Semi-Analytic Gaussian Assumed Density Filter

Marco F. Huber, Frederik Beutler, and Uwe D. Hanebeck

Abstract— For Gaussian Assumed Density Filtering based on moment matching, a framework for the efficient calculation of posterior moments is proposed that exploits the structure of the given nonlinear system. The key idea is a careful discretization of some dimensions of the state space only in order to decompose the system into a set of nonlinear subsystems that are conditionally integrable in closed form. This approach is more efficient than full discretization approaches. In addition, the new decomposition is far more general than known Rao-Blackwellization approaches relying on conditionally linear subsystems. As a result, the new framework is applicable to a much larger class of nonlinear systems.

I. INTRODUCTION

In Bayesian estimation, the hidden internal state of arbitrary systems has to be estimated based on measured input and output sequences that are typically corrupted by noise. For linear systems affected with Gaussian noise, the Kalman Filter [1] is the best estimator and allows a closed-form calculation. In case of nonlinearities in the system and measurement equation, however, estimation cannot be performed analytically in general. Instead, approximate state estimators have to be employed. Popular estimators for nonlinear systems make use of the Gaussian assumption for representing the posterior density of the state. Examples are sample-based "black box" approaches utilizing random sampling such as the Gaussian Particle Filter [2], deterministic sampling [3], or statistical linear regression [4], [5], [6], [7]. Alternatively to sampling, analytic approaches may be applicable, where analytic moment calculation [8] leads to closed form formulas of the required first two moments. However, this requires carefully performed derivations and thus, is impractical for high-dimensional systems. If the first two moments cannot be calculated in closed form, an nth-order Taylor series expansion can be used to approximate the underlying system and measurement equation for calculating the moments, which is used in the Extended Kalman Filter [9] or the Gaussian second-order filter [10]. On the other hand, if the system exhibits conditionally linear substructures, sample-based approaches can exploit this fact by using Rao-Blackwellization [11] in order to calculate the required moments more efficiently [12], [13].

In this paper, a combination of sample-based estimation and analytic moment calculation for Gaussian Assumed

M. F. Huber is with the Variable Image Acquisition and Processing Research Group, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany. `marco.huber@ieee.org`.

F. Beutler and U. D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany. `beutler@ieee.org`, `uwe.hanebeck@ieee.org`.

Density Filters is proposed by using a very general form of Rao-Blackwellization. Instead of merely exploiting linear substructures, the new approach relies on an intelligent decomposition of the system and measurement equation into *nonlinear* subsystems that are conditionally integrable in closed form by means of discretizing *some dimensions of the state space*. Thus, given a discretization of some dimensions, moment calculation of the remaining dimensions of the state space can be performed analytically and exactly, which improves the overall estimation accuracy and reduces the computational burden.

The structure of the paper is as follows. In Sec. II, a problem formulation is given, providing a brief introduction into Bayesian estimation and Gaussian Assumed Density Filtering. If the moments can be calculated in closed form, we come up with the Analytic Gaussian Assumed Density Filter (AGF) as described in Sec. III. The class of Sample-Based Gaussian Assumed Density Filters (SGFs) is explained in Sec. IV. The proposed combination of analytic moment calculation and the sample-based approach, named Semi-Analytic Gaussian Assumed Density Filter (SAGF), is part of Sec. V. The performance of the SAGF is shown in simulation examples in Sec. VI. A final discussion and an outlook to future research work are part of Sec. VII.

II. PROBLEM FORMULATION

A nonlinear discrete-time dynamic system is given by

$$\underline{x}_{k+1} = \underline{a}_k\left(\underline{x}_k, \underline{u}_k, \underline{w}_k\right) \ , \tag{1}$$

$$\underline{y}_k = \underline{h}_k\left(\underline{x}_k, \underline{v}_k\right) \ , \tag{2}$$

where the functions $\underline{a}_k\left(\cdot,\cdot,\cdot\right)$ and $\underline{h}_k\left(\cdot,\cdot\right)$ are known. The vector $\underline{x}_k$ is the state of the system, $\underline{y}_k$ is the measurement vector, and $\underline{u}_k$ is a known system input at the discrete time step k. The terms $\underline{v}_k$ and $\underline{w}_k$ represent zero-mean measurement and process noise, respectively. Measurement values $\hat{\underline{y}}_k$ are realizations of the measurement process (2).

A. Bayesian Estimation

In Bayesian estimation, two alternating steps, i.e., prediction and filtering, are performed for estimating the system state $\underline{x}_k$. In the prediction step, the result $f^e(\underline{x}_k) := f(\underline{x}_k|\hat{\underline{y}}_{1:k})$ of the previous filter step is propagated from time step k to $k+1$ by means of

$$f^p(\underline{x}_{k+1}) := f(\underline{x}_{k+1}|\hat{\underline{y}}_{1:k}) = \int f(\underline{x}_{k+1}|\underline{x}_k) \cdot f^e(\underline{x}_k)\,\mathrm{d}\underline{x}_k \ , \tag{3}$$

where $f(\underline{x}_{k+1}|\underline{x}_k)$ is the transition density defined by (1). $\hat{\underline{y}}_{1:k} = (\hat{\underline{y}}_1, \ldots, \hat{\underline{y}}_k)$ summarizes all measurement values up to and including time step k. In the filter step, the current

measurement value $\hat{\underline{y}}_k$ is used for updating the result of the prediction step $f^p(\underline{x}_k)$ according to Bayes' rule

$$f^e(\underline{x}_k) := f(\underline{x}_k | \hat{\underline{y}}_{1:k}) = \tfrac{1}{c_k} \cdot f(\hat{\underline{y}}_k | \underline{x}_k) \cdot f^p(\underline{x}_k) \ , \quad (4)$$

where $c_k = \int f(\hat{\underline{y}}_k | \underline{x}_k) \cdot f^p(\underline{x}_k) \, \mathrm{d}\underline{x}_k$ is a normalization constant and $f(\hat{\underline{y}}_k | \underline{x}_k)$ is the likelihood defined by (2).

For both prediction and filtering, closed-form solutions of the occurring integrals are not available in general and thus, appropriate approximations have to be applied. In this paper, we restrict attention to Gaussian Assumed Density Filters [14], i.e., the densities in (3) and (4) are assumed to be Gaussian. Furthermore, the parameters of these Gaussian densities are calculated by moment matching. In doing so, filtering and prediction boils down to the efficient calculation of the mean vector and the covariance matrix.

B. Gaussian Assumed Density Filter

For deriving a Gaussian Assumed Density Filter, it is sufficient to concentrate on the nonlinear transformation

$$\underline{y} = \underline{g}(\underline{x}) \quad (5)$$

of the Gaussian random vector $\underline{x}$ with density function $f(\underline{x}) = \mathcal{N}(\underline{x}; \underline{\mu}^x, \mathbf{C}^x)$ to the random vector $\underline{y}$ with density $f(\underline{y})$.

1) Forward Inference: In the forward inference, the random vector $\underline{x}$ is propagated through the nonlinear transformation (5) in order to calculate the first two moments of $\underline{y}$, i.e., mean $\underline{\mu}^y$ and covariance $\mathbf{C}^y$. This type of inference occurs in the prediction step, where the nonlinear transformation (5) corresponds to the system function (1) and the density of $\underline{x}$ is given by $f^e(\underline{x})$ with mean $\underline{\mu}_k^e$ and covariance $\mathbf{C}_k^e$. Furthermore, $\underline{\mu}^y$ and $\mathbf{C}^y$ correspond to the predicted mean $\underline{\mu}_{k+1}^p$ and predicted covariance $\mathbf{C}_{k+1}^p$, respectively. Analogously, forward inference is employed for calculating the predicted measurement, which is required for backward inference as described next.

2) Backward Inference: The goal of the backward inference is to determine the conditional density $f(\underline{x} | \underline{y})$, i.e., the conditional mean $\underline{\mu}^{x|y}$ and covariance $\mathbf{C}^{x|y}$. For this purpose, the joint density $f(\underline{x}, \underline{y})$ is determined first. By additionally assuming that this joint density is Gaussian[1], we need to compute the cross-covariance $\mathbf{C}^{x,y}$ of the joint covariance block matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^x & \mathbf{C}^{x,y} \\ (\mathbf{C}^{x,y})^{\mathrm{T}} & \mathbf{C}^y \end{bmatrix} \ .$$

Given the cross-covariance, the desired mean and covariance are then given by

$$\begin{aligned} \underline{\mu}^{x|y} &= \underline{\mu}^x + \mathbf{C}^{x,y} \cdot (\mathbf{C}^y)^{-1} \cdot \left(\hat{\underline{y}} - \underline{\mu}^y \right) \ , \\ \mathbf{C}^{x|y} &= \mathbf{C}^x - \mathbf{C}^{x,y} \cdot (\mathbf{C}^y)^{-1} \cdot (\mathbf{C}^{x,y})^{\mathrm{T}} \ , \end{aligned} \quad (6)$$

where $\underline{\mu}^y$ and $\mathbf{C}^y$ result from a forward inference step and $\hat{\underline{y}}$ is a realization of $\underline{y}$.

[1]This assumption is common in Gaussian Assumed Density Filtering (as in the EKF or the UKF) and is only true for linear systems affected with Gaussian noise. Otherwise it is an approximation.

III. ANALYTIC GAUSSIAN ASSUMED DENSITY FILTER (AGF)

At first, we demonstrate that the mean $\underline{\mu}^y$ and covariance $\mathbf{C}^y$ for forward inference as well as the cross-covariance $\mathbf{C}^{x,y}$ for backward inference can be determined analytically and exactly in special cases. Thus, the prediction step (3) and filter step (4) can be calculated in closed form. In order to calculate the mean $\underline{\mu}^y$, it can be utilized that

$$\underline{\mu}^y = \mathrm{E}\{\underline{y}\} = \int \underline{y} \cdot f(\underline{y}) \, \mathrm{d}\underline{y} = \int\int \underline{y} \cdot f(\underline{x}, \underline{y}) \, \mathrm{d}\underline{x} \, \mathrm{d}\underline{y} \quad (7)$$

holds. With (5) and Bayes' rule, the joint density function $f(\underline{x}, \underline{y})$ of $\underline{x}$ and $\underline{y}$ can be written according to

$$f(\underline{x}, \underline{y}) = f(\underline{y}|\underline{x}) \cdot f(\underline{x}) = \delta(\underline{y} - \underline{g}(\underline{x})) \cdot f(\underline{x}) \ , \quad (8)$$

where $\delta(\underline{x} - \underline{\mu})$ is the Dirac Delta distribution at position $\underline{\mu}$. Plugging (8) into (7) and utilizing the sifting property of the Dirac Delta distribution results in

$$\underline{\mu}^y = \int \underline{g}(\underline{x}) \cdot f(\underline{x}) \, \mathrm{d}\underline{x} \ . \quad (9)$$

Hence, the mean of $\underline{y}$ can be calculated directly based on the nonlinear function $\underline{g}(\cdot)$ and the density of $\underline{x}$. With (8), the covariance of $\underline{y}$ can be derived in a similar manner, which leads to

$$\begin{aligned} \mathbf{C}^y &= \int\int (\underline{y} - \underline{\mu}^y) \cdot (\underline{y} - \underline{\mu}^y)^{\mathrm{T}} \cdot f(\underline{x}, \underline{y}) \, \mathrm{d}\underline{x} \, \mathrm{d}\underline{y} \\ &= \int \underline{g}(\underline{x}) \cdot \underline{g}(\underline{x})^{\mathrm{T}} \cdot f(\underline{x}) \, \mathrm{d}\underline{x} - \underline{\mu}^y \cdot (\underline{\mu}^y)^{\mathrm{T}} \ . \end{aligned} \quad (10)$$

For the filter step, the cross-covariance $\mathbf{C}^{x,y}$ is required. Similar to the covariance $\mathbf{C}^y$, the cross-covariance $\mathbf{C}^{x,y}$ is calculated by

$$\begin{aligned} \mathbf{C}^{x,y} &= \int\int (\underline{x} - \underline{\mu}^x) \cdot (\underline{y} - \underline{\mu}^y)^{\mathrm{T}} \cdot f(\underline{x}, \underline{y}) \, \mathrm{d}\underline{x} \, \mathrm{d}\underline{y} \\ &= \int \underline{x} \cdot \underline{g}(\underline{x})^{\mathrm{T}} \cdot f(\underline{x}) \, \mathrm{d}\underline{x} - \underline{\mu}^x \cdot (\underline{\mu}^y)^{\mathrm{T}} \ . \end{aligned} \quad (11)$$

Again, merely $\underline{g}(\cdot)$ and $f(\underline{x})$ are necessary for calculating the covariance. Unfortunately, analytically solving the integrals in (9), (10), and (11) is not possible in general. For special function types such as polynomials or trigonometric functions and their combination, however, a closed-form solution is available. If the moments can be calculated in closed form, the resulting estimator is called *A*nalytic *G*aussian Assumed Density *F*ilter (AGF).

Example 1 (Quadratic Transformation) In this example, the nonlinear transformation

$$\underline{y} = \underline{x}^2 \quad (12)$$

is considered. For this simple polynomial transformation, equations (9)–(11) can be solved analytically exactly. The mean μ^y, the variance C^y and the cross-variance $C^{x,y}$ are given by

$$\mu^y = (\mu^x)^2 + C^x \ , \quad (13)$$
$$C^y = 2 \cdot C^x \cdot (C^x + 2 \cdot (\mu^x)^2) \ , \text{ and } \ C^{x,y} = 2 \cdot C^x \cdot \mu^x \ ,$$

respectively.

(a) Analytic stochastic linearization (AGF). (b) Sample-Based stochastic linearization (SGF). (c) Taylor-series based linearization (EKF).
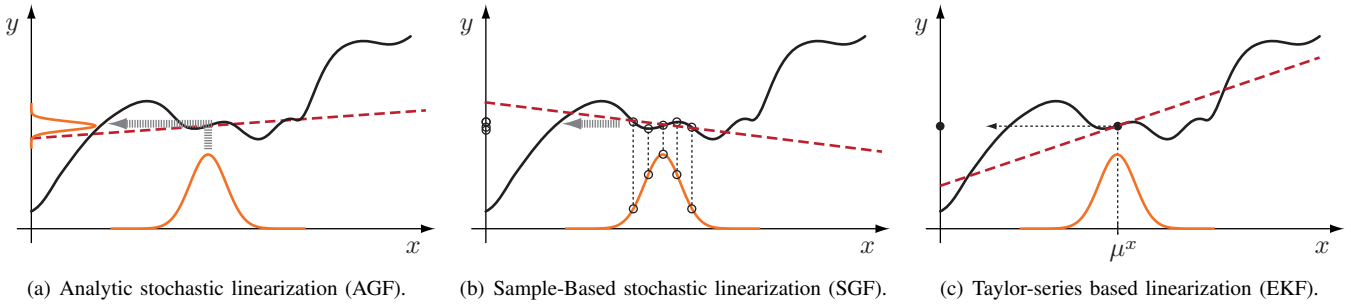
Fig. 1. Illustration of the different linearization approaches: the nonlinear function (black) and its linearized version (red dashed). (a) AGF propagates the entire density $f(\underline{x})$ for linearization. (b) The linearization of an SGF is based on an approximate sample representation of $f(\underline{x})$. (c) The EKF linearizes the nonlinear function around a single point.

It is worth mentioning that for linear transformations $\underline{y} = \mathbf{A} \cdot \underline{x}$, forward and backward inference correspond to the prediction and filter step of the well-known Kalman filter. That is, the Kalman filter is a special case of an AGF.

IV. SAMPLE-BASED GAUSSIAN ASSUMED DENSITY FILTER (SGF)

For nonlinear functions $g(\cdot)$ that prevent a closed-form solution, the so-called Linear Regression Kalman Filters (LRKFs) or Sample-Based Gaussian Assumed Density Filters (SGFs) allow approximately calculating the first two moments of $\underline{y}$. Examples for LRKFs/SGFs are the well-known Unscented Kalman Filter (UKF) [4], the Divided Difference Filter [6], or the Gaussian Filter [7]. These filters utilize a sample representation of $\underline{x}$ given by the Dirac mixture

$$f(\underline{x}) \approx \sum_{i=1}^{L} w_i \cdot \delta(\underline{x} - \underline{\mu}_i) \ , \qquad (14)$$

where L is the number of samples, w_i are non-negative weighting factors, and $\underline{\mu}_i$ are the sample positions. The Dirac mixture (14) exactly captures the mean $\underline{\mu}^x$ and the covariance $\mathbf{C}^x$ of $\underline{x}$. Propagating the samples through (5) corresponds to approximating (9) and (10) by a weighted sample mean and sample covariance, respectively. In case of the mean $\underline{\mu}^y$, replacing the density $f(\underline{x})$ in (9) by the Dirac mixture (14) leads to

$$\underline{\mu}^y \approx \int \underline{g}(\underline{x}) \cdot \left(\sum_i w_i \cdot \delta(\underline{x} - \underline{\mu}_i) \right) \mathrm{d}\underline{x} = \sum_i w_i \cdot \underline{g}(\underline{\mu}_i) \ . \quad (15)$$

The quality of this mean approximation (and similarly of the covariance approximation) depends on both the function $g(\underline{x})$ and the quality of the sample representation of $f(\underline{x})$ in (14). Especially in cases of strong nonlinearities and a small number L of sample points, (15) provides a poor approximation of the true mean $\underline{\mu}^y$.

Example 2 (Quadratic Transformation (cont'd)) We again consider the quadratic transform (12) of Example 1. The density $f(x)$ is approximately represented by the samples

$$\mu_1 = \mu^x + \sqrt{C^x} \ , \ \mu_2 = \mu^x - \sqrt{C^x} \ , \ w_1 = w_2 = 0.5 \ ,$$

which exactly capture the mean μ^x and variance C^x. Propagating the samples through the quadratic transformation and calculating the weighted mean and variance yields

$$\mu^y = (\mu^x)^2 + C^x \ ,$$

$$C^y = 0.5 \sum_{i=1}^{2} ((\mu_i)^2 - \mu^y)^2 = 4 \cdot C^x \cdot (\mu^x)^2 \ ,$$

$$C^{x,y} = 2 \cdot C^x \cdot \mu^x \ .$$

Comparison with (13) shows that the sample-based variance C^y is not exact.

Calculating the mean and covariance of $\underline{y}$ corresponds to a linearization of the nonlinear transformation $g(\cdot)$ as the density functions $f(\underline{x})$ and $f(\underline{y})$ are both assumed to be Gaussian (Fig. 1). In case of the AGF, linearization is performed implicitly under consideration of the entire Gaussian density $f(\underline{x})$. In contrast to this, SGFs merely propagate a sample-based approximation of $f(\underline{x})$. Even if the mean and covariance of $\underline{x}$ are captured exactly by the samples, the same is not true for (all) higher-order moments due to the finite number of samples. But again, linearization is performed implicitly due to the calculation of the sample mean and sample covariance (see for example equation (15)). The Extended Kalman Filter (EKF, [9]) is an example of a Gaussian filter, where linearization is done explicitly by first-order Taylor-series expansion of $g(\cdot)$ around the mean $\underline{\mu}^x$. Here, linearization is only performed at a single point (the mean of $\underline{x}$) and no uncertainty information about $\underline{x}$ is considered. This typically leads to an inferior performance in comparison with SGFs or the AGF. Merely in case of a linear transformation, AGF, SGF, and EKF provide identical results. The same is true in case of AGF and SGF, if an infinite number of samples is used.

V. SEMI-ANALYTIC GAUSSIAN ASSUMED DENSITY FILTER (SAGF)

To attenuate the drawbacks of a purely sample-based approximation of the mean and covariance, the key idea of the proposed Semi-Analytic Gaussian Assumed Density Filter (SAGF) is to combine the Sample-Based Gaussian Assumed Density Filter (SGF) and the Analytic Gaussian Assumed Density Filter (AGF) such that only some dimensions of the random vector $\underline{x}$ are discretized by means of a

sample representation. Thus, only some parts of the nonlinear transformation (5) have to be evaluated approximately. For this purpose, we rearrange the nonlinear equation (5) to

$$\underline{\boldsymbol{y}} = \underline{g}(\underline{\boldsymbol{x}}^a, \underline{\boldsymbol{x}}^b) \ , \qquad (16)$$

where the Gaussian random vector $\underline{\boldsymbol{x}} = [(\underline{\boldsymbol{x}}^a)^{\mathrm{T}}, (\underline{\boldsymbol{x}}^b)^{\mathrm{T}}]^{\mathrm{T}}$ consists of the substates $\underline{\boldsymbol{x}}^a$ and $\underline{\boldsymbol{x}}^b$ with mean and covariance

$$\underline{\mu}^x = \begin{bmatrix} \underline{\mu}^a \\ \underline{\mu}^b \end{bmatrix} \ , \ \ \mathbf{C}^x = \begin{bmatrix} \mathbf{C}^a & \mathbf{C}^{a,b} \\ \mathbf{C}^{b,a} & \mathbf{C}^b \end{bmatrix} \ .$$

Generally, there exists no closed-form expression for the desired moments. However, the decomposition into $\underline{\boldsymbol{x}}^a$ and $\underline{\boldsymbol{x}}^b$ is chosen in such a way that the integrals in (9)–(11) can be calculated in closed form for any given fixed value of $\underline{\boldsymbol{x}}^b$. Hence, we say that $\underline{g}(\cdot, \cdot)$ is *conditionally integrable*. For determining a sample-based representation of $\underline{\boldsymbol{x}}^b$, the sampling techniques of the SGFs are applied.

Example 3 In the following, two example transformations are discussed. At first, a nonlinear transformation with conditionally linear function $g(\cdot, \cdot)$ is given by

$$y = \mathrm{e}^{-\boldsymbol{x}_1^b} \cdot (\boldsymbol{x}_2^b)^2 \cdot \boldsymbol{x}^a \ .$$

To see this, we replace the substate $\underline{\boldsymbol{x}}^b = [\boldsymbol{x}_1^b, \boldsymbol{x}_2^b]^{\mathrm{T}}$ with a single sample point $\underline{\mu} = [\mu_1, \mu_2]^{\mathrm{T}}$, which leads to $y = c \cdot \boldsymbol{x}^a$, where $c := \mathrm{e}^{-\mu_1} \cdot (\mu_2)^2$. Hence, given the sample point $\underline{\mu}$, the first two moments of y can be calculated via the Kalman predictor. The scope of this paper is even more general. Also decompositions into conditionally integrable *nonlinear* transformations $g(\cdot, \cdot)$ are covered. For the example equation

$$y = \mathrm{e}^{-\boldsymbol{x}^b} \cdot (\boldsymbol{x}_1^a)^2 \cdot \boldsymbol{x}_2^a \ ,$$

the function $g(\underline{\boldsymbol{x}}^a, \mu) = c \cdot (\boldsymbol{x}_1^a)^2 \cdot \boldsymbol{x}_2^a$ with $c := \mathrm{e}^{-\mu}$ is now nonlinear when replacing $\boldsymbol{x}^b$ by a sample point μ. However, $g(\cdot, \mu)$ is polynomial and thus, the moments can be calculated analytically since $\underline{\boldsymbol{x}}^a$ is Gaussian distributed.

The AGF (Section III) and SGF (Section IV) are extreme cases of the SAGF: if $\underline{\boldsymbol{x}}^b$ is an empty vector, SAGF becomes an AGF and if $\underline{\boldsymbol{x}}^a$ is an empty vector, SAGF degenerates to an SGF.

A. General Solution

For the general transformation given by (16), the desired moments of $\underline{\boldsymbol{y}}$ can be calculated as follows. At first, the joint density $f(\underline{x}, \underline{y})$ is separated by employing Bayes' rule

$$f(\underline{x}, \underline{y}) = \underbrace{\delta\Big(\underline{y} - \underline{g}\big(\underline{x}^a, \underline{x}^b\big)\Big)}_{= f(\underline{y}|\underline{x})} \cdot \underbrace{f(\underline{x}^a|\underline{x}^b) \cdot f(\underline{x}^b)}_{= f(\underline{x})} \ ,$$

where the density $f(\underline{x})$ is replaced by $f(\underline{x}^a|\underline{x}^b) \cdot f(\underline{x}^b)$ and the conditional density $f(\underline{x}^a|\underline{x}^b) = \mathcal{N}\big(\underline{x}^a; \underline{\mu}^{a|b}, \mathbf{C}^{a|b}\big)$ is (conditionally) Gaussian with mean and covariance

$$\underline{\mu}^{a|b} = \underline{\mu}^a + \mathbf{C}^{a,b} \cdot \big(\mathbf{C}^b\big)^{-1} \cdot \big(\underline{x}^b - \underline{\mu}^b\big) \ , \\ \mathbf{C}^{a|b} = \mathbf{C}^a - \mathbf{C}^{a,b} \cdot \big(\mathbf{C}^b\big)^{-1} \cdot \mathbf{C}^{b,a} \ . \qquad (17)$$

For determining the mean $\underline{\mu}^y$ in (9), the density $f(\underline{x}^b)$ of the substate $\underline{x}^b$ is represented by means of a Dirac mixture as

in (14) in order to allow applying an SGF. To integrate over $\underline{x}^b$, the sifting property of the Dirac Delta distribution is exploited. Hence, the mean of $\underline{\boldsymbol{y}}$ is given by

$$\underline{\mu}^y \approx \sum_i w_i \cdot \underline{\mu}_i^y \ \ \text{with} \qquad (18)$$

$$\underline{\mu}_i^y = \int \underline{g}(\underline{x}^a, \underline{\mu}_i) \cdot f(\underline{x}^a|\underline{\mu}_i) \, \mathrm{d}\underline{x}^a \ . \qquad (19)$$

Analogously, the covariance of $\underline{\boldsymbol{y}}$ results in

$$\mathbf{C}^y \approx \sum_i w_i \cdot \big(\mathbf{C}_i^y - \underline{\mu}_i^y(\underline{\mu}^y)^{\mathrm{T}} - \underline{\mu}^y(\underline{\mu}_i^y)^{\mathrm{T}} + \underline{\mu}^y(\underline{\mu}^y)^{\mathrm{T}}\big) \ ,$$
$$\qquad\qquad (20)$$
$$\mathbf{C}_i^y = \int \underline{g}\big(\underline{x}^a, \underline{\mu}_i\big) \cdot \underline{g}\big(\underline{x}^a, \underline{\mu}_i\big)^{\mathrm{T}} \cdot f\big(\underline{x}^a|\underline{\mu}_i\big) \, \mathrm{d}\underline{x}^a \ .$$

It is important to note that the integrals in (19) and (20) can be evaluated analytically as the function $\underline{g}(\cdot, \cdot)$ is chosen to be conditionally integrable. Furthermore, solving these integrals is an off-line task and the solution is characterized by a parametric representation for efficient on-line evaluation.

B. Estimation

With the results of the previous section, a complete SAGF consisting of a prediction and a filter step is now derived.

1) Prediction Step: In the prediction step, the predicted mean $\underline{\mu}_{k+1}^p$ and covariance $\mathbf{C}_{k+1}^p$ of $f^p(\underline{x}_{k+1})$ for time step $k + 1$ have to be calculated. For this purpose, the system function (1) can be directly mapped to the nonlinear transformation (16) according to

$$\underline{\boldsymbol{x}}_{k+1} = \underline{a}_k\big(\underline{\boldsymbol{x}}_k, \underline{u}_k, \underline{\boldsymbol{w}}_k\big) = \underline{g}\big(\underline{\boldsymbol{x}}_k^a, \underline{\boldsymbol{x}}_k^b\big) \ .$$

Here, the (deterministic) system input $\underline{u}_k$ becomes a part of the function $\underline{g}(\cdot, \cdot)$ and the substates $\underline{\boldsymbol{x}}_k^a, \underline{\boldsymbol{x}}_k^b$ are augmented with the noise variables $\underline{\boldsymbol{w}}_k^a, \underline{\boldsymbol{w}}_k^b$, where $\underline{\boldsymbol{w}}_k = [(\underline{\boldsymbol{w}}_k^a)^{\mathrm{T}}, (\underline{\boldsymbol{w}}_k^b)^{\mathrm{T}}]^{\mathrm{T}}$, in order to consider additive and/or multiplicative noise.

The sample points of substate $\underline{\boldsymbol{x}}^b$ are calculated based on the sampling scheme of the used SGF. For the mean $\underline{\mu}_{k+1}^p$ and covariance $\mathbf{C}_{k+1}^p$ of $\underline{\boldsymbol{x}}_{k+1}$, (18) and (20) are employed.

2) Filter Step: The measurement equation (2) is mapped to the nonlinear transformation (16) according to

$$\underline{\boldsymbol{y}}_k = \underline{h}_k(\underline{\boldsymbol{x}}_k, \underline{\boldsymbol{v}}_k) = \underline{g}\big(\underline{\boldsymbol{x}}_k^a, \underline{\boldsymbol{x}}_k^b\big) \ ,$$

where the measurement noise $\underline{\boldsymbol{v}}_k$ is part of the substates $\underline{\boldsymbol{x}}_k^a, \underline{\boldsymbol{x}}_k^b$. It is worth mentioning that the decomposition of $\underline{\boldsymbol{x}}_k$ into the substates for the filter step is independent of the decomposition of the prediction step.

The goal of the filter step is to determine the mean $\underline{\mu}_k^e$ and covariance $\mathbf{C}_k^e$ of the estimated density $f^e(\underline{x}_k) = \mathcal{N}(\underline{x}_k; \underline{\mu}_k^e, \mathbf{C}_k^e)$ by using (6), where $\underline{\mu}_k^y$ and $\mathbf{C}_k^y$ of $\underline{\boldsymbol{y}}_k$ correspond to (18) and (20), respectively. The cross-covariance $\mathbf{C}^{x,y} = [\mathbf{C}^{a,y}, \mathbf{C}^{b,y}]^{\mathrm{T}}$ in (6) consists of

$$\mathbf{C}^{a,y} = \sum_i w_i \cdot \left(\mathbf{C}_i^{a,y} - \underline{\mu}_i^{a|b}\big(\underline{\mu}^y\big)^{\mathrm{T}} + \underline{\mu}^a\left(\underline{\mu}^y - \underline{\mu}_i^y\right)^{\mathrm{T}} \right) \ ,$$

$$\mathbf{C}^{b,y} = \sum_i w_i \cdot \left(\underline{\mu}_i - \underline{\mu}^b \right) \cdot \left(\underline{\mu}_i^y - \underline{\mu}^y \right)^{\mathrm{T}} \ ,$$

with

$$\mathbf{C}_i^{a,y} = \int \underline{x}^a \cdot \left(\underline{g}\left(\underline{x}^a, \underline{\mu}_i\right) \right)^{\mathrm{T}} \cdot f\left(\underline{x}^a|\underline{\mu}_i\right) \, \mathrm{d}\underline{x}^a .$$

Here, $\underline{\mu}_i^y$ results from (19) and $\underline{\mu}_i^{a|b}$ is calculated according to (17) by replacing $\underline{x}^b$ with $\underline{\mu}_i$.

VI. SIMULATION EXAMPLES

The proposed approach is now compared with two sample-based estimators, i.e., the Unscented Kalman Filter (UKF, [4]) and the Gaussian Particle Filter (GPF, [2]). The UKF makes use of a deterministic sampling method and assumes that state and measurement are jointly Gaussian. The GPF is a special sequential importance sampling Particle Filter. Here, after each prediction and filtering, the *randomly* drawn samples are used for determining the mean and the covariance of the state. Thus, a Gaussian representation is provided after each step and no resampling is necessary. As any Particle Filter, the GPF does not require the assumption of a jointly Gaussian state and measurement.

A. System Equation

In the simulations, the estimation of the altitude $\boldsymbol{\alpha}_k$, velocity $\boldsymbol{\beta}_k$, and constant ballistic coefficient $\boldsymbol{\gamma}_k$ of a falling body is considered [9], [15]. The system equation is given by

$$\underline{x}_{k+1} = \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{\gamma}_k \end{bmatrix} + \Delta_t \begin{bmatrix} -\boldsymbol{\beta}_k \\ -\mathrm{e}^{-\rho\cdot\boldsymbol{\alpha}_k} \cdot (\boldsymbol{\beta}_k)^2 \cdot \boldsymbol{\gamma}_k \\ 0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{w}_k^\alpha \\ \boldsymbol{w}_k^\beta \\ \boldsymbol{w}_k^\gamma \end{bmatrix} ,$$

where $\underline{x}_k = [\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k, \boldsymbol{\gamma}_k]^{\mathrm{T}}$ is the state vector, Δ_t the discretization constant, ρ a constant factor, and $\boldsymbol{w}_k^\alpha$, $\boldsymbol{w}_k^\beta$, $\boldsymbol{w}_k^\gamma$ are process noise. The discretization constant Δ_t is set to 1 and the constant factor ρ is $5 \cdot 10^{-5}$. The noises $\boldsymbol{w}_k^\alpha$, $\boldsymbol{w}_k^\beta$, $\boldsymbol{w}_k^\gamma$ are zero-mean Gaussian with joint covariance matrix $\mathbf{Q} = 0.1\cdot\mathbf{I}_3$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix. The initial state of the falling body is $\underline{x}_0 = [3 \cdot 10^5, 2 \cdot 10^4, 10^{-3}]^{\mathrm{T}}$. The initial mean and covariance of the estimators for all simulation runs is set to

$$\underline{\mu}^x = \begin{bmatrix} 3 \cdot 10^5 \\ 2 \cdot 10^4 \\ 10^{-5} \end{bmatrix} \quad , \quad \mathbf{C}^x = \begin{bmatrix} 10^6 & 0 & 0 \\ 0 & 4 \cdot 10^6 & 0 \\ 0 & 0 & 20 \end{bmatrix} .$$

System Decomposition for the SAGF: The state variables can be decomposed into $\underline{x}_k^a = [\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k]^{\mathrm{T}}$ and $\boldsymbol{x}_k^b = \boldsymbol{\alpha}_k$. If the density of the variable $\boldsymbol{x}_k^b$ is represented by a Dirac mixture, the remaining moment integrals in (19) and (20) can be calculated in closed form due to the remaining polynomial system function (see Example 3).

B. Case I: Linear Measurement Equation

In the first case, a linear measurement equation is considered, where the altitude is measured directly according to

$$\boldsymbol{r}_k = \boldsymbol{\alpha}_k + \boldsymbol{v}_k ,$$

where $\boldsymbol{v}_k$ is zero-mean Gaussian measurement noise with variance $R = 10^3$. Due to the linearity of the measurement equation, all three estimators (SAGF, UKF, and GPF) are solving the filter step via the Kalman corrector equations.

TABLE I

AVERAGE RMSE AND ITS STANDARD DEVIATION OVER ALL TEST RUNS FOR CASE I.

	Alt.	Vel.	Coef.
SAGF	12.6 ± 8.3	59.3 ± 143.7	0.016 ± 0.063
UKF	14.2 ± 7.9	100.1 ± 212.4	0.016 ± 0.058
GPF 100 p.	13.0 ± 8.0	60.2 ± 134.9	0.029 ± 0.098
GPF 1000 p.	12.7 ± 8.2	59.3 ± 142.1	0.019 ± 0.066

TABLE II

AVERAGE RMSE AND ITS STANDARD DEVIATION OVER ALL TEST RUNS FOR CASE II.

	Alt.	Vel.	Coef.
SAGF	0.7 ± 2.2	36.5 ± 128.5	0.0126 ± 0.052
UKF	11.7 ± 47.7	102.5 ± 258.9	0.0127 ± 0.051

C. Case II: Nonlinear Measurement Equation

In the second case, a radar measures the altitude, where the measurement equation is given by

$$\boldsymbol{r}_k = \sqrt{M^2 + (\boldsymbol{\alpha}_k - H)^2 + (\boldsymbol{v}_k)^2} . \qquad (21)$$

$M = 10^4$ is the horizontal range and $H = 10^4$ is the altitude of the radar. The measurement noise $\boldsymbol{v}_k$ is Gaussian distributed with the variance $R = 10^3$.

Modified Measurement Equation for SAGF: In order to allow a closed-form solution of the filter step in case of the SAGF, squared ranges are considered. Thus, the measurement equation becomes polynomial according to

$$\boldsymbol{r}_k^2 = M^2 + (\boldsymbol{\alpha}_k - H)^2 + (\boldsymbol{v}_k)^2 , \qquad (22)$$

and can now be used in the AGF formalism. In doing so, the measurement $\hat{y}$ in equation (5) is given by $\hat{y} = \hat{r}_k^2$, where $\hat{r}_k$ is the measured altitude value.

D. Simulation Results

For each case, 1000 simulation runs are performed. In the first case (linear measurement equation), the SAGF is compared with the UKF as well as with the GPF with 100 and 1000 particles. In the second case (nonlinear measurement equation), the SAGF is compared with the UKF. The GPF is omitted, as it provides no reliable estimates in this scenario due to the non-additive noise term in (21).

1) Case I: In Fig. 2, the root mean square error (RMSE) over the 1000 simulation runs is shown. It is obvious that the proposed SAGF converges faster than the UKF. Compared to the GPF with 100 particles, the SAGF has a smaller error in the altitude. Furthermore, the average RMSEs of the SAGF and GPF are smaller than the error of the UKF (see Tab. I). In terms of run time, the SAGF is two times faster than the PF with 100 particles and four times faster than the UKF.

2) Case II: The differences between the two estimators, SAGF and UKF, significantly increase if the altitude is measured according to the nonlinear measurement equation (21). In Fig. 3 the RMSEs are shown. At 10 seconds, the error for the altitude and the velocity is increasing due to the drag of the nonlinear motion [15]. This can be also seen in the average RMSE in Tab. II. It is important to note
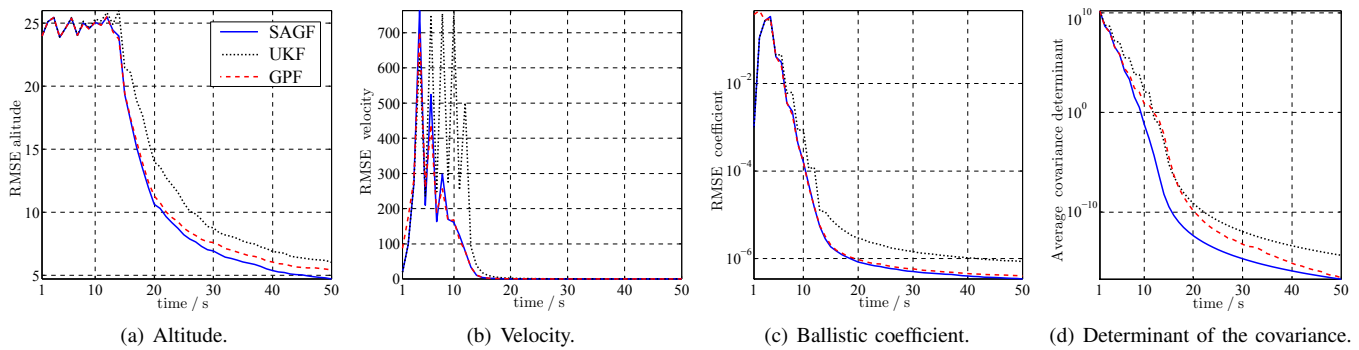
(a) Altitude.   (b) Velocity.   (c) Ballistic coefficient.   (d) Determinant of the covariance.

Fig. 2. The RMSE over the 1000 test runs for case I. For the GPF 100 particles are used.

(a) Altitude   (b) Velocity.   (c) Ballistic coefficient.   (d) Determinant of the covariance.
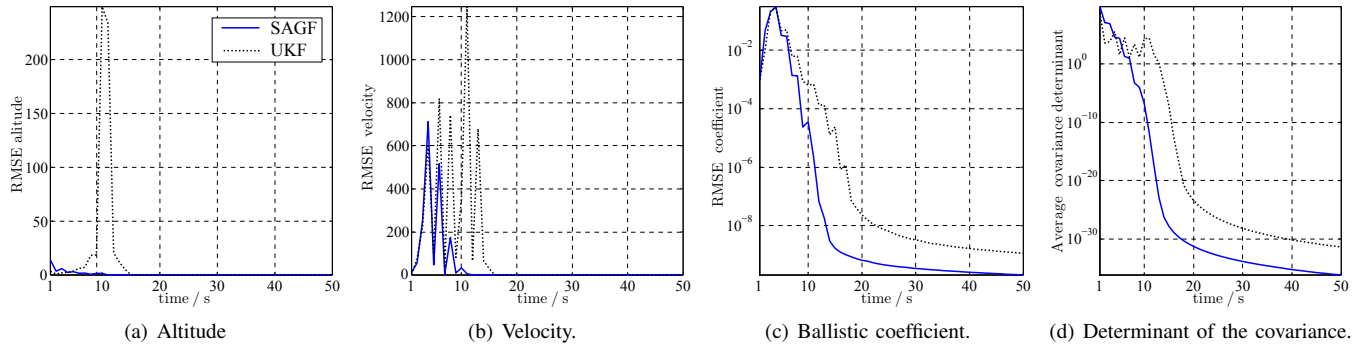
Fig. 3. The RMSE over the 1000 test runs for case II.

that applying the UKF to measurement equation (21) or its modified version (22) has no notable impact on the results.

VII. DISCUSSION AND FUTURE WORK

A new framework for the efficient calculation of posterior moments in the context of Gaussian Assumed Density Filtering based on moment matching has been proposed, which exploits the structure of the given nonlinear system. For this purpose, the system is decomposed into a set of nonlinear subsystems that are conditionally integrable in closed form by means of discretizing *some dimensions of the state space*.

For systems of moderate complexity, a suitable decomposition can typically be found by inspection. For large systems, however, automatic methods for the optimal decomposition according to some predefined quality measure are required and will be pursued in future research. Quality measures might include i) the minimum number of samples or ii) the minimum total number of computations for a given estimation accuracy.

REFERENCES

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

[2] J. Kotecha and P. Djuric, "Gaussian Particle Filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2592–2601, Oct. 2003.

[3] 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.

[4] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[5] T. Lefebvre, H. Bruyninckx, and J. D. Schutter, *Nonlinear Kalman Filtering for Force-Controlled Robot Tasks*. Springer, 2005.

[6] M. Norgaard, N. Poulsen, and O. Ravn, "New Developments in State Estimation for Nonlinear Systems," *Automatica*, vol. 36, no. 11, pp. 1627–1638, Nov. 2000.

[7] 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 2008)*, vol. 17, no. 2, Seoul, Korea, July 2008.

[8] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Science/Engineering/Math; 4. Edition, 2002.

[9] D. Simon, *Optimal State Estimation*. John Wiley & Sons Inc., 2006.

[10] A. Jazwinski, "Filtering for Nonlinear Dynamical Systems," *IEEE Transactions on Automatic Control*, vol. 11, no. 4, pp. 765–766, Oct. 1966.

[11] T. Schon, F. Gustafsson, and P.-J. Nordlund, "Marginalized Particle Filters for Mixed Linear/Nonlinear State-Space Models," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2279–2289, July 2005.

[12] F. Beutler, M. F. Huber, and U. D. Hanebeck, "Gaussian Filtering using State Decomposition Methods," in *Proceedings of the 12th International Conference on Information Fusion (Fusion 2009)*, Seattle, Washington, July 2009.

[13] M. Morelande and B. Moran, "An Unscented Transformation for Conditionally Linear Models," *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007 (ICASSP 2007)*, vol. 3, pp. 1417–1420, April 2007.

[14] K. Ito and K. Xiong, "Gaussian Filters for Nonlinear Filtering Problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[15] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A New Method for the Nonlinear Trannsformation of Means and Covariances in Filters and Estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, 2000.