Optimal Dirac Approximation by Exploiting Independencies

Henning Eberhardt, Vesa Klumpp, and Uwe D. Hanebeck

Abstract— The sample-based recursive prediction of discrete-time nonlinear stochastic dynamic systems requires a regular reapproximation of the Dirac mixture densities characterizing the state estimate with an exponentially increasing number of components. For that purpose, a systematic approximation method is proposed that is deterministic and guaranteed to minimize a new type distance measure, the so called modified Cramér-von Mises distance. A huge increase in approximation performance is achieved by exploiting structural independencies usually occurring between the random variables used as input to the system. The corresponding prediction step achieves optimal performance when no further assumptions can be made about the system function. In addition, the proposed approach shows a much better convergence compared to the prediction step of the particle filter and by far fewer Dirac components are required for achieving a given approximation quality. As a result, the new approximation method opens the way for the development of new fully deterministic and optimal stochastic state estimators for nonlinear dynamic systems.

I. INTRODUCTION

The estimation of a system state from noisy measurements is a common engineering problem and various techniques have been proposed for solving it. Probably the best established state estimator is the Kalman filter, which is optimal for estimating the first two moments of a linear system [1]. As most systems are nonlinear, several extensions to the Kalman filter have been published [2], but still the degree of nonlinearity and especially the shape of the densities to be estimated remain somehow limited [3], [4]. There are several other approaches, like the SGMF [5] or Fourier densities [6] and the family of particle filters, which are much less limited in this context. Particle filters [7], from the family of Monte Carlo methods, do approximate the moments of any given density by drawing random samples from it. They have reached great popularity because of their ease of implementation. While particle filters greatly differ [8] in the insertion of new measurements, the prediction step is the same for most variants.

The sample-based density representation used in the particle filter implies several disadvantages. Since the samples are drawn randomly, a relatively poor convergence is achieved and therefore, the number of samples required for an adequate approximation quality is usually high.

Dirac mixture densities [9] on the other hand are generated by systematically minimizing a distance measure. The

H. Eberhardt, V. Klumpp, and U. D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics, Karlsruhe Institute of Technology (KIT), Germany, henning.eberhardt@kit.edu, vesa.klumpp@kit.edu, uwe.hanebeck@ieee.org
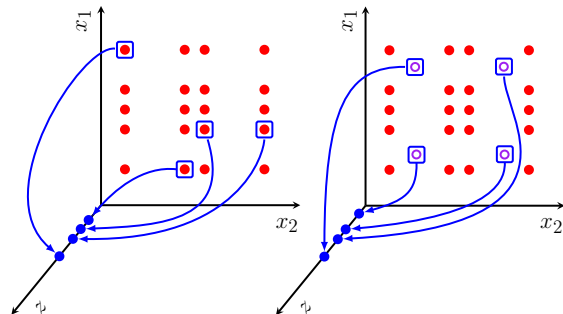
Fig. 1: The particle filter *(left)* selects random pairs of particles from the densities $f(x_1)$ and $f(x_2)$, while the proposed method *(right)* deterministically approximates the joint state space of the two densities. The system function is then applied to the result of each.

resulting approximation is deterministic and of high quality, so that only few Dirac mixture components are necessary. In addition to that, a discrete density representation has several advantages, such as the ability to solve the stochastic prediction analytically for most cases, which make this density representation beneficial for the use in nonlinear filters. Major problems that may arise in sample-based Bayesian estimators are the approximation of continuous densities by means of sample-based densities [10], [11] and the reduction of the number of samples used [12], while maintaining an appropriate representation quality of the state estimate.

The deterministic approximation to be deduced is derived from the Localized Cumulative Distribution and the modified Cramér-von Mises distance measure presented in [13], [14]. This technique allows for arbitrary approximations of continuous and discrete densities. The focus of the presented work is on the special case where the density to be approximated is completely described by several stochastically independent densities of lower dimension. Furthermore, the work is limited to Dirac mixture densities for simplification, although this limitation can be easily lifted.

II. PROBLEM FORMULATION

In this paper, a novel approach for systematically calculating an approximation of the joint density of several independent random variables is presented. Such an approximation can be used in the prediction step of a Bayesian estimator. The prediction step is defined by the Chapman-Kolmogorov equation

$$f^p(\underline{z}) = \int \delta\left(\underline{z} - a(\underline{x})\right) f(\underline{x}) \, \mathrm{d}\underline{x} \;, \tag{1}$$

with a system function $a(\cdot)$, a posterior density $f^p(\underline{z})$, and a prior density $f(\underline{x})$. Here, the focus lies on the prior density $f(\underline{x})$ assumed to be composed of independent lower-dimensional Dirac mixture densities

$$f(\underline{x}) = {}^1f({}^1\underline{x}) \cdot {}^2f({}^2\underline{x}) \cdot \ldots \cdot {}^Rf({}^R\underline{x}) \ ,$$

where the lower-dimensional densities ${}^rf({}^r\underline{x})$ are represented as Dirac mixtures and $\underline{x} = \left[{}^1\underline{x}^T, {}^2\underline{x}^T, \ldots, {}^R\underline{x}^T\right]^T$. Each lower-dimensional density might for example represent a noise density or the density of a system state that is stochastically independent from the rest. A Dirac mixture density with L components is given by

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

with positive weights ω_i, which sum up to 1, and locations $\underline{x}_i = \left[x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(N)}\right]^T$, where x_i^j is the jth component at $\underline{x}_i$ for $i = 1, \ldots, L$ Dirac components and N dimensions.

The exact solution of Equation (2) is a Dirac mixture, where the number of components is the product of the number of the components of the lower-dimensional densities. For several consecutive prediction steps, e.g., a combination of a prior state estimate and a stochastic noise term, this leads to an exponential growth in the number of Dirac components. As a result of this, the application of the system function to those densities quickly gets computationally unfeasible. Thus, a reduction of Dirac mixture components is needed.

In this work, the reduction of the joint density $f(\underline{x})$, which is composed of several independent Dirac mixture densities, is presented. Here, the approximated density is calculated directly from the densities ${}^1f({}^1\underline{x}), \ldots, {}^Rf({}^R\underline{x})$ without explicit calculation of the joint density $f(\underline{x})$ by exploiting its separability.

In Figure 1, the approximation is shown in the context of the prediction step, where the system function maps two independent densities to the predicted density $f(\underline{z})$. On the left hand side, the particle filter is shown. Here, the prediction step (2) is solved by selecting pairs of samples from the prior and the noise density and applying the system function on each pair. This process can be interpreted as an approximation of the joint density ${}^1f({}^1x) \cdot {}^2f({}^2x)$. While the particle filter does give reasonably good results for large numbers of Dirac components, it may perform poorly when only a few samples are available as a lot of information is not accounted for in the random selection. On the right hand side, the proposed method is visualized. The algorithm reduces the joint Dirac mixture density to a representation with fewer components. This approximation is then propagated through the system function.

The approximation method proposed in this paper is based on the minimization of a global distance measure between a generalized definition of the cumulative distribution of the joint density and its approximation. The result is calculated deterministically and is optimal according to a specific distance measure, the modified Cramér-von Mises distance,

which will be presented in the next section. The use of a global distance measure guarantees that all information in the joint density is taken into account.

III. Localized Cumulative Distribution and the Modified Cramér–von Mises Distance

For comparing the given true Dirac mixture density with its approximation, a suitable distance measure is needed. As the cumulative distribution is ambiguous for multi-dimensional densities, a unique generalization of the cumulative distribution like the one described in [13] is defined for later employment in a distance measure:

Definition III.1 (Localized Cumulative Distribution)
Let $\underline{\mathbf{x}}$ be a random vector with $\underline{\mathbf{x}} \in \mathbb{R}^N$, which is characterized by an N–dimensional probability density function $f(\underline{x}) : \mathbb{R}^N \to \mathbb{R}_+$. The corresponding Localized Cumulative Distribution (LCD) is defined as

$$F(\underline{m}, \underline{b}) = \int_{\mathbb{R}^N} f(\underline{x}) \, K(\underline{x} - \underline{m}, \underline{b}) \, \mathrm{d}\underline{x} \ , \qquad (2)$$

with $\underline{b} \in \mathbb{R}_+^N$ and $F(.,.) : \Omega \to [0,1], \ \Omega \subset \mathbb{R}^N \times \mathbb{R}_+^N$.

In this paper, the kernels to be used are of Gaussian type defined by

$$K(\underline{x} - \underline{m}, b) = \prod_{k=1}^{N} \exp\left(-\frac{1}{2} \frac{\left(x^{(k)} - m^{(k)}\right)^2}{b^2}\right)$$

with a one-dimensional width parameter b.

Based on these Gaussian kernels, the LCD of the Dirac mixture in (3) is given by

$$F(\underline{m}, b) = \int_{\mathbb{R}^N} f(\underline{x}) \, K(\underline{x} - \underline{m}, b) \, \mathrm{d}\underline{x}$$

$$= \sum_{i=1}^{L} w_i \prod_{k=1}^{N} \exp\left(-\frac{1}{2} \frac{\left(x_i^{(k)} - m^{(k)}\right)^2}{b^2}\right) \ .$$

This expression can be extended to the LCD of the product $\tilde{f}(\underline{x}) = {}^1\tilde{f}({}^1\underline{x}) \cdot {}^2\tilde{f}({}^2\underline{x}) \cdot \ldots \cdot {}^R\tilde{f}({}^R\underline{x})$ of R Dirac mixtures, where the tilde denotes the density before approximation. Due to the separability of the kernels, we get the LCD $\tilde{F}$ for $\tilde{f}$, which is just the product of the individual LCDs

$$\tilde{F}(\underline{m}, b) = \prod_{r=1}^{R} {}^r\tilde{F}({}^r\underline{m}, b) \ ,$$

where $\underline{m} = [{}^1\underline{m}^T, {}^2\underline{m}^T, \ldots, {}^R\underline{m}^T]^T$.

Based on the definition of the Localized Cumulative Distribution and the Cramér–von Mises Distance, the distance measure required for the calculation of the approximation of a joint density is presented in the following definition.

Definition III.2 (Modified Cramér–von Mises Distance)
The distance D between two densities $\tilde{f}(\underline{x}) : \mathbb{R}^N \to \mathbb{R}_+$ and

$f(\underline{x}) : \mathbb{R}^N \to \mathbb{R}_+$ is given in terms of their corresponding LCDs $\tilde{F}(\underline{m}, b)$ and $F(\underline{m}, b)$ as

$$D = \int_{\mathbb{R}_+} w(b) \int_{\mathbb{R}^N} \left(\tilde{F}(\underline{m}, b) - F(\underline{m}, b) \right)^2 \, \mathrm{d}\underline{m} \, \mathrm{d}b \ , \quad (3)$$

where $w(b) : \mathbb{R}_+ \to [0, 1]$ is a suitable weighting function.

Intuitively, the LCD represents the mass of a density in a region described by the kernel's shape and position. This distance measure compares two densities by the representation of their masses for every possible kernel width and position and allows, due to the smoothing implied through the use of Gaussian kernels, for the comparison of Dirac mixtures. For that purpose, the integral of the squared difference between the LCD of the true density $\tilde{f}(\underline{x})$ and the LCD of its approximation $f(\underline{x})$ is employed. In order to calculate the distance measure, an integration over the kernel position $\underline{m}$ as well as over the kernel width b has to be performed.

IV. EVALUATION AND MINIMIZATION OF THE DISTANCE MEASURE

Expanding the expression in (4) and then using the integral over the LCDs of two one-dimensional Dirac components

$$\int\limits_{-\infty}^{\infty} F_x(m, b) F_y(m, b) \, \mathrm{d}m$$

$$= \sqrt{\pi} \, b \, w^x \, w^y \, \exp\left(-\frac{(x - y)^2}{4b^2} \right)$$

leads to the following theorem:

Theorem IV.1 *For the LCDs of the product of R Dirac mixtures $\tilde{F}(\underline{m}, b) = \prod\limits_{r=1}^{R} {}^r\tilde{F}({}^r\underline{m}, b)$ and the approximating Dirac mixture $F(\underline{m}, b)$ with the weighting function*

$$w(b) = \begin{cases} \frac{1}{b^{N-1}} & b \in [0, b_{max}] \\ 0 & elsewhere \end{cases} , \quad (4)$$

the following expression for the distance D holds

$$D = \int\limits_{0}^{b_{max}} b \, \pi^{\frac{N}{2}} \left(\prod_{r=1}^{R} {}^rP_1 - 2 \prod_{r=1}^{R} {}^rP_2 + P_3 \right) \mathrm{d}b \ , \quad (5)$$

with

$${}^rP_1 = \sum_{i=1}^{{}^rM} \sum_{j=1}^{{}^rM} {}^rw_i \, {}^rw_j \, \exp\left(\frac{-1}{4 b^2} \sum_{k=1}^{{}^rN} \left({}^ry_i^{(k)} - {}^ry_j^{(k)} \right)^2 \right) ,$$

$${}^rP_2 = \sum_{i=1}^{{}^rM} \sum_{j=1}^{L} {}^rw_i \, v_j \, \exp\left(\frac{-1}{4 b^2} \sum_{k=1}^{{}^rN} \left({}^ry_i^{(k)} - x_j^{(k+K_k)} \right)^2 \right) ,$$

$$P_3 = \sum_{i=1}^{L} \sum_{j=1}^{L} v_i \, v_j \, \exp\left(\frac{-1}{4 b^2} \sum_{k=1}^{N} \left(x_i^{(k)} - x_j^{(k)} \right)^2 \right) ,$$

where ${}^r\underline{y}_i$ denotes the locations of the ${}^rM$ Diracs weighted by ${}^r\underline{w}_i$ in the ${}^rN$-dimensional Dirac Mixture ${}^r\tilde{f}$. $\underline{x}_i$ are the

Dirac positions in the N-dimensional Dirac mixture f being weighted by $\underline{v}_i$. The expression

$$K_k = \sum_{l=1}^{k-1} {}^lN$$

describes the position, respectively the axes, of the r-th Dirac mixture in the N-dimensional space of the complete density.

PROOF. The theorem can be proven by first expanding the square in Definition (III.2) and then integrating over $\underline{m}$. $\square$

In order to calculate the optimal Dirac mixture regarding Theorem IV.1, the Dirac locations and weights have to be determined. For simplification, the weights of the approximating Dirac mixture are assumed to be equal, i.e., $w_i = \frac{1}{L}$, so only the Dirac locations remain to be computed. Given this simplification, there is still a high-dimensional optimization problem left. In this paper, a quasi-Newton optimization called limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [15] is applied. In order to perform the optimization, a gradient has to be provided to the optimizer.

Theorem IV.2 *The gradient of the distance measure in Theorem IV.1 with respect to the locations $\underline{x}$ of the Dirac components is given by*

$$\frac{\partial D}{\partial x_{\xi}^{(\eta)}} = \int\limits_{0}^{b_{max}} b \, \pi^{\frac{N}{2}} \left(\frac{\partial P_3}{\partial x_{\xi}^{(\eta)}} - 2 \frac{\partial \left({}^{K_\eta}P_2 \right)}{\partial x_{\xi}^{(\eta)}} \prod_{\substack{r=1 \\ r \neq K_\eta}}^{R} {}^rP_2 \right) \mathrm{d}b \ ,$$

where K_η is defined as

$$K_\eta = \max\left\{ k : \left(\sum_{r=1}^{k} {}^rN \right) < \eta \right\} \ .$$

PROOF. As ${}^rP_1$ is constant, only ${}^rP_2$ and P_3 have to be differentiated, which is straightforward. $\square$

V. COMPUTATIONAL COMPLEXITY

The formulae for the distance measure in Theorem IV.1 and its gradient in Theorem IV.2 still require a *one-dimensional* integration over the variable b. This integration could be performed analytically by simply expanding the product of sums in ${}^rP_1$ and ${}^rP_2$, which basically results in a summation of terms of the form $\int b \exp\left(\frac{-1}{4b^2} \right) \mathrm{d}b$. Unfortunately, the expansion causes an exponential growth in the number of summands. This would give an implementation of the analytic solution of the integral over b a computational complexity of

$$\mathbf{O}(N \cdot \prod_{r=1}^{R} {}^rM^2) \ \text{for} \ \prod_{r=1}^{R} {}^rP_1 \,,$$

$$\mathbf{O}(N \cdot L \cdot \prod_{r=1}^{R} {}^rM) \ \text{for} \ \prod_{r=1}^{R} {}^rP_2 \,,$$

$$\mathbf{O}(N \cdot L^2) \ \text{for} \ P_3 \,,$$

which equals the complexity that would occur if the complete joint density was first created in order to be reduced
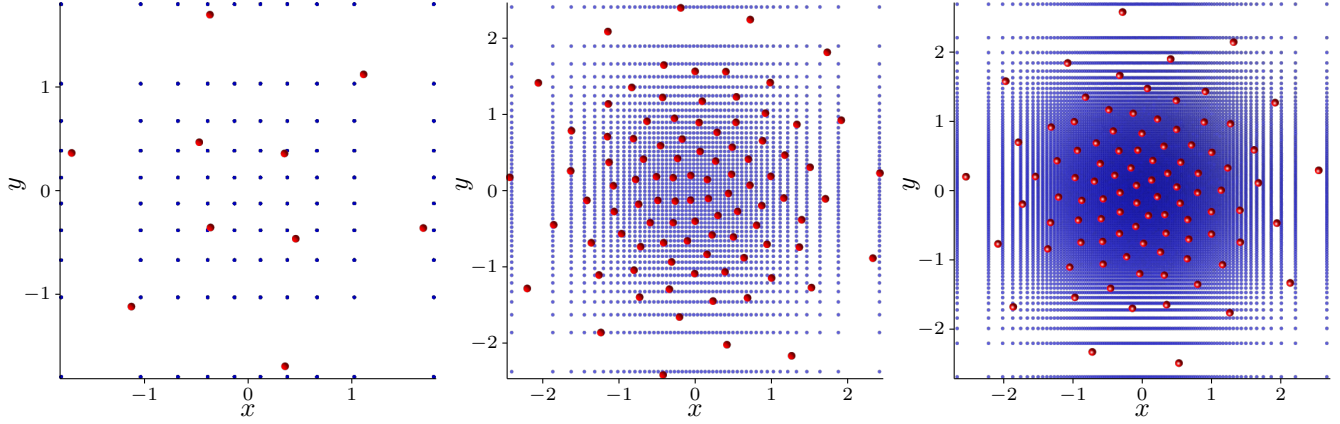
Fig. 2: A reduction of the joint density of two one-dimensional Dirac mixtures. *(a)* from 10×10 Diracs *(blue)* to 10 Diracs *(red)*, *(b)* from 50×50 Diracs *(blue)* to 97 Diracs *(red)*, *(c)* from 110×110 Diracs *(blue)* to 97 Diracs *(red)*.

afterwards. It is obvious that although the computation of $^rP_1$ can be omitted, the computational complexity of this solution is challenging for a reasonable number of components in the Dirac mixtures ^rM and any large R. A more promising approach is the implicit evaluation of the distance measure that avoids the expansion that causes the high complexity. A downside of such an approach is that a numerical integration for D becomes necessary. Fortunately, $b\exp\left(\frac{-1}{4b^2}\right)$ is bounded and smooth on the interval $[0, b_{max}]$, which also holds for the complete expressions $^rP_1$, $^rP_2$, P_3, and the required products. This makes D in Equation (6) suitable for numerical integration. The advantage of the implicit solution is demonstrated in the following. While expanding the terms results in an exponential growth of complexity, a simple function evaluation for the terms in Theorem IV.1 and Theorem IV.2 can be performed in polynomial time. So, given the use of q integration points, the total complexity for the computation of the numerically solved integral dramatically drops to

$$\mathbf{O}\left(q \cdot N \cdot \sum_{r=1}^{R} {}^rM^2\right) \text{ for } \prod_{r=1}^{R} {}^rP_1 \,,$$

$$\mathbf{O}\left(q \cdot N \cdot L \cdot \sum_{r=1}^{R} {}^rM\right) \text{ for } \prod_{r=1}^{R} {}^rP_2 \,,$$

$$\mathbf{O}\left(q \cdot N \cdot L^2\right) \text{ for } P_3 \,,$$

which enables an efficient evaluation by the optimizer.

VI. SIMULATIONS

In this section, several simulations will be presented, demonstrating the functionality and performance of the proposed Dirac mixture approximation algorithm.

A. Dirac Mixture Approximation

The first simulation displays the result of the reduction of a joint density composed of two one-dimensional Dirac mixture densities. For the two densities, a Dirac approximation of two one-dimensional Gaussian densities with $10, 50,$ and

110 Dirac components is used. The resulting joint density is reduced to 10 Dirac components for the first and 97 equally weighted Dirac components for the other two runs. Figure 2 displays the joint density in blue and its reduction in red. As expected, the approximation is similar to the approximation of a two-dimensional normal distribution shown in [14].

B. Runtime Measurements

To demonstrate the speedup achieved by the implicit evaluation of the distance measure, a second simulation is performed. The simulation consists of a MATLAB™ implementation of Theorems IV.1 and IV.2. A reduction to $L = 97$ Dirac components is performed for different one-dimensional Dirac mixtures spanning a two- to fourteen-dimensional space. The Dirac mixtures are chosen to have 10 to 110 components per dimension with a step size of 20. The Dirac components of all one-dimensional densities are equally weighted and their locations are chosen to minimize the distance to a standardized normal distribution as proposed in [14]. A total of 12 simulation runs is performed on an Intel(R) Core(TM)2 Quad CPU Q9300 @ 2.5GHz running Linux 64 bit and MATLAB 2009a™, computing four runs at a time with one run on each core. The averaged results are depicted in Figure 4. For numerical integration, an adaptive Lobatto quadrature, which uses less integration points for smoother integrands, with an absolute maximum error of 10^{-6} is applied [16]. The number of integration points used varies from about 130 for all dimensions at ten Dirac components to about twelve integration points for all dimensions at 110 Dirac components as shown in Figure 3. The reduction of integration points for larger numbers of Dirac components is a result of the increasing smoothness of the integrands in IV.1 and IV.2 for the given Dirac mixtures and will not hold for every possible Dirac mixture. It may be regarded as a beneficial side effect of this reduction that for a large number of Dirac components the displayed simulation does even slightly surpass the run times predicted in Section V. Finally, the optimizer itself has to be taken into account as
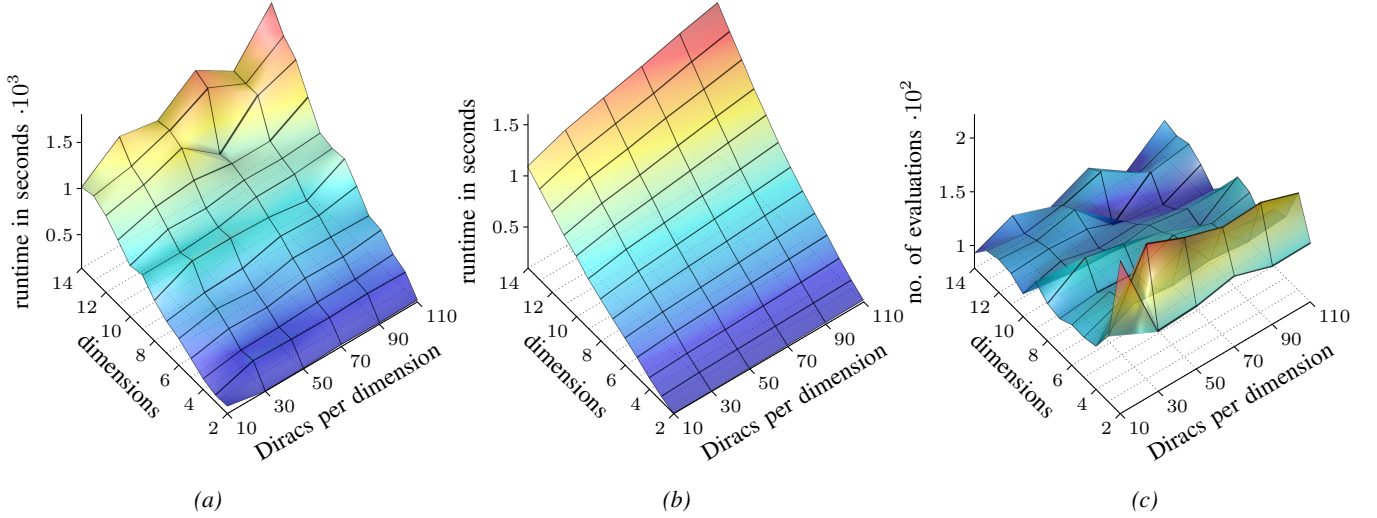
Fig. 4: Runtime measurements for a reduction to $L = 97$ Dirac components, averaged over 12 simulation runs. *(a)* Runtime of the complete reduction, *(b)* Averaged runtime of the numerical integration ($> 10\,000$ runs), *(c)* Number of optimizer steps executed.
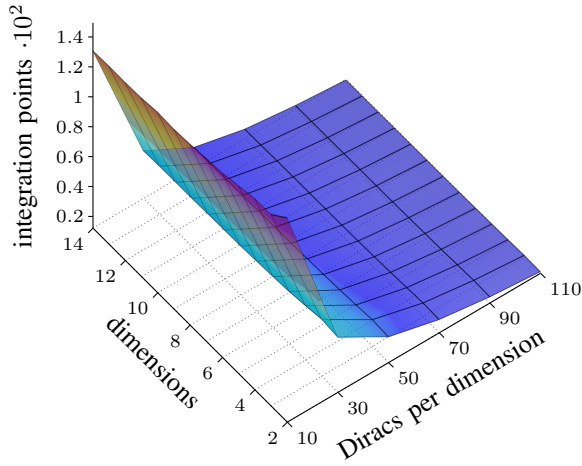

Fig. 3: The number of integration points required in the runtime simulation decreases, as the number of Dirac components increases due to the increasing smoothness of the integrand in the Theorems IV.1 and IV.2.

the optimization process does take about 100 to 200 iterations for the given simulation. This factor of the runtime is still a major caveat.

C. Stochastic Prediction

The third simulation uses the presented results to implement the prediction step for a simple system. The system function represents an omnidirectional vehicle moving at constant speed. The three-dimensional system state consists of an x,y-position and a movement direction α. The system function

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \\ \alpha_{k+1} \end{bmatrix} = \begin{bmatrix} x_k + v_k^x & + & 10\sin(\alpha_k + v_k^\alpha) \\ y_k + v_k^y & + & 10\cos(\alpha_k + v_k^\alpha) \\ \alpha_k + v_k^\alpha \end{bmatrix} \quad (6)$$

is disturbed by noise $\underline{v}_k$, represented with 27 sample points generated from a Gaussian density with covariance matrix $\mathrm{diag}(0.3, 0.3, 0.15)$ using the scheme described in [14].

The simulation compares the particle filter and the exact solution of the prediction to the proposed method by calculating four time steps. The initial system state estimate is given by a single Dirac component placed at the coordinates $[0, 0, 0]^T$. This state estimate is, as described by the system function (7), first disturbed by the system noise and then predicted through the system function. The resulting system state estimate at the second time step consists of 27 Dirac components and the process is repeated for each of those until the fourth time step, when the total number of Dirac components is $27^4 = 531\,441$ is reached. The position components of the exact solution are depicted in Figure 5 *(c)*. For each prediction step, the mean and the covariance of the systems state are calculated. A comparison is then employed by calculating the absolute error on a selection of those moments for the proposed method for $10, \ldots, 110$ Dirac components. In order to get a meaningful comparison, the particle filter is then used to calculate the same simulation with an increasing number of particles until the same RMSE, averaged over $1\,000$ runs, is reached as for the proposed method. While two exemplary runs of the simulation are displayed in Figures 5 *(a)* and *(b)*, the result of the comparison is depicted in Figure 6 and 7. A maximum of $1\,000\,000$ particles has been chosen in order to limit the runtime of the simulation, as that amount of particles is reasonably large for a three-dimensional state space for the given simulation. The figures show that for different moments and time steps, the number of particles required can be as low as 42 for C_{xx} which makes a factor of 4.2 to more than the cap of $1\,000\,000$ for several other moments. It is a noticable fact that the result of the proposed method seems to degrade from time steps three to four compared to the
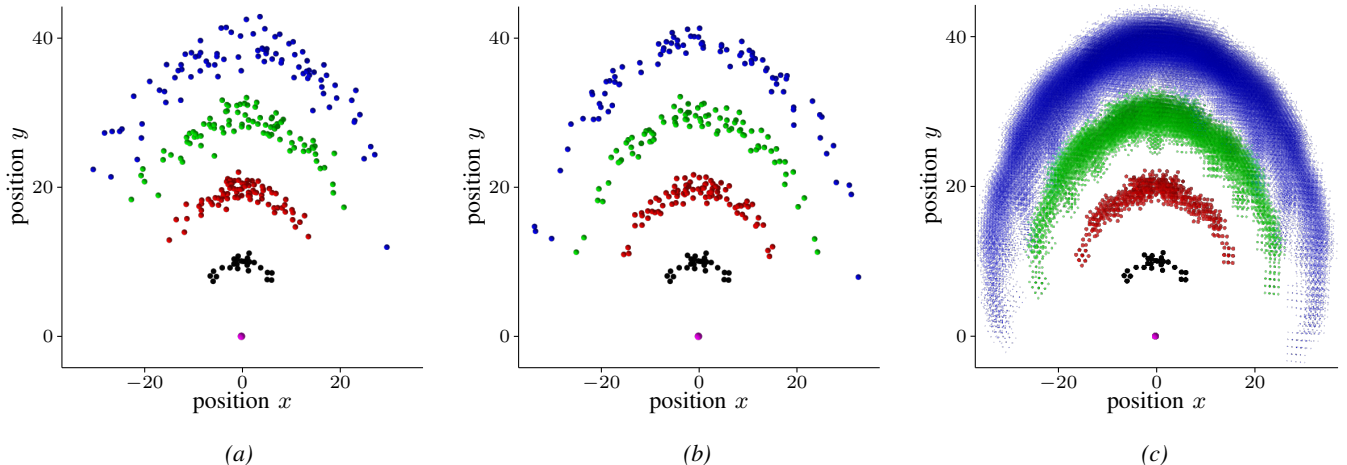
Fig. 5: Position component of the third simulation for 90 Dirac components: The five colors are the starting value (purple) and the four time steps (black, red, green, and blue). *(a)* Exemplary particle filter run, *(b)* Proposed prediction step, *(c)* Exact solution.

particle filter. This is due to a systematic error introduced by the proposed method that does not affect the particle filter, which will be referred to in the conclusion. Nevertheless, the number of components saved by using the proposed method is significant.

VII. DISCUSSION

In this paper, a novel approach to the deterministic reduction of the joint density of stochastically independent random variables represented by Dirac mixtures and the application in the prediction step of a stochastic state estimator is presented. The systematic Dirac locations are obtained by minimizing a modified version of the Cramér-von Mises distance measure, which is the result of the use of an alternative to the classical cumulative distribution called Localized Cumulative Distribution (LCD). Due to the computationally intensive nature of the distance measure, a suitable approximation is found, which uses an implicit evaluation. As a result, a dramatic speedup is achieved and for the approximation, the curse of dimensionality on the number of Dirac components in the joint density is eliminated. It is then shown by two simulations that the implicit evaluation even slightly surpasses the predicted speedup for the calculation of the approximation. In addition, the sampling itself outperforms a particle filter in a comparison over the RMSE of the mean and the covariance matrix.

The proposed approach is, in contrast to the particle filter, capable of maintaining the shape of a density even if only few Dirac components are used. The capability to use a sampling of the system noise with only few samples available opens great possibilities for applications such as learning scenarios. In addition to this, there are many fields in science and engineering where the generation of samples is expensive or even impossible for a large amount of samples or the calculation of the prediction step is of high computational complexity, so only very few samples can be used [17], [18]. It is also possible to easily expand the presented work to continuous, e.g., Gaussian, densities and therefore to approximate combinations of continuous and discrete densities, like those used in the SGMF [5]. In order to achieve that expansion, the LCDs of the continuous densities, as presented in [14] for a Gaussian case, have to be used for the terms ${}^{r}P$ in Theorem IV.1.

In addition, the optimization procedure cannot only be extended to minimizing the distance measure under constraints such as equal covariances or other moments. It is also simple to exclude certain regions from the approximation or to include additional information about the system function.

Future work has to be directed towards the development of a similar filtering step. In addition to this, a further improvement of the prediction step regarding the systematic error implied by the systematic approximation has to be investigated. Of course, the systematic error cannot be eliminated completely when a systematic approximation is used. Still, a promising approach for reducing the systematic error is the inclusion of the system function in the optimization process. This would give the advantage of enabling the prediction of even strongly nonlinear systems and still maintain high-quality results like the ones shown in the simulations.

REFERENCES

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[2] S. J. Julier and J. K. Uhlmann, "A New Extension of the Kalman Filter to Nonlinear Systems," in *International Symposium on Aerospace/Defence Sensing, Simulation and Control*, 1997.

[3] E. A. Wan and R. van der Merwe, "The Unscented Kalman Filter for Nonlinear Estimation," in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000*, 2000, pp. 153–158.

[4] M. F. Huber and U. D. Hanebeck, "Gaussian Filter based on Deterministic Sampling for High Quality Nonlinear Estimation," in *Proceedings of the 17th IFAC World Congress (IFAC)*, Seoul, Korea, Jul. 2008.
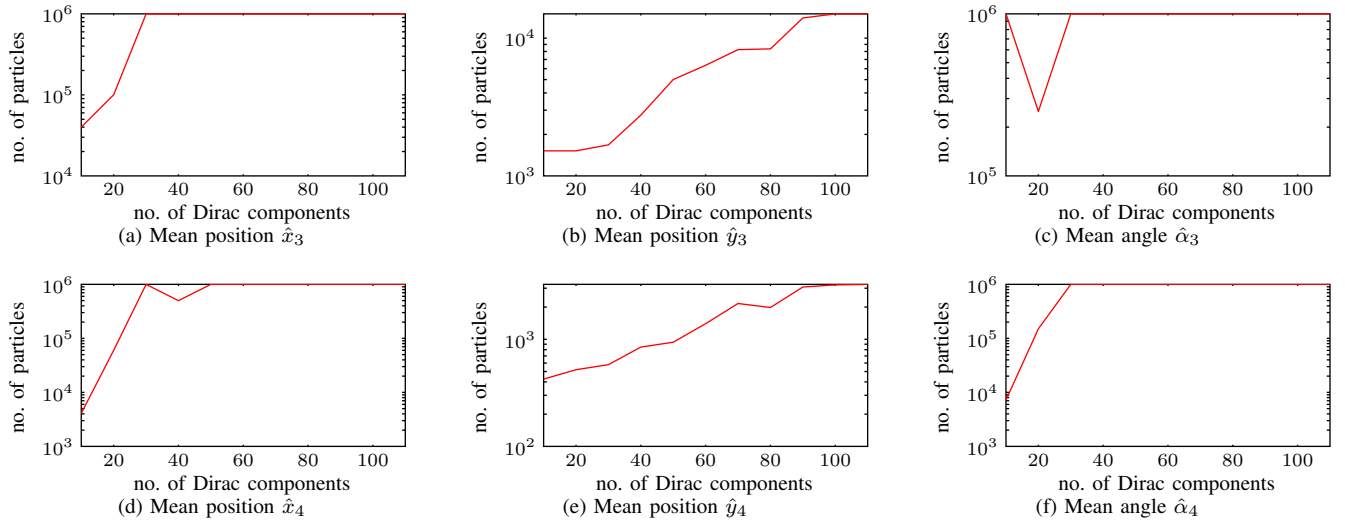
Fig. 6: Number of particles against number of Dirac components that yield the same mean RMSE for the first moment for the time steps $k = 3$ and $k = 4$.
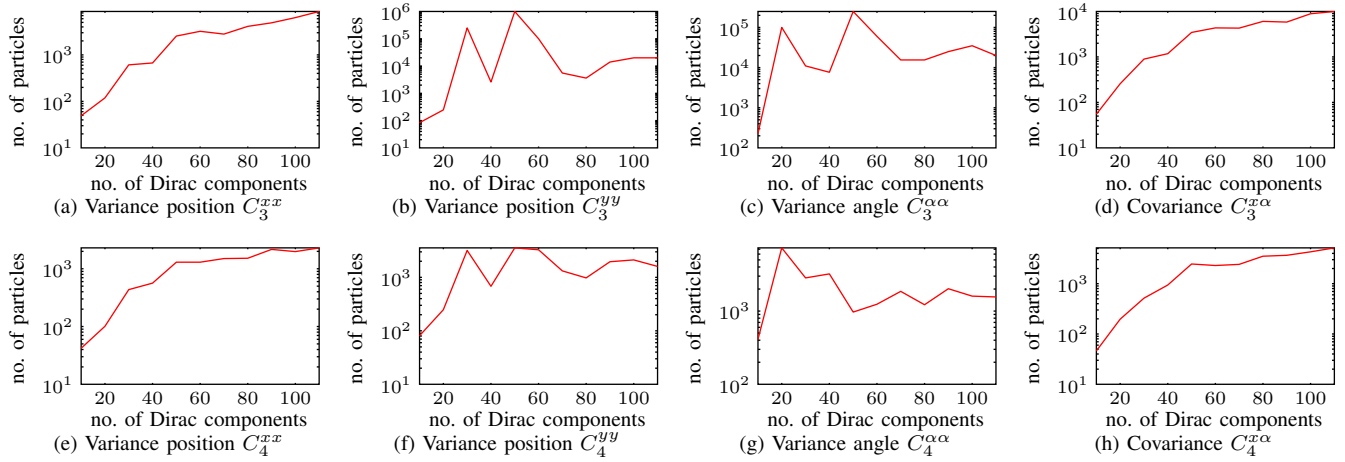

Fig. 7: Number of particles against number Dirac components that yield the same mean RMSE for selected covariance entries for the time steps $k = 3$ and $k = 4$.

[5] V. Klumpp, F. Sawo, U. D. Hanebeck, and D. Fränken, "The Sliced Gaussian Mixture Filter for Efficient Nonlinear Estimation," in *Proceedings of the 11th International Conference on Information Fusion (Fusion 2008)*, Cologne, Germany, Jul. 2008, pp. 1–8.

[6] D. Brunn, F. Sawo, and U. D. Hanebeck, "Nonlinear Multidimensional Bayesian Estimation with Fourier Densities," in *Proceedings of the 2006 IEEE Conference on Decision and Control (CDC 2006)*, San Diego, California, Dec. 2006, pp. 1303–1308.

[7] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[8] R. van der Merwe, N. de Freitas, A. Doucet, and E. Wan, "The Unscented Particle Filter," in *Advances in Neural Information Processing Systems 13*, Nov 2001.

[9] 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, Dec. 2006.

[10] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, 1992.

[11] U. D. Hanebeck and O. C. Schrempf, "Greedy Algorithms for Dirac Mixture Approximation of Arbitrary Probability Density Functions," in *Proceedings of the 2007 IEEE Conference on Decision and Control (CDC 2007)*, New Orleans, Louisiana, Dec. 2007, pp. 3065–3071.

[12] R. Xu and D. C. Wunsch II, "Survey of Clustering Algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

[13] U. D. Hanebeck and V. Klumpp, "Localized Cumulative Distributions and a Multivariate Generalization of the Cramér-von Mises Distance," 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. 33–39.

[14] U. D. Hanebeck, M. F. Huber, and V. Klumpp, "Dirac mixture approximation of multivariate gaussian densities," in *Proceedings of the 2009 IEEE Conference on Decision and Control (CDC 2009)*, Shanghai, China, Dec. 2009.

[15] P. Lu, J. Nocedal, C. Zhu, and R. H. Byrd, "A Limited-Memory Algorithm for Bound Constrained Optimization," *SIAM Journal on Scientific Computing*, vol. 16, pp. 1190–1208, 1994.

[16] W. Gander and W. Gautschi, "Adaptive Quadrature - Revisited," *Tidskrift för Informationsbehandling (BIT)*, vol. 40, pp. 84–101, 1998.

[17] G. Fasano and A. Franceschini, "A Multidimensional Version of the Kolmogorov-Smirnov Test," in *Royal Astronomical Society, Monthly Notices (ISSN 0035-8711), vol. 225, March 1, 1987, p. 155-170.*, 1987.

[18] D. Bacon, N. Ahmad, Z. Boybeyi, T. Dunn, M. Hall, P. Lee, R. Sarma, M. Turner, K. Waight, S. Young, and J. Zack, "A Dynamically Adapting Weather and Dispersion Model: The Operational Multiscale Environment Model with Grid Adaptivity (OMEGA)," *Monthly Weather Review*, vol. 128, no. 7, pp. 2044–2076, 2000.