Estimation Of The Conditional State And Covariance with Taylor Polynomials

Simone Servadio *The University of Texas at Austin*, Austin, Texas, simo_serva@utexas.edu
Renato Zanetti *The University of Texas at Austin*, Austin, Texas, renato@utexas.edu

***Abstract*—A novel estimator is presented that expands the typical state and covariance update laws of Kalman filters to polynomial updates in the measurement. The filter employs Taylor series approximations of the nonlinear dynamic and measurement functions. All polynomials (functions approximation, state update, and covariance update) can be selected up to an arbitrary order to trade between filter's accuracy and computational time. The performance of the algorithm is tested in numerical simulations.**

***Index Terms*—Differential Algebra, Nonlinear filtering, Polynomial update**

I. Introduction

Estimation is the process of inferring the value of a quantity of interest from indirect, inaccurate and noisy observations. When the quantity of interest is the (current) state of a dynamic system, the problem is often referred to as "filtering": the best estimate is obtained by "filtering out" the noise from noisy measurements. The estimate is the output given by an optimal estimator, which is a computational algorithm that processes measurements while maximizing a certain performance index. The optimal estimator makes the best use of the data, the knowledge of the system, and of the disturbances.

For the well-known linear and Gaussian case, the posterior distribution remains Gaussian and the Kalman Filter [1], [2] provides the mechanization to calculate its mean and covariance matrix. However, most practical problems are nonlinear in the dynamics and in the measurement equations, leading to non-Gaussian probability density functions (PDFs).

Many techniques have been developed to deal with the nonlinear estimation problem. A simple solution is based on the linearization of the dynamics and measurement equations around the most current estimate. The Extended Kalman Filter (EKF) [3] algorithm applies the Kalman filter mechanization to the linearized system. Another well-know technique to account for the system nonlinearities is the unscented transformation. The Unscented Kalman filter (UKF) [4], [5] is able to better handle the effects of nonlinearities in the dynamics and in the measurements and, typically, achieves higher accuracy and robustness levels when compared to the EKF. The UKF applies the unscented transformation to achieve a more accurate approximation of the predicted mean and covariance matrix. The UKF is a linear estimator, i.e. the estimate is a linear function of the current measurement.

The first order approximation of the EKF can be extended to higher order Taylor series [3], [6]. Generally, the higher the order of the Taylor series, the better the performance of the filter. The Gaussian Second Order Filter (GSOF) [7] truncates the Taylor series at second order to better account for the system's nonlinearities. Truncating the Taylor series to order c requires knowledge of the estimation error's central moments up to order $2c$ in order to calculate the Kalman gain. E.g., the EKF truncates at first order and it requires knowledge of the covariance matrices. Consequentially, the GSOF requires knowledge of the third and forth central moments of the state distribution. At each iteration, the GSOF approximates the prior PDF as Gaussian so that the third order central moment is zero and the fourth is easily calculated from the covariance matrix. The GSOF performs a linear update based on a second order approximation of the posterior estimation error. Linear Gaussian filters exist up to any arbitrary truncation order of the Taylor series appriximation of the dynamic/measurement functions [8].

Other linear filters make different types of approximations, such as Gaussian quadrature (QKF)[9], spherical cubature (CKF)[10], ensemble points (EnKF)[11], central differences (CDKF)[12], finite differences (DDKF) [13], etc..

All of the filters mentioned above are linear estimators, i.e. the estimate is a linear function of the current measurement. The conditional mean, which is the optimal Minimum Mean Square Error (MMSE) solution, is typically some unknown nonlinear function of the measurement whose exact computation is usually not feasible. A linear estimator, even when accounting for the nonlinearities of the measurement function, is typically outperformed by nonlinear estimators such as the Gaussian Sum Filter (GSF) [14], [15] or Particle Filters (PF) such as Bootstrap PF (BPF) [16], Marginalized PF (MPF) [17], Auxiliary PF (APF), Unscented PF (UPF) [18], Gaussian PF (GPF) [19], Monte Carlo Filter PF (MCFPF) [20].

In Ref. [21] derives the evolution of the conditional mean, covariance, and higher order moments of a dynamic system subject to continuous measurements. To make the solution practical, the nonlinear dynamic and measurement equations are approximated with Taylor series expansions.

Another, less studied, approach to nonlinear filtering is to expand the linear update structure to a polynomial update function of the measurement. De Santis *et al.* [22] propose an augmented state to obtain a polynomial update but preserving the linear update structure. Their work augments the measurement vector with its square to form a quadratic update [22] and was extended to polynomial updates [23]. Li et al. [24] propose to augment the measurement vector with uncorrelated nonlinear conversions. Similarly to [22], [23], Liu et al. [25] obtain a nonlinear estimator preserving the

linear structure of the measurement update. The mean square error (MSE) can be minimized by an optimal selection of the uncorrelated functions [26]. Later, Zhang and Lan merged [26] with the GSF mathematics [27]. Servadio and Zanetti [28] also implemented a quadratic update (extendable to polynomial update of any order) based on Taylor series expansions. The polynomial update requires knowledge of high order central moments, and [28] carries these moments, exactly like the EKF carries mean and covariance. The computational demand of carrying higher order central moments (propagating forward in time and updating with measurement data) grows quickly with the truncation order of the Taylor series, the size of the state vector, and the order of the polynomial update. Ref. [29] performs a polynomial update without carrying the higher order central moments and hence reduces overall computational cost by approximating non-Gaussian distributions as polynomial transformation of Gaussian random variables. In doing so, all high order central moments are easily and efficiently calculated in closed form. Consequently, in [29], polynomial updates can be performed much more efficiently than in [28].

The update methodologies presented in [22], [23], [24], [28], [29] produce a more precise state estimate than those produced by a linear state update. This work introduces a higher order update for the covariance matrix as well as for the state update, which results in a more accurate quantification of the uncertainty associated with the estimate. In turn, the more accurate uncertainty representation produces a more accurate estimator and hence a reduced estimation error.

The paper is structured in the following way. First a short background section highlights the novel contributions of the work. This is followed by the development of the new methodology and by applications to three numerical examples. Lastly, conclusions are drawn.

II. Background

The linear update rule for mean $\hat{\mathbf{x}}^+$ and covariance matrix $\mathbf{P}^+_{\mathbf{xx}}$ are given by

$$\hat{\mathbf{x}}^+ = \hat{\mathbf{x}}^- + \mathbf{K}(\tilde{\mathbf{y}} - \hat{\mathbf{y}}^-) \tag{1}$$

$$\mathbf{P}^+_{\mathbf{xx}} = \mathbf{P}^-_{\mathbf{xx}} - \mathbf{K}\mathbf{P}_{\mathbf{yy}}\mathbf{K}^T \tag{2}$$

where $\mathbf{K}$ is the Kalman gain, $\tilde{\mathbf{y}}$ is the measurement outcome, $\hat{\mathbf{y}}^-$ is the predicted measurement mean, $\hat{\mathbf{x}}^-$ is the prior mean, $\mathbf{P}^-_{\mathbf{xx}}$ is the covariance of the state and $\mathbf{P}_{\mathbf{yy}}$ is the covariance of the measurement. The above equations are optimal in a minimum mean square error (MMSE) only when the prior distribution and the measurement are jointly Gaussian (which implies a linear relation between the two). In general the MMSE estimate is the conditional mean, an unknown and typically nonlinear function of the measurement outcome; Equation (1) is the statistical linear regression of the conditional mean [30], that is to say: Equation (1) is the best linear fit of the conditional mean with respect to a mean-square error performance index:

$$\hat{\mathbf{x}}^+ \approx \mathbb{E}\left\{\mathbf{x}\middle|\mathbf{y} = \tilde{\mathbf{y}}\right\}$$

where the approximation holds to first order. Equation (2) on the other hand, is the total covariance of the estimation error:

$$\mathbf{P}^+_{\mathbf{xx}} = \mathbb{E}\left\{(\mathbf{x} - \hat{\mathbf{x}}^+)\ (\mathbf{x} - \hat{\mathbf{x}}^+)^T\right\}$$

but it is also the best constant approximation of the conditional covariance of the state given the measurement, also in a mean-square error sense.

$$\mathbf{P}^+_{\mathbf{xx}} \approx \mathbb{E}\left\{(\mathbf{x} - \mathbb{E}\{\mathbf{x}\})\ (\mathbf{x} - \mathbb{E}\{\mathbf{x}\})^T\middle|\mathbf{y} = \tilde{\mathbf{y}}\right\}$$

where the approximation holds to zeroth order.

For nonlinear dynamics/measurements, the linear update equations above are not fully recursive, processing nonlinear measurements as a batch is more accurate than processing them individually [28]. For nonlinear systems, Bayes' rule can be applied recursively to obtain an optimal estimator, that is to say: the quantity to be calculated recursively is the conditional PDF given the measurements outcome. Hence, a linear recursive filter can be interpreted as an approximated filter where the distribution of the state given the measurements is approximately Gaussian with mean $\hat{\mathbf{x}}^+$ and covariance matrix $\mathbf{P}^+_{\mathbf{xx}}$.

Experience has shown that the order of the statistical regression approximation of the covariance needs to be lower than that of the mean in order to obtain for good numerical performance of the algorithm. A zeroth order covariance approximation, therefore, has endured as a companion of a linear mean update rule, but it is also used in higher order update methodologies [22], [23], [24], [28], [29]. Our prior work, HOPUF-ℓ-c [29], presents a high order polynomial state update, i.e. a higher-than-linear polynomial approximation of the conditional mean. This paper presents a novel higher order polynomial covariance update to better approximate the conditional covariance than the standard zeroth order approach.

A. The Polynomial Estimator

Gaussian filters are linear filters that approximate the distribution of the state given the measurements as Gaussian with mean $\hat{\mathbf{x}}^+$ and covariance matrix $\mathbf{P}^+_{\mathbf{xx}}$. This is equivalent to approximating the distribution of the state given the measurements as a linear transformation of a standard normal. This linear transformation is given by a shift of $\hat{\mathbf{x}}^+$ and a scale of $\sqrt{\mathbf{P}^+_{\mathbf{xx}}}$.

Our previous work (HOPUF-ℓ-c) expanded this concept by introducing a filter that approximates the distribution of the state given the measurements as a polynomial transformations of standard normal random variables and uses a higher-than-linear polynomial update function. This work introduces a novel covariance update technique and uses the HOPUF-ℓ-c state update which is summarized here.

Let $\mathbf{x}$ be the state of the dynamic system which is desired to be estimated, and let $\mathbf{y}$ be another random vector, sampleable, related to $\mathbf{x}$. Estimators are functions $\mathbf{g}(\mathbf{y})$ that infer the unknown value of $\mathbf{x}$ based on the know outcome of $\mathbf{y}$. Polynomial estimators are a subset of all estimators which, using the Kronecker operator, can be written as

$$\mathbf{g}(\mathbf{y}) = \mathbf{a} + \mathbf{K}_1\mathbf{y} + \mathbf{K}_2\mathbf{y}^{[2]} + \mathbf{K}_3\mathbf{y}^{[3]} + \mathbf{K}_4\mathbf{y}^{[4]} + \dots \tag{3}$$

where $\mathbf{a}$ is a constant, each $\mathbf{K}_i$ is a constant matrix of appropriate dimensions, and each $\mathbf{y}^{[i]}$ is calculated using the Kronecker product

$$\mathbf{y}^{[i]} = \mathbf{y} \otimes \mathbf{y} \otimes \mathbf{y} \otimes \dots \tag{4}$$

In order to avoid redundancy, each repeated component of Eq. (4) generated by the Kronecker product is eliminated, which means that, as an example, only one term between $y_i y_j$ and $y_j y_i$ is kept. It is convenient to derive the estimator's constants by working with deviation vectors. Deviation vectors are defined as

$$d\mathbf{x} = \mathbf{x} - \mathbb{E}\{\mathbf{x}\} \tag{5}$$

$$d\mathbf{y}^{\{i\}} = \mathbf{y}^{[i]} - \mathbb{E}\left\{\mathbf{y}^{[i]}\right\} \tag{6}$$

Deviations have zero mean by construction. The family of polynomial estimators defined by Equation (3) is redefined by adding and subtracting constants, in order to obtain a new, but theoretically equivalent, polynomial estimator family

$$\begin{aligned}\mathbf{g}(\mathbf{y}) &= \mathbf{a} + \mathbb{E}\{\mathbf{x}\} + \mathbf{K}_1(\mathbf{y} - \mathbb{E}\{\mathbf{y}\}) + \\ &\quad + \mathbf{K}_2\left(\mathbf{y}^{[2]} - \mathbb{E}\left\{\mathbf{y}^{[2]}\right\}\right) + \\ &\quad + \mathbf{K}_3\left(\mathbf{y}^{[3]} - \mathbb{E}\left\{\mathbf{y}^{[3]}\right\}\right) + \dots \\ &= \mathbf{a} + \mathbb{E}\{\mathbf{x}\} + \mathbf{K}_1 d\mathbf{y} + \mathbf{K}_2\, d\mathbf{y}^{\{2\}} + \mathbf{K}_3\, d\mathbf{y}^{\{3\}} + \dots \end{aligned} \tag{7}$$

$$= \mathbf{a} + \mathbb{E}\{\mathbf{x}\} + \mathcal{K} d\mathcal{Y} \tag{8}$$

where both the measurement residual with its powers, $d\mathcal{Y}$, and the matrices $\mathbf{K}_i$ are stacked

$$\mathcal{K} = \begin{bmatrix}\mathbf{K}_1 & \mathbf{K}_2 & \mathbf{K}_3 & \dots\end{bmatrix} \tag{9}$$

$$d\mathcal{Y} = \begin{bmatrix}d\mathbf{y}^T & d\mathbf{y}^{\{2\}T} & d\mathbf{y}^{\{2\}T} & \dots\end{bmatrix}^T \tag{10}$$

The optimal estimator, in a Minimum Mean Square Error (MMSE) sense, satisfies the orthogonality principle, from which it follows that the optimal polynomial update estimator becomes

$$\hat{\mathbf{x}} = \mathbb{E}\{\mathbf{x}\} + \mathbf{P}_{\mathbf{x}\mathcal{Y}}\mathbf{P}_{\mathcal{Y}\mathcal{Y}}^{-1} d\mathcal{Y} \tag{11}$$

Matrices $\mathbf{P}_{\mathbf{x}\mathcal{Y}}$ and $\mathbf{P}_{\mathcal{Y}\mathcal{Y}}$ are the augmented state-measurement cross-covariance matrix and the augmented measurement covariance matrix, respectively. These matrices are constructed blockwise by using covariances $\mathbf{P}_{\mathbf{x}\mathbf{y}^{[j]}}$ and $\mathbf{P}_{\mathbf{y}^{[i]}\mathbf{y}^{[j]}}$, for any combination of i and j. As an example, $\mathbf{P}_{\mathbf{y}^{[3]}\mathbf{y}^{[4]}}$ indicates the covariance between the third order measurement vector $\mathbf{y}^{[3]}$ and the forth order $\mathbf{y}^{[4]}$. Since deviations have zero mean by construction, the identities $\mathbf{P}_{\mathbf{y}^{[i]}\mathbf{y}^{[j]}} = \mathbf{P}_{d\mathbf{y}^{\{i\}}d\mathbf{y}^{\{j\}}}$ and $\mathbf{P}_{\mathbf{x}\mathbf{y}^{[j]}} = \mathbf{P}_{d\mathbf{x}d\mathbf{y}^{\{j\}}}$ are valid $\forall i, j \in \mathbb{N}_0$.

B. Differential Algebra

In this work, Gaussian random vector undergo nonlinear (polynomial) transformations. The methodology used here to approximate these transformation is differential algebra (DA) via the Differential Algebra Core Engine (DACE2.0) software program. DA is used as a tool to implement the polynomial filter. Other approximations of nonlinear transformations are also possible but not considered here, [24] for example used the Unscented Transformation.

The theory of DA has been developed by Martin Berz in the late 1980's [31]. The DA framework is an algebra of Taylor polynomials. All functions are represented through a matrix of coefficients and exponents rather than the classical representation with an array of floating points (FP) numbers. The DACE2.0 [32] software has a hard-coded library of the Taylor series expansion of elementary functions. As a consequence, derivatives are not computed numerically (e.g finite differences), but evaluated directly from the Taylor polynomials. DA offers a way of working in a computer environment where the algebra of polynomials is endowed of composition of function, function inversions, explicit system solving, etc., as in the standard FP arithmetic.

Differential algebra has been proven to reduce computational costs in solving ordinary differential equations (ODE) [33]. Once the maximum truncation order of the polynomial is selected, DA creates the Taylor polynomial expansion of the flow of ODEs as a function of the provided initial conditions. This approach can replace thousands of integrations with the computationally faster evaluation of the Taylor expansion [34]. As a result, the computational burden reduces considerably [11]. In the filtering problem, DA techniques have been used for the development of an efficient mapping of uncertainties [35] and for the evaluation of high-order moments [36]. Wittig *et al.* [37] developed a domain splitting technique that improves the state propagation when initial uncertainties are large by creating multiple polynomials.

The main concept of DA is that each function $f(\mathbf{x})$ can be expressed as a polynomials $p(\delta\mathbf{x})$; where the new variable $\delta\mathbf{x}$ is the deviation from the expansion center $\hat{\mathbf{x}}$. The polynomial $p(\delta\mathbf{x})$ is the Taylor series expansion of $f(\mathbf{x})$, centered at $\hat{\mathbf{x}}$, and truncated up to a user-selected order c.

For a detailed description of DA, its techniques, and how the DACE2.0 works in a computer environment, the reader is referred to the references.

III. The State And Covariance Estimation Filter

A new filtering technique, based on a double polynomial estimator, is proposed in the DA framework. The double nature of the filter refers to the sequential estimation of the state and the covariance, where, at each time step, the same measurement outcome is used twice to achieve matching between the conditioned state mean and its relative uncertainty spread.

Consider the generic dynamic system described by the following equations of motion and measurement equations:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) + \mathbf{v}_k \tag{12}$$

$$\mathbf{y}_{k+1} = \mathbf{h}(\mathbf{x}_{k+1}) + \mathbf{w}_{k+1} \tag{13}$$

where $\mathbf{f}(\cdot)$ is the dynamics function, $\mathbf{x}_k$ is the n-dimensional state of the system at time-step k, $\mathbf{y}_{k+1}$ is the m-dimensional measurement vector at time-step $k+1$, and $\mathbf{h}(\cdot)$ is the measurement function. The noises are assumed to be zero mean Gaussians and uncorrelated, such that their distribution

is fully described by the first two moments. For all discrete time indexes i and j

$$\mathbb{E}\left\{\mathbf{v}_i\mathbf{w}_j^T\right\} = 0 \tag{14}$$

$$\mathbb{E}\left\{\mathbf{v}_i\mathbf{v}_j^T\right\} = \mathbf{Q}_i\delta_{ij} \tag{15}$$

$$\mathbb{E}\left\{\mathbf{w}_i\mathbf{w}_j^T\right\} = \mathbf{R}_i\delta_{ij} \tag{16}$$

where $\mathbf{Q}_i$ is the process noise autocovariance function while $\mathbf{R}_i$ is for the measurement noise. The initial condition of the state of the system is assumed to be Gaussian as well $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_0, \mathbf{P}_0)$; however, for all other time steps $k > 0$, the state distribution will be non-Gaussian due the nonlinearities in the dynamics.

The main result of this paper, the State And Covariance Estimation Filter (SACE-c-η-μ) shares the prediction step with our previous work [29] and introduces a new update technique. The single distribution used in SACE-c-η-μ is expanded using Gaussian Multiple Models (GMM) theory [38] to create the Multiple Models State And Covariance Estimation Filter (SACEMM-c-η-μ).

SACE-c-η-μ, is composed of three different parts: the prediction, the state update, and the covariance update. The three integers c, η and μ in SACE-c-η-μ refer to the tuning parameters of the filter: they are, respectively, the order of the Taylor polynomial approximation of $\mathbf{f}(\cdot)$ and $\mathbf{h}(\cdot)$, (c), the order of the state polynomial update, (η), and the order of the covariance polynomial update, (μ).

A. Prediction

At the beginning of each time step, the state distribution is assumed to be Gaussian $\mathbf{x}_k \sim \mathcal{N}(\hat{\mathbf{x}}_\mathbf{k}, \mathbf{P}_k)$. The state can therefore be initialized in the DA framework as a first order polynomial

$$\mathbf{x}_k = \mathbf{x}_k(\delta\mathbf{x}_k) = \hat{\mathbf{x}}_\mathbf{k} + \mathbf{S}_k\delta\mathbf{x}_k \tag{17}$$

where $\mathbf{S}_k\mathbf{S}_k^T = \mathbf{P}_k$ and the DA variable $\delta\mathbf{x}_k = \mathbf{x}_k - \hat{\mathbf{x}}_\mathbf{k}$ expresses the deviation from the expansion center and it is interpreted as a Gaussian with zero mean and identity covariance matrix. Therefore, matrix $\mathbf{S}_k$ (here calculated through Cholesky Decomposition), scales the coefficients of the state polynomial and results in the moments of $\mathbf{x}_k$ easily calculated from the moments of $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

The propagation function is applied directly to the state polynomial, such that the predicted state vector is

$$\mathbf{x}_{k+1}^- = \mathbf{x}_{k+1}^-(\delta\mathbf{x}_k) = \mathbf{f}\big(\mathbf{x}_k(\delta\mathbf{x}_k)\big) \tag{18}$$

where $\mathbf{x}_{k+1}^-$ indicates the Taylor series expansion of the dynamics centered at $\hat{\mathbf{x}}_k$ and truncated at the user-defined integer order c. Equation (18) is carried out in the DA framework. Each component of $\mathbf{x}_{k+1}^-$ is a polynomial map (centered at $\hat{\mathbf{x}}_k$) that maps deviations $(\delta\mathbf{x}_k)$ from time step k to time step $k+1$ and describes how the state PDF evolves in time. The predicted polynomials are lacking the influence of the process noise. Process noise can be mapped in the DA framework with the same representation reserved for the state of the system. Thus, a new DA variable $\delta\mathbf{v}_k$, interpreted again as a standard normal random vector, is introduced

$$\mathbf{x}_{k+1}^-(\delta\mathbf{x}_k, \delta\mathbf{v}_k) \coloneqq \mathbf{x}_{k+1}^-(\delta\mathbf{x}_k) + \mathbf{T}_k\delta\mathbf{v}_k \tag{19}$$

where $\mathbf{v}_k = \mathbf{T}_k\delta\mathbf{v}_k$ and $\mathbf{T}_k\mathbf{T}_k^T = \mathbf{Q}_k$.

Analogously, the predicted measurement is expressed as a Taylor polynomial expansion in the DA framework

$$\mathbf{y}_{k+1} = \mathbf{y}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k) = \mathbf{h}\big(\mathbf{x}_{k+1}^-(\delta\mathbf{x}_k, \delta\mathbf{v}_k)\big) \tag{20}$$

where $\mathbf{y}_{k+1}$ is, again, a polynomial centered at $\hat{\mathbf{x}}_k$ with maximum order c. In Equation (20), the expansion is now w.r.t. both the state deviation vector $(\delta\mathbf{x}_k)$ and the process noise $(\delta\mathbf{v}_k)$. The influence of the measurement noise is added to the polynomials like in Equation (19). A new DA variable $\delta\mathbf{w}_{k+1}$ is introduced

$$\mathbf{y}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1}) \coloneqq \mathbf{y}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k) + \mathbf{U}_{k+1}\delta\mathbf{w}_{k+1} \tag{21}$$

where $\mathbf{w}_k = \mathbf{U}_k\delta\mathbf{w}_k$ and $\mathbf{U}_k\mathbf{U}_k^T = \mathbf{R}_k$. Once again, $\delta\mathbf{w}_{k+1}$ is interpreted as a standard normal random vector.

All the predicted quantities have been calculated and they are represented as polynomial functions of standard random vectors. The number of variables is $2n+m$: n deviations map the state behavior, n map the process noise, and the remaining m map the measurement noise. The Gaussian nature of the random vectors leads to a fast evaluation of all expectation operations since, for a Gaussian PDF, central moments can be easily computed using the Isserlis' formulation [39].

B. The State Polynomial Update

The second part of SACE-c-η-μ is the state polynomial update. After selecting the integer c in the prediction step, the user defines a second integer, η, which selects the order of the polynomial estimator dedicated to the state of the system.

The polynomial update evaluates the augmented Kalman gain and for high powers of the measurement polynomials. Starting from the latter,

$$\mathbf{y}_{k+1}^{[2]} = \mathbf{y}_{k+1} \otimes \mathbf{y}_{k+1} \tag{22}$$

$$\mathbf{y}_{k+1}^{[i]} = \mathbf{y}_{k+1} \otimes \mathbf{y}_{k+1} \otimes \dots \tag{23}$$

with $i = 1, \dots, \eta$ and, once again, the redundant components are eliminated, in order to have independent measurements.

The means of the predicted state polynomials are now evaluated. Each polynomial undergoes the expectation operator which, being a linear operator, works directly on the single monomials of the expansion [8].

$$\hat{\mathbf{x}}^- = \mathbb{E}\left\{\mathbf{x}_{k+1}^-\right\} \tag{24}$$

The deviations have a Gaussian distribution with zero mean and identity covariance matrix, therefore the expected value substitutes the relative Isserlis' moment in for each monomial, according to Table I.

exponent	0	1	2	3	4	5	6	7	8	…
coefficient	1	0	1	0	3	0	15	0	105	…

TABLE I
ISSERLIS' MOMENTS OF GAUSSIAN $\mathcal{N}(0,1)$

For example: $\mathbb{E}\left\{\alpha\delta x_1^8\delta x_2^4\delta x_4^6\delta v_2^2\delta w_3^4\right\} = 4725\alpha$. The pre-

dicted means of the measurement polynomials are similarly evaluated with Equation (24)

$$\hat{\mathbf{y}}_{k+1} = \mathbb{E}\left\{\mathbf{y}_{k+1}\right\} \tag{25}$$

$$\hat{\mathbf{y}}^{[2]}_{k+1} = \mathbb{E}\left\{\mathbf{y}^{[2]}_{k+1}\right\} \tag{26}$$

$$\hat{\mathbf{y}}^{[i]}_{k+1} = \mathbb{E}\left\{\mathbf{y}^{[i]}_{k+1}\right\} \tag{27}$$

where, once again, $i = 1, \ldots, \eta$.

The augmented measurement covariance $\mathbf{P}_{\mathcal{Y}\mathcal{Y},[\eta]}$ is evaluated blockwise. The matrix is guaranteed to be nonsingular because redundant rows and columns have been eliminated. The matrix is symmetric and each block is evaluated as

$$\mathbf{P}_{\mathbf{y}^{[i]}\mathbf{y}^{[j]}} = \mathbb{E}\left\{(\mathbf{y}^{[i]}_{k+1} - \hat{\mathbf{y}}^{[i]}_{k+1})(\mathbf{y}^{[j]}_{k+1} - \hat{\mathbf{y}}^{[j]}_{k+1})^{\mathrm{T}}\right\} \tag{28}$$

$\forall i, j = 1, \ldots, \eta$. Every time a polynomial multiplies itself, the maximum truncation order of the Taylor series doubles. For example, the evaluation of $\mathbf{P}_{\mathbf{y}^{[5]}\mathbf{y}^{[3]}}$ applies the expectation operator to a polynomial with monomials up to order $8c$. The augmented state-measurement cross covariance matrix $\mathbf{P}_{\mathbf{x}\mathcal{Y},[\eta]}$ is evaluated blockwise, each block is evaluated as

$$\mathbf{P}_{\mathbf{x}\mathbf{y}^{[i]}} = \mathbb{E}\left\{(\mathbf{x}^{-}_{k+1} - \hat{\mathbf{x}}^{-}_{k+1})(\mathbf{y}^{[i]}_{k+1} - \hat{\mathbf{y}}^{[i]}_{k+1})^{\mathrm{T}}\right\} \tag{29}$$

$\forall i = 1, \ldots, \eta$. The subscript $[\eta]$ specifies that the covariance matrices are created with measurement powers up to order η. From these covariances it is now possible to evaluate the augmented Kalman gain

$$\mathcal{K} = \mathbf{P}_{\mathbf{x}\mathcal{Y},[\eta]}\mathbf{P}^{-1}_{\mathcal{Y}\mathcal{Y},[\eta]} \tag{30}$$

Denote with $\tilde{\mathbf{y}}_{k+1}$ the numerical outcome of the random vector $\mathbf{y}_{k+1}$, its powers are evaluated using the Kronecker product

$$\tilde{\mathbf{y}}^{[2]}_{k+1} = \tilde{\mathbf{y}}_{k+1} \otimes \tilde{\mathbf{y}}_{k+1} \tag{31}$$

$$\tilde{\mathbf{y}}^{[i]}_{k+1} = \tilde{\mathbf{y}}_{k+1} \otimes \tilde{\mathbf{y}}_{k+1} \otimes \ldots \tag{32}$$

with $i = 1, \ldots, \eta$ and, once again, the redundant components are eliminated. The polynomial update exploits the influence of high powers from the measurement outcome. The measurement residual is developed to create the augmented innovation vector

$$d\tilde{\mathcal{Y}}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1}) = \begin{bmatrix} \tilde{\mathbf{y}}_{k+1} - \mathbf{y}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1}) \\ \tilde{\mathbf{y}}^{[2]}_{k+1} - \mathbf{y}^{[2]}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1}) \\ \ldots \\ \tilde{\mathbf{y}}^{[\eta]}_{k+1} - \mathbf{y}^{[\eta]}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1}) \end{bmatrix} \tag{33}$$

The updated distribution (polynomial) of the state is given by

$$\mathbf{x}^{+}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1}) = \mathbf{x}^{-}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k) + \mathcal{K}d\tilde{\mathcal{Y}}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1}) \tag{34}$$

and the posterior estimate is its mean

$$\hat{\mathbf{x}}^{+}_{k+1} = \mathbb{E}\left\{\mathbf{x}^{+}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1})\right\} \tag{35}$$

evaluated, through Isserlis's moments, monomial by monomial using Table I.

Equation (34) shows that the state polynomials are function of the three different deviations: the state deviation, the process noise, and the measurement noise. Furthermore, the new order of the polynomial as increased by a factor η, dictated by the order of the polynomial update. If the order of the polynomial approximation of the prior distribution ($\mathbf{x}^{-}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k)$) is c, then, the order of the posterior polynomial ($\mathbf{x}^{+}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1})$) is ηc. The higher the polynomial order, the higher the number of moments to be calculated by Table I, which leads to a higher computational burden.

C. The Covariance Polynomial Update

The third, and last, part of SACE-c-η-μ is the covariance polynomial update. After having estimated the state of the system, SACE-c-η-μ applies a second polynomial estimator to identify the value of the state covariance conditioned to the measurements. Therefore, the user defines one last integer parameter, μ, that specifies the order of the covariance polynomial update. Unlike previous tuning parameters, μ cannot be freely chosen but it has to respect the inequality $\mu < \eta$. The covariance cannot have an higher update order than the state.

The covariance matrix is obtained as

$$\mathbf{P}_{\mathbf{x}\mathbf{x},k+1} = \mathbb{E}\left\{(\mathbf{x}^{+}_{k+1} - \hat{\mathbf{x}}^{+}_{k+1})(\mathbf{x}^{+}_{k+1} - \hat{\mathbf{x}}^{+}_{k+1})^{\mathrm{T}}\right\} \tag{36}$$

This value shows the average spread of the posterior distribution among all different possible outcomes, $\tilde{\mathbf{y}}$, of the random variable $\mathbf{y}$. Equation (36) is the equivalent of the classical covariance update formulation, Equation (2), that is used in the most common filters such as EKF, UKF, QKF, CBF, Central Difference Filter, GSOF, etc. Therefore, even if correct, using the average error covariance does not extract all the possible information from the measurement outcome. Similar to the polynomial formulation for estimating the state presented in Equation (7), Equation (36) can be seen as a zeroth order polynomial estimator of the covariance matrix.

A new approach is therefore presented in which the estimate of the covariance is performed to order higher than zero. Define a polynomial vector, $\boldsymbol{\rho}_{k+1}$, as the covariance polynomial

$$\boldsymbol{\rho}^{-}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1}) = (\mathbf{x}^{+}_{k+1} - \hat{\mathbf{x}}^{+}_{k+1}) \otimes (\mathbf{x}^{+}_{k+1} - \hat{\mathbf{x}}^{+}_{k+1}) \tag{37}$$

where, in order to reduce the computational burden, the redundant terms of the symmetric covariance matrix have been eliminated, e.g. the upper diagonal terms are removed. The covariance polynomial maximum order is $2\eta c$, being the square of the posterior distribution. The mean of $\boldsymbol{\rho}_{k+1}$ is exactly the vectorized version of the covariance matrix expressed in Equation (36)

$$\hat{\boldsymbol{\rho}}^{-}_{k+1} = \mathbb{E}\left\{\boldsymbol{\rho}^{-}_{k+1}(\delta\mathbf{x}_k, \delta\mathbf{v}_k, \delta\mathbf{w}_{k+1})\right\} \tag{38}$$

$$= \mathrm{stack}(\mathbf{P}_{\mathbf{x}\mathbf{x},k+1}) \tag{39}$$

where the stack() operator indicates the vectorization of a matrix, performed by stacking columns on top of each other. The covariance update is treated in the same mannar as the state vector: adding to a known prior a polynomial function of the measurement outcome $\tilde{\mathbf{y}}_{k+1}$. This second polynomial update provides an updated covariance value that better represents the state estimate's uncertainty.

The starting point is the already computed augmented measurement covariance matrix $\mathbf{P}_{\mathcal{Y}\mathcal{Y},[\mu]}$. The constrain $\mu < \eta$ makes $\mathbf{P}_{\mathcal{Y}\mathcal{Y},[\mu]}$ a subset of $\mathbf{P}_{\mathcal{Y}\mathcal{Y},[\eta]}$, obtained by selecting the first μ rows and columns. The cross covariance matrix $\mathbf{P}_{\boldsymbol{\rho}\mathcal{Y},[\mu]}$ is evaluated block-wise

$$\mathbf{P}_{\boldsymbol{\rho}\mathcal{Y},[\mu]} = \begin{bmatrix}\mathbf{P}_{\boldsymbol{\rho}\mathbf{y}} & \mathbf{P}_{\boldsymbol{\rho}\mathbf{y}^{[2]}} & \mathbf{P}_{\boldsymbol{\rho}\mathbf{y}^{[3]}} & \dots\end{bmatrix} \tag{40}$$

similarly to $\mathbf{P}_{\mathbf{x}\mathcal{Y},[\eta]}$. Each block is obtained as

$$\mathbf{P}_{\boldsymbol{\rho}\mathbf{y}^{[i]}} = \mathbb{E}\left\{(\boldsymbol{\rho}^-_{k+1} - \hat{\boldsymbol{\rho}}^-_{k+1})(\mathbf{y}^{[i]}_{k+1} - \hat{\mathbf{y}}^{[i]}_{k+1})^{\mathrm{T}}\right\} \tag{41}$$

with $i = 1, \dots, \mu$. The Kalman gain associated to the covariance correction is calculated as

$$\mathcal{G} = \mathbf{P}_{\boldsymbol{\rho}\mathcal{Y},[\mu]}\mathbf{P}^{-1}_{\mathcal{Y}\mathcal{Y},[\mu]} \tag{42}$$

The covariance is updated to its posterior estimate as

$$\hat{\boldsymbol{\rho}}^+_{k+1} = \hat{\boldsymbol{\rho}}^-_{k+1} + \mathcal{G}\begin{bmatrix}\tilde{\mathbf{y}}_{k+1} - \hat{\mathbf{y}}_{k+1} \\ \tilde{\mathbf{y}}^{[2]}_{k+1} - \hat{\mathbf{y}}^{[2]}_{k+1} \\ \dots \\ \tilde{\mathbf{y}}^{[\mu]}_{k+1} - \hat{\mathbf{y}}^{[\mu]}_{k+1}\end{bmatrix} \tag{43}$$

where the influence of the measurement is weighted by the augmented Kalman gain. Before starting the next iteration, vector $\hat{\boldsymbol{\rho}}^+_{k+1}$ is brough back to its matrix formulation

$$\hat{\mathbf{P}}_{\mathbf{xx},k+1} = \text{matrix}(\hat{\boldsymbol{\rho}}^+_{k+1}) \tag{44}$$

where the matrix() operator is the inverse of the stack() operator.

The updated posterior distribution can be approximated as Gaussian with mean $\hat{\mathbf{x}}^+_{k+1}$ and covariance matrix $\hat{\mathbf{P}}_{\mathbf{xx},k+1}$ to start the next iteration from Equation (17), where the DA variables related to the noises are discarded and a new state deviation vector is initialized.

SACE-c-η-μ contain three tuning parameters to enhance the performances of classic estimators. In fact, SACE-1-1-0 reduces to the extended Kalman filter and SACE-2-1-0 is the Gaussian Second Order Filter. The polynomial estimator better weights the information from the measurement by computing high order central moments. The increase in accuracy is paid by an increase in computational effort, which practically limits the filter's order selection. The highest polynomial order the filter has to compute (in the evaluation of $\mathbf{P}_{\boldsymbol{\rho}\mathbf{y}^{[\mu]}}$) is $(2\eta+\mu)c$.

The computational time required by the filter depends on the selection of its three tuning parameters and on the dimension of the state vector. SACE-c-η-μ is not suitable for extremely large systems because of the exponential grow in the number of monomials in the Taylor expansion [40]. An in-depth analysis of the computational time of filters developed in the DACE2.0 framework is presented in [41]. The reference portraits an exhaustive analysis of the execution time on the BeagleBone Black (BBB) Single Board Computer, with particular focus on the duty cycles of filter execution on BBB and its dependency on the Taylor truncation order.

IV. The Multiple Models Space And Covariance Estimation Filter

SACE-c-η-μ approximates the time propagation of the state with one single polynomial representation of the flow. However, as the Taylor polynomial series gets further away from the expansion center, it becomes less accurate. Therefore, when the initial uncertainties of the state distribution are extremely large, a single polynomial map may not be sufficient to truthfully describe the predicted PDF [37]. Splitting the initial uncertainties in multiple (smaller) sub-domains aids the filter in reaching convergence. Thus, a second filter called Multiple Models Space And Covariance Estimation Filter, SACEMM-c-η-μ, merges SACE-c-η-μ with the GMM formulation. In the DA framework, multiple models translate to multiple polynomials.

A. Initialization

The initial state distribution is assumed to be Gaussian $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_\mathbf{0}, \mathbf{P}_0)$. The initialization of the models follows an analogy with the unscented transformation [5]. Therefore, the initial domain is divided into $\theta = 2n + 1$ models, where n is the number of states. Each ith model is a Gaussian with mean $\hat{\mathbf{x}}_{0,\{i\}}$ and covariance $\mathbf{P}_{0,\{i\}}$. Being symmetric, the state covariance matrix can be elaborated into its eigenvalue decomposition

$$\mathbf{P}_0 = \mathbf{V}\mathbf{D}\mathbf{V}^T \tag{45}$$

where $\mathbf{V}$ is the matrix of eigenvectors that describes the orientation of the uncertainty ellipsoids, and the diagonal matrix of eigenvalues $\mathbf{D}$ describes the magnitude of the uncertainties. The mean of each Gaussian kernel is selected as

$$\hat{\mathbf{x}}_{0,\{0\}} = \hat{\mathbf{x}}_0 \tag{46}$$

$$\hat{\mathbf{x}}_{0,\{j\}} = \hat{\mathbf{x}}_0 + \mathbf{V}\mathbf{D}_j \qquad j = 1, \dots, n \tag{47}$$

$$\hat{\mathbf{x}}_{0,\{j\}} = \hat{\mathbf{x}}_0 + \mathbf{V}\mathbf{D}_{j-n} \qquad j = n+1, \dots, 2n \tag{48}$$

where $\mathbf{D}_j$ indicates the jth column of the matrix. The centers of the models lie on the principal axes and their initial weights are proportional to their probability w.r.t. the initial distribution

$$\omega_{0,\{i\}} = \frac{(2\pi)^{-n/2}}{\mathcal{W}_0\sqrt{\det \mathbf{P}_0}} \exp\left(-\frac{1}{2}\left(\hat{\mathbf{x}}_{0,\{i\}} - \hat{\mathbf{x}}_0\right)^T \mathbf{P}_0^{-1}\left(\hat{\mathbf{x}}_{0,\{i\}} - \hat{\mathbf{x}}_0\right)\right) \tag{49}$$

$$\mathcal{W}_0 = \sum_{i=0}^{\theta-1} \omega_{0,\{i\}} \tag{50}$$

where $\mathcal{W}_0$ normalizes the weights such that their sum is unity. The models are assumed to share the same covariance, they all have the same initial level of uncertainty

$$\mathbf{P}_{0,\{j\}} = \mathbf{P}_0 + \mathbf{x}_0\mathbf{x}_0^T - \sum_{i=0}^{\theta-1} \omega_{0,\{i\}}\mathbf{x}_{0,\{i\}}\mathbf{x}_{0,\{i\}}^T \tag{51}$$

with $j = 0, \dots, \theta - 1$. Therefore, at the beginning of the first iteration, the initial Gaussian distribution has been divided in θ smaller Gaussian kernels $\mathbf{x}_{0,\{i\}} \sim \mathcal{N}(\hat{\mathbf{x}}_{0,\{i\}}, \mathbf{P}_{0,\{i\}})$ with the same covariance matrix and means on the principal axes, selected as sigma points from the unscented transformation.

B. Prediction

The models have been initialized as Gaussian Kernels. SACEMM-c-η-μ applies SACE-c-η-μ on each kernel like it were operating by its own. As a consequence, θ different polynomials are created in the DA frameworks and θ polynomial maps of the flow describe the time propagation of the state.

$$\mathbf{P}_{k,\{i\}} = \mathbf{S}_{k,\{i\}}\mathbf{S}_{k,\{i\}}^T \tag{52}$$

$$\delta\mathbf{x}_{k,\{i\}} = \mathbf{x}_{k,\{i\}} - \hat{\mathbf{x}}_{k,\{i\}} \tag{53}$$

$$\mathbf{x}_{k,\{i\}} = \mathbf{x}_{k,\{i\}}(\delta\mathbf{x}_{k,\{i\}}) = \hat{\mathbf{x}}_{k,\{i\}} + \mathbf{S}_{k,\{i\}}\delta\mathbf{x}_{k,\{i\}} \tag{54}$$

$$\mathbf{x}^-_{k+1,\{i\}} = \mathbf{x}^-_{k+1,\{i\}}(\delta\mathbf{x}_{k,\{i\}}) = \mathbf{f}\big(\mathbf{x}_{k,\{i\}}(\delta\mathbf{x}_{k,\{i\}})\big) \tag{55}$$

with $i = 0, \dots, \theta - 1$. Multiple Taylor series expansions improve the approximation accuracy of the polynomial maps since, at the boundaries, deviations are closer to their relative centers. Following SACE-c-η-μ for each model, the process noise is mapped on each polynomial expansion

$$\mathbf{v}_{k,\{i\}} = \mathbf{T}_k\delta\mathbf{v}_{k,\{i\}} \tag{56}$$

$$\mathbf{x}^-_{k+1,\{i\}}(\delta\mathbf{x}_{k,\{i\}}, \delta\mathbf{v}_{k,\{i\}}) \coloneqq \mathbf{x}^-_{k+1,\{i\}}(\delta\mathbf{x}_{k,\{i\}}) + \mathbf{T}_k\delta\mathbf{v}_{k,\{i\}} \tag{57}$$

and a measurement polynomial is evaluated for each kernel

$$\mathbf{w}_{k+1,\{i\}} = \mathbf{U}_{k+1}\delta\mathbf{w}_{k+1,\{i\}} \tag{58}$$

$$\begin{aligned}\mathbf{y}_{k+1,\{i\}} &= \mathbf{y}_{k+1,\{i\}}(\delta\mathbf{x}_{k,\{i\}}, \delta\mathbf{v}_{k,\{i\}}) \\ &= \mathbf{h}\big(\mathbf{x}^-_{k+1,\{i\}}(\delta\mathbf{x}_{k,\{i\}}, \delta\mathbf{v}_{k,\{i\}})\big)\end{aligned} \tag{59}$$

Measurement noise is added in the DA framework

$$\begin{aligned}&\mathbf{y}_{k+1,\{i\}}(\delta\mathbf{x}_{k,\{i\}}, \delta\mathbf{v}_{k,\{i\}}, \delta\mathbf{w}_{k+1,\{i\}}) \coloneqq \\ &\qquad \mathbf{y}_{k+1,\{i\}}(\delta\mathbf{x}_{k,\{i\}}, \delta\mathbf{v}_{k,\{i\}}) + \mathbf{U}_{k+1}\delta\mathbf{w}_{k+1,\{i\}}\end{aligned} \tag{60}$$

such that the prediction step is completed for each Gaussian kernel.

C. The State and Covariance Polynomial Update

The prediction step has been exploited by the introduction of multiple polynomials. In the update step, each kernel undergoes the polynomial update for the state and for the covariance described by SACE-c-η-μ. Therefore, after having selected η and μ as the orders for the polynomial estimators, the state posterior estimate and the conditional covariance of each model are evaluated as

$$\mathbf{x}^+_{k+1,\{i\}}(\delta\mathbf{x}_{k,\{i\}}, \delta\mathbf{v}_{k,\{i\}}, \delta\mathbf{w}_{k+1,\{i\}}) = \mathbf{x}^-_{k+1,\{i\}} + \mathcal{K}_{\{i\}}d\tilde{\mathcal{Y}}_{\{i\}} \tag{61}$$

$$\hat{\mathbf{x}}^+_{k+1,\{i\}} = \mathbb{E}\left\{\mathbf{x}^+_{k+1,\{i\}}(\delta\mathbf{x}_{k,\{i\}}, \delta\mathbf{v}_{k,\{i\}}, \delta\mathbf{w}_{k+1,\{i\}}\right\} \tag{62}$$

and

$$\hat{\boldsymbol{\rho}}^+_{k+1,\{i\}} = \hat{\boldsymbol{\rho}}^-_{k+1,\{i\}} + \mathcal{G}_{\{i\}}\begin{bmatrix}\tilde{\mathbf{y}}_{k+1,\{i\}} - \hat{\mathbf{y}}_{k+1,\{i\}} \\ \tilde{\mathbf{y}}^{[2]}_{k+1,\{i\}} - \hat{\mathbf{y}}^{[2]}_{k+1,\{i\}} \\ \cdots \\ \tilde{\mathbf{y}}^{[\mu]}_{k+1,\{i\}} - \hat{\mathbf{y}}^{[\mu]}_{k+1,\{i\}}\end{bmatrix} \tag{63}$$

$$\hat{\mathbf{P}}_{\mathbf{xx},k+1,\{i\}} = \text{matrix}(\hat{\boldsymbol{\rho}}^+_{k+1,\{i\}}) \tag{64}$$

with $i = 0, \dots, \theta - 1$. Every Kalman gain and expectation has been calculated according to the polynomial estimator theory and using Table I, since each deviation is interpreted as a standard normal random vector.

The influence of each ith Gaussian to the posterior PDF needs to be updated as well. The posterior distribution of the probability of each Gaussian given the measurements can be evaluated using Bayes' rule. Therefore, the updated weight of each model is proportional to its measurement likelihood. Let us define with $P(\mathbf{y}_{k+1}|i, \mathbf{Y}_k)$ the probability of $\tilde{\mathbf{y}}_{k+1}$ to be the outcome from the ith Gaussian:

$$\begin{aligned}P(\tilde{\mathbf{y}}_{k+1}|i, \mathbf{Y}_k) &= \frac{(2\pi)^{-m/2}}{\sqrt{\det \mathbf{P}_{\mathbf{yy},\{i\}}}} \\ \exp&\left(-\frac{1}{2}\left(\tilde{\mathbf{y}}_{k+1} - \hat{\mathbf{y}}_{k+1,\{i\}}\right)\mathbf{P}^{-1}_{\mathbf{yy},\{i\}}\left(\tilde{\mathbf{y}}_{k+1} - \hat{\mathbf{y}}_{k+1,\{i\}}\right)\right)\end{aligned} \tag{65}$$

where $\mathbf{Y}_k$ indicates all the measurements realizations up to time step k. The weight update formulation is derived, for the ith kernel, as

$$\begin{aligned}\omega_{k+1,\{i\}} &= P(i|\mathbf{Y}_{k+1}) = \\ &= P(i|\tilde{\mathbf{y}}_{k+1}, \mathbf{Y}_k) \\ &= \frac{P(i, \tilde{\mathbf{y}}_{k+1}|\mathbf{Y}_k)}{P(\tilde{\mathbf{y}}_{k+1}|\mathbf{Y}_k)} \\ &= \frac{P(i, \tilde{\mathbf{y}}_{k+1}|\mathbf{Y}_k)}{\sum_{j=0}^{\theta-1} P(j, \tilde{\mathbf{y}}_{k+1}|\mathbf{Y}_k)} \\ &= \frac{P(\tilde{\mathbf{y}}_{k+1}|i, \mathbf{Y}_k)P(i|\mathbf{Y}_k)}{\sum_{j=0}^{\theta-1} P(j, \tilde{\mathbf{y}}_{k+1}|\mathbf{Y}_k)} \\ &= \frac{P(\tilde{\mathbf{y}}_{k+1}|i, \mathbf{Y}_k)}{\sum_{j=0}^{\theta-1} \omega_{k,\{j\}}P(\tilde{\mathbf{y}}_{k+1}|j, \mathbf{Y}_k)}\omega_{k,\{i\}}\end{aligned} \tag{66}$$

where the denominator normalizes the weights such that they sum to unity. Equation (66) is recursive and modifies the importance of each model based on how likelihood it could have generated the measurement outcome.

The filtering algorithm has ended and it can start the following iteration from $\hat{\mathbf{x}}^+_{k+1,\{i\}}$, $\hat{\mathbf{P}}_{\mathbf{xx},k+1,\{i\}}$, and $\omega_{k+1,\{i\}}$ for each model. However, the weighted state estimate, $\bar{\mathbf{x}}$, and covariance, $\bar{\mathbf{P}}$, are calculated for downstream users and they are used to assess the performance of the filtering technique.

$$\bar{\mathbf{x}} = \sum_{i=0}^{\theta} \omega_{k+1,\{i\}}\hat{\mathbf{x}}^+_{k+1,\{i\}} \tag{67}$$

$$\bar{\mathbf{P}} = -\bar{\mathbf{x}}\bar{\mathbf{x}}^T + \sum_{i=0}^{\theta} \omega_{k+1,\{i\}}\left(\hat{\mathbf{P}}_{\mathbf{xx},k+1,\{i\}} + \hat{\mathbf{x}}^+_{k+1,\{i\}}\hat{\mathbf{x}}^{+T}_{k+1,\{i\}}\right) \tag{68}$$

Once again, for basic parameters, SACEMM-c-η-μ reduces to well-know filters: in fact, picking SACEMM-1-1-0 reduces to the GSF. The computational complexity of SACEMM-c-η-μ is approximately θ times bigger when compared to SACE-c-η-μ. Therefore, it is advised to operate the multiple models technique when the initial state uncertainties are particularly large, or when the time step is long enough that one polynomial approximation is not sufficient to adequately represents the flow of the dynamics. Therefore, for problems with high initial uncertainty, SACEMM-c-η-μ can be used for the first

few iteration steps and then replaced with SACE-c-η-μ once the sate error covariance has decreased.

V. Numerical Examples

The proposed filtering techniques have been applied to three different scenarios. First, a scalar application gives a visual representation of how the new update algorithm works and highlights the innovative features as compared to other estimators. The second problem consists in a tracking application where the system undergoes the highly nonlinear dynamics of a Lorenz96 system. The third application uses Lorenz63 dynamics to underline the benefits of the multiple model filtering technique.

A. Scalar Problem

A simple scalar problem is presented here to highlight the improvements of the new filtering technique by estimating the conditional covariance. It has already been proven that high order polynomial estimators are a better approximation of the true MMSE [29]. However, the presented example underlines the matching between state and covariance for each different realization of the measurement.

Define a normal prior state distribution $x \sim \mathcal{N}(1, 0.02)$ and a measurement

$$y = 1/x + \nu \tag{69}$$

where $\nu \sim \mathcal{N}(0, 0.003)$ is independent from x and represents the measurement noise.

Figure 1 shows the true joint distribution of x and y represented using 10^5 points (gray dots in the figure). The figure compares SACE-c-η-μ and SACEMM-c-η-μ with a few common estimators: the EKF, the UKF, the GSF, the Iterated Extended Kalamn Filter (IKF) [42], the Particle Filter (PF), and the high-order extended Kalman filter (DAHO-k) [35]. The first row of graphs (EKF, UKF, DAHO-3) contains linear estimators, therefore their representation on the (x, y) plane is a straight line, shown in red. The slope of the red line is the Kalman gain, whose optimal value is $\mathbf{P_{xy}}\mathbf{P_{yy}^{-1}}$. The different slope shown by the different linear estimators is due to the different approximation each linear filter employs to evaluate the moments. The EKF applies basic linearization (Jacobians), the UKF uses the unscented transformation, and DAHO-3 uses Taylor polynomials up to the third order. The green lines depict the filter's own assessment of the estimation error uncertainty as a $\pm 3\sigma$ boundary. The different evaluation of the moments leads to a different value on the estimation of the variance, as it follows Equation (2). The green lines share the same slope of the corresponding red line: they are just translated left (and right) by 3σ. These linear filters estimate the same uncertainty level regardless the measurement outcome and the predicted covariance value is the mean among all the possible different realizations.

The second row of graphs in Figure 1 shows nonlinear estimators. The GSF has been implemented with 3 models, which allows the estimator function, red line, to follow the curved shape of the posterior distribution. However, when the likelihood of one model becomes predominant with respect to the others, the GSF behaves similarly to the EKF: this aspect is mostly evident near the tails of the distribution. The estimated covariance of the GSF is a function of the measurement because it is evaluated as a weighted mean among all the models, whose importance weight is based on their likelihood. However, the $\pm 3\sigma$ green lines show the same problems connected with linear estimators: the lines are able to change slope when the models have approximately the same weight, otherwise they are straight. Furthermore, since the GSF can be intended as multiple EKFs with reduced subdomains, the filter shows the same behavior of the linear estimator at the edges of the posterior PDF. The IKF performs multiple updates to repeatedly calculate the measurement Jacobian each time linearizing with respect to the most current estimate. The IKF minimizes a nonlinear Least Square performance index that, for appropriate probability distributions functions, approximates the Maximum A Posteriori (MAP) estimate. As such, the IKF is a nonlinear estimator, and its red line follows the bend of the posterior distribution, setting on the most likely value of x for each measurement outcome y. The $\pm 3\sigma$ green lines correctly bound the distribution; however, the IKF is not necessary an unbiased filter and choosing the peak of the posterior distribution does not necessarily minimize the mean-square error. Hence the IKF's mean square error is often larger than filters based on the MMSE principle [43]. The third nonlinear estimator presented in Figure 1 is the particle filter. Particle filters are accurate nonlinear estimators that use an ensemble of weighted particles to calculate the state estimate. The weight of each particle depends on its measurement likelihood. Both the state estimate and the predicted error covariance are (nonlinear) functions of the measurements. The graph shows that the PF estimates do not form well defined lines, but the state and covariance estimate values depend on the randomness of the data. In other words, while in the EKF the state estimates from two separate updates with the same measurement outcome give exactly the same value, two PF estimates depend on the randomness of the initial ensemble used to generate them. Consequently, the green and red "lines" of the PF become thicker as moving towards the tails of the posterior distribution.

In the third row SACE-3-5-2 and SACEMM-3-5-2 are reported. The 5th order polynomial estimator is able to follow the curved shape of the joint distribution and it accurately approximates the true MMSE. The optimal MMSE is the conditional mean, which visually is the line that divides in half the distribution of y, as horizontal spread of points, for each value of x. Therefore, while EKF, UKF and DAHO-3 can be interpreted as different linear approximation of the true MMSE, SACE-3-5-2 represents a 5th order approximation, which shows a more accurate result. By increasing the estimator order η to infinity, SACE-c-η-μ would asymptotically reach the true MMSE. The green lines related to SACE-3-5-2 show how the uncertainty level has become a (nonlinear) function of the measurement. The $\pm 3\sigma$ boundary increases and tightens depending on the horizontal spread of samples around the estimator function. For example, when the current measurement is $y = 1$, SACE-3-5-2 gives its estimate with a level of uncertainty that matches the spread of the gray points

on the line $y = 1$. When the sensor gives $y = 2$, SACE-3-5-2 outputs a level of confidence in its estimate higher than in the previous case, since the spread of the gray samples around its estimate at $y = 2$ is tighter. Therefore, the estimated covariance of the filter is a function of the measurement and the performance improves drastically because the uncertainties level always matches the estimate, providing a more reliable outcome. There appears to be no influential benefits in applying the multiple model polynomial estimator: SACEMM-3-5-2 behaves similarly to its single model counterpart and shares the same features. However, at the tails of the distribution, SACEMM-3-5-2 estimated conditional covariance better follows the distribution of the samples.

The accuracy level reached by each filter is compared in Figure 2, where the results of a RMSE analysis is reported.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{samples}} (x_i - \hat{x}_i^+)^2}{N_{samples}}} \tag{70}$$

The RMSE of each estimator is evaluated using the entire set of 10^5 points. The bars show that SACE-3-5-2 is the most accurate filter while the linear estimators are the least. However, a more precise approximation of the measurement equation leads to a smaller RMSE and to a more precise estimate, as proven by DAHO-3 (3rd order Taylor polynomial) being the most accurate among the other linear estimators. The IKF shares the same accuracy level as DAHO-3, while the other nonlinear estimators have lower RMSE. Two PF implementations are shown with different number of particles; 10^3 and 10^4. It has error levels comparable with SACE-3-5-2, and the PF with 10^4 particles has a heavier computational burden. Figure 2 reports, in orange, the average GPU time of each estimator, evaluated among all the 10^5 runs shown in Figure 1. As expected, PF-1e4 has the highest computational time, while linear estimators have the lowest. SACE-3-5-2 achieves the best accuracy levels comparable to sample-based filters in a shorter amount of time, although the performance of all nonlinear filter is very similar in this simple motivating example.

The proposed scalar problem shows no significant difference between SACE-c-η-μ and SACEMM-c-η-μ. Let us increase the prior uncertainty level to $x \sim \mathcal{N}(1, 0.03)$ in order to underline the benefits of having multiple models. Figure 3 shows the estimator function and confidence level of the two filters, among with the joint distribution. SACE-3-5-2 outputs an unphysical results for the predicted conditional covariance of the state: a negative value of σ^2. The initial prior uncertainties are excessively large to allow the filter to work properly. On the left tail of the join distribution, the variance becomes negative and that is represented by the green lines overlapping the red one, tp show that the filtering algorithm is not functioning correctly. SACEMM-3-5-2, on the other hand, has no issues in estimating correctly both the state and the covariance for all possible outcomes of the measurement. The green lines bound the samples of the joint distribution narrowing and widening as needed. The correct result is connected to the reduced initial covariance associated to each model, which increases the filter robustness and performance.

The proposed problem underlines a couple of characteristics of the proposed algorithms. Unlike the linear and Gaussian case, the conditional covariance and the estimation error covariance are different. Linear filters employ the estimation error covariance which expresses the average spread of the estimation error over all possible measurement realizations. This is a good metric, but once a measurement is actually available to process, the covariance conditioned on the actual measurement outcome is a more informative quantity, because it provides the spread of the estimation error for the actual value of y. In fact, the conditional covariance is a (nonlinear) function of the measurement whose evaluation is usually non feasible. SACE-c-η-μ and SACEMM-c-η-μ use a polynomial estimator to approximate the function, achieving better results with respect to filters that do not.

B. Lorenz96 System

The performance of the proposed filter is tested on a Lorenz96 example [38] where the state dynamics are

$$\frac{dx_i(t)}{dt} = x_{i-1}(t)(x_{i+1}(t) - x_{i-2}(t)) - x_i(t) + F + \nu_i(t) \tag{71}$$

with $i = 1, \ldots, 4$, since $\mathbf{x}(t)$ is selected to be four-dimensional. The following conventions are used: $x_{-1}(t) = x_{n-1}(t)$, $x_0(t) = x_n(t)$, and $x_1(t) = x_{n+1}(t)$. The term F is a constant external force with value chosen equal to eight, since it introduces a chaotic behavior in the system. The initial condition is assumed to be Gaussian, with mean $\hat{\mathbf{x}} = \begin{bmatrix} F & F & F + 0.01 & F \end{bmatrix}^T$ and diagonal covariance matrix, with the same standard deviation for each component of the state: $\sigma_{\mathbf{x}} = 10^{-3}$. The process noise is assumed to be Gaussian and uncorrelated among states, with known standard deviation $\sigma_{\boldsymbol{\nu}} = 10^{-3}$. The dynamics are propagated at 2 Hz for a total of 20 seconds. The measurement are obtained each time step according to the following model

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}(t_k) + \boldsymbol{\mu}_k, \quad \mathbf{H}_{i,j} = \begin{cases} 1 & j = 2i - 1 \\ 0 & \text{otherwise} \end{cases} \tag{72}$$

with $i = \{1, 2\}$ and $j = \{1, 2, 3, 4\}$. In other words, the sensors observe the components of the state with odd indices. Measurement noises are assumed to be Gaussian and uncorrelated within each other and with the process noise. The standard deviation is selected as $\sigma_{\boldsymbol{\mu}} = 0.5$: this value is particularly high and filters based on linear estimators are not able to track the state of the system and achieve convergence [29].

Figure 4 shows the Monte Carlo analysis results performed with SACE-2-3-2 on the presented application. The figure shows, for each ith component of the state, the estimation error of each realization (gray lines), calculated as

$$\epsilon_{j,i} = x_{j,i} - \hat{x}_{j,i} \tag{73}$$

for each jth time step. A total of 100 realizations are reported. Figure 4 describes the error means, in black, and the error standard deviations, as 3σ values, in blue. The black lines show that SACE-c-η-μ is an unbiased filter, as expected from the theory of MMSE estimation. The predicted error standard

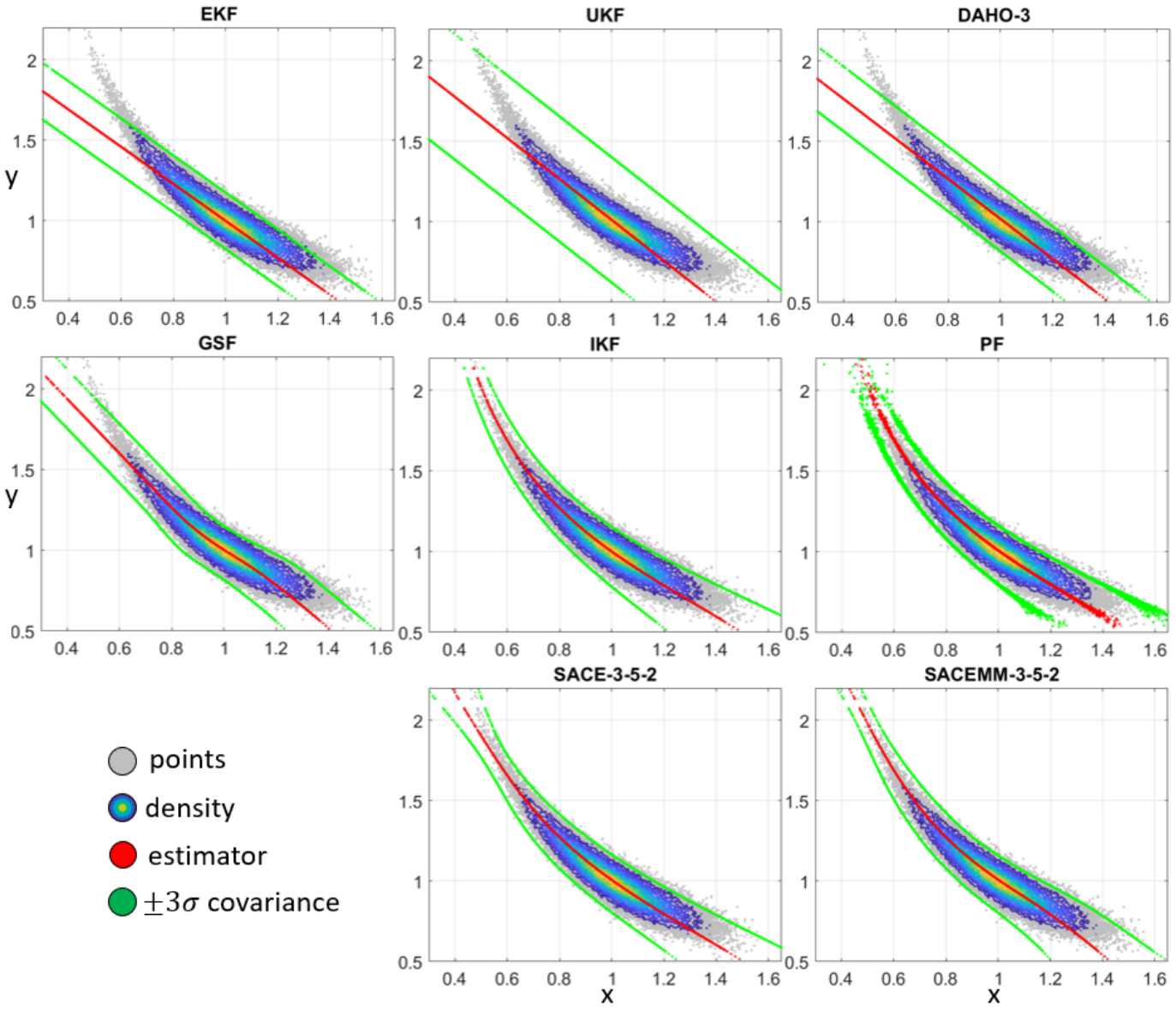

Fig. 1. Comparison among different estimators. Posterior distribution (gray), the estimator functions (red), and their confidence levels (green).

deviation, continuous blue line, is evaluated directly from the updated covariance matrix, by taking the square root of the diagonal terms. The effective performance of the filter is assessed by the sample standard deviation of the Monte Carlo estimation errors, dashed blue lines. At each time step, the actual error covariance of the filter is evaluated by working directly on the samples. The consistency of SACE-2-3-2 is established by the overlapping of the dashed and continuous blue lines, which proves that the filter can correctly predict its own uncertainty levels.

The performance comparison among different filters is shown in Figure 5 through another Monte Carlo analysis conducted with 100 runs. The figure shows, for each filter, the comparison between the effective and predicted error covariance. The continuous lines represent the filter own estimate of the error standard deviation, calculated directly form the updated covariance matrix as the square root of its trace:

$$\bar{\sigma} = \sqrt{\mathrm{tr}(\hat{\mathbf{P}}_{\mathbf{xx}})} \tag{74}$$

The dashed lines represent the effective error standard deviation derived from the Monte Carlo analysis. A consistent filter has the matching between its dashed and continuous line, meaning that the estimated uncertainty level reflects the actual error standard deviation. The top graph in Figure 5 shows how linear estimators, the EKF, UKF, and DAHO-2, diverge (and break down) while trying to track the state of the system. The measurement noise level is excessively large and a linear dependence on the measurement outcome is not sufficient to achieve a correct estimate. The EKF lines also represent the behavior of the IKF: since measurement is linear measurement the IKF reduces to the EKF. The UKF and DAHO-2 use, respectively, the unscented transformation and second order Taylor polynomial to improve the prediction step of the filter

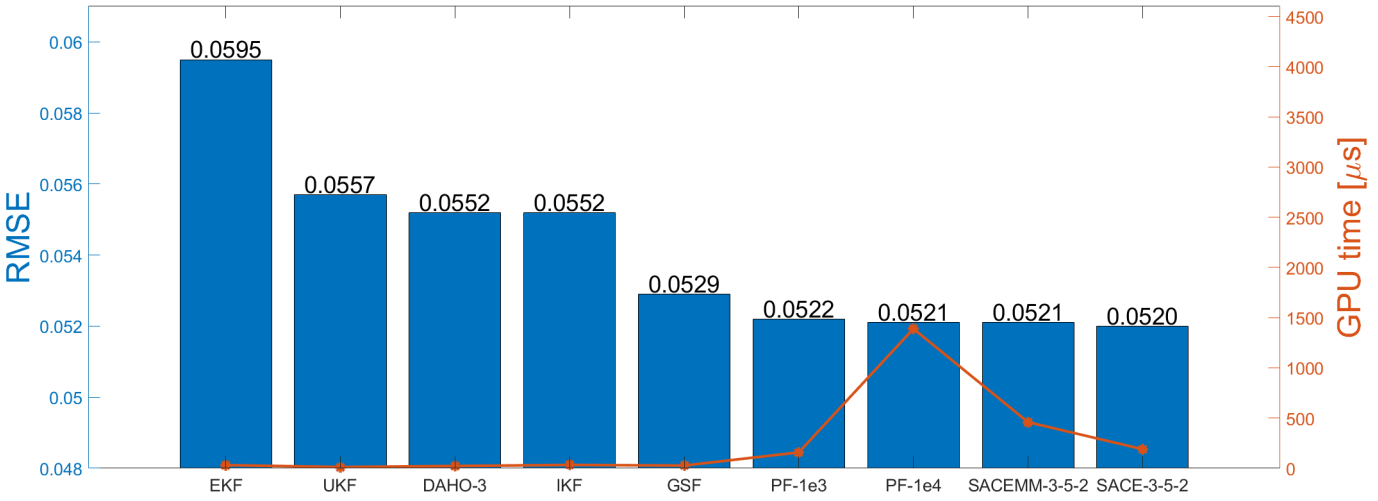

Fig. 2. Comparison of RMSE and computational time among different estimators.

and have a more accurate propagated state prior distribution. However, the update step is still linear and highly influenced by the noise standard deviation that prevents the evaluation of a reliable Kalman gain. The polynomial estimator better weights the information from the measurements using high order moments and it achieves convergence and consistency. Therefore, SACE-2-3-0, in blue, and SACE-2-3-2, in red, correctly estimate the state of the system along the whole simulation. The bottom graph in Figure 5 zooms in for the performance of SACE-c-η-μ for the two different sets of parameters. SACE-2-3-0 shows a filter whose estimate is a polynomial function of the measurement and its estimated covariance is evaluated as a mean among all possible resolutions; it is not influenced by the measurement outcome. SACE-2-3-2, on the other hand, improves accuracy by estimating the covariance giving it the same importance reserved to the state. Thus, the red lines settle below the blue ones for the whole simulation, since the predicted error standard deviation better matches the conditional mean.

C. Lorenz63 System

The performance of the proposed algorithms is also tested on a Lorenz63 application [44], [38], a challenging nonlinear system without process noise. The absence of process noise causes impoverishments in particle filters typically resulting in unsatisfactory performance. The state of the system undergoes the following dynamics

$$\frac{dx_1(t)}{dt} = \alpha(x_2(t) - x_1(t)) \tag{75}$$

$$\frac{dx_2(t)}{dt} = x_1(t)(\gamma - x_3(t)) - x_2(t) \tag{76}$$

$$\frac{dx_3(t)}{dt} = x_1(t)x_2(t) - \beta x_3(t) \tag{77}$$

where $\alpha = 10$, $\beta = 8/3$, and $\gamma = 28$. For this selection of parameters, the Lorenz system has chaotic solutions. Almost all initial points will tend to the invariant set; the Lorenz attractor. In the presented application, the initial condition is assumed Gaussian with mean $\hat{\mathbf{x}} = \begin{bmatrix}10 & 10 & 10\end{bmatrix}^T$ and diagonal covariance matrix, with the same standard deviation for each component of the state: $\sigma_{\mathbf{x}} = 2.5$. The state is integrated in time at 30 Hz, with observations taken each time step. The measurement model consists in the range of the state from origin

$$y_k = \sqrt{x_1(t_k)^2 + x_2(t_k)^2 + x_3(t_k)^2} + \mu_k \tag{78}$$

where measurement noise is assumed Gaussian with zero mean and standard deviation $\sigma_\mu = 1$.

Figure 6 shows, on the top, one of the trajectories described by the state of the system, in its three components. The Lorenz attractor has two main lobes symmetric with respect to the x_3 axis: the resulting pathway has been labeled as a "butterfly" shape. A Monte Carlo analysis with 1000 realizations with SACEMM-2-5-2 is reported on the bottom of Figure 6. For each ith component of the state, the estimation error of each realization is calculated according to Equation (73), and reported in gray. Analogously with the previous application, the continuous blue lines represent the predicted error standard deviations, as 3σ values, of each component, while the dashed blue lines are the effective error standard deviations, again as 3σ values, calculated directly from the Monte Carlo realizations at each time step. The overlapping between the dashed and the continuous lines indicates that SACEMM-2-3-2 is a consistent filter able to correctly estimate its own uncertainties. The black lines are the error means and they prove the unbiased nature of the proposed filtering technique, as expected from the MMSE theory.

The performance of the filters have been assessed through a covariance comparison carried out with multiple Monte Carlo analysis, each performed with 1000 runs. Figure 7 reports, for each filter, the effective and the predicted error standard deviation. As shown in previous analysis, the dashed lines represent the actual uncertainty level of the filter, while the continuous line are the filter's own uncertainty estimate, evaluated according to Equation (74). Figure 7 reports SACE-c-η-μ and SACEMM-c-η-μ with different sets of parameters.

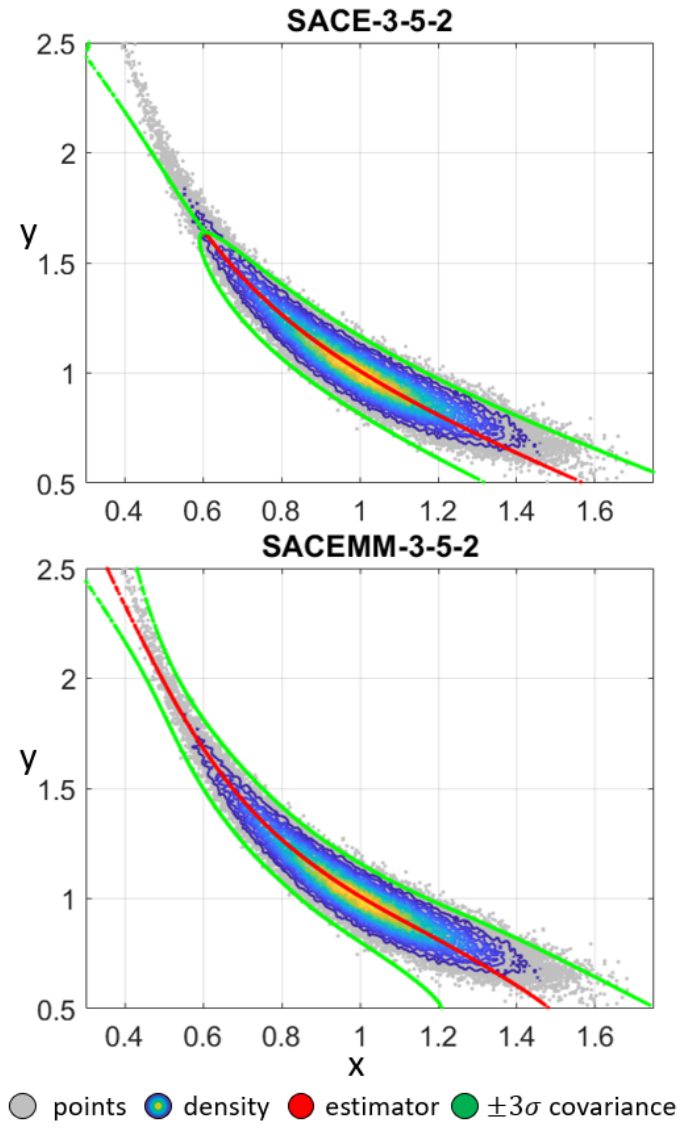

Fig. 3. SACE-3-5-2 vs. SACEMM-3-5-2. Posterior distribution (gray), the estimator functions (red), and their confidence levels (green).

For the basic selection of SACEMM-1-1-0, the filter reduces to the GSF, where the dynamics are linearized around the current center of each model and the update is a linear estimator. The GSF is reported with black lines and it fails to estimate the state of the system. The effective covariance indicates divergence and goes out of scale with respect to the predicted standard deviation.

The state of the system is also estimated with a 10^4 particles BPF, shown in orange, and the IKF. The IKF diverges rapidly and is not reported in the figure since the errors quickly reaches out-of-scale large values. The linearization of the dynamics employed by the IKF is not sufficient to correctly propagate the state covariance forward in time. The divergence of the IKF might be connected to the poor time propagation. However, this issue might be alleviated by using the Levenberg-Marquardt algorithm [45]. The BPF performs better than the GSF but shows convergency problems and it is not able to achieve an accurate estimate of the state.

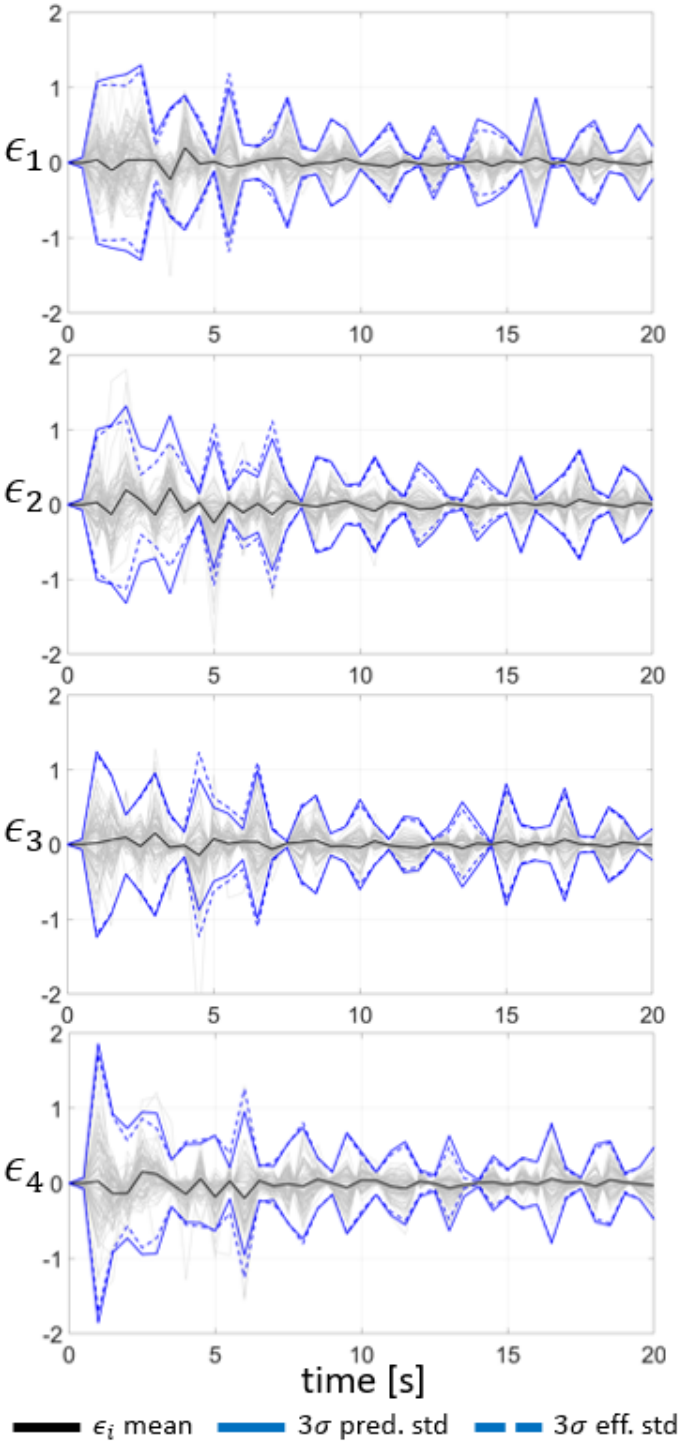

Fig. 4. Monte Carlo performance analysis with SACE-2-3-2: 100 runs.

The BPF has issues due to the lack of process noise in the dynamics. After resampling, the propagated particles are not spread enough to be an appropriate representation of the prior uncertainty in order to accurately perform the measurement update.

SACE-c-η-μ is analyzed with the traditional zeroth order covariance estimation, SACE-2-5-0 shown in green, and with a second order covariance polynomial estimator, SACE-2-5-2 in blue. The two filters behave similarly: they both

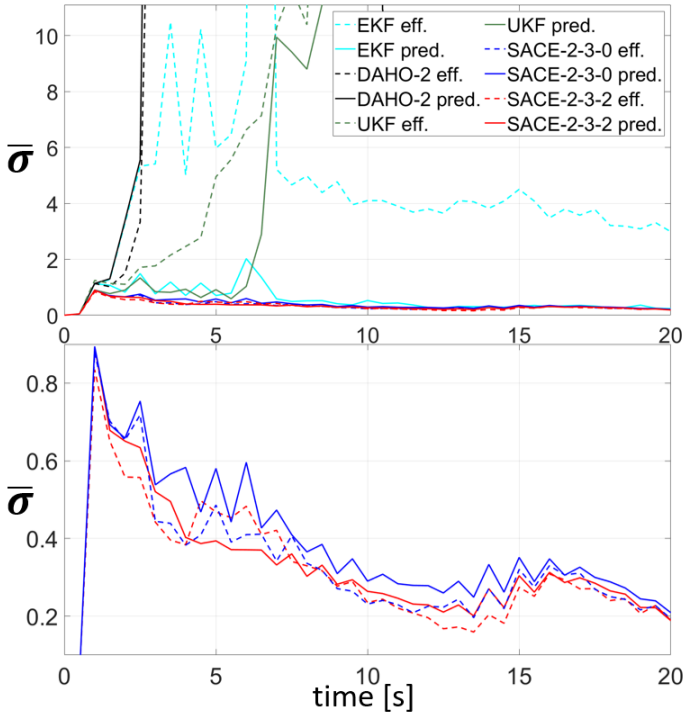

Fig. 5. Lorenz96: covariance comparison among different filters.

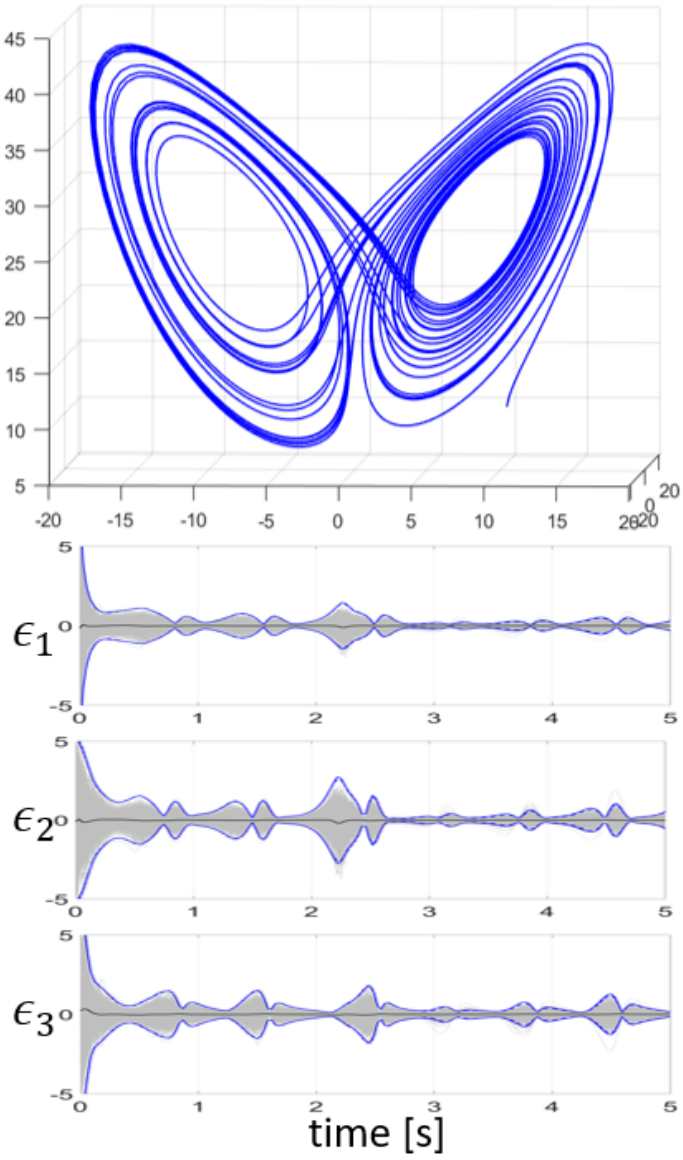

Fig. 6. Trajectory and SACEMM-2-3-2 Monte Carlo analysis results. 1000 runs.

show convergence with consistency for the first half of the simulation, and they diverge for the remaining half. At time step $t = 2.2$ seconds, the state of the system is near the origin, in between the two lobes of the Lorenz attractor. This point is critical because, due to uncertainty, the estimated state may select the incorrect lobe, while the true state follows the other. The measurement model, consisting solely in the range, gives no beneficial information regarding the lobe selection: thus, the correction terms in the update step do not help tracking the state along the correct path. Consequently, in some realizations of the Monte Carlo analysis, the filter is tracking the state of the system as if it were on the incorrect lobe. The radial nature of the range measurement provides no information to the estimator into correcting the estimated state because of the symmetric nature of the"butterfly" trajectory. Therefore, both SACE-2-5-0 and SACE-2-5-2 show inconsistency after the critical point, and the effective standard deviation is bigger than the predicted one. However, it is worth noticing that the dashed blue line settles below the dashed green line, indicating an increase in accuracy achieved due to the estimated covariance being connected with the measurement outcome. Lastly, SACEMM-2-3-2 is reported in red and it is the only filter that shows convergency and consistency during the whole length of the simulation. The introduction of multiple models improves accuracy especially around the critical point, where smaller subdomains make it easier for the filter to follow the right path along the correct lobe. If a model separates from the others, following the incorrect lobe, it is weighted down in order to ensure a correct estimation. The division of the system uncertainties in smaller subdomains helps the filter track the correct trajectory while the high order polynomial update ensures excellent accuracy levels. SACEMM-c-η-μ has better performance than SACE-c-η-μ when the initial uncertainties of the state of the system are exceptionally high and when the propagated state PDF is multi-modal.

The second part of Figure 7 reports an analysis on the computational time requested by each filter. The parameter τ is evaluated as

$$\tau = \frac{\mathcal{T}_i}{\mathcal{T}_{GSF}} \tag{79}$$

where $\mathcal{T}_i$ is the computational time of the ith filter, with i={GSF, BPF, SACE-2-5-0, SACE-2-5-2, SACEMM-2-3-2}. Therefore, the τ bar expresses the relative computational effort among the different filters for theis application. The τ analysis shows that the BPF is the computational heaviest filter, while the computational time requested by SACE-c-η-μ and SACEMM-c-η-μ changes depending on the selection of their parameters.

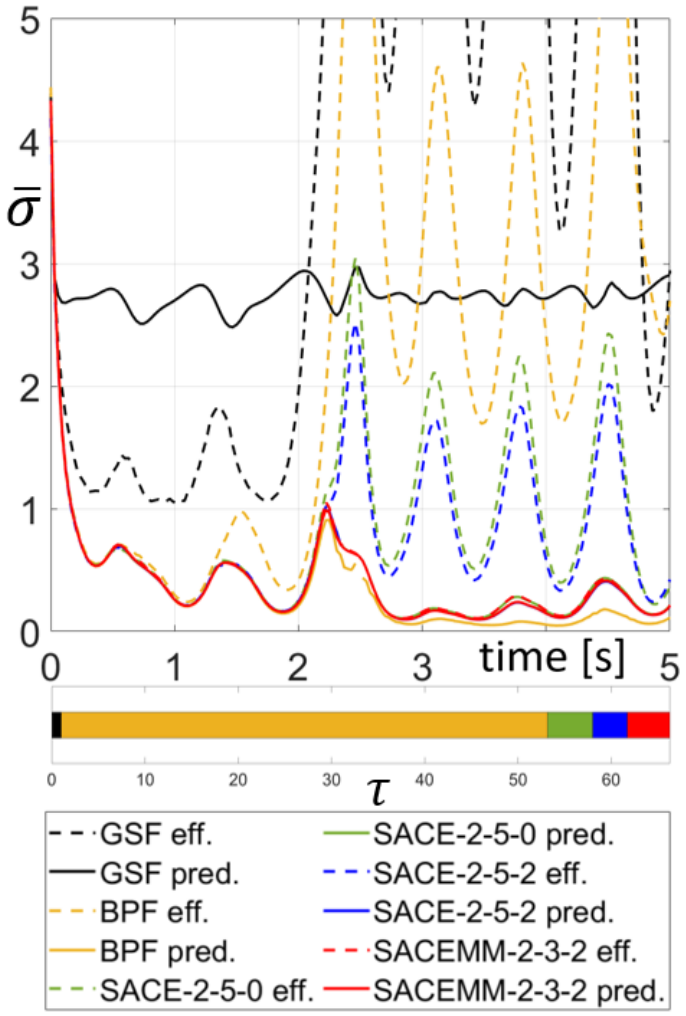

Fig. 7. Lorenz63: covariance comparison and time analysis among different filters.

VI. Conclusions

A novel filter based on a double estimator has been presented. The new technique estimates the conditional mean and the conditional covariance of the posterior distribution by applying, sequentially, two polynomial estimators, using the same measurement outcome. The new approach better matches the estimated state with its error standard deviation, which is now a polynomial function of the measurement. Therefore, the newly proposed filter is able to reduce the error uncertainty when the posterior distribution gets narrower around a low probability realization of the measurement. In turn, the better representation of the uncertainty produces a better estimate of the state during the subsequent measurement updates.

Three numerical examples have been reported. The scalar application gives a visual representation of the benefits of the polynomial approximation of the true MMSE and its covariance. Thus, the higher the order of the updates, the more precise the relative state estimate and its covariance. The vectorial application underlines the benefits of predicting the covariance by considering its estimation as working with an augmented state. The new state estimate improves in accuracy and a smaller error standard deviation is obtained. The multiple model filter is more robust against high initial standard deviations and multi-modal distributions.

VII. Acknowledgment

This work was sponsored in part by the Air Force Office of Scientific Research under grant number FA9550-18-1-0351

References

[1] Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, Vol. 82, No. Series D, March 1960, pp. 35–45, doi:10.1115/1.3662552.

[2] Kalman, R. E. and Bucy, R. S., "New Results in Linear Filtering and Prediction," *Journal of Basic Engineering*, Vol. 83, No. Series D, March 1961, pp. 95–108, doi:10.1115/1.3658902.

[3] Gelb, A., editor, *Applied Optimal Estimation*, The MIT press, Cambridge, MA, 1974, ISBN:9780262200271.

[4] Julier, S. J. and Uhlmann, J. K., "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, Vol. 92, No. 3, March 2004, pp. 401–422, doi: 10.1109/JPROC.2003.823141.

[5] Julier, S. J., Uhlmann, J. K., and Durrant-Whyte, H. F., "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, Vol. 45, No. 3, March 2000, pp. 477–482, doi: 10.1109/9.847726.

[6] Cavenago, F., Di Lizia, P., Massari, M., Servadio, S., and Wittig, A., "DA-based nonlinear filters for spacecraft relative state estimation," *2018 Space Flight Mechanics Meeting*, 2018, p. 1964, doi: 10.2514/6.2018-1964.

[7] Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Vol. 64 of *Mathematics in Sciences and Engineering*, Academic Press, New York, New York 10003, 1970, ISBN: 9780486318196.

[8] Servadio, S., *High Order Filters For Relative Pose Estimation Of An Uncooperative Target*, Master's thesis, Politecnico di Milano, Milan, Italy, 20156, 2017.

[9] Arasaratnam, I., Haykin, S., and Elliot, R. J., "Discrete-Time Nonlinear Filtering Algorithms using Gauss-Hermite Quadrature," *Proceedings of the IEEE*, Vol. 95, No. 5, May 2007, pp. 953–977, doi: 10.1019/JPROC.2007.894705.

[10] Arasaratnam, I. and Haykin, S., "Cubature Kalman Filters," *IEEE Transactions on Automatic Control*, Vol. 54, No. 6, 2009, pp. 1254–1269, doi: 10.1109/TAC.2009.2019800.

[11] Valli, M., Armellin, R., Di Lizia, P., and Lavagna, M., *International Astronautical Federation,*, ISBN: 9781622769797.

[12] Schei, T. S., "A finite-difference method for linearization in nonlinear estimation algorithms," *Automatica*, Vol. 33, November 1997, pp. 2053–2058, doi: 10.1016/s005-1098(97)00127-1.

[13] Ravn, O., Norgaard, M., and Poulsen, N. K., "New developments in state estimations for nonlinear systems," *Automatica*, Vol. 36, No. 11, November 2000, pp. 1627–1638, doi: 10.1016/s005-1098(00)000089-3.

[14] Sorenson, H. W. and Alspach, D. L., "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, Vol. 7, No. 4, July 1971, pp. 465–479, doi:10.1016/0005-1098(71)90097-5.

[15] Alspach, D. and Sorenson, H., "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Transactions on Automatic Control*, Vol. 17, No. 4, August 1972, pp. 439–448, doi: 10.1109/SAP.1970.270017.

[16] Gordon, N. J., Salmond, D. J., and Smith, A. F., "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE proceedings F (radar and signal processing)*, Vol. 140, IET, 1993, pp. 107–113.

[17] Schon, T., Gustafsson, F., and Nordlund, P.-J., "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Transactions on signal processing*, Vol. 53, No. 7, 2005, pp. 2279–2289, doi: 10.1109/TSP.2005.849151.

[18] Van Der Merwe, R., Doucet, A., De Freitas, N., and Wan, E. A., "The unscented particle filter," *Advances in neural information processing systems*, 2001, pp. 584–590.

[19] Hutter, F. and Dearden, R., "The gaussian particle filter for diagnosis of non-linear systems," *IFAC Proceedings Volumes*, Vol. 36, No. 5, 2003, pp. 909–914.

[20] Murata, M., Nagano, H., and Kashino, K., "Monte Carlo filter particle filter," *2015 European Control Conference (ECC)*, IEEE, 2015, pp. 2836–2841.

[21] Fisher, J., "Optimal nonlinear filtering," *Advances in Control Systems*, Vol. 5, Elsevier, 1967, pp. 197–300.

[22] De Santis, A., Germani, A., and Raimondi, M., "Optimal quadratic filtering of linear discrete-time non-Gaussian systems," *IEEE Transactions on Automatic Control*, Vol. 40, No. 7, 1995, pp. 1274–1278.

[23] Carravetta, F., Germani, A., and Raimondi, N., "Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise," *IEEE Transactions on Automatic Control*, Vol. 42, No. 8, 1997, pp. 1106–1126, doi:10.1109/9.618240.

[24] Lan, J. and Li, X. R., "Nonlinear estimation based on conversion-sample optimization," *Automatica*, Vol. 121, 2020, pp. 109160, https://doi.org/10.1016/j.automatica.2020.109160.

[25] Liu, Y., Li, X. R., and Chen, H., "Generalized linear minimum mean-square error estimation with application to space-object tracking," *2013 Asilomar Conference on Signals, Systems and Computers*, IEEE, 2013, pp. 2133–2137.

[26] Lan, J. and Li, X. R., "Nonlinear estimation by LMMSE-based estimation with optimized uncorrelated augmentation," *IEEE Transactions on Signal Processing*, Vol. 63, No. 16, 2015, pp. 4270–4283.

[27] Zhang, Y. and Lan, J., "Gaussian sum filtering using uncorrelated conversion for nonlinear estimation," *2017 20th International Conference on Information Fusion (Fusion)*, IEEE, 2017, pp. 1–8.

[28] Servadio, S. and Zanetti, R., "Recursive Polynomial Minimum Mean-Square Error Estimation with Applications to Orbit Determination," *Journal of Guidance, Control, and Dynamics*, Vol. 43, No. 5, 2020, pp. 939–954, https://doi.org/10.2514/1.G004544.

[29] Servadio, S., Zanetti, R., and Jones, B. A., "Nonlinear Filtering with a Polynomial Series of Gaussian Random Variables," *IEEE Transactions on Aerospace and Electronic Systems*, 2020, doi: 10.1109/TAES.2020.3028487.

[30] Lefebvre, T., Bruyninckx, H., and Schutter, J. D., "Comment on "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators"," *IEEE Transactions on Automatic Control*, Vol. 47, No. 8, 2002, pp. 1406–1408, doi: 10.1109/TAC.2002.800742.

[31] M., B., *Modern Map Methods in Particle Beam Physics*, Academic Press, 1999.

[32] Makino, K. and Berz, M., "Cosy infinity version 9," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Vol. 558, No. 1, 2006, pp. 346–350, doi: 10.1016/j.nima.2005.11.109.

[33] Massari, M., Di Lizia, P., Cavenago, F., and Wittig, A., "Differential Algebra software library with automatic code generation for space embedded applications," *2018 AIAA Information Systems-AIAA Infotech@ Aerospace*, 2018, p. 0398, doi: 10.2514/6.2018-0398.

[34] Armellin, R., Di Lizia, P., Bernelli-Zazzera, F., and Berz, M., "Asteroid close encounters characterization using differential algebra: the case of Apophis," *Celestial Mechanics and Dynamical Astronomy*, Vol. 107, No. 4, 2010, pp. 451–470, doi: 10.1007/s10569-010-9283-5.

[35] Valli, M., Armellin, R., Di Lizia, P., and Lavagna, M., "Nonlinear mapping of uncertainties in celestial mechanics," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 1, 2012, pp. 48–63, doi: 10.2514/1.58068.

[36] Armellin, R. and Di Lizia, P., "Probabilistic optical and radar initial orbit determination," *Journal of Guidance, Control, and Dynamics*, Vol. 41, No. 1, 2018, pp. 101–118, https://doi.org/10.2514/1.G002217.

[37] Wittig, A., Di Lizia, P., Armellin, R., Makino, K., Bernelli-Zazzera, F., and Berz, M., "Propagation of large uncertainty sets in orbital dynamics by automatic domain splitting," *Celestial Mechanics and Dynamical Astronomy*, Vol. 122, No. 3, 2015, pp. 239–261, https://doi.org/10.1007/s10569-015-9618-3.

[38] Raihan, D. and Chakravorty, S., "Particle Gaussian mixture filters-I," *Automatica*, Vol. 98, 2018, pp. 331–340, https://doi.org/10.1016/j.automatica.2018.07.023.

[39] Isserlis, L., "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, Vol. 12, No. 1/2, 1918, pp. 134–139.

[40] Cavenago, F., Di Lizia, P., Massari, M., and Wittig, A., "On-board spacecraft relative pose estimation with high-order extended Kalman filter," *Acta Astronautica*, Vol. 158, 2019, pp. 55–67.

[41] Di Lizia, P., Massari, M., and Cavenago, F., "Assessment of onboard DA state estimation for spacecraft relative navigation," 2017, Final report, ESA .

[42] Bell, B. M. and Cathey, F. W., "The iterated Kalman filter update as a Gauss-Newton method," *IEEE Transactions on Automatic Control*, Vol. 38, No. 2, 1993, pp. 294–297.

[43] Servadio, S. and Zanetti, R., "Maximum A Posteriori Estimation of Hamiltonian Systems with High Order Series Expansions," .

[44] Terejanu, G., Singla, P., Singh, T., and Scott, P. D., "Adaptive Gaussian Sum Filter for Nonlinear Bayesian Estimation," *IEEE Transactions on Automatic Control*, Vol. 56, No. 9, 2011, pp. 2151–2156, doi: 10.1109/TAC.2011.2141550.

[45] Skoglund, M. A., Hendeby, G., and Axehill, D., "Extended Kalman filter modifications based on an optimization view point," *2015 18th International Conference on Information Fusion (Fusion)*, IEEE, 2015, pp. 1856–1861.