

Particle Filter Based Entropy

Yvo Boers and Hans Driessen

Thales Nederland B.V.
Hengelo, The Netherlands

Arunabha Bagchi and Pranab Mandal

Department of Applied Mathematics
University of Twente
Enschede, The Netherlands

Abstract – *For many problems in the field of tracking or even the wider area of filtering the a posteriori description of the uncertainty can oftentimes not be described by a simple Gaussian density function. In such situations the characterization of the uncertainty by a mean and a covariance does not capture the true extent of the uncertainty at hand. For example, when the posterior is multi-modal with well separated narrow modes. Such descriptions naturally occur in applications like target tracking with terrain constraints or tracking of closely spaced multiple objects, where one cannot keep track of the objects identities. In such situations a covariance measure as a description of the uncertainty is not appropriate anymore. In this paper we look at the use of entropy as an uncertainty description. We show how to calculate the entropy based on a running particle filter. We will verify the particle based approximation of the entropy numerically. We also discuss theoretical convergence properties and provide some motivating examples.*

Keywords: Particle filters, entropy, multi-object tracking

1 Introduction

For an array of problems in the field of target tracking or even in the wider area of general filtering problems the a posteriori descriptions of the uncertainty can become quite cumbersome, e.g. multi modal and not even with connected support. Examples of such situations in target tracking are to be found in tracking in binary sensor networks with limited sensor coverage, terrain aided tracking or closely spaced multi-object tracking, see e.g. [1, 2, 3], just to name a few. In such applications we can run into problems if we adhere strictly to using covariance as a measure of uncertainty. In this paper we take a closer look at the use of entropy as an uncertainty description, especially in combination with the use of particle filters. The use of entropy, even in combination with particle filtering is not entirely new, in the excellent paper, [4], entropy has been used for sensor path planning. In this paper we will discuss several

aspects of the use of entropy and verify experimentally and theoretically that the entropy of the a posteriori density can be approximated well by means of a running particle filter.

The contributions of this paper are:

- A detailed particle based approximation of the entropy for general nonlinear dynamical systems
- A numerical verification of the approximation
- A numerical verification of the quality of the particle based entropy approximation vs. the number of particles
- Theoretical verification of the particle based entropy converging to the true entropy
- Different considerations on the use of entropy as an optimization criterion in filtering and sensor management problems

A considerable part of this paper is devoted to examples and discussion w.r.t. this last point, that is probably of most interest to the practical user.

2 System Setup & Problem Formulation

Let us consider the following general discrete time dynamical system

$$s_k = f(s_{k-1}, w_{k-1}), \quad k \in \mathbb{N} \quad (1)$$

$$z_k = h(s_k, v_k), \quad k \in \mathbb{N} \quad (2)$$

$$s_0 \sim p_0(s_0) \quad (3)$$

where $s_k \in \mathcal{S} \subset \mathbb{R}^n$ is the state of the system, $z_k \in \mathbb{R}^p$ is the measurement, w_k is the process noise with probability density $p_{w_k}(w_k)$, v_k is the measurement noise with probability density $p_{v_k}(v_k)$ and $p_0(s_0)$ is the density of the initial state.

The filtering problem amounts to finding the a posteriori probability density function:

$$p(s_k | Z_k) \quad (4)$$

where $Z_k = \{z_0, \dots, z_k\}$ is referred to as the measurement history.

The standing assumption is that we have a running particle filter, see e.g. [5]. Such a filter can generally be characterized by the set of pairs

$$\{(s_k^i, q_k^i)\}_{i=1 \dots N} \quad (5)$$

where s_k^i is the state part of particle i at time step k , q_k^i the corresponding weight for this particle and N is the total number of particles. Furthermore, it is assumed that the weights q_k^i have been normalized and thus sum up to one.

The following distributional representation is frequently associated with (5):

$$\sum_{i=1}^N q_k^i \delta(s - s_k^i) \quad (6)$$

where $\delta(s)$ is the Dirac delta function. Also the following notation is often being used:

$$p(s_k | Z_k) \approx \sum_{i=1}^N q_k^i \delta(s - s_k^i) \quad (7)$$

or even

$$p(s_k | Z_k) = \sum_{i=1}^N q_k^i \delta(s - s_k^i) \quad (8)$$

Care must be taken with the last two notational representations, because they can be a source of error and/or confusion. In fact they only hold in a certain sense with a lot of restrictions attached. Moreover, in the context of particle filters (7) often merely means that:

$$\lim_{N \rightarrow \infty} \sum_{i=1}^N g(s_k^i) q_k^i = \int_S g(s_k) p(s_k | Z_k) ds_k \quad (9)$$

holds for functions g that are continuous and bounded. This is called convergence in law, or weak convergence, see [6] and holds for particle filters, see [7]. Note that already for a simple mean, where g would be the identity, weak convergence cannot be proven on the grounds of the result from [7], the reason being that the identity function is not bounded.

A central problem that is tackled in this paper is to find a particle based approximation for the entropy of the dynamical system, or equivalently the entropy of the filtering density. For a given filtering density, $p(s_k | Z_k)$, the entropy is:

$$H(p(s_k | Z_k)) = - \int_S \log(p(s_k | Z_k)) p(s_k | Z_k) ds_k \quad (10)$$

Following the discussion above it should be clear that we cannot just 'blindly' substitute $\sum_{i=1}^N q_k^i \delta(s - s_k^i)$ for $p(s_k | Z_k)$ in this expression. This has also been discussed in [8] and [9] for the case of the particle based MAP estimator.

In this case if we were substitute to $\sum_{i=1}^N q_k^i \delta(s - s_k^i)$ for $p(s_k | Z_k)$ into equation (10), apart from the fact that mathematically this is not allowed, see also the discussion above, we also get a strange result. Namely, working out the expressions would lead to:

$$H(p(s_k | Z_k)) \approx - \sum_{i=1}^N q_k^i \log(q_k^i)$$

Which is the entropy of a discrete distribution. However, looking at this expression, it should be clear intuitively that this approximation cannot be correct. One reason for this is that the location and local density of the particles has been totally lost in this approximation, as it only depends on the weights and not on the particles or particle locations. The result becomes even more cumbersome if we were to consider the particle representation after re-sampling. In this case the entropy approximation would become a constant, as all weights are equal to $\frac{1}{N}$. This constant would even be independent of the system and resulting filtering density at hand.

3 Entropy approximation

In this section we will provide an expression for the entropy of a filtering density in terms of particles and weights. The approach taken here is heavily motivated by the approach provided in [8] and [9]. Similar derivations have recently and independently also been provided in [4].

Let us assume we have a system description, as provided in section 2, characterized by the equations: (1), (2) and (3). Let us furthermore assume that there is a running particle filter, characterized by (5) that approximately describes the filtering solution (4).

Now using Bayes' rule: $p(s_k | Z_k) = \frac{p(z_k | s_k) p(s_k | Z_{k-1})}{p(z_k | Z_{k-1})}$ and performing some manipulations, for the entropy, see equation (10), we can write:

$$\begin{aligned} H(p(s_k | Z_k)) = & \quad (11) \\ & - \int_S \log(p(z_k | s_k) p(s_k | Z_{k-1})) p(s_k | Z_k) ds_k + \\ & + \log(p(z_k | Z_{k-1})) \end{aligned}$$

The two terms in (11) will be further expanded next. The term $\log(p(z_k | Z_{k-1}))$ can be written as:

$$\begin{aligned} \log(p(z_k | Z_{k-1})) &= \log\left(\int_S p(z_k | s_k) p(s_k | Z_{k-1}) ds_k\right) \approx \\ &\approx \log\left(\sum_{i=1}^N p(z_k | s_k^i) q_{k-1}^i\right) \quad (12) \end{aligned}$$

The term $\int_{\mathcal{S}} \log(p(z_k | s_k)p(s_k | Z_{k-1}))p(s_k | Z_k)ds_k$ can be approximated by:

$$\begin{aligned} & \int_{\mathcal{S}} \log(p(z_k | s_k)p(s_k | Z_{k-1}))p(s_k | Z_k)ds_k \approx \\ & \approx \sum_{i=1}^N \log\{p(z_k | s_k^i)(\sum_{j=1}^N p(s_k^i | s_{k-1}^j)q_{k-1}^j)\}q_k^i \end{aligned} \quad (13)$$

Now taking the terms together we obtain:

$$\begin{aligned} H(p(s_k | Z_k)) & \approx \log(\sum_{i=1}^N p(z_k | s_k^i)q_{k-1}^i) - \\ & - \sum_{i=1}^N \log(p(z_k | s_k^i)(\sum_{j=1}^N p(s_k^i | s_{k-1}^j)q_{k-1}^j))q_k^i \end{aligned} \quad (14)$$

Thus, we have obtained a valid expression for the entropy in terms of quantities available in a running particle filter.

4 Empirical Convergence

In this section we will provide an example to illustrate the proposed entropy calculation and to experimentally verify the proposed particle based calculation.

We consider a simple one dimensional linear Gaussian system. The reason for this is that for this system we can analytically compute the a posteriori probability density function as a Gaussian density of which the mean and covariance are calculated by the Kalman filter as \hat{s}_k and σ_k^2 respectively. We can also analytically calculate the entropy associated with this a posteriori probability density, see also [10] or [11].

The entropy is given by:

$$\log \sqrt{2\pi e \sigma_k^2} \quad (15)$$

Or more generally for an n-dimensional Gaussian variable with covariance matrix C :

$$\log \sqrt{(2\pi e)^n |C|} \quad (16)$$

The quantity σ_k^2 , is provided by a running Kalman filter and the *exact* entropy of the system can be calculated, using (15). Consider the linear first order system, also complying with the general formulation (1).

$$\begin{aligned} s_{k+1} &= s_k + w_k \\ z_k &= s_k + v_k \end{aligned} \quad (17)$$

where w_k and v_k have a standard, i.e $\mathcal{N}(\cdot; 0, 1)$, Gaussian density.

The initial condition, $s_0 \sim \mathcal{N}(s_0; 10, 1)$, is assumed to be distributed according to a Gaussian distribution with mean 10 and variance 1. Data has been generated according to this model for 15 (time) steps.

We have generated measurements based on this system description. For different numbers of particles we

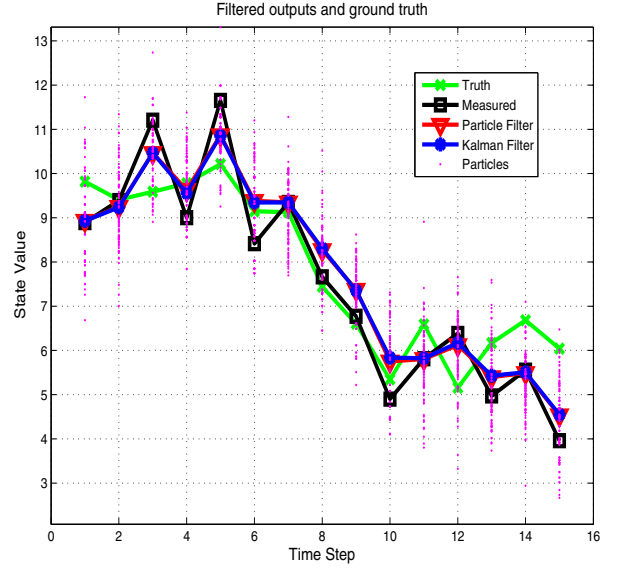


Figure 1: Different filter outputs compared

Entropy Approximation errors		
N	Error STD	Rel Error (%)
100	0.15	13
500	0.070	6
2500	0.031	3
25000	0.0093	1

Table 1: Errors in PF based entropy calculation as a function of N , the number of particles used. Results are based on 50 MC runs and are averaged over all time steps

have calculated the PF based entropy. We have also calculated the filter outputs. The results are displayed in the figures 1 and 2. As expected the accuracy of the approximation improves with a growing number of particles.

Furthermore, we have performed Monte Carlo simulations to assess the accuracy of PF based approximation in more statistical manner. We have performed 50 MC runs for the each of the different numbers of particles. Every run also based on different measurement realization. The results are displayed in table 1.

One would expect the standard deviation of the approximation error to be proportionally equal to $\frac{1}{\sqrt{N}}$. Closer inspection of table 1 confirms this expectation.

5 Theoretical convergence

In section 3 formulas for the calculation of the entropy of the a posteriori distribution have been provided. Naturally, this is merely an approximation of the true entropy, as the number of particles is always finite in an application. The quality of this approximation has been evaluated on a simulation basis for

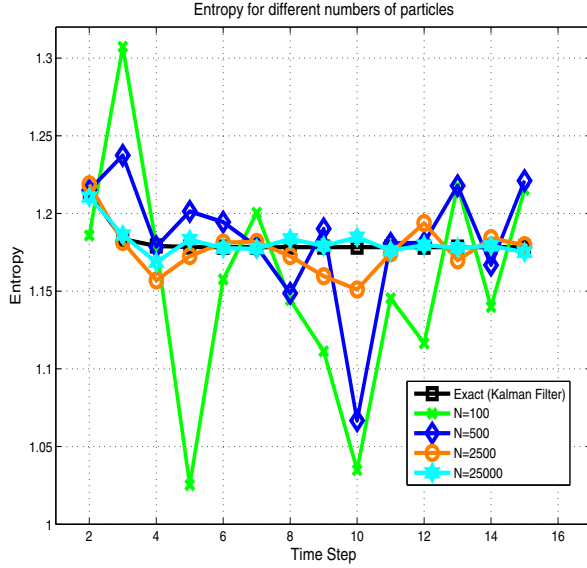


Figure 2: Particle based calculated entropy for different numbers of particles and the exact entropy value of the a posteriori distribution

a specific example in section 4. From that example it seems that the PF based entropy converges to the true entropy for a growing number of particles. However, this numerical verification is no proof or guarantee for general convergence in whatever sense.

Below we will discuss convergence of the particle based entropy to the true entropy. In deriving conditions for convergence, we follow more or less a similar line of reasoning as we did in e.g. [9], [12] or [13]. Where the goal was to prove convergence of the particle based approximate MAP estimate to the true MAP estimate. In addition to this we have to provide some additional conditions and results to make our claim.

If we look at expression for the entropy, see equation (11), we see two terms. For the second term, $\log(p(z_k | Z_{k-1}))$, convergence of its particle approximation is guaranteed under the assumption that the likelihood function $p(z | s)$ is bounded, see [7]. This is the case because:

$$\log(p(z_k | Z_{k-1})) = \log\left(\int_{\mathcal{S}} p(z_k | s_k) p(s_k | Z_{k-1}) ds_k\right) \approx \quad (18)$$

$$\approx \log\left(\sum_{i=1}^N p(z_k | s_k^i) q_{k-1}^i\right) \quad (19)$$

Thus, we have a particle based approximation of a bounded function and thus weak convergence is guaranteed.

For the first term, dropping the minus sign, we have;

$$\int_{\mathcal{S}} \log(p(z_k | s_k) p(s_k | Z_{k-1})) p(s_k | Z_k) ds_k \quad (20)$$

We cannot directly apply the standard result that guarantees convergence for a particle based approximation of an integral of a bounded function. The reason being that $\log(p(z_k | s_k) p(s_k | Z_{k-1}))$ is generally not bounded (think of the likelihood approaching zero or actually being equal to zero).

What we do is we split the integral in (20) into two parts. Also we define $\phi(s_k) = p(z_k | s_k) p(s_k | Z_{k-1})$ and we use the fact that $p(s_k | Z_k) = c\phi(s_k)$, c being a normalizing constant. This follows immediately from Bayes' rule applied to the filtering density $p(s_k | Z_k)$. Thus we can rewrite (20) as:

$$c \int_{\mathcal{B}} \log(\phi(s_k)) \phi(s_k) ds_k + c \int_{\mathcal{U}} \log(\phi(s_k)) \phi(s_k) ds_k \quad (21)$$

where:

$$\mathcal{U} = \{s_k \in \mathcal{S} \mid 0 \leq \phi(s_k) < \epsilon < 1\} \quad (22)$$

and $\mathcal{B} = \mathcal{S}/\mathcal{U}$.

If now we impose the condition that for ϵ tending to zero, the second term in (21) goes to zero we have convergence. The reason for this being that in that case effectively we only need to consider the first term, which then becomes the integral of a bounded function and therefore its particle approximation converges to its true value.

The interesting question or open problem that remains now is to characterize those densities for which the following holds:

$$\int_{\mathcal{U}} \log(\phi(s_k)) \phi(s_k) ds_k \rightarrow 0 \quad \text{for } \epsilon \downarrow 0 \quad (23)$$

We do not provide a full characterization of the class of densities for which this holds here (this remains an open problem). But, for e.g. for a Gaussian density (23) holds. Also for any density that is strictly greater than zero and has a compact support, think e.g. of a uniform density on a bounded subset of the state space, (23) holds.

6 Entropy vs. Minimum Variance: Some Motivating Examples

In this section we provide some examples that illustrate an important fundamental difference between a minimum covariance based (equivalently minimum variance or minimum mean squared error) approach and an entropy based approach. We will, amongst others, show that for bi-modal densities the variance does not decrease anymore, although the intrinsic uncertainty in the system is still decreasing. We also show that in a realistic application of closely spaced target tracking a covariance measure would result in an awkward uncertainty description, due to the so called mixed labelling problem, whereas an entropy measure would not.

In a minimum variance (MV) approach the following criterion is minimized w.r.t. the estimator \hat{s}_k

$$E_{p(s_k|Z_k)}(\hat{s}_k - s_k)^T(\hat{s}_k - s_k) \quad (24)$$

It can be shown, see e.g. [14], that the MV estimator, i.e. the estimator minimizing the criterion in (24), is equal to the conditional expectation of the state:

$$\hat{s}_k^{MV} = E_{p(s_k|Z_k)} s_k \quad (25)$$

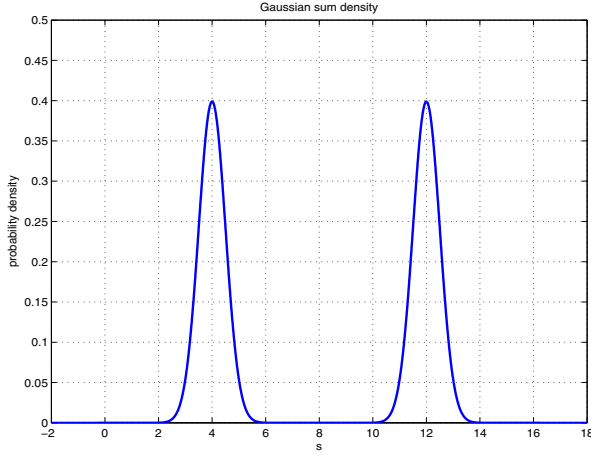


Figure 3: bi-modal a posteriori density (Gaussian sum pdf)

Suppose we have a filtering problem of some kind, for which we have an a posteriori density function $p(s_k | Z_k)$ that equals the following Gaussian sum density (dropping the index k):

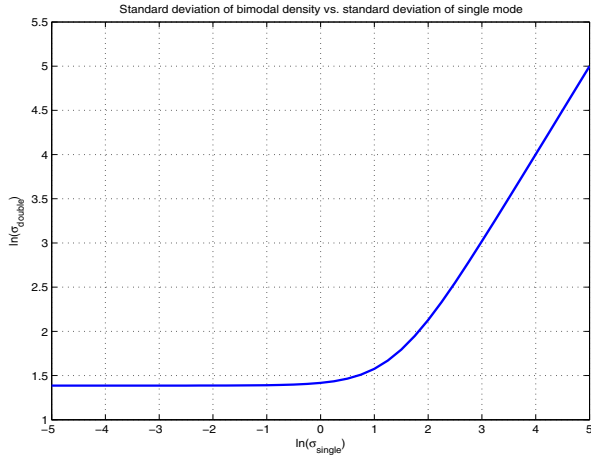


Figure 4: Natural logarithm of the standard deviation of the Gaussian sum density as a function of the natural logarithm of the standard deviation of a single mode in the Gaussian sum density

$$p(s) = 0.5\mathcal{N}(s; 4, 0.5) + 0.5\mathcal{N}(s; 12, 0.5) \quad (26)$$

where $\mathcal{N}(s; \mu, \sigma)$ is a Gaussian density function, parametrized by a mean μ and an standard deviation σ . The density function (26) has been plotted in figure 3. One could relate this to say a 1D target tracking problem where the a posteriori density is a density on the target position.

Now we are going to modify this (a posteriori) density function by changing the standard deviation of the individual components. Thus, we are looking at:

$$p(s) = 0.5\mathcal{N}(s; 4, \sigma) + 0.5\mathcal{N}(s; 12, \sigma) \quad (27)$$

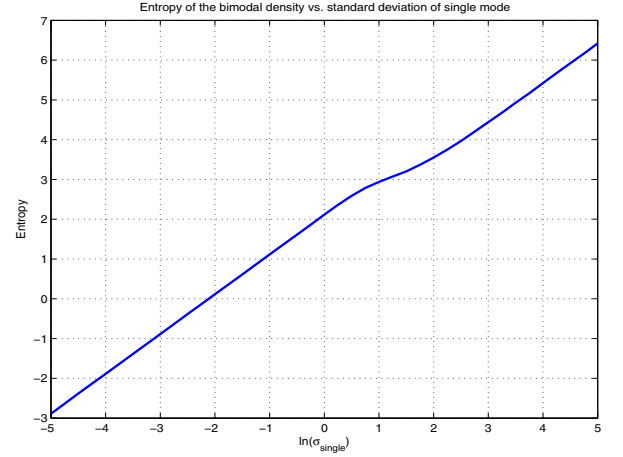


Figure 5: Entropy as a function of the natural logarithm of the standard deviation of a single mode in the Gaussian sum density

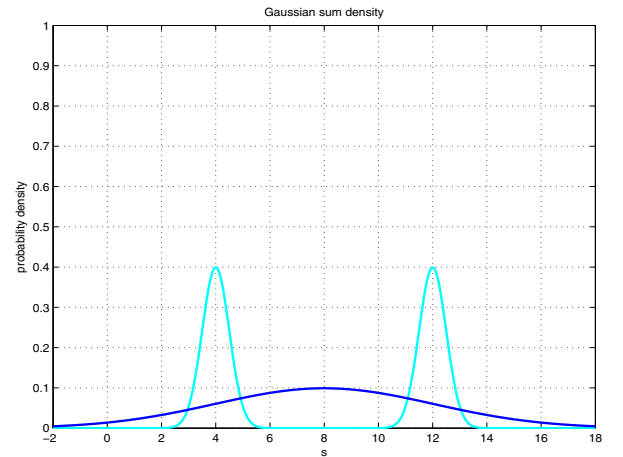


Figure 6: bi-modal a posteriori density (Gaussian sum pdf) and Gaussian approximation

The resulting standard deviation of the stochastic variable s with density function (27) has been plotted in figure 4. What can be seen is that for large values of the standard deviation of the individual components the resulting standard deviation, or equivalently

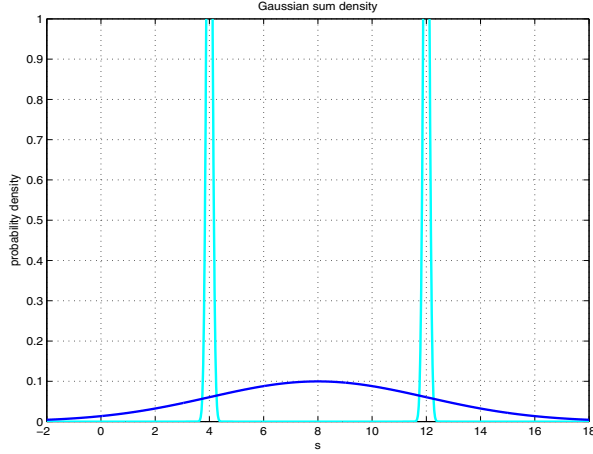


Figure 7: bi-modal a posteriori density (Gaussian sum pdf) and Gaussian approximation - decreased standard deviation of a single mode

the covariance, is growing as a function of the standard deviation of the individual component. However, for smaller values of the standard deviation of the individual component, the resulting standard deviation will assume a constant value. The constant value is determined only by the difference in location of the two peaks, again see also figure 3. In fact if the difference between the peaks is 2Δ then the limiting value of the standard deviation of the Gaussian sum stochastic variable is Δ . In our example $\Delta = 4$. Taking the natural logarithm results in 1.3863, which is the limiting value shown in figure 4. This also shows that the uncertainty measured in terms of covariance grows with a growing distance between the peaks. Also if we look at the figures 6 and 7, we see quite a difference in the uncertainty description of the underlying stochastic variable, nevertheless, they result in identical standard deviations or variances. This is also reflected by the fact that the resulting Gaussian approximations are equal for these two bi-modal densities, this is seen also in these figures. Thus, from a minimum variance standpoint these descriptions are 'equally good' (or bad), where it is quite obvious that if the aim is to reduce uncertainty in the system, the situation in figure 7 is preferred.

Let us now take a look at the entropy of the stochastic variable having a Gaussian sum distribution like the one in (27). In figure 5 we show this entropy as a function of the natural logarithm of the standard deviation of the individual components. What we can see from this figure is that the entropy as a function of the (natural logarithm of the) standard deviation of an individual component is a monotonically increasing function. Thus, there is no limiting value for small values of the standard deviation of the individual components, unlike for the covariance! Furthermore, it is also easily verified that for small values of the standard deviation of the individual components, or stated otherwise, well

separated modes, the entropy is invariant with respect of the distance between the peaks.

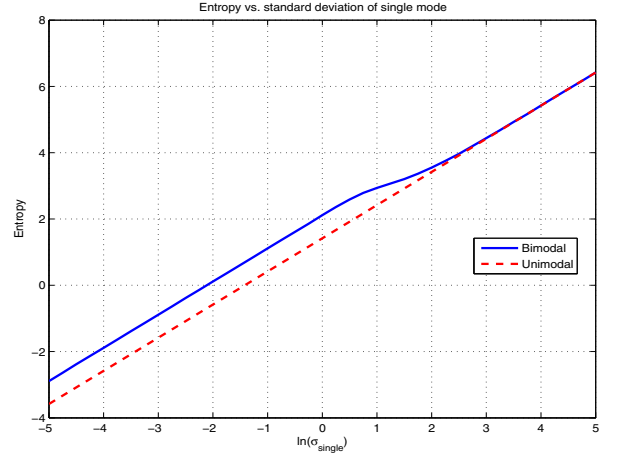


Figure 8: Entropy as a function of the natural logarithm of the standard deviation of a single mode in the Gaussian sum density - modes separated (blue-solid) and modes on top of each other, thus uni-modal (red-dashed)

Another interesting observation is that obviously for either a uni-modal density or a multi-modal density the entropy tends to negative infinity ($-\infty$) whenever the (component) standard deviations tend to zero. This could lead one to think that from an entropy standpoint for small values of the component standard deviation there is no preference for a uni-modal density over a multi-modal one. This is not the case as can be seen in figure 8. In this figure the following densities were considered:

For the bi-modal density:

$$p(s) = 0.5\mathcal{N}(s; 4, \sigma) + 0.5\mathcal{N}(s; 12, \sigma) \quad (28)$$

For the uni-modal density:

$$p(s) = \mathcal{N}(s; 12, \sigma) \quad (29)$$

Thus, although for both densities the limit of the entropy is minus infinity for small values of the standard deviations the uni-modal density is favored from an entropy standpoint.

An implication of the above discussion for the case of tracking closely spaced objects, see also [13], is discussed now.

We adopt the example of [13]. The exact application details are not that important. What is important is that the a posteriori density is a density over two objects. Initially there is no confusion about which object is which, this is reflected by the color coding of the two sub-clouds in figure 9, one sub-cloud is red and one is green. We emphasize that in this example one particle consists of one part of the one sub-cloud and one part of the other, i.e. a joint multi-object particle filter. In this

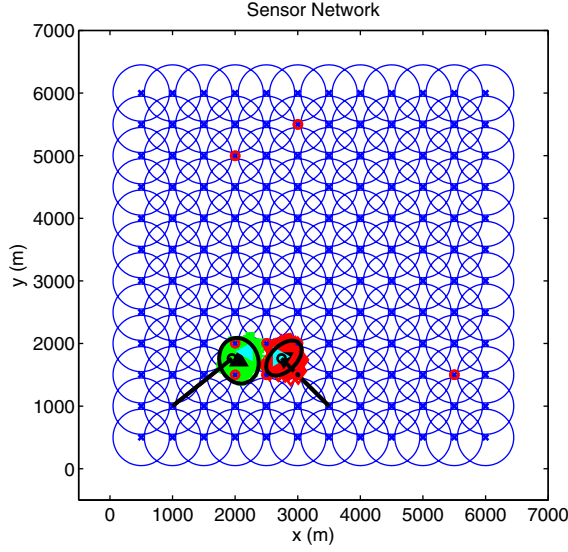


Figure 9: Situation at time step 20

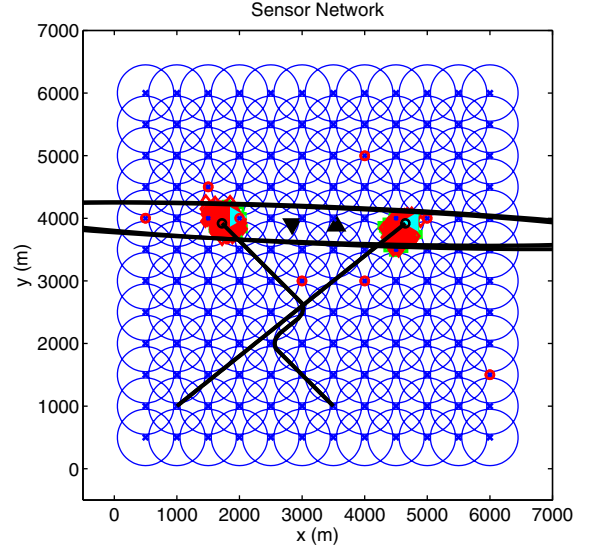


Figure 11: Situation at time step 74

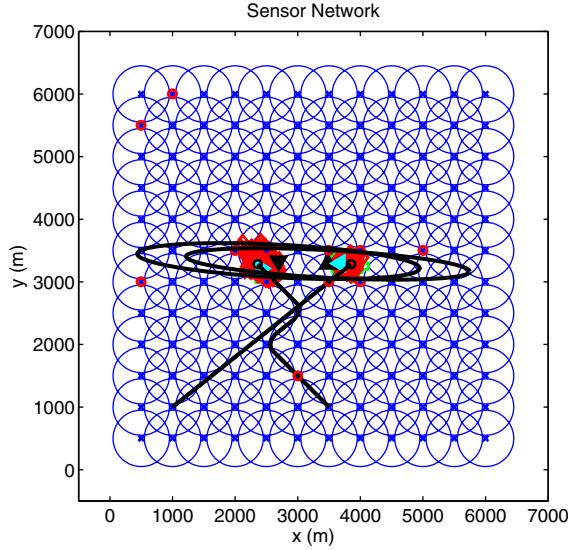


Figure 10: Situation at time step 58

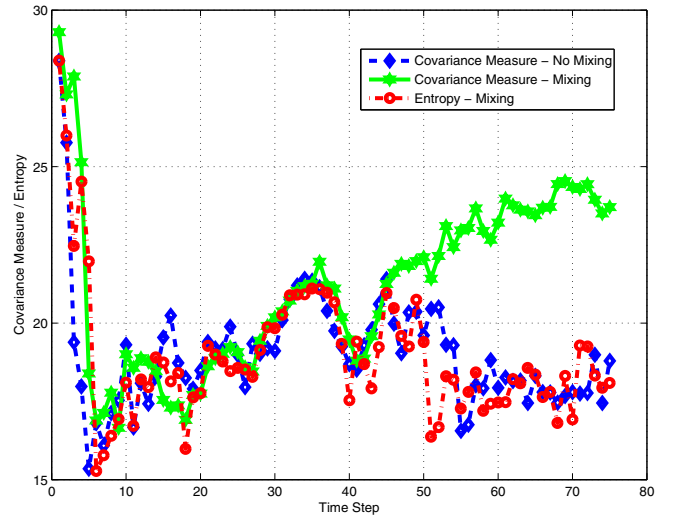


Figure 12: Covariance and entropy measures for the multi object example

case, although there are two sub-clouds, we do not have a multi-modal situation. Also in calculating both the covariance as well as the entropy for this multi-object state, they both will not increase if e.g. the sub-clouds are more separated, for the covariance on the position of the objects this is also already indicated in figure 9, as well as figure 12.

We note here that in figure 12 two of the three lines are covariance measures, see also the legend in the figure itself. Where we have used formula (16), with C being the covariance for the two objects joint position (thus a 4×4 covariance matrix) for the calculation of these lines. The third line is the actual entropy for the multi-object position, calculated on the basis of a run-

ning particle filter and according to formula (14). Furthermore, what is important to note is that in figure 12, one line, the blue one, is actually based on a run for which mixed labelling did not occur after the objects separated again, again see also [13] for more detail.

The objects will move along and move along closely together for quite a while, until they separate again. After this separation, see figures 10 and 11, the sub-clouds are mixed, reflecting the fact that the filter no longer knows which object is which. This is reflected by the fact that the sub-clouds are no longer uniquely color coded. In this case now, we also lost the uni-modality. Now, if the objects are separating the covariance will grow with the growing distance between the objects,

whereas the entropy will not, see figures 10, 11 and 12. Thus, the entropy measure in this case is reflecting much more what would be in line with our intuition. Namely that there is not more uncertainty when the objects move away from each other. This observation can be very important when e.g. deciding on the question to employ a sensor management strategy based on either a covariance measure or an entropy measure. At least for this example it would be at least very tricky to perform sensor management based on the covariance.

7 Conclusions

In this paper we have discussed several aspects of entropy and its use as a measure of uncertainty, especially for tracking problems. We have provided means to calculate the entropy based on a running particle filter and verified its numerical and theoretical correctness.

Furthermore, we have discussed several examples, showing that entropy can be a very good alternative as a measure of uncertainty. This is especially the case when posterior densities become multi-modal.

We have seen in section 6 that the (co)variance of a multi modal density will not decrease more beyond a certain level, even if the variance of the individual components is decreasing. All of this provided that the modes of the multi-modal density are well separated. This limiting behavior for stochastic variables with a multi-modal density does not hold for the entropy of a stochastic variable.

Some future challenges are to find more efficient particle based entropy approximations as well as finding a full rigorous proof for the convergence of the particle based entropy to the true entropy for the broadest class of densities possible.

8 Acknowledgements

This research has been partially financially supported by the Netherlands Organisation for Scientific Research (NWO) under the Casimir program, contract number 018.003.004. Under this grant the first author also holds a part-time position at the Stochastic Systems and Signals group of the Applied Mathematics Department at the University of Twente.

Also discussions with and remarks from Dr. Rienk Bakker from Thales Nederland B.V. are greatly acknowledged.

References

- [1] F. Gustafsson. Particle Filter Theory and Practice. To appear in *IEEE Aerospace and Electronic Systems Magazine*.
- [2] H.A.P. Blom, E. A. Bloem, Y. Boers and J.N. Driessen. Tracking Closely Spaced Targets: Bayes Outperformed by an Approximation? In Proceedings of *FUSION 2008*, Cologne, Germany, July, 2008.
- [3] Y. Boers, E. Sviestins and J.N. Driessen. The Mixed Labeling Problem in Multi Target Particle Filtering. To appear in *IEEE Transactions on Aerospace and Electronic Systems*, 2010.
- [4] P. Skoglar, U. Orguner and F. Gustafsson. On Information Measures Based on Particle Mixture for Optimal Bearings-Only Tracking. In Proceedings of *IEEE Aerospace Conference*, Big Sky, MT, 2009.
- [5] B. Ristic, S. Arulampalam and N. Gordon, *Beyond the Kalman Filter - Particle Filters for Tracking Applications*, Artech House, Boston - London, 2004.
- [6] P. Brémaud. *An Introduction to Probabilistic Modeling*, Springer-Verlag, New York, 1980.
- [7] D. Crisan and A. Doucet, A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, , vol. 50, no. 3, pp. 736-746, 2002.
- [8] J.N. Driessen and Y. Boers. MAP Estimation in Particle Filter Tracking. In *Proceedings of the IET Seminar on Target Tracking and Data Fusion: Algorithms and Applications*, Birmingham, UK, April 15-16, 2008, pp. 41-45.
- [9] J.N. Driessen and Y. Boers. The Particle Based MAP Estimator. Submitted to *IEEE Transactions on Aerospace and Electronic Systems*.
- [10] S. Goldman, *Information Theory*, Prentice-Hall, New York, 1953.
- [11] M. F. Huber, T. Bailey, H. Durrant-Whyte and U. D. Hanebeck, On Entropy Approximation for Gaussian Mixture Random Vectors, In *Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008)*, Seoul, Republic of Korea, 2008.
- [12] Y. Boers, J.N. Driessen and A. Bagchi. Particle Filter Based MAP Estimation for Jump Markov Systems. Under preparation for submission.
- [13] Y. Boers, J.N. Driessen and A. Bagchi. Point Estimation for Jump Markov Systems: Various MAP estimators. In proceedings of the *FUSION 2009 Conference*. Seattle, WA, July 2009.
- [14] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*, Mathematics in Science and Engineering, Vol. 64, Academic Press, 1970.