldsymbol{x}_{1\|1}$	?	Gauss
process noise	$\boldsymbol{v}_0$	?	Gauss	$\boldsymbol{v}_1$	?	Gauss
observ. noise	$\boldsymbol{w}_1$	Gauss	Gauss	$\boldsymbol{w}_2$	Gauss	Gauss
updated state	$\boldsymbol{x}_{1\|1}$	?	Gauss	$\boldsymbol{x}_{2\|2}$	?	Gauss
predicted state	$\boldsymbol{x}_{2\|1}$	VBF	-	$\boldsymbol{x}_{3\|2}$	?	Gauss

Table 1: Bba and $Betf$ computed at times t_1 and t_2.

5.3 Predicted state for t_2: bba on $\boldsymbol{x}_{2|1}$ induced by $\boldsymbol{x}_{1|1}$

The input is $\boldsymbol{x}_{1|1} \sim NB(\hat{\boldsymbol{x}}_{1|1}, \boldsymbol{\Sigma}_{1|1})$ and $\boldsymbol{v}_1$ for which we ignore its bbd but we know its pignistic probabilities. Thus we know the pignistic probabilities of both variables, and thus we know the pignistic probability function on their joint space, out of which we compute (by convolution) the pignistic probabilities relative to their sum. For what concerns the pignistic probabilities, probability theory applies and we have: $Betf\{\boldsymbol{x}_{2|1}\} = N(\hat{\boldsymbol{x}}_{2|1}, \boldsymbol{\Sigma}_{2|1})$ where $\hat{\boldsymbol{x}}_{2|1} = F\hat{\boldsymbol{x}}_{1|1}$ and

$\boldsymbol{\Sigma}_{2|1} = \boldsymbol{F}\boldsymbol{\Sigma}_{1|1}\boldsymbol{F}' + \boldsymbol{Q}$. Still, the bbd of $\boldsymbol{x}_{2|1}$ is unknown because the bbd of $\boldsymbol{v}_1$ is unknown.

5.4 Updated state for t_2: bba on $\boldsymbol{x}_{2|2}$ induced by $\boldsymbol{x}_{2|1}$ and $\boldsymbol{z}_2$

The input is $\boldsymbol{x}_{2|1}$ which bbd is unknown but which pignistic probability function is Gaussian and $\boldsymbol{w}_1$ which bbd is is a Gaussian Bayesian belief function.

For $\boldsymbol{z}_2$, we ignore its bbd but we know its pignistic probability function: $Betf\{\boldsymbol{z}_2\} = N(\hat{\boldsymbol{z}}_{2|1}, \boldsymbol{S}_2)$ where $\hat{\boldsymbol{z}}_{2|1} = \boldsymbol{H}\hat{\boldsymbol{x}}_{2|1}$ and $\boldsymbol{S}_2 = \boldsymbol{H}\boldsymbol{\Sigma}_{2|1}\boldsymbol{H}' + \boldsymbol{R}$.

For $\boldsymbol{x}_{2|2}$, if we knew the bbd $m^{R^n}\{\boldsymbol{x}_{2|1}\}$, we would vacuously extend it on the R^{2n} space. We would also vacuously extend the belief about $\boldsymbol{w}_2$ on the R^{2n} space, and combine these two bbds with the conjunctive combination rule. Then we condition the result on the observation $\boldsymbol{z}_2$, and marginalize the result on $\boldsymbol{x}_{2|2}$ domain. This last bbd would be the bbd on $\boldsymbol{x}_{2|2}$ induced by $\boldsymbol{x}_{2|1}$ and $\boldsymbol{z}_2$.

Unfortunately, we do not know $m^{R^n}\{\boldsymbol{x}_{2|1}\}$. We can then invoke the principle of maximum commitment that states: keep as much information as possible, and select for $\boldsymbol{x}_{2|1}$ the most committed normalized bbd among those which pignistic transformation is $Betf\{\boldsymbol{x}_{2|1}\}$. The solution is the Bayesian belief function corresponding to the pignistic probability function. Thus $\boldsymbol{x}_{2|1} \sim NB(\hat{\boldsymbol{x}}_{2|1}, \boldsymbol{\Sigma}_{2|1})$. In that case, both belief functions on R^{2n} are bbfs, and probability theory just applies.

We get $\boldsymbol{x}_{2|2} \sim NB(\hat{\boldsymbol{x}}_{2|2}, \boldsymbol{\Sigma}_{2|2})$ where $\hat{\boldsymbol{x}}_{2|2}$ and $\boldsymbol{\Sigma}_{2|2}$ are given by relations (9) and (20), respectively.

During this derivation, it became clear that once all $Betf$ are Gaussian, the relations for $Betf$ are those of the KF, what we just summarize in the next lemma.

Lemma 5.1 *In a Kalman filter with Gaussian noises, once the state $\boldsymbol{x}_t$ is Gaussian, all successive states satisfy the relations described in the classical Kalman filter theory.*

Being back into the classical KF relations setting, we can proceed for $t_3, t_4, \ldots$ just as with a classical KF for the prediction phases using relations (11)–(20). The only particularity is that the pignistic probability function on $\boldsymbol{z}_t$ is known but not its bbd. So for the classification phase, we will have to reconsider the procedure.

5.5 Diffuse prior and TBM solution

It is worth mentioning that the TBM solution is the same solution one would have obtained if one used the probabilistic approach with a diffuse (also called improper or uninformative) prior on $\boldsymbol{x}_0$. This should not be understood as the fat that the TBM is just a particular probabilist solution. What it shows is that, given our assumptions, the TBM solution produces the same relations as those derived by probability analysis when assuming a diffuse prior. This should be considered essentially as an anecdotical observation.

Still one might ask why bother with the TBM, if all it achieves is what probability theory produces with a diffuse prior.

A possible answer is that a 'diffuse pdf' is not a pdf, and its use violates the foundation of probability theory, even thought its users claim to be using probability theory. The TBM also violates the foundation of probability theory, but purposely and indeed we never claim to be using probability theory.

Another answer is that even though a diffuse prior applied to a probability analysis produces the TBM solution, we can also consider other priors, and the flexibility of the TBM allows us to use any prior on $\boldsymbol{x}_0$, as well in fact as for $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$. Our presentation focuses on finding the KF relations within the TBM framework so we could apply the alternate classification method. Therefore we limit ourselves to a context very similar to the one used in probability theory, but the TBM solution can be generalized and that will be studied in future works.

6 The classification phase in the TBM framework

The JTC in the TBM framework is done conceptually in a similar manner as in the Bayesian framework. Again we have a bank of tracking filters matched to target behavior or class, which (according to the previous section) can be based on classical Kalman filters. The main difference, however, is in the way TBM performs classification.

As in probability theory, classification is based on the likelihood functions (as in eq.(21)). Within the TBM, the likelihood of an hypothesis is equated to the conditional plausibility of the observation given the hypothesis. Let us denote by $l_i = pl(\boldsymbol{Z}_t|b_i)$ the measurement likelihoods which are output from filter i matched to behavior b_i. Furthermore, we consider a vacuous a priori on B. Then the General Bayesian Theorem (GBT) permits us to compute the posterior belief on B. A convenient formulation of the GBT is [11, 12]:

$$m^B[\boldsymbol{Z}_t](b) = \prod_{i:b_i \in b} l_i \prod_{j:b_j \notin b} (1 - l_j), \qquad \forall b \subseteq B.$$

In our study, the bbd of the credal variable $\boldsymbol{z}_t$ is unknown, hence the likelihood cannot be directly derived. But we know the pignistic probability function of $\boldsymbol{z}_t$: it is computed from relation (21). In order to compute the needed likelihoods we need the bbd of $\boldsymbol{Z}_t$. But as we only know its pignistic transformation and there are many bbds that share the same pignistic transformation, we apply the least committed principle that states: never allocate more belief than necessary. It means we select the q-least committed belief functions among those which pignistic transformation is the known one derived from the relation (21). The solution is presented in Appendix A.2.

The relation between B and C can be established in a precise manner on the power set. This is explained using the example of section 2.2. In this example the relation between B and C is described by three conditional belief functions:

$$
\begin{aligned}
m^C[b_1](\{c_1, c_2, c_3\}) &= 1, && (29)\\
m^C[b_2](\{c_2, c_3\}) &= 1, && (30)\\
m^C[b_3](\{c_3\}) &= 1. && (31)
\end{aligned}
$$

Let us represent the bbas as vectors whose elements are ordered as follows (for m^B):

$$
\emptyset, \{b_1\}, \{b_2\}, \{b_1, b_2\}, \{b_3\}, \{b_1, b_3\}, \{b_2, b_3\}, \{b_1, b_2, b_3\},
$$

and similarly for m^C. Then, the derivation of the bba on C given the bba on B and the three conditional belief functions (29)–(31) is achieved using matrix $\overline{\boldsymbol{M}}$ as follows:

$$
m^C = \overline{\boldsymbol{M}} \cdot m^B,
$$

where:

$$
\overline{\boldsymbol{M}} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0\\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1
\end{bmatrix}.
$$

This transformation means that

- $m^B(\emptyset)$ is transferred to $m^C(\emptyset)$
- $m^B(\{b_1\})$ is transferred to $m^C(\{c_1, c_2, c_3\})$ as b_1 implies $\{c_1, c_2, c_3\}$
- $m^B(\{b_2\})$ is transferred to $m^C(\{c_2, c_3\})$ as b_2 implies $\{c_2, c_3\}$

- $m^B(\{b_1, b_2\})$ is transferred to $m^C(\{c_1, c_2, c_3\})$ as b_1 or b_2 implies $\{c_1, c_2, c_3\}$
- $m^B(\{b_3\})$ is transferred to $m^C(\{c_3\})$ as b_3 implies $\{c_3\}$
- $m^B(\{b_1, b_3\})$ is transferred to $m^C(\{c_1, c_2, c_3\})$ as b_1 or b_3 implies $\{c_1, c_2, c_3\}$
- $m^B(\{b_2, b_3\})$ is transferred to $m^C(\{c_2, c_3\})$ as b_2 or b_3 implies $\{c_2, c_3\}$
- $m^B(\{b_1, b_2, b_3\})$ is transferred to $m^C(\{c_1, c_2, c_3\})$ as b_1 or b_2 or b_3 implies $\{c_1, c_2, c_3\}$

These rules are based on the logical property: if a implies x and b implies y, then a or b implies x or y.

Next we reconsider the example of section 2.2. Let $l_1 = 1$, $l_2 = l_3 = 0$. It means that the behavior is b_1. Then $m^B[\boldsymbol{Z}_t](b_1) = 1$ and $m^C[\boldsymbol{Z}_t](C) = 1$. Thus the a posteriori on C is vacuous, as it should be as under b_1, classes are not distinguishable.

Let $l_1 = l_2 = l_3 = 1$. It means that none of the behaviors is better supported than the other two. Then $m^B[\boldsymbol{Z}_t](B) = 1$ and $m^C[\boldsymbol{Z}_t](C) = 1$. Thus the a posteriori on C is also vacuous, as it should be as none of the behaviors is supported.

To further illustrate the theory, we repeat the same experimental setup from section 2.2, with the KF, IMM-3 and IMM-5 filters tuned to the target behavior. We show the resulting pignistic class probabilities, obtained using the described TBM classifier, in figure 3 (obtained by averaging over 20 Monte Carlo runs). The classification results appear reasonable: before the maneuver all three target classes are equally probable, while after the maneuver (which can be performed only by class c_2 and c_3 targets), the probability of class c_1 drops to zero while the probability of class c_2 and c_3 target jumps to $1/2$. The TBM classifier is thus capable of resolving the issues raised in section 2.2, due to its higher flexibility in representing the belief states.

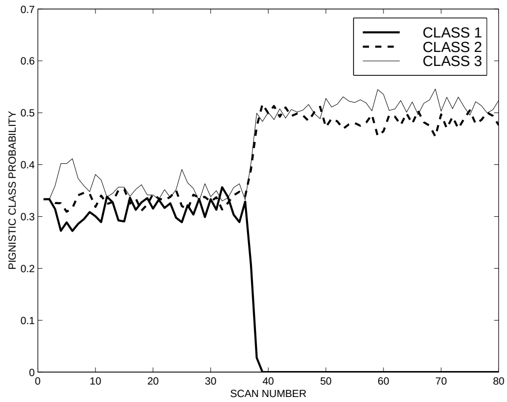


Figure 4: Classification results using JTC scheme in the TBM framework

7 Conclusions

We have shown that it is possible to derive the Kalman filter within the TBM framework. The TBM solution for the tracking (filtering) phase of JTC is essentially the same as the one achieved within the probabilistic framework. For its derivation, however, several assumptions have been already seriously relaxed, and it can still be generalized to more complex contexts (if necessary). For the classification phase of JTC, where we feel the probabilistic approach essentially fails, the TBM offers a solution which is intuitively satisfactory. The overall TBM solution is theoretically sound and coherent, and provides a seemingly better framework for joint tracking and classification.

Acknowledgements

B. Ristic would like to thank IRIDIA, Université libre de Bruxelles, for being his host while the research presented in this paper was carried out.

Appendix A. Belief Functions

A.1 Belief functions on R

This section summarizes results presented in [15].

Consider the real axis R. Let $\mathcal{I}$ be the set of closed intervals in R. We assume that the 'masses' of the bba defined on R are only allocated to the closed intervals of R. As we work on continuous spaces, 'masses' become densities and the basic belief masses are now called basic belief densities.

Let $m^{\mathcal{I}}([a,b])$ be a basic belief density (bbd). The value of $bel^{\mathcal{I}}([a,b])$ is obtained by 'adding' (integrating) all the 'masses' (densities) given to subsets of $[a,b]$. Similarly for $pl^{\mathcal{I}}([a,b])$ where we integrate all the densities which have a non empty intersection with $[a,b]$. The limits for the integrations are the shaded areas in figure 5.

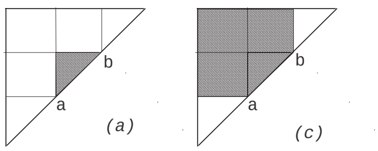


Figure 5: *Graphical representation of (a) belief; (c) plausibility*

The pignistic density function $Betf$ can be derived from $m^{\mathcal{I}}([x,y])$ as follows:

$$Betf(a) = \lim_{\epsilon \to 0} \int_{x=-\infty}^{x=a} \int_{y=a+\epsilon}^{y=\infty} \frac{m^{\mathcal{I}}([x,y])}{y-x}\, dx\, dy, \tag{32}$$

for $a \in R$.

These concepts can be generalized to R^n.

A.2 Ordering bbas and the least committed bbas

Suppose your domain knowledge is partial and based only on some potential betting behaviour, represented by the pignistic density function $Betf(a)$. Since the pignistic transform is many-to-one transform, an infinite number of bbds can induce the same $Betf$. These belief funtions are said to be isopignistic. In order to apply the belief function theory, one needs to formulate a method of building a bbd from the pignistic density. The least commitment principle [6],[24] suggests to choose among all the isopignistic bbds, the one which maximizes the commonality function q. The q-least committed belief density is a consonant bbd. On the real axis R this means that all focal elements on $\mathcal{I}$ are nested, i.e. can be ordered in such a way that each focal interval is contained by the following one [15]. One of the essential properties of consonant plausibility functions is $pl([a,b]) = \max_{x\in[a,b]} pl(x)$.

When the pignistic density $Betf$ is Gaussian ($Betf(x) = \mathcal{N}(x;\mu,\Sigma)$), the plausibility function related to the q-least committed bbd isopignistic with $Betf$ is given by

$$pl([x,y]) = 2(x-\mu)\mathcal{N}(x;\mu,\Sigma) + \int_{t=x}^{t=\infty} 2\mathcal{N}(t;\mu,\Sigma)dt, \qquad y \geq x > \mu \tag{33}$$

$$= 2(\mu-y)\mathcal{N}(y;\mu,\Sigma) + \int_{t=-\infty}^{t=y} 2\mathcal{N}(t;\mu,\Sigma)dt, \qquad \mu > y \geq x \geq \tag{34}$$

$$= 1 \qquad y \geq \mu \geq x \tag{35}$$

References

[1] P. Smets and B. Ristic, "Kalman filters and joint tracking and classification in the TBM framework," in *Proceedings of the Seventh International Conference on Information Fusion* (Fusion04, ed.), (Stockholm, Sweden), pp. IF04–0046, 2004.

[2] S. Challa and G. W. Pulford, "Joint target tracking and classification using radar and ESM sensors," *IEEE Trans. Aerospace and Electronic Systems*, vol. 37, no. 3, pp. 1039–1055, 2001.

[3] A. Farina, P. Lombardo, and M. Marsella, "Joint tracking and identification algorithms for multisensor data," *IEE Proc. Radar Sonar Navig.*, vol. 149, no. 6, pp. 271–280, 2002.

[4] N. Gordon, S. Maskell, and T. Kirubarajan, "Efficient particle filters for joint tracking and classification," in *Proc. SPIE*, vol. 4728, pp. 439–449, 2002.

[5] B. Ristic, N. Gordon, and A. Bessell, "On target classification using kinematic data," *Information Fusion*, pp. 0–0, 2004.

[6] P. Smets, "The transferable belief model for quantified belief representation," in *Handbook of Defeasible Reasoning and Uncertainty Management Systems* (D. M. Gabbay and P. Smets, eds.), vol. 1, pp. 267–301, Kluwer, Doordrecht, The Netherlands, 1998.

[7] P. Smets, "Practical uses of belief functions," in *Uncertainty in Artificial Intelligence 99* (K. B. Laskey and H. Prade, eds.), pp. 612–621, Morgan Kaufman, San Francisco, Ca, 1999.

[8] P. Smets, "Data fusion in the transferable belief model.," in *Proc. 3rd Int. Conf. Information Fusion*, pp. PS 21–33, FUSION 2000, Paris, France, 2000.

[9] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, pp. 191–234, 1994.

[10] V. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation*. New York: Wiley, 2001.

[11] F. Delmotte and P. Smets, "Target identification based on the transferable belief model interpretation of Dempster-Shafer model. pars 1: Methodology.," technical report, IRIDIA, Université Libre de Bruxelles, Belgium., 50 av. Roosevelt, CP194/6, 1050 Brussels, Belgium, 2002.

[12] P. Smets, "Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem," *Int. J. Approximate Reasoning*, vol. 9, pp. 1–35, 1993.

[13] A. Dempster, "Construction and local computation aspects of network belief functions," in *Influencce diagrams, belief nets and decision analysis* (R. M. Oliver and J. Q. Smith, eds.), pp. 121–141, Chichester: John Wiley and Sons, 1990.

[14] J. Kohlas and P. A. Monney, *A Mathematical Theory of Hints: An Approach to Dempster-Shafer Theory of Evidence*. Lecture Notes in Economics and Mathematical Systems No. 425. Springer-Verlag, 1995.

[15] P. Smets, "Belief functions on real numbers," *Submitted: see http://iridia.ulb.ac.be/∼psmets*, 2004.

[16] B. Ristic and P. Smets, "Belief function theory on the continuous space with an application to model based classification," in *Information Processing and Management of Uncertainty* (IPMU-2004, ed.), pp. 1119–1126, 2004.

[17] P. Smets, "What is Dempster-Shafer's model?," in *Advances in the Dempster-Shafer Theory of Evidence* (R. R. Yager, J. Kacprzyk, and M. Fedrizzi, eds.), pp. 5–34, Wiley, New York, 1994.

[18] G. Shafer, *A Mathematical Theory of Evidence*. Princeton Univ. Press. Princeton, NJ, 1976.

[19] A. P. Dempster, "Upper and lower probabilities induced by a multiple valued mapping," *Ann. Math. Statistics*, vol. 38, pp. 325–339, 1967.

[20] P. Smets, "Decision making in a context where uncertainty is represented by belief functions," in *Belief Functions in Business Decisions* (R. P. Srivastava and T. Mock, eds.), pp. 17–61, Physica-Verlag, Heidelberg, Germany, 2002.

[21] P. Smets, "Decision making in the TBM: the necessity of the pignistic transformation," *Int. J. Approximate Reasoning*, vol. 38, pp. 133–147, 2005.

[22] B. Ben Yaghlane, P. Smets, and K. Mellouli, "Belief function independence: I. the marginal case," *Int. J. Approximate Reasoning*, vol. 29, pp. 47–70, 2001.

[23] B. Ben Yaghlane, P. Smets, and K. Mellouli, "Belief function independence: II. the conditional case," *Int. J. Approximate Reasoning*, vol. 31, pp. 31–75, 2002.

[24] D. Dubois and H. Prade, "The principle of minimum specificity as a basis for evidential reasoning," in *Uncertainty in Knowledge-based Systems* (B. Bouchon and R. R. Yager, eds.), pp. 75–84, Springer Verlag, Berlin, 1987.