Progressive Correction for Deterministic Dirac Mixture Approximations

Patrick Ruoff, Peter Krauthausen, and Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory (ISAS),
Institute for Anthropomatics,
Karlsruhe Institute of Technology (KIT), Germany.
Email: patrick.ruoff@kit.edu, peter.krauthausen@kit.edu, uwe.hanebeck@ieee.org

***Abstract*—Since the advent of Monte-Carlo particle filtering, particle representations of densities have become increasingly popular due to their flexibility and implicit adaptive resolution.**

In this paper, an algorithm for the multiplication of a systematic Dirac mixture (DM) approximation with a continuous likelihood function is presented, which applies a progressive correction scheme, in order to avoid the particle degeneration problem. The preservation of sample regularity and therefore, representation quality of the underlying smooth density, is ensured by including a new measure of smoothness for Dirac mixtures, the DM energy, into the distance measure. A comparison to common correction schemes in Monte-Carlo methods reveals large improvements especially in cases of small overlap between the likelihood and prior density, as well as for multi-modal likelihoods.

Keywords: state estimation, nonlinear filtering.

I. Introduction

State estimation as the inference of a hidden system state based on noisy observations is a fundamental problem in engineering. Due to the uncertainty inherent in the problem, Bayesian estimation methods are commonly employed. The solution to the state estimation problem is then obtained by recursive prediction and correction steps. Even though analytic solutions exist for some special type of systems, e.g., linear systems [1], in general no analytic solutions to the prediction and correction equation are obtainable. Therefore, suitable approximations of the systems and/or the densities are required. The rest of this paper is concerned with deriving an approximation of the correction step as needed in a Bayesian state estimator for the case of systematic Dirac mixture approximations of the underlying densities.

For nonlinear systems, the existing approaches may be categorized according to the employed density approximation schemes, as the specific type of approximation governs the treatment of the correction step. For example, the extended Kalman filter (EKF) [2] uses a Gaussian density approximation and linearization of the likelihood around the current state estimate. Gaussian mixture filters [3] offer a more flexible density representation, but in general suffer from an exponential increase in components. The conceptually simplest density representation is an approximation by a Dirac mixture (DM) of equally weighted particles, as used in canonical particle filters. The particle representation is not only very advantageous for simulations, but also offers an implicit adaptive resolution of the density approximation. As pure Monte-Carlo particle representations suffer from slow convergence rates to the true density in the number of particles used, different approximation schemes were developed, such as quasi-Monte-Carlo approximations [4]. However, as particle representations do not capture the smoothness of the original density, naïve algorithms for the correction step in a state estimator lead to the typical particle degeneration problem, where eventually only a few different particles contribute to the posterior distribution.

This paper introduces a regularized progressive correction step in the case that a certain distance measure to the underlying density shall be minimized by deterministically placing particles. In contrast to a progressive correction for Monte-Carlo methods, this problem is especially challenging as the benefit of deterministic approximation schemes, i.e., a representation with fewer particles, is even more likely to degrade in the correction step. Many different schemes have been proposed to handle the correction step for Monte-Carlo particle filters, cf. [5] for an overview. The approach resembling the present work most is [6], which is a progressive correction step for Monte Carlo methods. We extend this scheme to deterministic Dirac mixture approximations, splitting the likelihood adaptively into several factors, each of which is easier to process. In the course of the progression, each factor of the likelihood is multiplied with the current deterministic DM approximation and subsequently reapproximated with equally weighted Diracs. As will be shown, this effectively relocates the Dirac components to regions with higher posterior density and thereby avoids the typical degeneration of the representation. In order to ensure that the DM representation maintains the smoothness/regularity properties of the underlying density throughout the progression, a novel energy term is added to the distance measure used for the reapproximation. This regularization of the DM representation results in a reduced approximation error as a comparison to Monte-Carlo filters reveals for a likelihood and prior density with small "overlap" as well as for multimodal likelihoods.

The rest of this paper is structured as follows: After giving the mathematical problem formulation in Sec. II, the required distance measure for Dirac mixtures is introduced in Sec. III and the progressive correction algorithm is stated in Sec. IV.

The benefits of the proposed approach are discussed in an experimental validation in Sec. V.

II. Problem Formulation

The Bayesian estimation framework consists of two recursively alternating steps: a prediction and a correction step. The prediction step corresponds to calculating the next state estimate $\underline{\boldsymbol{x}}_{k+1}$ based on the current estimate $\underline{\boldsymbol{x}}_k$ and the system dynamics $\underline{a}_k(.)$. The correction step integrates a measurement $\underline{\hat{y}}$ according to the measurement function $\underline{h}_k(.)$ into the current state estimate and thereby corrects the predicted state.

This paper is concerned with the correction step only, i.e.,

$$p(\underline{x}_{k+1}|\underline{y}_{1:k+1}) \propto p(\underline{x}_{k+1}|\underline{y}_{1:k})\, p(\underline{y}_{k+1}|\underline{x}_{k+1}) \ ,$$

where $\underline{y}_{1:k}$ denotes the set of measurements from timestep 1 to k and focuses on a prior density approximated by a DM with $n \in \mathbb{N}$ equally weighted components minimizing the distance measure described in the Sec. III,

$$p(\underline{x}_{k+1}|\underline{y}_{1:k}) \approx \frac{1}{n}\sum_{i=1}^{n} \delta(\underline{x} - \underline{x}_i) \ , \tag{1}$$

where $\delta(\underline{x} - \underline{x}_i)$ denotes the Dirac measure at position $\underline{x}_i \in \mathbb{R}^N$. The goal is to find the optimal DM resulting from the multiplication of the prior density with a continuous likelihood function $l(.)$

$$l(\underline{x}_{k+1}) \equiv p(\underline{\hat{y}}_{k+1}|\underline{x}_{k+1}) \ , \tag{2}$$

where $\underline{\hat{y}}_{k+1}$ denotes the current observation and renormalization was omitted. As we focus on a single correction step, we will omit the time index k from now on. The correction step (2) may be trivially "solved" by directly multiplying the (1) with the likelihood and subsequent renormalization of the new weights

$$w_i = \frac{l(\underline{x}_i)}{\sum_{j=1}^{n} l(\underline{x}_j)} \ . \tag{3}$$

However, this approach will lead to weight decay over time and thus, reduces the effective resolution per Dirac component. Eventually, all but very few components will have nearly zero weights and could be neglected. In order to avoid this degeneracy problem, the basic SIR algorithm [7] draws new samples from the reweighted Diracs in each correction step. This is problematic for systems with low noise, where the propagated components will mainly concentrate around the old particle positions. A possible remedy that is used in "post regularized particle filters", e.g., [8], is to regularize the posterior measure by convolving the particle distribution - after sampling from the reweighted DM - with a Gaussian kernel. Sampling from this regularized density then effectively broadens the distribution, but erodes representation details especially for multimodal densities.

The origin of these problems is illustrated in Fig. 1: Let $R_n(\tilde{f})$ denote the optimal representation in the set of DM with n components and equal weights for a given density $\tilde{f}$. The goal is to find the optimal approximation of the underlying true density $\tilde{f}$ multiplied with the likelihood l as a DM, $R_n(\tilde{f} \cdot l)$, given the optimal approximation of the true density $f = R_n(\tilde{f})$. This problem is obviously ill-posed, as

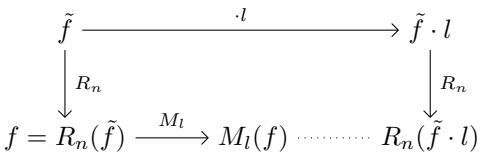


Figure 1. Diagram of the problem to be solved, see text for explanation.

the reduction R_n (drawing samples for Monte-Carlo-based methods or minimization of a certain distance measure here) is not invertible. In order to solve this problem, additional prior information about the smoothness properties of l and $\tilde{f}$ has to be employed. Roughly speaking, what one assumes is that the true density has its probability mass "smeared out" uniformly between Dirac components. The goal is then to use an initial DM approximation and the smoothness assumptions in order to construct an algorithm M_l producing results as close to the optimal solution $R_n(\tilde{f} \cdot l)$ as possible.

There are two challenges when trying to incorporate the smoothness assumption of the underlying densities. First, the likelihood evaluation at the original Dirac positions will yield only a high resolution in regions where the prior density is high. What is sought, however, is a resolution that is adapted to the posterior density instead. Fig. 2 illustrates this problem.

Based on the insight that this problem is less problematic for likelihoods with low variation, the idea of the proposed approach is to split the likelihood into factors with lower variation and perform the multiplication progressively. The proposed approach resembles [6] and can be understood as an extension of this approach to systematic DM approximations. The factors of the likelihood function are each multiplied pointwise with the current intermediate density and subsequently reapproximated with an equally weighted DM. This effectively moves the Dirac components to regions with higher posterior probability, thereby sampling the likelihood with a higher resolution in these regions.

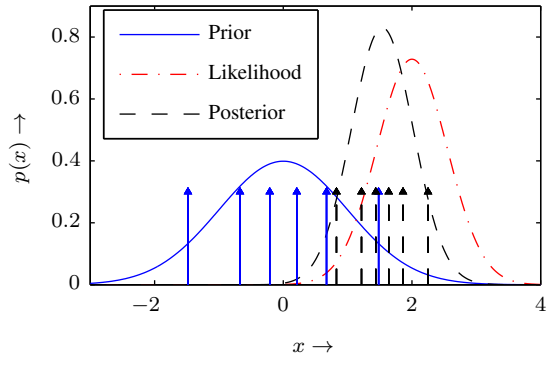


Figure 2. Illustration of the resolution problem: The blue Dirac components approximate the prior Gaussian well, but offer low resolution in regions with high posterior density. The optimal evaluation points for the likelihood would correspond to the points that optimally approximate the posterior density (dashed Dirac components).

A second challenge is to guarantee consistency with the original continuous density, in the sense that the processed samples should still represent a smooth function and therefore, must not form local clusters throughout the progression. Instead of only working with a general purpose distance measures on the lower path of Fig. 1, one has to incorporate another term that connects to the underlying true density $\tilde{f}$ by enforcing inter-Dirac distance preservation in the reapproximation step, which will be described in Sec. IV-B. First, in Sec. III, the employed distance measure is introduced and in Sec. IV, the progressive correction step is explained.

III. Localized Cumulative Distribution and Modified Cramér-von-Mises distance

In order to compare two given Dirac mixture distributions, a suitable distance measure has to be defined. As pointwise evaluation and integral distance measures are not defined for Dirac distributions, this distance measure has to be based on probability mass differences under a subset of possible kernel functions, which motivates the following concepts, as originally defined in [9].

Definition 1 (Localized Cumulative Distribution, [9])
Let $f : \mathbb{R}^N \to \mathbb{R}_+$ be a probability density function. The corresponding Localized Cumulative Distribution (LCD) is defined as

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

where $\star$ denotes the convolution product, $\underline{b} \in \mathbb{R}_+^N$ is a width parameter, and $K : \mathbb{R}^N \times \mathbb{R}^N \to [0, 1]$ a kernel function.

In this paper, only Gaussian kernels of the form

$$K_b(\underline{x}) \equiv K(\underline{x}, b) = \exp\left(-\frac{\|x\|^2}{2b^2}\right) \ ,$$

with a single width parameter $b \in \mathbb{R}$ and the Euclidean distance $\|.\|$ are used. In analogy to the Cramér-von Mises distance, a distance between two generic probability distributions is defined in [9] via the squared integral distance of their corresponding LCDs.

Definition 2 (Modified Cramér-von Mises distance, [9])
Let $f : \mathbb{R}^N \to \mathbb{R}_+$ and $g : \mathbb{R}^N \to \mathbb{R}_+$ be two probability distributions and F_f / F_g their corresponding LCDs. The modified Cramér-von Mises distance (mCvMD) is defined as

$$\begin{aligned} D(f, g) &\equiv \int_{\mathbb{R}} w(b) \int_{\mathbb{R}^N} \left(F_f(\underline{m}, b) - F_g(\underline{m}, b)\right)^2 \mathrm{d}\underline{m}\, \mathrm{d}b \\ &= \int_{\mathbb{R}} w(b)\, \|K_b \star f - K_b \star g\|_{L^2}^2\, \mathrm{d}b \ , \end{aligned}$$

with a suitable weighting function $w : \mathbb{R}_+ \to \mathbb{R}_+$.

As defined above, the mCvMD compares two densities by their differences in probability mass under smooth localized kernels. As a weighted mean over all possible kernel positions and kernel widths is performed, the resulting measure is translation-invariant and does not favor the equivalence of the given densities at a particular scale.

The use of isotropic Gaussian kernels and the squared integral distance measure allows for an analytic solution of all integrals in the mCvMD of the LCDs of two DMs.

Theorem 1 (mCvMD for Dirac Mixture Densities)
Let $f(\underline{x}) = \sum_{i=1}^{n} w_i\, \delta(\underline{x} - \underline{x}_i)$ and $g(\underline{x}) = \sum_{i=1}^{m} \tilde{w}_i\, \delta(\underline{x} - \tilde{\underline{x}}_i)$ denote two N-dimensional Dirac mixture densities. If f and g possess equal means, the mCvMD can be expressed as

$$D(f, g) = \frac{\pi^{N/2}}{8} \left(\underline{w}^T \mathbf{\Gamma}^{\mathbf{xx}} \underline{w} - 2\underline{w}^T \mathbf{\Gamma}^{\mathbf{x\tilde{x}}} \underline{\tilde{w}} + \underline{\tilde{w}}^T \mathbf{\Gamma}^{\mathbf{\tilde{x}\tilde{x}}} \underline{\tilde{w}}\right) \ ,$$

where $(\underline{w})_i \equiv w_i$ denote the weight vectors and the matrices $\mathbf{\Gamma}^{\mathbf{xy}}$ are defined via

$$\mathbf{\Gamma}^{\mathbf{xy}}{}_{i,j} \equiv \gamma(\|\underline{x}_i - \underline{y}_j\|^2)$$

with the function $\gamma(s) \equiv s \ln s$.

A proof of this theorem is given in [10].

IV. Progressive Correction

In this section, the progressive correction step is derived. The likelihood decomposition and the regularization term are introduced, as well as the overall algorithm is stated.

A. Likelihood Decomposition

As motivated, the likelihood decomposition addresses the problem of misaligned likelihood evaluation points by the introduction of sub-steps. To this end, the likelihood function l is dynamically decomposed into a product of the form

$$l(\underline{x}) = \prod_i l(\underline{x})^{\lambda_i} \ , \tag{4}$$

with the constraints

$$\lambda_i > 0 \ , \qquad \sum_{i=1} \lambda_i = 1 \ . \tag{5}$$

At a given intermediate progression step k, the DM from the previous progression step

$$f_{k-1}(\underline{x}) \equiv \frac{1}{n} \sum_{i=1}^{n} \delta(\underline{x}_i - \underline{x}) \tag{6}$$

approximates the prior density multiplied with the pseudo-likelihoods from all preceding steps, i.e.,

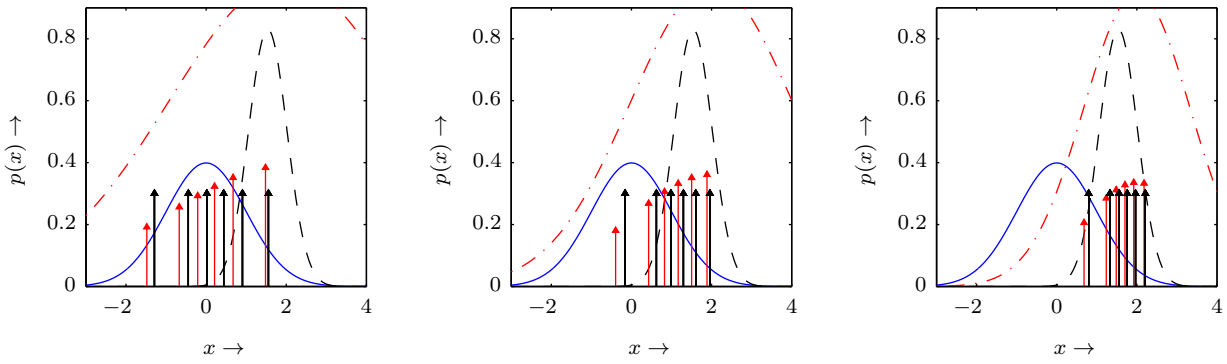
$$f_{k-1} \approx \tilde{f} \cdot l^{\lambda_1} \cdot \ldots \cdot l^{\lambda_{k-1}} \ .$$

Note, that the DM in (6) has equal weights.

The next exponent λ_k is determined in the following way: In order to control the local error in this progression step due to misaligned likelihood evaluation points, we demand that the maximum and minimum Dirac weights after pointwise multiplication with the current pseudo-likelihood l^{λ_k} must not differ by a factor greater than α. The maximum exponent λ_k meeting this requirement may be calculated by

$$\frac{{l_{\max}}^{\lambda_k}}{{l_{\min}}^{\lambda_k}} = \alpha \iff \lambda_k = \frac{\ln \alpha}{\ln l_{\max} - \ln l_{\min}} \ , \tag{7}$$

where $l_{\max} \equiv \max_{i \in \{1,..,n\}} l(\underline{x}_i)$ and $l_{\min} \equiv \min_{i \in \{1,..,n\}} l(\underline{x}_i)$ denote the maximum and minimum likelihood coefficients. The subproblem in this progression step is then to find a good DM representation of $f_{k-1} \cdot l^{\lambda_k}$ given a DM representation of f_{k-1}. As the weights of f_{k-1} multiplied with l^{λ_k} differ



(a) Step 1. (b) Step 6. (c) Step 11.

Figure 3. Progressive correction with $\alpha = 2$ applied to the example in Fig. 2, leading to 11 correction steps. The plots show the current Dirac components before the respective progression step in red, which are reweighted with the current pseudo-likelihood (dash-dotted curve), a broader version of the original likelihood. The resulting Diracs after reapproximation are shown in black. As the progression continues, the Dirac positions move into regions with higher posterior density.

only by a factor of α, the intermediate product $f_{k-1} \cdot l^{\lambda_k}$ is close to f_{k-1} and, hence, the likelihood evaluation points are close to optimal, cf. Fig. 2. The resulting reweighted DM is then reapproximated by an equally weighted Dirac mixture f_k. As a result, the DM components move into areas with higher posterior probability and consequently increase the likelihood sampling resolution in those areas during the next progression step. This procedure is repeated until the whole likelihood is processed. Fig. 3 shows an illustration of different stages in the progression.

B. Regularization

As motivated before, processing the DM progressively solely with a generic distance measure may produce solutions that are close to the original Dirac mixture multiplied with the likelihood $f \cdot l$ but not necessarily close to the real product $\tilde{f} \cdot l$ that one wants to approximate, cf. Fig. 1. In order to incorporate the smoothness assumption of the underlying density during the progression's reapproximation step, a measure for the regularity and smoothness of the underlying density has to be defined. The *energy* of a DM provides such a measure.

Definition 3 (Energy of a Dirac Mixture)
Let $f \equiv \sum_{i=1}^{n} w_i \delta(\underline{x}_i - .)$ denote a DM density. The mixture's energy is defined as

$$E \equiv \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{w_i \, w_j}{\|\underline{x}_i - \underline{x}_j\|} \quad . \tag{8}$$

The definition of the energy in (8) is an analogy from physics, resembling the energy of an electric field of charged particles at positions $\underline{x}_i$ carrying the charge w_i. This new measure of regularity or smoothness of the underlying density is then employed in the reapproximation step as an additional penalty term: In addition to a low mCvMD to the reweighted DM, the reapproximating DM should also have similar smoothness properties and hence similar energy values. The proportion β between approximation quality and smoothness preservation quantifies the importance of the representation quality for the problem at hand and may be determined by generic model selection algorithms [11].

C. Algorithm

In each progression step, the algorithm is composed of two main parts: the likelihood factorization with the weight computation based on the multiplication of the preceding DM with the pseudo-likelihood and the penalized reapproximation of the obtained DM.

First, the new weights are calculated based on the current likelihood exponent as described in Sec. IV-A. As the constraints (5) have to be met, some additional bookkeeping needs to be performed in order to obtain a valid decomposition.

Second, the subprogram `REAPPROX` in Alg. 1, reapproximates the reweighted Dirac mixture $f = \sum_{i=1}^{n} w_j \, \delta(\tilde{\underline{x}}_i - .)$ with an equally weighted DM $g[\mathbf{X}] = \frac{1}{n} \sum_{i=1}^{n} \delta(\underline{x}_i - .)$, where $\mathbf{X}_{i,j} = \underline{x}_i^{(j)}$ denotes the matrix of all Dirac positions, by minimizing the distance measure presented in Sec. III penalized by the energy difference as defined in Sec. IV-B. In summary, the following optimization problem is solved

$$\begin{aligned} \mathbf{X} = \operatorname{argmin} \; & D(f, g[\mathbf{X}]) \; + \; \beta \left(E(f) - E(g[\mathbf{X}]) \right)^2 \\ \text{s.t. } & \underline{1}^T \mathbf{X} / n = \underline{w}^T \tilde{\mathbf{X}} \; . \end{aligned} \tag{9}$$

This problem may be solved with standard solvers, e.g., with a limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [12]. The required gradient of the objective function is provided in the appendix. The algorithm summarizing both main parts is given in Alg. 1.

V. Experiments

In this section, we perform an experimental comparison of the proposed algorithm with standard methods in Monte-Carlo particle filtering. To this end, we set up two experiments, which allow for an analytic solution in order to compute different performance metrics on the results. First, the experimental setup of both experiments is explained. Then, the different

Algorithm 1 Progressive Correction

$t \leftarrow 0$
while $t < 1$ **do**
  $l_i \leftarrow l(\underline{x}_i) \ \forall i$ ▷ Calc. next exponent
  $\lambda \leftarrow \ln \alpha / (\ln l_{\text{max}} - \ln l_{\text{min}})$
  if $\lambda + t > 1$ **then**
    $\lambda \leftarrow 1 - t$
  end if
  $w_i \leftarrow {l_i}^{\lambda} / (\sum_{j=1}^{n} {l_j}^{\lambda}) \ \forall i$ ▷ Calc. new weights
  $\underline{x} \leftarrow \text{REAPPROX}(\underline{x}, \underline{w})$ ▷ Reapproximation
  $t \leftarrow t + \lambda$
end while

performance metrics are explained. Finally, the results are shown and discussed.

A. Experimental Setup

Experiment 1 (Unimodal likelihood): The first experiment considers a Gaussian prior distribution $f = \mathcal{N}\left(\underline{\mu}; \mathbf{\Sigma}\right)$ with zero mean $\underline{\mu} = [0, 0]^T$ and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \ ,$$

which is to be multiplied by a single Gaussian likelihood $l = \mathcal{N}\left(\underline{\mu}_l; \mathbf{\Sigma}_l\right)$ with mean $\underline{\mu}_l = [3, 0]^T$ and covariance matrix

$$\Sigma_l = \begin{bmatrix} 0.4 & 0.3 \\ 0.3 & 0.4 \end{bmatrix} \ .$$

The difficulties for particle algorithms in this experiment lie in the relatively small overlap between prior density and likelihood and, as the likelihood covariance is not axis-aligned, to capture the rotation of the posterior distribution's covariance ellipse in the right way.

Experiment 2 (Bimodal likelihood): The second experiment consists of a zero-mean Gaussian prior distribution with unit covariance, which is to be multiplied with a bimodal likelihood of the form

$$l \equiv \mathcal{N}\left([-3, 0]^T; \mathbf{\Sigma}_l\right) + \mathcal{N}\left([3, 0]^T; \mathbf{\Sigma}_l\right) \ ,$$

with

$$\Sigma_l = \begin{bmatrix} 0.2 & 0 \\ 0 & 10^3 \end{bmatrix} \ ,$$

i.e., a Gaussian mixture with two components that form two narrow strips in y-direction at $x = \pm 3$. The challenges for particle algorithms in this experiment stem from the bimodal likelihood that also yields a bimodal posterior distribution. Hence, the Gaussian assumption usually made in additional Monte-Carlo regularization steps is inappropriate.

B. Initial Approximation and Performance Metrics

First, the Gaussian prior density has to be approximated by a DM with equal weights. For the proposed deterministic algorithm this is performed by minimizing the mCvMD [13]. For the Monte-Carlo algorithms, the initial density is approximated by random sampling.

In order to compare the results of the different algorithms quantitatively, suitable performance metrics have to be defined. As the processed particle density is still a DM, integral distance measures cannot be employed, so we limit ourselves to compare the first and second moment of the resulting densities as well as the mean modified Cramér-von Mises distance (MMCvMD). In the case of Monte-Carlo methods, an average over $n_{\text{run}} = 100$ runs is performed, whereas the proposed deterministic algorithms are run only once. If we define the residual probability in the k^{th} run as $r_k \equiv l \cdot \tilde{f} - M_l(f)_k$, the employed error metrics are defined via

$$\text{RMSE}_m \equiv \left(\frac{1}{n_{\text{run}}} \sum_{k=1}^{n_{\text{run}}} \sum_{i=1}^{N} E[r_k]_i{}^2 \right)^{1/2} \ ,$$

$$\text{RMSE}_C \equiv \left(\frac{1}{n_{\text{run}}} \sum_{k=1}^{n_{\text{run}}} \sum_{i,j=1}^{N} C[r_k]_{i,j}{}^2 \right)^{1/2} \ ,$$

$$\text{MMCvMD} \equiv \frac{1}{n_{\text{run}}} \sum_{k=1}^{n_{\text{run}}} D(l \cdot \tilde{f}, M_l(f)_k) \ ,$$

with E / C denoting the mean / covariance and N the dimension of the system.

C. Results and Discussion

Figure 4 shows a comparison of exemplary runs of the two experiments with 256 particles. In the following, the results of the different algorithms will be discussed.

- **COND:** As expected, due to the only weakly "overlapping" likelihood and prior density, the naïve *condensation algorithm* selects only a few different particles which then have higher weight, thereby reducing the effective number of particles.
- **PR:** The problem of degenerating particle approximations is often addressed by adding a *post-regularization* step, i.e., convolving the given particle density with a Gaussian kernel with a scaled version of the DM's covariance. The optimal scaling parameter
  $$h_{opt} = (4/(N+2))^{\frac{1}{N+4}} \ n^{-\frac{1}{N+4}}$$
  with n the number of particles can be derived in a density estimation context [14], but is only optimal for Gaussian densities. This explains the bad performance of post-regularization in the second example: As the true density is bimodal, with larger extent in x-direction, the smoothing kernels' variance in this direction is too large, which results in a large particle variation in this direction.
- **PPR:** The same problem arises in *progressive post-regularization* [6], which uses $h_{opt}/2$ as scaling parameter but yields comparable covariances to PR, as the error of multiple post-regularization steps accumulate.
- **SRA:** The naïve approach for a systematic multiplication in the mCvMD-minimization context would be to multiply the given prior DM pointwise with the likelihood and then reapproximate the resulting DM by equally weighted Diracs minimizing the mCvMD. This *systematic reapproximation* does not produce a high representation quality as not enough information of the likelihood is taken into account.

- **PSRA w/o energy:** Applying the naïve systematic reapproximation procedure progressively without the additional energy distance term surprisingly does not improve the situation much. As the progression continues, particles tend to form local clusters. The problem here is that the reduction steps minimizing the mCvMD alone introduce small errors that accumulate. Minimization of the mCvMD as a general concept is not specifically tailored to preserve the smoothness properties of a DM, i.e., to assure that the approximated DM is still the reduction of an underlying smooth function.
- **PSRA:** In order to incorporate the assumption that the underlying densities possess strong smoothness properties, another distance term based on the DM energy has to be introduced. The proposed *progressive systematic reapproximation* then produces high-quality results as the resulting particles cover the posterior density more uniformly.

Fig. 5 shows a quantitative evaluation of the different algorithms as a function of the number of particles. One can clearly see that the proposed algorithm outperforms standard Monte-Carlo techniques and the naïve systematic approximation in almost all cases. However, the main advantage of the proposed algorithm is the regular coverage of the posterior distribution with particles, which increases the sampling resolution for further processing and cannot be really quantified by the employed error metrics.

VI. Conclusion

In this paper, we have proposed a progressive filter step for deterministic Dirac mixture approximations, which avoids particle degeneration by preserving an additional energy term throughout the progression. The experimental comparison with current (progressive) Monte-Carlo methods for the correction step, such as the method proposed by [6], reveals a much higher representation quality per particle, as the particles are placed deterministically to cover the posterior density homogeneously. In contrast, Monte-Carlo methods treat each sample individually, which on the one hand reduces the computational complexity to grow only linearly with the number of particles. On the other hand, this independence of particles leads to wasted samples in terms of the posterior sampling resolution and therefore to a worse representation quality per sample.

The higher computational costs involved with the nonlinear optimization in this approach can become negligible if complex system and measurement models are employed. In such cases, the number of samples and thereby evaluations of the system model, needed for Monte-Carlo methods to reach the same representation quality as the proposed method, becomes computationally unfeasible.

VII. Appendix

Theorem 2 (Gradient of Modified Distance Measure) *Let $f = \sum_{i=1}^{n} \tilde{w}_j \, \delta(\tilde{\underline{x}}_i - .)$ and $g[\underline{x}] = \sum_{i=1}^{n} w_i \delta(\underline{x}_i - .)$ denote two Dirac mixture densities. The gradient of the modified distance measure in (9) with respect to the k^{th} component of the l^{th} Dirac position is then given as*

$$\partial x_l^{(k)} \left(D(f, g[\underline{x}]) \; + \; \beta \left(E(f) - E(g[\underline{x}]) \right)^2 \right) =$$
$$\partial x_l^{(k)} D(f, g[\underline{x}]) \; + \; 2\beta \left(E(f) - E(g[\underline{x}]) \right) \partial x_l^{(k)} E(g[\underline{x}]) \;\; ,$$

with

$$\partial x_l^{(k)} D(f, g[\underline{x}]) = \underline{w}^T \left(\partial x_l^{(k)} \mathbf{\Gamma}^{\mathbf{x}\mathbf{x}} \right) \underline{w} - 2 \underline{w}^T \left(\partial x_l^{(k)} \mathbf{\Gamma}^{\mathbf{x}\tilde{\mathbf{x}}} \right) \tilde{\underline{w}}$$
$$= 4 w_l \sum_{j=1}^{n} (w_j \dot{\gamma}(\|\underline{x}_l - \underline{x}_j\|^2)(x_l^{(k)} - x_j^{(k)})$$
$$- \tilde{w}_j \dot{\gamma}(\|\underline{x}_l - \tilde{\underline{x}}_j\|^2)(x_l^{(k)} - \tilde{x}_j^{(k)})) \;\; ,$$

where $\dot{\gamma}(s) = \ln s + 1$ and

$$\partial x_l^{(k)} E(g[\underline{x}]) = w_l \sum_{\substack{j=1 \\ j \neq l}}^{n} w_j \frac{x_j^{(k)} - x_l^{(k)}}{\|\underline{x}_l - \underline{x}_j\|^3} \;\; .$$

References

[1] R. Kalman *et al.*, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[2] S. Julier and J. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," in *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, vol. 3. Citeseer, 1997, p. 26.

[3] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *Automatic Control, IEEE Transactions on*, vol. 17, no. 4, pp. 439–448, 1972.

[4] D. Guo and X. Wang, "Quasi-Monte Carlo filtering in nonlinear dynamic systems," *Signal Processing, IEEE Transactions on*, vol. 54, no. 6, pp. 2087–2098, 2006.

[5] 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, 2002.

[6] N. Oudjane and C. Musso, "Progressive correction for regularized particle filters," in *Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on*, vol. 2. IEEE, 2002.

[7] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *Radar and Signal Processing, IEE Proceedings F*, vol. 140, no. 2. IET, 1993, pp. 107–113.

[8] F. LeGland, C. Musso, and N. Oudjane, "An analysis of regularized interacting particle methods for nonlinear filtering," in *Proceedings of the 3rd IEEE European Workshop on Computer-Intensive Methods in Control and Signal Processing*. Citeseer, 1998.

[9] 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.

[10] H. Eberhardt, V. Klumpp, and U. D. Hanebeck, "Optimal Dirac Approximation by Exploiting Independencies," in *Proceedings of the 2010 American Control Conference (ACC 2010)*, Baltimore, Maryland, Jun. 2010.

[11] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Verlag, 2001.

[12] R. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.

[13] 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.

[14] B. Silverman, *Density estimation for statistics and data analysis*. Chapman & Hall/CRC, 1986.

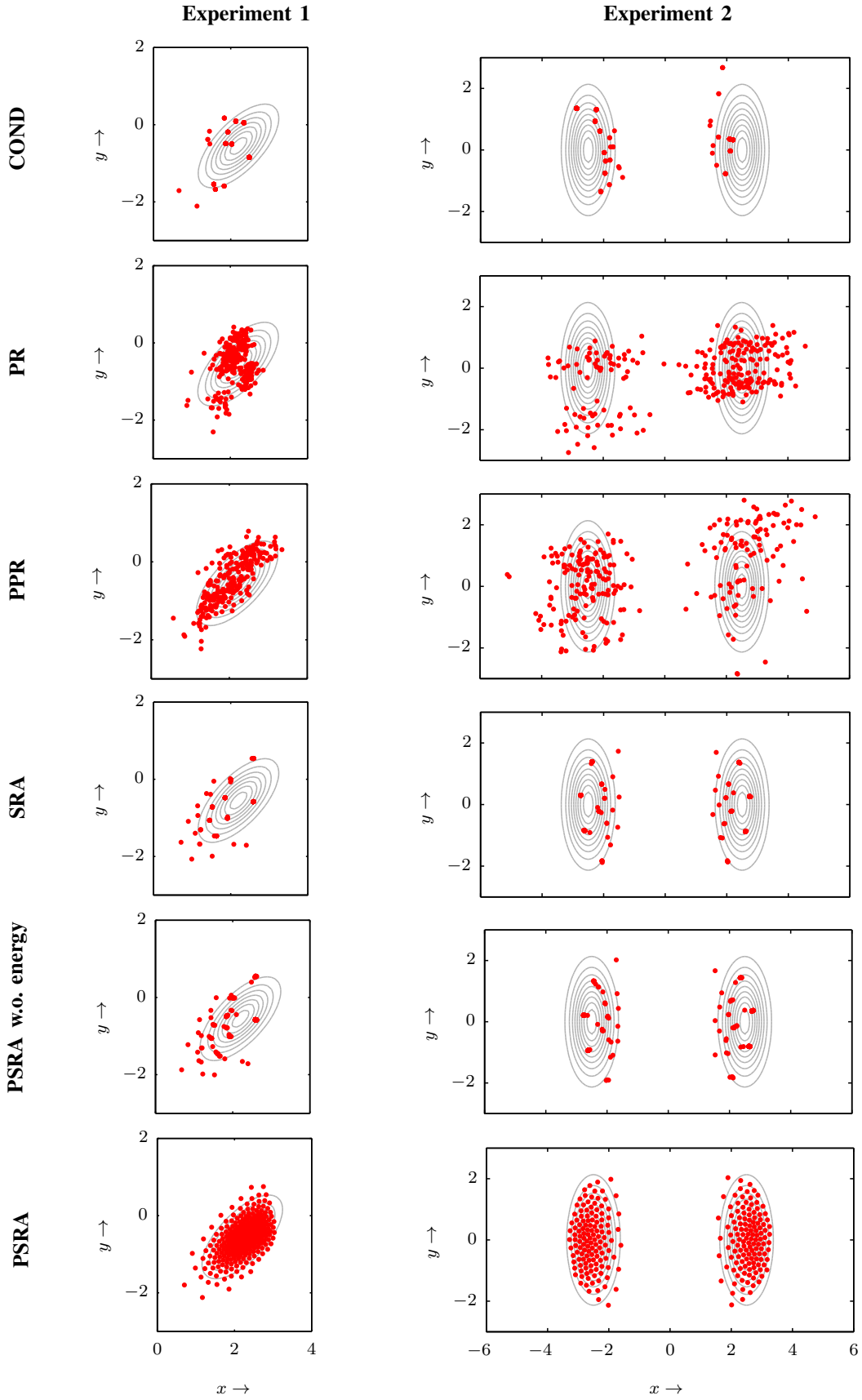


Figure 4. Exemplary runs of typical particle filter methods and the proposed deterministic algorithm for 256 particles. (COND) denotes the simple condensation algorithm, (PR) adds a post-regularization step, (PPR) is the progressive post-regularization method as in Oudjane et al. [6] with $\alpha = 64$, (SRA) is the naive systematic reapproximation, (PSRA w.o. energy) the proposed algorithm with $\alpha = 64$ but $\beta = 0$, and (PSRA) the proposed algorithm with $\alpha = 64$ and $\beta = 100$, cf. Sec. V-C for further details.

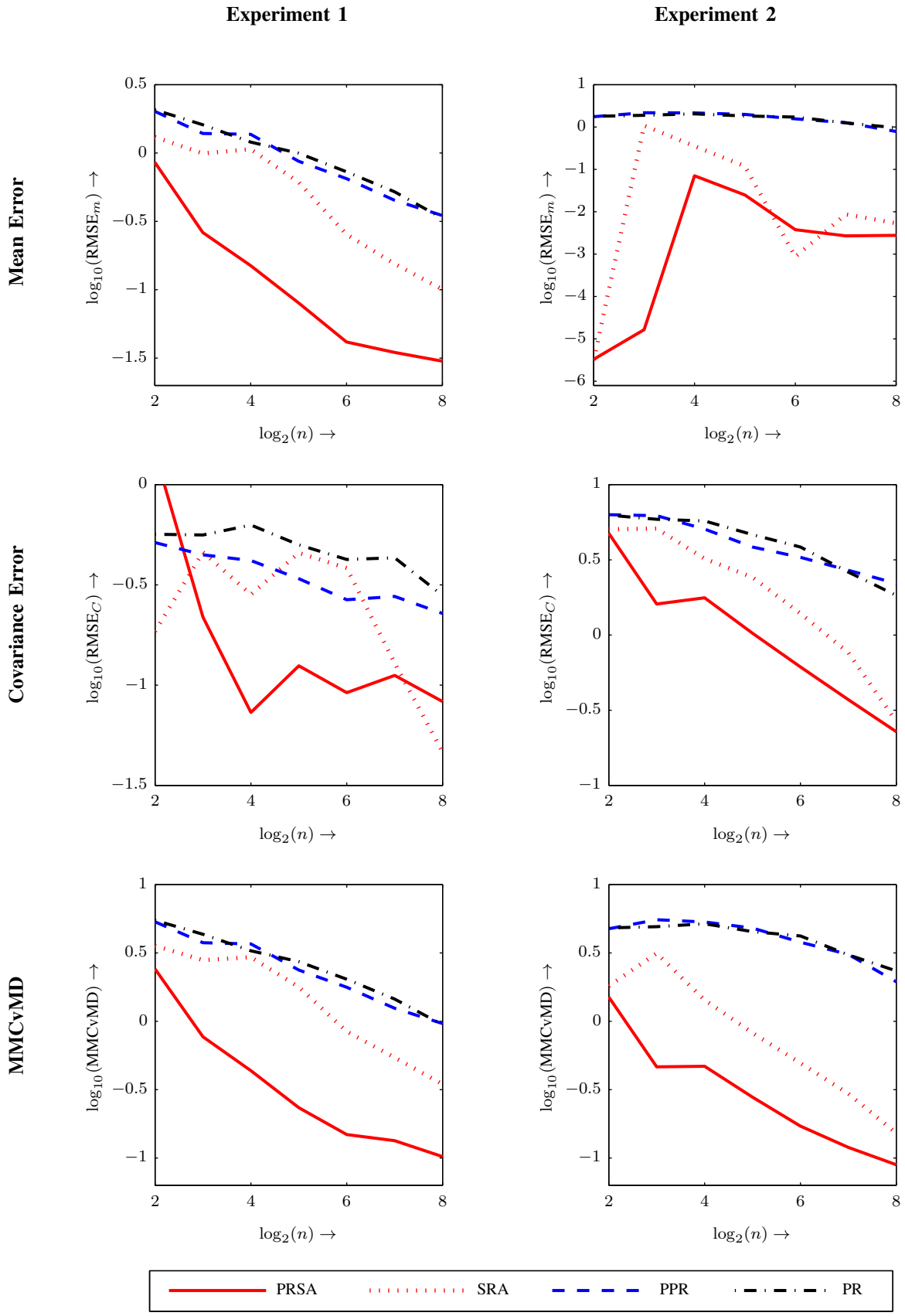


Figure 5. Performance of the proposed method with $\alpha = 64$ and $\beta = 100$ (PRSA) versus the naïve systematic reapproximation (SRA), condensation and post-regularization (PR) and progressive post-regularization [6] with $\alpha = 64$ (PPR) in different error metrics: RMSE of the mean and covariance matrix, as well as the (shifted) mean mCvMD. More explanation can be found in the text.