

Bayesian Fusion of Empirical Distributions Based on Local Density Reconstruction

Uwe D. Hanebeck

Abstract—Fusing two random vectors is simple, when they are characterized by continuous probability density functions. According to Bayes’ law, fusion then consists of multiplying the two densities. When only empirical distributions are given and a resulting empirical distribution is desired, Bayes’ law is no longer applicable. Obviously, fusion could now be performed by reconstructing the underlying continuous densities, subsequent multiplication, and sampling of the result. As this is overly complicated, our goal is to perform a *direct Bayesian fusion* of the two given empirical distributions. We devise a generalized multiplication procedure that mutually reweights appropriate points of one density by *local density values* of the other density. The density values are efficiently estimated locally by nearest neighbor operations. The method is symmetric in the sense that it uses points from both densities.

I. INTRODUCTION

For a continuous random vector $\underline{x} \in \mathbb{R}^N$, the uncertainty is described by a continuous probability density function $\tilde{f}(\underline{x})$. Summary statistics such as mean or covariance matrix can be derived from this density. In many cases, the true density $\tilde{f}(\underline{x})$ is not known and only data points, say \underline{x}_i , $i = 1, \dots, L$, are given. These data points can be used to find a simple discrete density estimate $f(\underline{x})$ of the unknown continuous density $\tilde{f}(\underline{x})$ by

$$f(\underline{x}) = \frac{1}{L} \sum_{i=1}^L \delta(\underline{x} - \underline{x}_i) ,$$

where $\delta(\cdot)$ is the Dirac delta function and equal weights on the data points are assumed.

Empirical distributions are a useful representation of complicated multi-modal probability density functions and are used, e.g., as the density representation in particle filters [1].

When given two random vectors $\underline{x}^{p1} \in \mathbb{R}^N$ and $\underline{x}^{p2} \in \mathbb{R}^N$ constituting uncertain estimates of an underlying state \underline{x} , we might want to fuse the two random variables in order to get a better estimate of \underline{x} , see Fig. 1. Here, we consider \underline{x}^{p1} and \underline{x}^{p2} to be characterized by empirical distributions only. The underlying true distributions are unknown.

This problem occurs, e.g., in hierarchical fusion when two particle filters independently estimate the state of a system and their outputs in the form of empirical distributions need to be fused.

Uwe D. Hanebeck is with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany. Uwe.Hanebeck@ieee.org

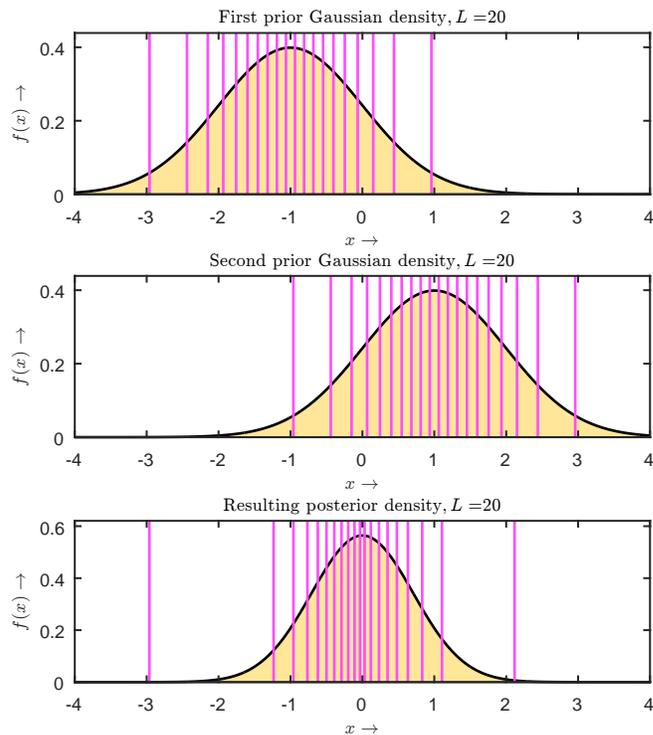


Fig. 1. Direct Bayesian fusion of empirical densities corresponding to two Gaussian densities. The graphs show the true continuous densities and the empirical densities.

Let us now take a look at the state of the art. Linear fusion can simply be performed by first calculating sample means and sample covariance matrices of the two empirical densities and then combining the samples linearly based on corresponding Kalman gains. Various methods are discussed in [2, p. 5]. This only gives satisfactory results when the given empirical distributions stem from underlying Gaussian distributions.

The ensemble Kalman filter [3] directly updates a given empirical distribution. There is only a given single measurement, which is typically assumed to be corrupted by Gaussian noise. The noise density centered around the measurement is intentionally sampled and used for updating the prior samples. However, for updating, a Gaussian assumption between measurements and states is made in some way or another, leading to a linear Kalman-like update rule.

The reconstruction of intermediate densities for fusing two particle sets is considered in [4]. Two methods are used for reconstructing continuous densities. The first method estimates a general Gaussian mixture. The second method

uses a simple Parzen kernel density estimator resulting in a Gaussian mixture with components of equal covariance. Both methods use a component reduction method for post-processing.

In [5], [6] efficient approaches for sampling the output of the product of input measures, e.g., typically two Gaussian mixture inputs, are studied, which do not require an explicit computation of the product itself. In [7], [8], a method for direct (non-Bayesian) information fusion of two empirical (discrete) measures is studied. This protocol operates solely on the sample points and is based on finding a discrete measure that is closest in the sense of the Wasserstein metric to the two empirical input measures. This method is shown to be consistent [7] given two Gaussian input with unknown correlation.

In [9], the problem is treated in the joint state space of the two random vectors to be fused. For continuous densities, Bayes' formula is recovered from the joint state space representation by only considering the density on the equality constraint $\underline{x}^{p1} = \underline{x}^{p2}$. For empirical distributions, fusion can be performed by considering a tube around the equality constraint. Hence, it is not necessary to consider the combination of all prior samples, but only those combinations that fall into the tube.

In this paper, we perform a direct Bayesian fusion of empirical distributions. The key idea is a generalized multiplication of the densities based on the reconstruction of local density values only, which can be done efficiently. The resulting non-equally weighted empirical distribution is then re-approximated by an equally weighted one, which allows recursive processing.

The paper is organized as follows. In the next section, the problem of direct Bayesian fusion is explained in more detail. The estimation of density values of an empirical density at a given evaluation point is discussed in Sec. III. Based on the local density reconstruction, a basic method for direct Bayesian fusion is derived in Sec. IV that only uses points from one of the prior densities. An enhanced method is then derived that is symmetric and uses points from both prior densities. The proposed fusion method is evaluated in Sec. V. Sec. VI concludes the paper.

II. PROBLEM FORMULATION

Our goal is to find a posterior estimate \underline{x}^e of an unknown state vector $\underline{x} \in \mathbb{R}^N$ by fusing two prior estimates \underline{x}^{p1} and \underline{x}^{p2} . The prior estimates are characterized by density functions $f^{p1}(\underline{x})$ and $f^{p2}(\underline{x})$ and we want to calculate $f^e(\underline{x})$ characterizing the posterior estimate.

For the case of continuous prior densities $\tilde{f}^{p1}(\underline{x})$ and $\tilde{f}^{p2}(\underline{x})$, calculating the posterior density is simply performed by Bayes' law

$$\tilde{f}^e(\underline{x}) = c \cdot \tilde{f}^{p1}(\underline{x}) \cdot \tilde{f}^{p2}(\underline{x}) , \quad (1)$$

where c is a normalization constant.

However, here we consider empirical densities, i.e., discrete densities on a continuous domain, also called Dirac mixture densities. The prior densities are defined by

$$f^{pk}(\underline{x}) = \sum_{i=1}^{L^{pk}} w_i^{pk} \delta(\underline{x} - \underline{x}_i^{pk})$$

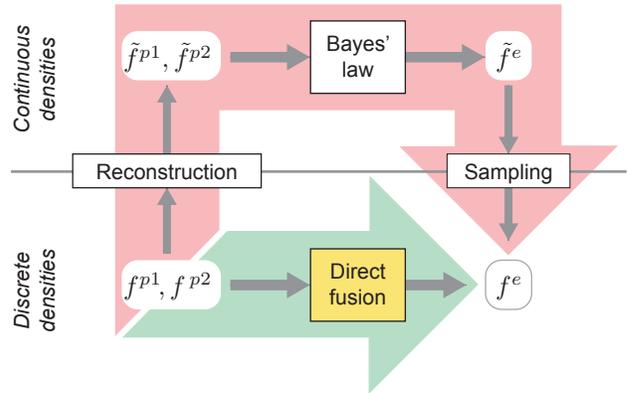


Fig. 2. Performing the fusion of two prior empirical densities f^{p1}, f^{p2} . **Red path:** Reconstruction of underlying continuous true densities $\tilde{f}^{p1}, \tilde{f}^{p2}$, multiplication according to Bayes' law to obtain true continuous posterior \tilde{f}^e , and sampling of \tilde{f}^e to obtain posterior empirical density f^e . **Green path:** Direct Bayesian fusion of the two prior empirical densities f^{p1}, f^{p2} to obtain posterior empirical density f^e .

for $k = 1, 2$, where $\delta(\cdot)$ is the Dirac delta function, \underline{x}_i^{pk} are the component locations, w_i^{pk} are the component weights with $w_i^{pk} > 0$ and $\sum_{i=1}^{L^{pk}} w_i^{pk} = 1$, and L^{pk} the number of components.

For empirical densities, Bayes' law according to (1) obviously cannot be applied. This is due to the fact that Bayes' law requires density values to be multiplied and there are simply no density values available for $f^{p1}(\underline{x})$ and $f^{p2}(\underline{x})$ as the "density" is coded by the number of components per unit area. So, the problem is not only the typical lack of a joint support. Even for two identical empirical densities, which obviously have the same support, a multiplication according to (1) is not well defined.

When we understand the given discrete densities as some approximation of underlying continuous densities, e.g., samples from those densities, an immediate idea would be to reconstruct these continuous densities $\tilde{f}^{p1}(\underline{x})$, $\tilde{f}^{p2}(\underline{x})$ from the discrete densities $f^{p1}(\underline{x})$, $f^{p2}(\underline{x})$ and multiply those in (1), see the red path in Fig. 2. The desired discrete posterior density $f^e(\underline{x})$ would be obtained by sampling $\tilde{f}^e(\underline{x})$. There are several problems associated with this procedure. First, the complexity is high as we need to perform two density estimations for the prior densities, where density estimation by itself is known to be a tedious and complex process. The density resulting from the multiplication then has to be re-approximated by a Dirac mixture density, which is also computationally intensive when we are not talking about Gaussian densities. Second, the procedure is not only complex in absolute terms, it is simply too complex for the result we want to achieve. Reconstructing the full densities is overly complex as we just require the densities at a finite number of points. Third, the selection of the type of continuous density, e.g., Gaussian mixture, introduces unwanted artifacts in the final posterior density.

Here in this paper, we are interested in the direct Bayesian fusion, see the green path in Fig. 2.

III. POINTWISE ESTIMATION OF DENSITY

We now review methods that aim at estimating the true underlying density of an empirical density at certain points. Again, we focus on efficient local methods that do not require the reconstruction of the full underlying continuous density.

In the 1D-case, a so called sample-spacings estimator can be used, see [10] or the first part of [11]. Considering sorted samples, a density estimate at a certain point is proportional to the inverse of the distance between adjacent points. In order to increase robustness to noise, the spacing can be increased to cover $k+1$ samples instead of two. As these density estimates are not continuously differentiable, a kernel-based method is devised in [12].

The general multi-dimensional case is based on k -nearest neighbor considerations [13], which results in non-smooth estimates. Weighting functions for ensuring smooth estimates are characterized in [14]. These concepts can, of course, also be used for one-dimensional problems as an alternative to sample-spacings estimators. The choice of an optimal k depending on sample size and dimensionality is discussed in [15], [16].

Both the k -nearest neighbor density estimators and their smoothed cousins were found *not to work* satisfactorily in our problem context (both for the one-dimensional and the multi-dimensional case). The reason is that for both types of estimators, the tails of the estimated densities die away very slowly. The 1D sample-spacings estimators exhibit a similar problem.

Here, we pursue a different approach. First, for a given empirical distribution, we estimate the density at the component location only. A k -nearest neighbor method is used for that purpose, where we use $k = 1$ in the evaluations. Then, we basically invert the method from [17] to smooth this pointwise estimate in order to calculate the density at points different from the component locations.

We now consider a single Dirac mixture density, say the second prior density $f^{p2}(\underline{x})$ with

$$f^{p2}(\underline{x}) = \sum_{i=1}^{L^{p2}} w_i^{p2} \delta(\underline{x} - \underline{x}_i^{p2})$$

and a given evaluation point, say some component \underline{x}_i^{p1} from the first prior density $f^{p1}(\underline{x})$. Our goal is to estimate the value of the underlying density $\hat{f}^{p2}(\underline{x})$ at \underline{x}_i^{p1} , i.e., $\hat{f}^{p2}(\underline{x}_i^{p1})$.

For estimating the density $\hat{f}^{p2}(\underline{x})$ at one of its own component locations, say \underline{x}_i^{p2} , i.e., $\hat{f}^{p2}(\underline{x}_j^{p2})$, we calculate the k -nearest neighbors of \underline{x}_i^{p2} in the set \underline{x}_j^{p2} , $j = 1, \dots, L^{p2}$, $j \neq i$. The density at the component location \underline{x}_i^{p2} is then given by the probability mass within an N -dimensional hypersphere containing the k components divided by the volume of the hypersphere. The probability mass is the sum of the weights of the k -nearest neighbors that we denote by w_k . The volume of the hypersphere with a radius equal to the distance between the component location \underline{x}_i^{p2} and the k -th (farthest) neighbor is denoted by V_k . Then the density estimate is given by $\hat{f}^{p2}(\underline{x}_j^{p2}) = w_k/V_k$.

The values of the true density function $\tilde{f}^{p2}(\underline{x})$ are so far only reconstructed at the component locations. For evaluating the density at arbitrary locations, we place Gaussian kernels at some components that are close to the evaluation point. The height is set equal to the estimated density. At component \underline{x}_i^{p2} , we obtain

$$K_i(\underline{x}) = \tilde{f}^{p2}(\underline{x}_i^{p2}) \exp\left(-\frac{1}{2} \frac{1}{\tau_i^2} (\underline{x} - \underline{x}_i^{p2})^T (\underline{x} - \underline{x}_i^{p2})\right) \quad (2)$$

for $i = 1, \dots, L^{p2}$, where the τ_i are individual spread parameters yet to be determined, see [17, p. 2].

The spread parameters τ_i in (2) are determined by maintaining the mass w_i^{p2} associated with Dirac component \underline{x}_i^{p2} , so we obtain

$$\int_{\mathbb{R}^N} K_i(\underline{x}) d\underline{x} = w_i^{p2} ,$$

which gives $(\sqrt{2\pi} \tau_i)^N \tilde{f}^{p2}(\underline{x}_i^{p2}) = w_i^{p2}$ or

$$\tau_i = \frac{1}{\sqrt{2\pi}} \left(\frac{w_i^{p2}}{\tilde{f}^{p2}(\underline{x}_i^{p2})} \right)^{\frac{1}{N}} .$$

The resulting kernels have a larger width in low-density regions of the underlying continuous density. In high-density areas, the kernels have smaller widths.

Given the evaluation point \underline{x}_i^{p1} , we calculate the m nearest neighbors of $f^{p2}(\underline{x})$ and call the index set \mathcal{I} , where we use $m = 4$ in the evaluations. By inserting τ_i into the respective kernels in (2), we obtain a smoothed estimate of the true density function $\hat{f}^{p2}(\underline{x})$ as

$$\hat{f}^{p2}(\underline{x}) \approx \sum_{\substack{i=1 \\ i \in \mathcal{I}}}^{L^{p2}} K_i(\underline{x}) .$$

IV. DIRECT FUSION

For performing a direct Bayesian fusion of the two prior densities $f^{p1}(\underline{x})$ and $f^{p2}(\underline{x})$, we now use the density estimate for given evaluation points from the previous section.

A. Basic Method

The basic method cycles through the components of one of the given prior densities and uses each component as an evaluation point to estimate the density value of the other prior density at that point. W.l.o.g., we cycle through the components \underline{x}_i^{p1} of the first prior density $f^{p1}(\underline{x})$ and estimate the value of the second prior density at the component locations of the first density $\hat{f}^{p2}(\underline{x}_i^{p1})$. We obtain the posterior density $f^e(\underline{x})$ by reweighting the components of the first prior density $f^{p1}(\underline{x})$ according to

$$f^e(\underline{x}) = c \sum_{i=1}^{L_1} w_i^{p1} \hat{f}^{p2}(\underline{x}_i^{p1}) \delta(\underline{x} - \underline{x}_i^{p1}) , \quad (3)$$

where c is again a normalization constant.

As a result of the reweighting process, we end up with a set of unequally weighted components for the posterior $f^e(\underline{x})$. For recursive estimation, we would like to perform a closed operation in the sense that the result of the fusion

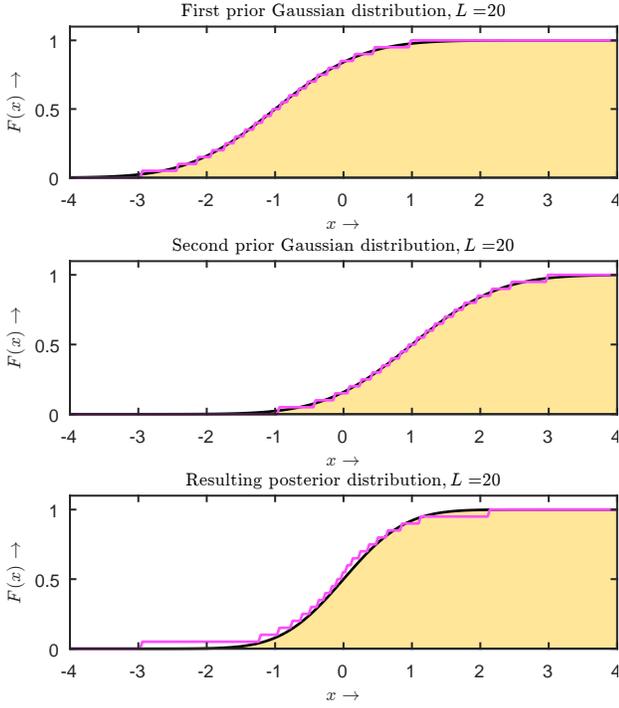


Fig. 3. Direct Bayesian fusion of empirical densities corresponding to two Gaussian densities. The graphs show the true cumulative distributions and the empirical cumulative distributions.

process is of the same type as the input densities. Hence, we have to approximate the unequally weighted posterior by an equally weighted posterior. For performing this approximation, the method described in [18] is used.

B. Enhanced Method

This basic method already gives nice results. However, the method so far is not elegant as it is not symmetric. Asymmetry is caused by selecting one of the prior density for reweighting purposes. We would expect a symmetric method as Bayes' law in (1) for continuous densities is symmetric. Besides not being elegant, we do not really make use of the points of the second prior density as these are only used for reweighting the points of the first one.

For deriving a modified fusion method that considers both prior densities in a symmetric way, we take a look at this trivial rewriting of Bayes' law in (1)

$$\tilde{f}^e(\underline{x}) = c \cdot \frac{1}{2} (\tilde{f}^{p1}(\underline{x}) \cdot \tilde{f}^{p2}(\underline{x}) + \tilde{f}^{p2}(\underline{x}) \cdot \tilde{f}^{p1}(\underline{x})) .$$

We now translate the multiplication of two densities to discrete densities by adding (3) and its dual version

$$f^e(\underline{x}) = c \sum_{i=1}^{L_2} w_i^{p2} \hat{f}^{p1}(\hat{\underline{x}}_i^{p2}) \delta(\underline{x} - \underline{x}_i^{p2}) .$$

To avoid covering the same part of the state space with points from both prior densities, we use selected components from both densities. Again, w.l.o.g., we cycle through the components \underline{x}_i^{p1} , $i = 1, \dots, L_1$ of the first prior density $f^{p1}(\underline{x})$. At each component \underline{x}_i^{p1} , we estimate

the value of the true first prior density $\tilde{f}^{p1}(\underline{x}_i^{p1})$ and also the value of the true second prior density $\tilde{f}^{p2}(\underline{x}_i^{p1})$.

When the value of the true first prior density $\tilde{f}^{p1}(\underline{x}_i^{p1})$ at component location \underline{x}_i^{p1} is larger than the value of the true second prior density $\tilde{f}^{p2}(\underline{x}_i^{p1})$ at that same location, we place the index i into the index set \mathcal{I}_1 . The reason is that this part of the state space is better covered by the first density. Otherwise, the index i will not be placed into the index set \mathcal{I}_1 .

The same procedure is performed for the second prior density $f^{p2}(\underline{x})$ resulting in an index set \mathcal{I}_2 .

Using the two index sets \mathcal{I}_1 and \mathcal{I}_2 results in a symmetric direct Bayesian fusion method as

$$f^e(\underline{x}) = c \cdot \frac{1}{2} \left(\sum_{\substack{i=1 \\ i \in \mathcal{I}_1}}^{L_1} w_i^{p1} \hat{f}^{p2}(\underline{x}_i^{p1}) \delta(\underline{x} - \hat{\underline{x}}_i^{p1}) + \sum_{\substack{i=1 \\ i \in \mathcal{I}_2}}^{L_2} w_i^{p2} \hat{f}^{p1}(\underline{x}_i^{p2}) \delta(\underline{x} - \hat{\underline{x}}_i^{p2}) \right) .$$

Again, for obtaining an equally weighted empirical density, the method described in [18] is subsequently used.

V. NUMERICAL EXAMPLES

The proposed fusion method for empirical density is evaluated by taking samples from given continuous densities, so that the true fusion result is known as ground truth.

We start with two Gaussian prior densities, where $\tilde{f}^{p1}(x)$ has mean -1 and standard deviation 1 , while $\tilde{f}^{p2}(x)$ has mean 1 and also standard deviation 1 . Both densities are sampled deterministically with equal weights based on the closed-form solution from [19]. $L^{p1} = L^{p2} = 20$ are used. The prior densities and their corresponding samples are shown in the two top graphs of Fig. 1. The bottom graph shows the true posterior with mean 0 and standard deviation $1/\sqrt{2}$ and the posterior empirical density resulting from the direct Bayesian fusion of the two priors. In Fig. 3, the prior cumulative distributions are shown in the two top graphs. The bottom graph shows a comparison of the true posterior cumulative distribution and the posterior empirical cumulative distribution resulting for the proposed fusion method, which shows a good agreement.

We now focus on two prior Gaussian mixtures with two components each of the form

$$\tilde{f}^{pk}(x) = \frac{1}{2} \sum_{i=1}^2 \mathcal{N}(x - m_i^{pk}, \sigma) ,$$

for $k = 1, 2$, with component means m_i^{pk} , equal weights, and equal standard deviations σ for all components. For component 1, we have $m_1^{p1} = -3$, $m_2^{p1} = 1$, for component 2, we have $m_1^{p2} = -1$, $m_2^{p2} = 3$.

The prior densities are now sampled to obtain empirical densities. A deterministic sampling procedure according to [20] is used. Here, equal weights are enforced and the number of sample points is set to $L^{p1} = L^{p2} = 200$.

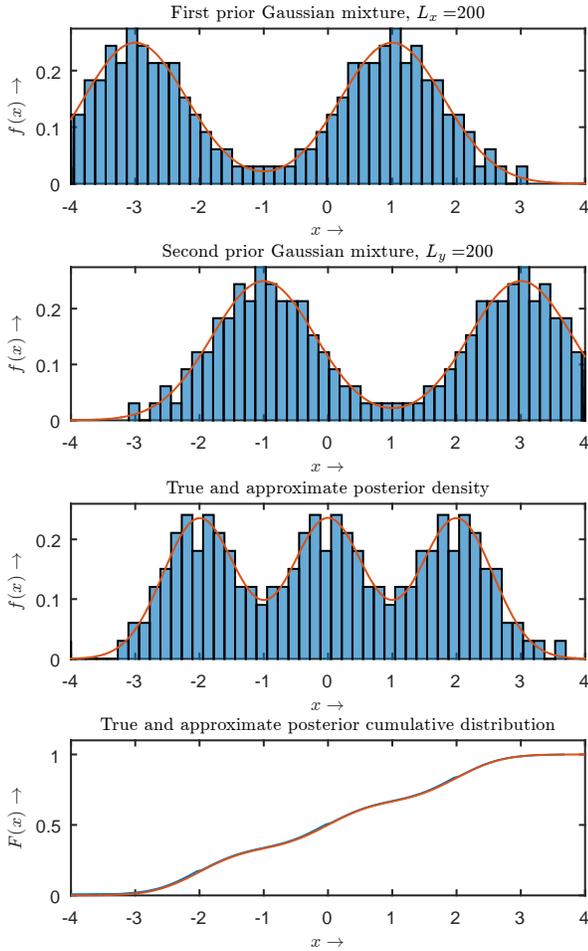


Fig. 4. Direct Bayesian fusion of empirical densities corresponding to two Gaussian mixtures with two components each. Standard deviation of each component is $\sigma = 0.8$.

The true posterior resulting from applying Bayes' law to the two prior Gaussian mixtures $f^{p1}(x)$ and $f^{p2}(x)$ is another Gaussian mixture with four components. The most complicated setup appears when σ is selected in the range $\sigma \in (0, 1]$ as the posterior then exhibits three distinct peaks. Of course, smaller σ make direct Bayesian fusion more difficult as the overlap of the prior densities becomes smaller.

We now take a look at the result of direct Bayesian fusion for different values of σ . For $\sigma = 0.8$, the prior Gaussian mixtures and histograms of their corresponding empirical densities are shown in the top two graphs of Fig. 4. The third graph shows the true posterior and a histogram of the empirical posterior resulting from direct Bayesian fusion. It is obvious that the empirical posterior is a good approximation of the true posterior. The bottom graph shows a comparison of the true posterior cumulative distribution and the empirical cumulative distribution resulting from direct Bayesian fusion. This also underlines the good quality of the approximation. Fig. 5 shows the results for $\sigma = 0.6$, which also leads to a good agreement of the posteriors. When σ is selected much smaller, see Fig. 6 for the case of $\sigma = 0.4$, the overlap

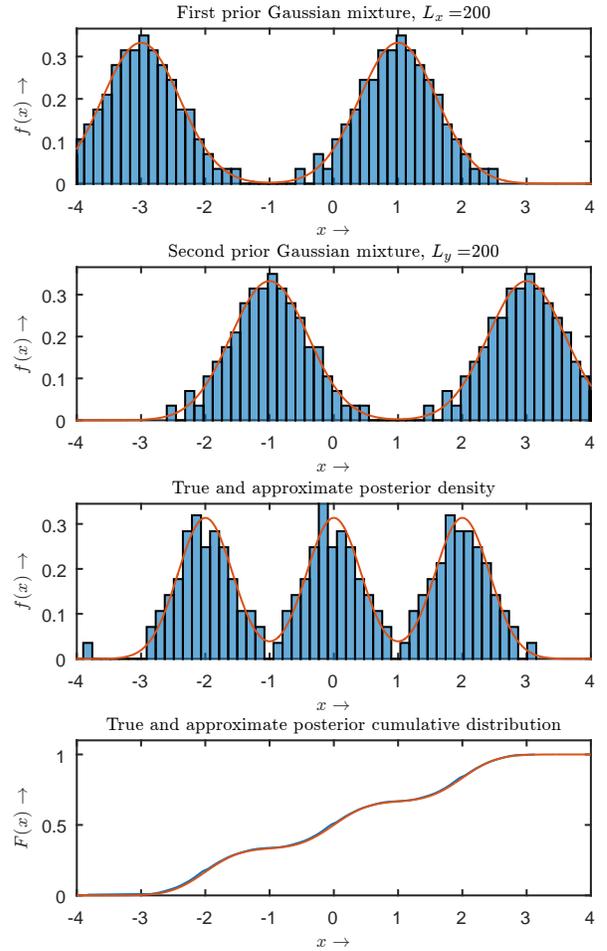


Fig. 5. Direct Bayesian fusion of empirical densities corresponding to two Gaussian mixtures with two components each. Standard deviation of each component is $\sigma = 0.6$.

between the true prior densities is already very small and the empirical prior densities have almost no overlap anymore. The proposed fusion method still manages to produce an acceptable result although the deviations are clearly visible in both the density domain as well as in the cumulative distribution domain.

VI. CONCLUSIONS

A simple and efficient method for direct Bayesian fusion of two random vectors characterized by empirical distributions was derived. The paper contains three main contributions: The first contribution is to reweight one of the given prior empirical distributions based on locally reconstructed density values of the other prior empirical distribution in order to directly obtain a result of the same density type. The second contribution is to perform a mutual update that uses the most appropriate points from both prior empirical distributions. The third contribution is an efficient new method for the reconstruction of the underlying continuous densities based on nearest neighbors.

The method is simple to implement and works in arbitrary dimensions. So far, one-dimensional deterministic samples have been considered in the evaluations.

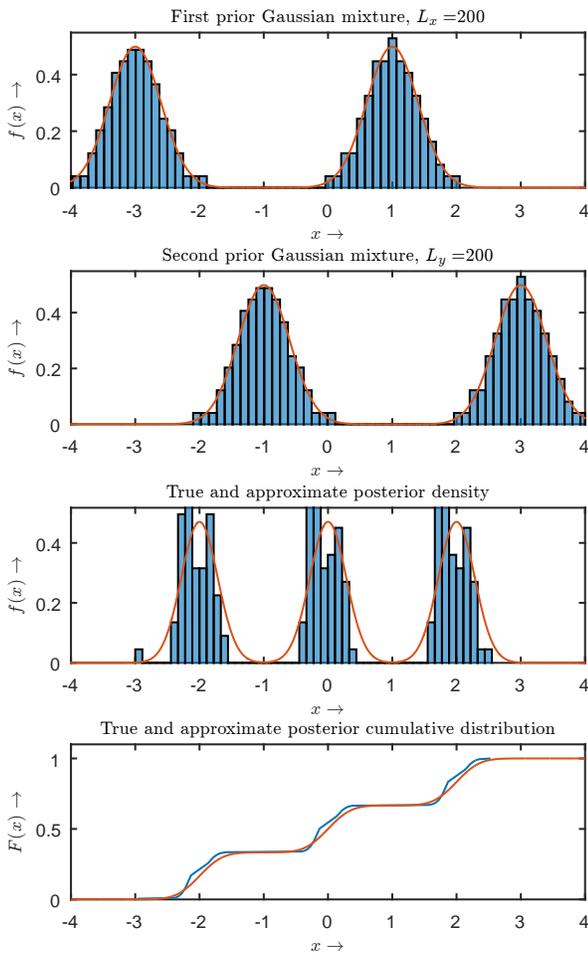


Fig. 6. Direct Bayesian fusion of empirical densities corresponding to two Gaussian mixtures with two components each. Standard deviation of each component is $\sigma = 0.4$.

The evaluation will be extended to higher dimensions in a follow-up paper. When noisy samples are given, the reduction method in [18] can be used to calculate cleaner sample sets then used for fusion. This is similar to the idea in [21].

REFERENCES

- [1] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
- [2] D. Danu, T. Kirubarajan, T. Lang, and M. McDonald, "Multisensor particle filter cloud fusion for multitarget tracking," in *2008 11th International Conference on Information Fusion*, Jun. 2008, pp. 1–8.
- [3] G. Evensen, "The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, 2003.
- [4] L.-L. Ong, B. Ucroft, M. Ridley, T. Bailey, S. Sukkarieh, and H. Durrant-Whyte, "Consistent Methods for Decentralised Data Fusion using Particle Filters," in *2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Sep. 2006, pp. 85–91.

- [5] A. Ihler, E. Sudderth, W. Freeman, and A. Willsky, "Efficient Multiscale Sampling from Products of Gaussian Mixtures," in *Proc. of 17th Annual Conf. on Neural Information Processing Systems*, 2003.
- [6] E. Sudderth, A. Ihler, M. Isard, W. Freeman, and A. Willsky, "Nonparametric belief propagation," *Communications of the ACM*, vol. 53, no. 10, pp. 95–103, 2010.
- [7] A. Bishop, "Information fusion via the Wasserstein barycenter in the space of probability measures: Direct fusion of empirical measures and Gaussian fusion with unknown correlation," in *Proc. of the 17th International Conference on Information Fusion*, Salamanca, Spain, 2014, pp. 1–7.
- [8] —, "Gossip-based distributed data fusion of empirical probability measures," in *Proc. of the 2014 IEEE Workshop on Statistical Signal Processing*, Gold Coast, Australia, 2014, pp. 372–375.
- [9] V. Klumpp and U. D. Hanebeck, "Direct Fusion of Dirac Mixture Densities using an Efficient Approximation in Joint State Space," in *Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008)*, Seoul, Republic of Korea, Aug. 2008, pp. 168–174.
- [10] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. van der Meulen, "Nonparametric Entropy Estimation: An Overview," *International Journal of Mathematical and Statistical Sciences*, vol. 6, 1997.
- [11] E. Miller, "A New Class of Entropy Estimators for Multi-Dimensional Densities," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)*, vol. 3, Apr. 2003, pp. III–297–300 vol.3.
- [12] U. Ozertem, I. Uysal, and D. Erdogmus, "Continuously Differentiable Sample-Spacing Entropy Estimation," *IEEE Transactions on Neural Networks*, vol. 19, no. 11, pp. 1978–1984, Nov. 2008.
- [13] D. O. Loftsgaarden and C. P. Quesenberry, "A Nonparametric Estimate of a Multivariate Density Function," *The Annals of Mathematical Statistics*, vol. 36, no. 3, pp. 1049–1051, Jun. 1965. [Online]. Available: <http://www.jstor.org/stable/2238216>
- [14] Y. P. Mack and M. Rosenblatt, "Multivariate K-Nearest Neighbor Density Estimates," *Journal of Multivariate Analysis*, vol. 9, no. 1, pp. 1–15, Mar. 1979.
- [15] K. Fukunaga and L. Hostetler, "Optimization of K Nearest Neighbor Density Estimates," *IEEE Transactions on Information Theory*, vol. 19, no. 3, pp. 320–326, May 1973.
- [16] J. Orava, "K-Nearest Neighbour Kernel Density Estimation, the Choice of Optimal K," *Tatra Mountains Mathematical Publications*, vol. 50, no. 1, pp. 39–50, Dec. 2011. [Online]. Available: <http://www.degruyter.com/view/j/tmmp.2011.50.issue-1/v10127-011-0035-z/v10127-011-0035-z.xml>
- [17] U. D. Hanebeck, "Kernel-based Deterministic Blue-noise Sampling of Arbitrary Probability Density Functions," in *Proceedings of the 48th Annual Conference on Information Sciences and Systems (CISS 2014)*, Princeton, New Jersey, USA, Mar. 2014.
- [18] —, "Optimal Reduction of Multivariate Dirac Mixture Densities," *at - Automatisierungstechnik, Oldenbourg Verlag*, vol. 63, no. 4, pp. 265–278, Apr. 2015. [Online]. Available: <http://dx.doi.org/10.1515/auto-2015-0005>
- [19] O. C. Schrempf, D. Brunn, and U. D. Hanebeck, "Density Approximation Based on Dirac Mixtures with Regard to Nonlinear Estimation and Filtering," in *Proceedings of the 2006 IEEE Conference on Decision and Control (CDC 2006)*, San Diego, California, USA, Dec. 2006.
- [20] —, "Dirac Mixture Density Approximation Based on Minimization of the Weighted Cramér-von Mises Distance," in *Proceedings of the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, Heidelberg, Germany, Sep. 2006, pp. 512–517.
- [21] F. Chazal, D. Cohen-Steiner, and Q. Mérigot, "Geometric Inference for Probability Measures," *Foundations of Computational Mathematics*, vol. 11, no. 6, pp. 733–751, Dec. 2011.