Modal Trajectory Estimation using Maximum Gaussian Mixture

André Monin

Modal Trajectory Estimation using Maximum Gaussian Mixture

André Monin
CNRS, LAAS, 7 ave du Colonel Roche, F-31077 Toulouse, France.
Univ de Toulouse, LAAS, F-31400 Toulouse, France.

***Abstract*—This paper deals with the estimation of the whole trajectory of a stochastic dynamic system with highest probability, conditionally upon the past observation process, using a maximum Gaussian mixture. We first recall the Gaussian sum technique applied to minimum variance filtering. It is then shown that the same concept of Gaussian mixture can be applied in that context, provided we replace the Sum operator by the Max operator.**

***Index Terms*—Filtering, Smoothing, Gaussian sum, Gaussian mixture.**

I. Introduction

The Gaussian mixture approximation introduced in [1] has become a very popular way to approximate many filtering issues. It is defined classically as a weighted sum of Gaussian probability density functions (pdf) as follows:

$$p\left(x\right)=\sum_{i=1}^{m}\rho_{i}\Gamma\left(x-\bar{x}_{i},P_{i}\right)$$

where $x\in\mathbb{R}^{n}$, and Γ stands for the Gaussian pdf

$$\Gamma\left(x,P\right)=\frac{1}{\sqrt{\left(2\pi\right)^{n}\left|P\right|}}\exp\left(-\frac{1}{2}x^{T}P^{-1}x\right)$$

Nowadays, computation facilities are allowed to implement algorithms based on this approximation. Indeed, in practice, the computational load is equivalent to N Kalman Filters or Extended Kalman Filters (EKFs) in parallel. The value of N depends obviously on the application, but it rarely exceeds a few hundreds. In practice, it may be more efficient than the so-called particle filtering technique which often needs thousands of particles [2] (convergence rate of $1/\sqrt{N}$ versus $1/N$ for deterministic algorithms).

Classically, the Gaussian mixture approximation consists of developing the transition or the observation pdf, or both, as weighted sums of Gaussian densities [1]. Such an approximation permits to deal, for example, with linear systems with non-Gaussian noises and/or non-Gaussian initial pdf, nonlinear systems with Gaussian noises but such that the standard deviation of noises is large with respect to the field of validity of the linearization [1], nonlinear systems with non-Gaussian white noises and multi-modal systems with Markovian commutations [3].

In this paper, the problem of the modal trajectory estimation (MTE) [4] is addressed. The goal is to find the whole trajectory over the horizon $[0,t]$ with the highest probability, conditionally upon the past observation process:

$$x_{0:t}^{*}=\arg\max_{x_{0:t}}\ p\left(x_{0:t}|y_{0:t}\right)$$

where $x_{0:t}\triangleq\{x_{0},\ldots,x_{t}\}$ and $y_{0:t}\triangleq\{y_{0},\ldots,y_{t}\}$. Note that the outcome x_{t}^{*} of the optimal trajectory $x_{0:t}^{*}$ is, in general, not the same as the state obtained by maximizing the marginal density $p\left(x_{t}|y_{0:t}\right)$, the Max a posteriori (MAP) filter, except for the linear Gaussian case (the Kalman filter).

Curiously, there are only few studies on the MTE issue even if it is of real interest, specifically when the conditional probability is multi-modal. In [5], the author states the general problem and then limits itself to the Gaussian case. In [4], the authors address the general case using the dynamic programming approach, as we do, but limiting themselves to the regular case (all pdf are functions but not distributions).

Our main contribution is to show that the same concept of Gaussian mixture filters, first devoted to minimum variance filtering, occurs in the MTE issue provided by replacing the Sum operator by a Max operator (section III). First, we recall in section II-A how the marginal conditional density and in section II-B the marginal Bayesian likelihood propagate in the general Markovian case [4]. Both minimum variance and modal trajectory smoothers are derived using backward computations in the general case. We then show in section II-B that the MTE issue is in most cases ill-posed and then we suggest a solution to regularize this.

II. General minimum variance and modal trajectory estimates

The goal of general optimal filtering theory is to compute the estimate of the internal state $x_{t}\in\mathbb{R}^{n}$ of a stochastic dynamic system partially observed by the process $y_{t}\in\mathbb{R}^{p}$ over the interval $[0,t]$. Generally, the Markovian system model reduced to the transition pdf $p\left(x_{t}|x_{t-1}\right)$ and the observation pdf $p\left(y_{t}|x_{t}\right)$. The optimal filtering concept refers to a particular criterion. There are two main criteria commonly used. The first one is the "minimum variance filter". This estimate is intended to minimize the expectation of the quadratic norm of the error between the process and its estimate, using only the past observations $y_{0:t}$, that is

$$\hat{x}_{t|t}=\arg\min_{\mathcal{F}(\bullet)}\ \mathbb{E}\left[\left\|x_{t}-\mathcal{F}\left(y_{0:t}\right)\right\|^{2}\right]\tag{1}$$

where $\mathcal{F}$ stands for any function of the observation process. It leads to the expectation of the conditional pdf $p\left(x_t|y_{0:t}\right)$:

$$\hat{x}_{t|t} = \int x_t p\left(x_t|y_{0:t}\right) dx_t \tag{2}$$

The second one, named MTE, is intended to maximize the pdf of the state path $x_{0:t}$ conditionally upon the output past, that is [6]:

$$x^*_{0:t} = \arg\max_{x_{0:t}} \; p\left(x_{0:t}|y_{0:t}\right)$$

Note that in most cases, only the outcome of the optimal trajectory x^*_t is of interest as the computation of the whole trajectory $x^*_\tau, \tau < t$ needs a backward calculation which is not generally computable in real-time (as it is the case with the minimum variance smoother $\hat{x}_{\tau|t} = \mathbb{E}\left[x_\tau|y_{0:t}\right]$ with $\tau < t$).

A. Minimum variance estimate

The pdf $p\left(x_t|y_{0:t}\right)$ can be propagated using the following theorem.

Theorem 1: Beginning with the knowledge of the initial pdf $p\left(x_0\right)$ (a measurable function), the *a posteriori* pdf can be computed recursively by:

$$p\left(x_t|y_{0:t}\right) = \frac{p\left(y_t|x_t\right) \int p\left(x_t|x_{t-1}\right) p\left(x_{t-1}|y_{0:t-1}\right) dx_{t-1}}{\int p\left(y_t|x_t\right) p\left(x_t|y_{0:t-1}\right) dx_t} \tag{3}$$

Proof: See [6]. ∎

Note that even if the transition pdf $p\left(x_t|x_{t-1}\right)$ is a distribution, the pdf $p\left(x_t|y_{0:t}\right)$ is in most cases a measurable function due to the smoothing property of the integral operator. The filter can then be computed using 2.

This means that the general solution to optimal filtering, in the minimum variance sense, is achieved by recursively spreading one function of the state (the *a posteriori* pdf of the state) with one integration operation over the state space and one function multiplication.

With the knowledge of the set $\left\{p\left(x_\tau|y_{0:t}\right), \tau = 0, \ldots, t\right\}$, the whole optimal smoother $\hat{x}_{\tau|t} = \mathbb{E}\left[x_\tau|y_{0:t}\right]$, $\tau < t$ can then be computed backward using the following theorem.

Theorem 2: Beginning with the optimal filtering solution $p\left(x_t|y_{0:t}\right)$, the optimal smoother can be computed backward as follows: $\forall \tau = t-1, \ldots, 0$

$$\begin{aligned} &p\left(x_\tau|y_{0:t}\right) = p\left(x_\tau|y_{0:\tau}\right) \\ &\times \int \frac{p\left(x_{\tau+1}|y_{0:t}\right)}{\int p\left(x_{\tau+1}|x_\tau\right) p\left(x_\tau|y_{0:\tau}\right) dx_\tau} p\left(x_{\tau+1}|x_\tau\right) dx_{\tau+1} \end{aligned}$$

Proof: See [6]. ∎

B. Modal trajectory estimate

1) Regular case: Recall that the goal is to maximize the pdf $p\left(x_{0:t}|y_{0:t}\right)$. Assume that this pdf is a measurable function (not a distribution). Using the Bayes rule, this density can be rewritten as

$$p\left(x_{0:t}|y_{0:t}\right) = \frac{p\left(y_{0:t}|x_{0:t}\right) p\left(x_{0:t}\right)}{p\left(y_{0:t}\right)}$$

As the denominator does not depend on the variable to be optimized ($x_{0:t}$), it is equivalent to maximizing the numerator and then to compute

$$x^*_{0:t} = \arg\max_{x_{0:t}} \; J\left(x_{0:t}, y_{0:t}\right)$$

where

$$J\left(x_{0:t}, y_{0:t}\right) \triangleq p\left(y_{0:t}|x_{0:t}\right) p\left(x_{0:t}\right) \tag{4}$$

As it is done dealing with dynamic programming, let us define the marginal Bayesian likelihood:

$$J^*_t\left(x_t, y_{0:t}\right) \triangleq \max_{x_{0:t-1}} J\left(x_{0:t}, y_{0:t}\right) \tag{5}$$

Again, this optimal marginal Bayesian likelihood can be computed recursively.

Theorem 3: Beginning with the knowledge of the initial marginal Bayesian likelihood $J^*_0\left(x_{0,}y_0\right) = p\left(y_0|x_0\right) p\left(x_0\right)$ (a measurable function), the optimal marginal maximum likelihood can be computed recursively by the following equation

$$J^*_t\left(x_t, y_{0:t}\right) = \max_{x_{t-1}} \left(p\left(y_t|x_t\right) p\left(x_t|x_{t-1}\right) J^*_{t-1}\left(x_{t-1}, y_{0:t-1}\right)\right) \tag{6}$$

Proof: See [4]. ∎

The outcome x^*_t of the optimal trajectory $x^*_{0:t}$ can then be computed by

$$x^*_t = \arg\max_{x_t} \; J^*_t\left(x_t, y_{0:t}\right)$$

This means that the general solution to optimal filtering, in the MTE sense, is achieved recursively with one Max operation over the state space and one function multiplication. Note then the similarity between equations (3) and (6) where integration operator is just replaced by a Max operator.

With the knowledge of the set $\left\{J^*_\tau\left(x_\tau, y_{0:\tau}\right), \tau = 0, \ldots, t\right\}$, the whole optimal trajectory (the MTE smoother) can then be computed backward using the following theorem.

Theorem 4: The MTE $x^*_{0:t}$ can be computed backward according to

- The outcome of the optimal trajectory is defined by

$$x^*_t = \arg\max_{x_t} \; J^*_t\left(x_t, y_{0:t}\right)$$

- The whole optimal trajectory can be computed backward by $\forall \tau = t-1, \ldots, 0$

$$x^*_\tau = \arg\max_{x_\tau} \left(p\left(x^*_{\tau+1}|x_\tau\right) J^*_\tau\left(x_\tau, y_{0:\tau}\right)\right) \tag{7}$$

Proof: See [4]. ∎

2) Singular case: In most cases, the transition probability measure has no measurable density. Indeed, the dynamic noise dimension is often less than the state dimension. For example, consider vehicle tracking model where only speed and course are noisy when position is a deterministic function of speed and course. In such a case, the MTE is ill-posed. Indeed, there is no mathematical sense to consider the maximization of measures. The problem has then to be regularized. Note that, to the knowledge of the author, this point has never been addressed in the literature (authors often consider, "for simplicity", that the covariance of the dynamic noise Q is a regular matrix [6] [5]).

In most practical cases, the state can be split into two parts [6]: the first one is deterministic with respect to the past ($x_t^{(1)} \in \mathbb{R}^{n_1}$) and the second one is noisy ($x_t^{(2)} \in \mathbb{R}^{n_2}$). In these cases, the state representation takes the following form:

$$\begin{aligned} x_t^{(1)} &= f^{(1)}\left(x_{t-1}\right) && \text{(8a)} \\ x_t^{(2)} &= f^{(2)}\left(x_{t-1}, w_t\right) && \text{(8b)} \end{aligned}$$

where w_t is the dynamic white noise such that $p\left(x_t^{(2)}|x_{t-1}\right)$ is a measurable pdf (a function). The transition pdf is then replaced by a transition probability measure and takes then the following form:

$$\mathbb{P}\left(dx_t^{(1)}|x_{t-1}\right) = \delta_{f^{(1)}(x_{t-1})}\left(dx_t^{(1)}\right) p\left(x_t^{(2)}|x_{t-1}\right) dx_t^{(2)}$$

where $\delta_a\left(dx\right)$ stands for the Dirac measure at a. As the cost defined by (4) is not a function but a measure, it is equivalent to consider the new joint likelihood as follows:

$$\tilde{J}\left(x_{0:t}, y_{0:t}\right) \triangleq p\left(y_{0:t}|x_{0:t}\right) \prod_{\tau=1}^{t} \left(p\left(x_\tau^{(2)}|x_{\tau-1}\right)\right) p\left(x_0\right) dx_0 \tag{9}$$

adding the constraint set

$$\forall \tau = 1, \ldots t,\ x_\tau^{(1)} = f^{(1)}\left(x_{\tau-1}\right) \tag{10}$$

This modified optimal marginal likelihood defined by 9 can be then computed recursively using the following theorem.

Theorem 5: In the singular case, beginning with the knowledge of the initial marginal likelihood

$$\tilde{J}_0^*\left(x_0, y_0\right) \triangleq p\left(y_0|x_0\right) p\left(x_0\right)$$

the optimal marginal maximum likelihood can be computed recursively by the following equation

$$\begin{aligned} \tilde{J}_t^*\left(x_t, y_{0:t}\right) &= \max_{x_{t-1},\ \lambda_{t-1}} \Big(p\left(y_t|x_t\right) p\left(x_t^{(2)}|x_{t-1}\right) \\ &\quad \times \exp\left(\lambda_{t-1}^T\left(x_t^{(1)} - f^{(1)}\left(x_{t-1}\right)\right)\right) \\ &\quad \times \tilde{J}_{t-1}^*\left(x_{t-1}, y_{0:t-1}\right)\Big) \end{aligned} \tag{11}$$

where $\lambda_{t-1} \in \mathbb{R}^{n_1}$ is a Lagrange multiplier related to the constraint $x_t^{(1)} = f^{(1)}\left(x_{t-1}\right)$.

Proof: It is directly derived from classical maximization under constraints using Lagrange multipliers. See [7] for details. ∎

Again, with the knowledge of the set $\{J_\tau^*\left(x_\tau, y_{0:\tau}\right), \tau = 0, \ldots, t\}$, the whole optimal trajectory (the smoother) can then be computed backward.

Theorem 6: The whole MTE can be computed backward according to

- The outcome of the optimal trajectory is defined by

$$x_t^* = \arg\max_{x_t}\ \tilde{J}_t^*\left(x_t, y_{0:t}\right)$$

- The whole optimal trajectory can be computed backward by $\forall \tau = t-1, \ldots, 0$

$$\begin{aligned} x_\tau^* &= \arg\max_{x_\tau,\ \lambda_\tau} \Big(p\left(x_{\tau+1}^{(2)*}|x_\tau\right) \\ &\quad \times \exp\left(\lambda_\tau^T\left(x_{\tau+1}^{(1)*} - f^{(1)}\left(x_\tau\right)\right)\right) \tilde{J}_\tau^*\left(x_\tau, y_{0:\tau}\right)\Big) \end{aligned} \tag{12}$$

with the constraint $x_{\tau+1}^{(1)} = f^{(1)}\left(x_\tau\right)$.

Proof: See [7]. ∎

III. Modal trajectory estimate using Gaussian mixture

A. Gaussian mixture definition

For simplicity of presentation, only the linear non-Gaussian case is developed here. Moreover, we consider here the singular case as it is the most common in practice. According to section II-B2, the state space is split into two components $x_t^{(1)} \in \mathbb{R}^{n_1}$ and $x_t^{(2)} \in \mathbb{R}^{n_2}$. The transition of $x_t^{(1)}$ is deterministic and the transition of $x_t^{(2)}$ is defined by weighted point-wise maximization. Indeed, if one deals with a linear non Gaussian system, one can write

$$\begin{aligned} x_t^{(1)} &= F^{(1)} x_{t-1} \\ x_t^{(2)} &= F^{(2)} x_{t-1} + w_t \end{aligned}$$

where the non Gaussian white noise pdf can be approximated as follows:

$$p\left(w_t\right) = \max_{i=1,\ldots,m} \eta^i \Gamma\left(w_t - \bar{w}^i, Q^i\right)$$

Indeed, it is easy to imagine that many pdf could be approximated by a point-wise Max of Gaussian pdfs as it is the case for the Sum of Gaussian pdfs [8]. As a consequence, the transition pdf can be written as:

$$p\left(x_t^{(2)}|x_{t-1}\right) \simeq \max_{i=1,\ldots,m} \eta^i \Gamma\left(x_t^{(2)} - F^{(2)} x_{t-1} - \bar{w}^i, Q^i\right)$$

and where all Q^i and R are assumed to be non-singular matrices.

This choice is motivated by our estimation scheme. Indeed, unlike the sum operator, the Max operator is distributive with itself and associative with multiplication by positive variable: $\forall a, b, c \geq 0,\ \max\left(c \max\left(a, b\right)\right) = \max\left(ac, bc\right)$ but $\max\left(c\left(a+b\right)\right) \neq \max\left(ac, bc\right)$.

For simplicity of presentation with restrict ourselves to Gaussian observation pdf

$$p\left(y_t|x_t\right) \simeq \Gamma\left(y_t - Hx_t, R\right)$$

In fact, it is straightforward to generalize the observation pdf as a Gaussian mixture defined by

$$p\left(y_t|x_t\right) \simeq \max_{j=1,\ldots,q} \nu^j \Gamma\left(y_t - Hx_t - \bar{v}^j, R^j\right)$$

Figure 1 shows and example of such a Gaussian mixture.

B. Spreading the optimal marginal likelihood

Recall that the MTE requires the recursive computation of $\tilde{J}_t^*\left(x_t, y_{0:t}\right)$. We then have the following theorem:

Theorem 7: If at step $t-1$, the optimal marginal likelihood can be written as a Gaussian mixture of N_{t-1} terms as follows:

$$\begin{aligned} & J_{t-1}^*\left(x_{t-1}, y_{0:t-1}\right) \\ =\ & \max_{k=1,..,N_{t-1}} \rho_{t-1}^k \Gamma\left(x_{t-1} - \hat{x}_{t-1|t-1}^k, P_{t-1|t-1}^k\right) \end{aligned}$$

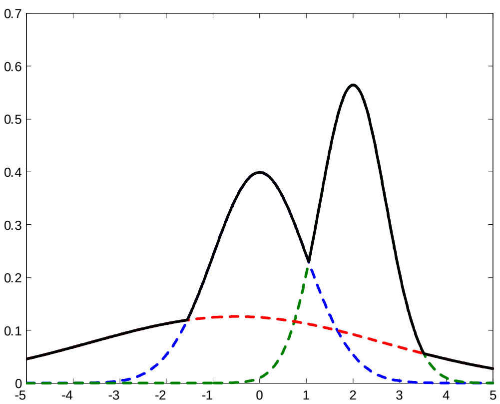

Fig. 1. Example of Gaussian mixture. The continuous black line represents the pointwise Max.

then, at step t, the marginal likelihood is again a Gaussian mixture with $N_t = N_{t-1} \times m \times q$ terms defined by

$$J_t^* \left(x_t, y_{0:t}\right) = \max_{\substack{k=1,..,N_{t-1} \\ i=1,...,m}} \rho_t^{k,i} \Gamma\left(x_t - \hat{x}_{t|t}^{k,i}, P_{t|t}^{k,i}\right)$$

where

$$\rho_t^{k,i} = \rho_{t-1}^k \frac{\eta^i \sqrt{\left|F^{(1)} B_t^{k,i} \left(F^{(1)}\right)^T\right|}}{\sqrt{(2\pi)^n \left|B_t^{k,i}\right|}} \Gamma\left(y_t - H\hat{x}_{t|t-1}^{k,i}, \Sigma_{t|t-1}^{k,i}\right)$$

$$\begin{aligned} B_t^{k,i} &= P_{t-1|t-1}^k - S_t^{k,i} F^{(2)} P_{t-1|t-1}^k \\ P_{t|t-1}^{(2),k,i} &= F^{(2)} P_{t-1|t-1}^k \left(F^{(2)}\right)^T + Q^i \\ S_t^{k,i} &= P_{t-1|t-1}^k \left(F^{(2)}\right)^T \left(P_{t|t-1}^{(2),k,i}\right)^{-1} \\ \hat{x}_{t|t-1}^{k,i} &\triangleq \begin{bmatrix} F^{(1)} \hat{x}_{t-1|t-1}^k \\ F^{(2)} \hat{x}_{t-1|t-1}^k + \bar{w}^i \end{bmatrix} \end{aligned}$$

$$P_{t|t-1}^{k,i} = \begin{bmatrix} F^{(1)} B_t^{k,i} \left(F^{(1)}\right)^T + F^{(1)} S_t^{k,i} P_{t|t-1}^{(2),k,i} S_t^{k,i} \left(F^{(1)} S_t^{k,i}\right)^T & F^{(1)} S_t^{k,i} P_{t|t-1}^{(2),k,i} \\ P_{t|t-1}^{(2),k,i} \left(F^{(1)} S_t^{k,i}\right)^T & P_{t|t-1}^{(2),k,i} \end{bmatrix} \quad (14)$$

$$\begin{aligned} \hat{x}_{t|t}^{k,i} &= \hat{x}_{t|t-1}^{k,i} + K_t^{k,i} \left(y_t - H\hat{x}_{t|t-1}^{k,i}\right) && (15a) \\ \Sigma_{t|t-1}^{k,i} &= H P_{t|t-1}^{k,i} H^T + R && (15b) \\ K_t^{k,i} &= P_{t|t-1}^{k,i} H^T \left(\Sigma_{t|t-1}^{k,i}\right)^{-1} && (15c) \\ P_{t|t}^{k,i} &= P_{t|t-1}^{k,i} - K_t^{k,i} P_{t|t-1}^{k,i} H^T && (15d) \end{aligned}$$

Proof: See [7]. ∎

As it was the case for minimum variance filtering, the conditional pdf at time t is similar to those defined at $t-1$ but the number of Gaussian pdfs grows exponentially with time as $N_t = N_{t-1} \times m \times q$. In practice, it is clear that such a filtering algorithm is not tractable. Many approximations have been proposed to reduce the exponential growing number of Gaussian pdfs. They mainly consist of maintaining only $N_{\max}$ pdfs, $N_{\max}$ being the maximum number of Kalman filters well-matched with the computational power allocated to the application. The most popular techniques are:

- maintaining only the $N_{\max}$ pdfs with highest weights, pruning others and then scaling [9];
- merging several Gaussian pdfs in one equivalent Gaussian pdf, according to some distance criterion, the new pdf having the same mean and variance as the Gaussian subset [1] [10] [11] [12] [13] [14] [15] [16] [17] [3] [18] [19].

If the conditional marginal likelihood is written at time t as follows

$$J_t^* \left(x_t, y_{0:t}\right) = \max_{k=1,..,N_t} \rho_t^k \Gamma\left(x_t - \hat{x}_{t|t}^k, P_{t|t}^k\right)$$

then the outcome of the maximum likelihood trajectory is computed by first looking for the Gaussian indice that leads the maximum value J_t^*:

$$k^* = \arg\max_{k=1,..,N_t} \frac{\rho_t^k}{\sqrt{(2\pi)^n \left|P_{t|t}^k\right|}} \quad (16)$$

The optimal estimate is then the expectation of this Gaussian pdf: $x_{t|t}^* = \hat{x}_{t|t}^{k^*}$. Note that in our case, the Max value of $J_t^* \left(x_t, y_{0:t}\right)$ coincide exactly with the mean of one Gaussian pdf (the k^*-th), unlike in the Gaussian sum approximation scheme [20].

C. Computing the modal trajectory estimate

The whole modal trajectory can be computed backward using the following theorem.

Theorem 8: The MTE over the interval $[0, t]$ can be computed backward as follows:

- Computation of the outcome of the optimal trajectory $x_{t|t}^*$ by 16.
- Backward computation of the whole trajectory: $\forall \tau = t-1, \ldots, 0$

$$\begin{aligned} \left(k_\tau^*, i_\tau^*\right) = \arg \max_{\substack{k=1,..,N_\tau \\ i=1,...,m}} \Bigg(& \rho_\tau^k \frac{\eta^i \sqrt{\left|F^{(1)} B_\tau^{k,i} \left(F^{(1)}\right)^T\right|}}{\sqrt{\left|B_\tau^{k,i}\right|}} \\ & \Gamma\left(x_{\tau+1}^{(2)*} - \hat{x}_{\tau+1|\tau}^{(2),k,i}, P_{\tau+1|\tau}^{(2),k,i}\right) \\ & \times \Gamma\left(x_{\tau+1}^{(1)*} - F^{(1)} \xi_{\tau+1|\tau}^{k,i}, F^{(1)} B_{t-1}^{k,i} \left(F^{(1)}\right)^T\right) \Bigg) \end{aligned} \quad (17)$$

$$\begin{aligned} x_\tau^* = \; & \xi_{\tau+1|\tau}^{k_\tau^*, i_\tau^*} + \\ & B_\tau^{k_\tau^*, i_\tau^*} \left(F^{(1)}\right)^T \left(F^{(1)} B_\tau^{k_\tau^*, i_\tau^*} \left(F^{(1)}\right)^T\right)^{-1} \\ & \times \left(x_{\tau+1}^{(1)*} - F^{(1)} \xi_{\tau+1|\tau}^{k_\tau^*, i_\tau^*}\right) \end{aligned} \quad (18)$$

where

$$\hat{x}^{(2),k,i}_{\tau+1|\tau} = F^{(2)}\hat{x}^{k}_{\tau|\tau} + \bar{w}^{i} \quad (19a)$$

$$P^{(2),k,i}_{\tau+1|\tau} = F^{(2)}P^{k}_{\tau|\tau}\left(F^{(2)}\right)^{T} + Q^{i} \quad (19b)$$

$$S^{k,i}_{\tau|\tau} = P^{(2),k,i}_{\tau|\tau}\left(F^{(2)}\right)^{T}\left(P^{(2),k,i}_{\tau+1|\tau}\right)^{-1} \quad (19c)$$

$$\xi^{k,i}_{\tau+1|\tau} = \hat{x}^{k}_{\tau|\tau} + S^{k,i}_{\tau|\tau}\left(x^{(2)*}_{\tau+1} - \hat{x}^{(2),k,i}_{\tau+1|\tau}\right) \quad (19d)$$

$$B^{k,i}_{\tau} = P^{k}_{\tau|\tau} - P^{k}_{\tau|\tau}\left(F^{(2)}\right)^{T}\left(P^{(2),k,i}_{\tau+1|\tau}\right)^{-1}F^{(2)}P^{k}_{\tau|\tau}$$

Proof: See [7]. ■

Note that it is generally necessary to compute again $\hat{x}^{k,i}_{\tau+1|\tau}$, $P^{k,i}_{\tau+1|\tau}$ and $B^{k,i}_{\tau}$ because the number of Gaussian pdfs is reduced to N_τ at each step with $N_{\tau+1} \neq N_\tau \times m \times q$ in practice. Moreover, note that one may be interested only by a sliding past horizon optimal trajectory limiting the backward computation $\forall\tau = t-1, \ldots, t-T$ where T stands for duration of this past horizon.

Remark 9: **Extension to the non linear case**. Recall that we have restricted ourselves to the linear non Gaussian case. If the dynamic and/or observation function are not linear, it is easy to extend this approach using classical approximations. Indeed, all computations made in this algorithm use classical Gaussian pdf multiplication formulae. In the non linear case, these products may be approximated using linearization [6] or unscented transformation [21], for example.

IV. Simulation results

The MTE algorithm has been tested with a simplified target tracking issue. Consider a target, for example a boat, with known speed (for simplicity) which can change its course at any time with an unknown rotation speed. Assume that the observer is static and has a remote access to the range r_t (accuracy of $10\,\mathrm{m}$) and the azimuth angle θ_t (accuracy of $13\,^\circ$) of the target. The state of the target can be defined by the vector $x_t = [r_t, \theta_t, \psi_t, \omega_t]$ where ψ_t stands for the target course and ω_t its rotation speed. The state equation may then be defined as follows

$$\begin{aligned} r_t &= r_{t-1} + V\cos(\psi_t - \theta_t)\,\Delta t \\ \theta_t &= \theta_{t-1} + \frac{V}{r_{t-1}}\sin(\psi_t - \theta_t)\,\Delta t \\ \psi_t &= \psi_{t-1} + \omega_{t-1}\Delta t \end{aligned}$$

where $V = 20$knot stands for the target speed and $\Delta t = 1\,\mathrm{s}$ for the sampling period. The random deviation of the rotation speed can be depicted as a Markovian stochastic process as follows

$$p(\omega_t|\omega_{t-1}) \propto \max\left(\eta_0\Gamma(\omega_t - \omega_{t-1}, Q_0), \eta_1\Gamma(\omega_t - \omega_{t-1}, Q_1), \eta_2\Gamma(\omega_t, Q_2)\right)$$

The first Gaussian density depicts the little drift of the rotation speed ($\sqrt{Q_0} = 3 \times 10^{-4}\,^\circ/s$), the second Gaussian density depicts a sudden change of the rotation speed ($\sqrt{Q_1} = 3\,^\circ/s$) and the third Gaussian density depicts the back-pulling to a rotation speed close to zero ($\sqrt{Q_2} = 3 \times 10^{-4}\,^\circ/s$). Thus, η_1 stands for the mean of the target steerage frequency ($\eta_1 = 1/500\,\mathrm{Hz}$) and $1/\eta_2$ for the mean of the steerage time ($1/\eta_2 = 60\,\mathrm{s}$). In our scenario, the observer is located at the origin of a Cartesian coordinate system. The initial location of the target is $(-1000\,\mathrm{m}, 1000\,\mathrm{m})$. Its initial course is equal to $70\,^\circ$. After $60\,\mathrm{s}$, the target starts a course steering fixing the rotation speed to $3\,^\circ/s$. This steering stops $40\,\mathrm{s}$ later.

As the transition pdf is singular, we used the algorithm of theorem 7. We have compared the MAP using a Gaussian sum approximation and the MTE. More precisely, we have extended both algorithms to the nonlinear case by linearization (extended Kalman filter). Both algorithms use $N_{\max} = 3^3$ Gaussian pdfs to represent the *a posteriori* density $p(x_t|y_{0:t})$ and the marginal maximum likelihood $\tilde{J}^*_t(x_t, y_{0:t})$. An example of target location tracking is shown on figure 2 where the modal trajectory (the smoother) is computed backward according to theorem **??**. The course tracking is illustrated on figure 3.

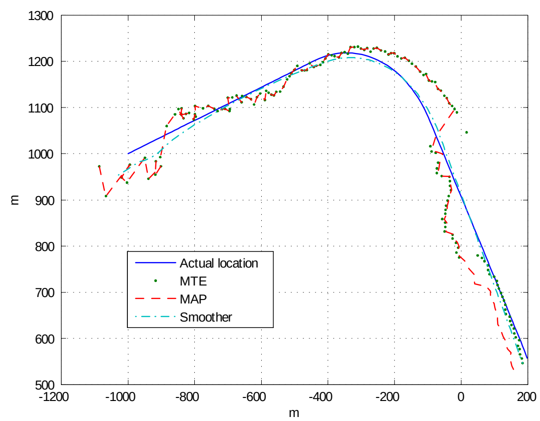

Fig. 2. Comparizon within MAP and MTE filters - Location of the target

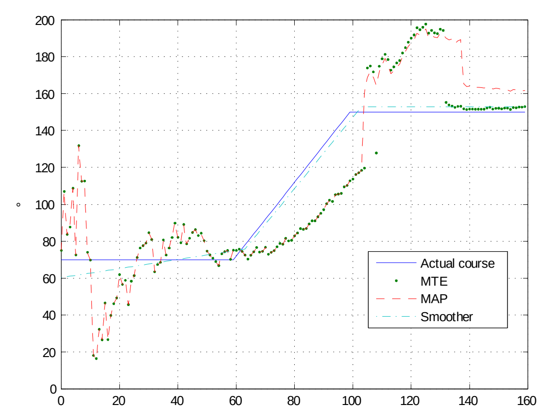

Fig. 3. Comparizon within the MAP and MTE filters - Course of the target

Although it is theoretically not legitimate to compare the mean square errors (MSE) of these filters (the only filter that minimizes the MSE is the minimum variance filter (MVF)), we did the comparison taking into account that neither the MVF using weighted Gaussian sum (WGS), the MAP filter using WSG or MTE using weighted Gaussian Max (WGM) is the exact filter associated to its criterion (they all use approximations). We did 100 runs with different noise trajectories. The results are illustrated on figures 4. One may conclude that the MSE of each filter are commensurate, which is not surprising. However, it appears that the MSE of the MTE location filter never exceeds $25\,\mathrm{m}$ whereas MAP can have a MSE greater that $45\,\mathrm{m}$ (see simulation number 16 and 49). The same phenomena appears for these runs concerning the course estimated.

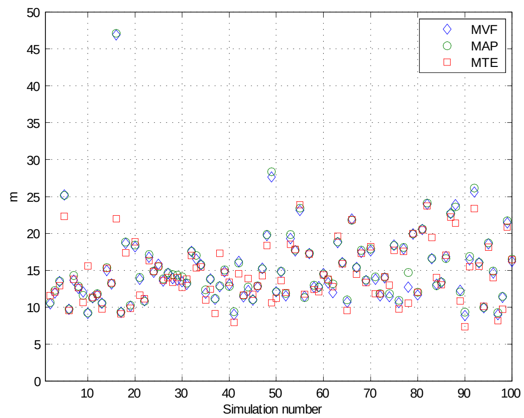

Fig. 4. Mean square errors on estimated target location

V. Conclusion

In many cases, the outcome of the MTE appears to be an interesting alternative to the minimum variance estimate and can be approximated by Gaussian mixtures in a similar way as for the minimum variance issue. The approximation facilities seem to be similar in both cases. As it is the case for WGS, its implementation leads to the computation of N Kalman filters in parallel (EKFs or UKFs in the nonlinear case). Moreover, this approach allows backward computation of a smoothing trajectory over any past interval under assumption that the past optimal marginal likelihood parameters have been memorized (Gaussian means and variances along with weights).

Note that this approach is particularly interesting when one deals with hybrid systems, it is easy to generalize this approach to such a case. Obviously, there is no sense to compute the mean of symbols. But recall that the marginal conditional pdf $p\,(x_t|y_{0:t})$ comes from the minimum variance estimator, the conditional mean. The MTE approach allows to be more coherent avoiding to mix means and Max.

References

[1] D. L. Alspach and H. W. Sorenson, “Nonlinear Bayesian estimation using Gaussian sum approximation,” *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, 1972.

[2] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, “Novel approach to Nonlinear/Non-Gaussian Bayesian state estimation,” *IEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, 1993.

[3] H. A. P. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with Markovian switching coefficients,” *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, 1988.

[4] R. E. Larson and J. Peschon, “A dynamic programming approach to trajectory estimation,” *IEEE Transactions on Automatic Control*, vol. 11, no. 3, pp. 537–540, 1966.

[5] H. Cox, “On the estimation of state variable and parameters for noisy dynamic systems,” *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 376–381, 1964.

[6] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, vol. 64. New York and London: Academic Press, 1970.

[7] A. Monin, “Joint maximum likelihood estimation with gaussian mixture,” Tech. Rep. 08304, LAAS-CNRS, 2008.

[8] H. W. Sorenson and D. L. Alspach, “Recursive Bayesian estimation using Gaussian sums,” *Automatica*, vol. 7, pp. 465–479, 1971.

[9] K. Turner and F. Fariqi, “A Gaussian sum filtering approach for phase ambiguity resolution in GPS attitude determination,” *IEEE Interational Conference on Acoustics, Speech and Signal Processing, ICASSP-97*, vol. 5, pp. 4093–4096, 1997.

[10] R. Singer, R. Sea, and K. Housewright, “Derivation and evaluation of improved tracking filters for use in dense multi-target environments,” *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 423–432, 1974.

[11] D. Reid, “An algorithm for tracking multiple targets,” *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.

[12] K. Turner and F. Faruqi, “Multiple GPS antenna attitude detremination using Gaussian sum filter,” *IEEE TENCON, Digital Signal Processing Applications*, vol. 1, pp. 323–328, 1996.

[13] G. Pulford and S. D.J, “A Gaussian mixture filter for near-far object tracking,” *8th International Conference on Information Fusion*, vol. 1, pp. 337–344, 2005.

[14] A. R. Runnalls, “Kullback-leibler approach to gaussian mixture kullback-leibler approach to gaussian mixture kullback-leibler approach to gaussian mixture reduction,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989–999, 2007.

[15] J. L. Williams and P. S. Maybeck, “Cost-function-based gaussian mixture reduction for target tracking,” in *Proceedings of the 6th International Conference on Information Fusion*, vol. 2, pp. 1047–1054, 2003.

[16] M. West, “Approximating posterior distributions by mixtures,” *Journal of the Royal Statistical Society: Series B*, vol. 55, no. 409-422, 1993.

[17] I. T. Wing and D. Hatzinakos, “An adaptive Gaussian sum algorithm for RADAR tracking,” *IEEE International Conference on Communications*, vol. 3, pp. 1351–1355, 1997.

[18] D. Schieferdecker and M. F. Huber, “Gaussian mixture reduction via clustering,” in *Proceedings of the 12th International Conference on Proceedings of the 12th International Conference on Information Fusion*, pp. 1536–1543, 2009.

[19] D. J. Salmond, “Mixture reduction algorithms for target tracking in clutter,” in *Proceedings of SPIE Signal and Data Processing of Small Targets*, vol. 1305, pp. 434–445, 1990.

[20] M. Carreira-Perpiñan, “Mode-finding for mixtures of gaussian distribution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1318–1323, 2000.

[21] E. Wan and R. Van Der Merwe, “The unscented kalman filter for nonlinear estimation,” in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pp. 153 –158, 2000.